ARIADNE

D2.3: Preliminary innovation agenda and action plan

Authors:

Guntram Geser, SRFG

Hannes Selhofer, SRFG

Ariadne is funded by the European Commission's 7th Framework Programme.

Version: 1.0 (*final*) **November 2015**

Authors: **Guntram Geser (SRFG)**

Hannes Selhofer (SRFG)

Contributing partners: **Michael Charno, Julian Richards and Holly Wright (UoY ADS)**

Panos Constantopoulos, Costis Dallas, Anastasia Michailidou, Christos Papatheodorou and Despoina Tsiafaki (ATHENA RC)

Martin Doerr and Maria Theodoridou (FORTH)

Achille Felicetti and Franco Niccolucci (PIN)

Elizabeth Fentress (AIAC)

César González-Pérez (Incipit CSIC)

Ulf Jakobsson and Karolina Kegel (SND)

Attila Kreiter (MNM-NÖK)

Martin Kuna (ARUP- CAS)

Carlo Meghini (CNR ISTI)

Douglas Tudhope (USW/Glamorgan)

Edeltraud Aspöck and Anja Masur (OEAW)

Nadezhda Kecheva (NIAM-BAS)

Sara Di Giorgio (ICCU)

Simona Simionescu (ARHEO)

Quality Review **Holly Wright (ADS) and Kate Fernie (PIN)**

ARIADNE is a project funded by the European Commission under the Community's Seventh Framework Programme, contract no. FP7-INFRASTRUCTURES-2012-1-313193. The views and opinions expressed in this report are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Table of contents

1 Introduction and overview

This document is a contractual deliverable (D2.3) of the ARIADNE - Advanced Research Infrastructure for Archaeological Dataset Networking in Europe project. The deliverable presents the Preliminary Innovation Agenda and Action Plan that has been produced under WP-2 "Community Building and Innovation".

1.1 Objectives, stakeholders and ARIADNE's role

Objectives

The objectives of the Innovation Agenda and Action Plan are directly related to the overall goals of the ARIADNE project: The project addresses the fragmentation of archaeological datasets in Europe by providing e-infrastructure and services that allow discovery and (re-)use of data held by different accessible and interoperable digital archives and collections. Thereby ARIADNE supports a culture of sharing and collaborative use of archaeological data across national, institutional and disciplinary boundaries.

In this context, the main objectives of the Innovation Agenda and Action Plan are:

- to point out innovation needs in archaeology with regard to open data sharing, digital archives and research e-infrastructure and services,
- to describe innovation potential within 5/10-year horizons, e.g. in the form of scenarios of transformative change in research and data management practices enabled or triggered by ARIADNE,
- to suggest actions ARIADNE and other sector stakeholders can take to meet the perceived innovation needs and enable favourable changes in research practices.

It is understood that innovation in the complex fields addressed will require orchestrated measures of research institutes, digital archives, funding agencies and other stakeholders.

Innovation needs

The archaeological research and data management communities face several challenges. There are innovation needs with regard to:

- Implementation of policies and incentives for open sharing of data
- Mobilisation of high-quality data from all relevant sources
- Wider uptake of common data standards and machine-readable semantics
- Provision of advanced services for data interoperability, discovery and access
- Coordination among e-infrastructure and service providers in and beyond the sector
- Development of environments and tools for innovative e-research

A major aspect of these challenges is that archaeology is an extensive and multi-disciplinary field of research that spans several domains of the humanities and the natural sciences, cultural heritage research and public administration, academic scholarship as well as commercial services.

Stakeholders and beneficiaries

The ARIADNE Innovation Agenda and Action Plan addresses all stakeholders in research e-infrastructure and services for sharing and (re-)using openly accessible archaeological data. The stakeholders include:

- The ARIADNE project partners
- Research policy makers and directors of funding agencies
- Directors of research institutes and boards of research associations
- Steering committees of collaborative research projects
- Heads of research groups and individual researchers
- University-based research libraries and institutional repositories
- Directors of national and international digital archives
- Cultural heritage agencies and museums
- Domain and cross-domain e-infrastructure and service providers
- Technology and software developers and providers

Because of the diversity of the stakeholders, mutual understanding, trusting relations and joint capacity building are crucial for the successful development of open data resources, common e-infrastructure and services. Most beneficiaries will be archaeological researchers and other end-users (e.g. cultural heritage administrators, interested citizens) for whom services such as a common data portal for searching and accessing data will become available. Also data archives will benefit from a common e-infrastructure and services, for example, through wider data access, recognition of curatorial work, and interlinking of data resources which can provide added value to the research community.

ARIADNE's roles

With the Innovation Agenda and Action Plan ARIADNE promotes coordination among many on-going but currently fragmented efforts in building e-infrastructures, data and knowledge resources for archaeological research in Europe and beyond. It is hoped that this will foster cohesion and synergies among initiatives, yield effective spending of available funds, and result in sustainable digital archives, e-infrastructures and services as needed by the research community.

How can ARIADNE help tackle the existing challenges? ARIADNE can play important roles in helping others to make a difference:

- *Energizer* – take the challenges as opportunities to innovate, suggest new approaches to archaeological data, invite organisations large and small to participate and contribute
- *Enabler* – establish a platform for community building, foster close cooperation of researchers, data managers and technology experts, offer guides to good practice and support in capacity development
- *Opener* – promote open access data, help open up 'data silos', and make shared data part of the scholarly record – persistent, available, citable and rewarded
- *Integrator* – align currently dispersed initiatives, provide interoperability services, help to embed the use of e-infrastructure into research culture

Stakeholders may ask if the e-infrastructure and services ARIADNE develops will replace or complement existing facilities and services. ARIADNE will not replace existing research e-infrastructures (e.g. digital archives and preservation services) but will provide integrating functionality and services on top of them. Thereby ARIADNE will help to make currently isolated archaeological data more accessible and useable for the research community and other groups, e.g. heritage management agencies and citizens. However, the ARIADNE e-infrastructure and services can only come into play when appropriate local information technology, data resources and principles of open data sharing are in place.

Stakeholders can take the ARIADNE initiative as an opportunity to organise and align their own efforts in open data mobilisation and provision. This may be stock taking (e.g. what data resources are available, how can they be made accessible, etc.), small-scale pilot projects, and participation in national and international initiatives. In particular we see good opportunities for community-building around national and other community data archives. Where digital archives are currently missing in European countries, acknowledged existing data centres in other countries can serve as role models.

Because of mandates by research funders to deposit and share the data produced in publicly funded research, there will be increasing demand for state-of-the-art archives. Overall there is a need for skilled data managers at all levels, including local/institutional repositories and community archives, as well as e-infrastructures and service providers. To support capacity building, ARIADNE provides guidance material, training offers, and trans-national access to partner facilities and expertise.

1.2 Promoting innovation over the next 5-10 years

The overall objectives of the ARIADNE Innovation Agenda and Action Plan are to point out the innovation needs of the archaeological research and data management community in Europe in the next 5-10 years, and to suggest actions for ARIADNE and other stakeholders on how to meet these needs. As the ARIADNE project centers on the development of e-infrastructure and services for data networking and access, the *preliminary* agenda and action plan for the five-year horizon focuses on open data sharing, data interoperability, access and re-use in further research.

The 10-year horizon looks toward potential innovations in ICT-enabled archaeological research ("e-archaeology"), partly based on the progress in open data sharing and access in the next few years. Consequently the *Final* Innovation Agenda and Action Plan will focus on road mapping towards "e-archaeology" in 2025. But the *preliminary* version already addresses some selected topics and methods considered for the road mapping.

The table below gives a schematic overview of the approach:

Agenda & Plan	***Horizon***	***Innovation focus***
Preliminary version	5 years (2020)	Open data sharing, data interoperability, and data access e-infrastructure & services
Final Version	10 years (2025)	Novel ICT-enabled archaeological research ("e-archaeology"), partly based on the progress achieved within the next 5 years

Table: Different innovation focus of 5-year and 10-year horizon

The ARIADNE Innovation Agenda and Action Plan does not promote a "technology-push" view of innovation and uptake of new digital practices in archaeological research. But Information and Communication Technology (ICT), systems/tools and services play an important role in enabling novel archaeological research methods and practices. Examples of impact and success achieved in recent decades include for instance the wide use of Geographic Information Systems/Services (GIS) and 3D methods in the representation and analysis of archaeological sites. Instead of a technology-push perspective, ARIADNE focuses on the middle ground between progress in data standards and technologies and potential users' readiness to adopt and benefit from these advances ("adopters pull"). This may also require exploring if solutions that have been proven to work in principle (e.g. small-scale Linked Data projects) can be implemented and sustained at a larger scale.

The ARIADNE project will run until January 2017, while the innovation horizons are 5 and 10 years. Therefore ARIADNE can mainly set an agenda, and together with other stakeholders take first steps towards achieving the innovation goals. Below we describe in greater detail the rationale and content of the different versions of the Innovation Agenda and Action Plan.

Preliminary innovation agenda and action plan

The Preliminary Innovation Agenda and Action Plan centres on objectives and challenges in the 5-year horizon (2020). This is about immediate needs, recommended actions and measures that should be taken in order to achieve progress in priority areas, with regard to open data and data interoperability, for instance.

ARIADNE develops data infrastructure and services for the archaeological research and data management communities, e.g. dataset registration, search, access and other services. The project is not charged to provide a virtual research or distributed computing solution for specific domains of archaeological research. There are innovation needs in these areas, but the development and uptake of solutions will more likely fall in the 10-year horizon. Project partners will explore approaches to e-archaeology as far as the ARIADNE infrastructure and data resources allow such use cases.

The immediate innovation needs are in the area of data sharing and access, and more of a social/institutional than technical character. Many archaeologists, like the researchers in most other disciplines, are not yet prepared to make data openly available to others outside a research project or organisation. Therefore it is vital that the ARIADNE project contributes to the emergence of a culture of open sharing of archaeological data, trusted data archives (where missing at present), and mobilisation of data resources that are interoperable and re-usable. The focus areas of the 5-year innovation agenda are:

- *Focus area 1 – Research e-infrastructures and digital resources:* data infrastructure for archaeological research, including coordination with infrastructures of related domains; mobilisation of high-quality data from relevant sources
- *Focus area 2 – Culture of open sharing and re-use of data:* open data policies/mandates, removal of barriers, and promotion of data sharing, re-use and citation
- *Focus area 3 – Data archives and curation of archaeological research data:* reliable and cost-effective community archives for long-term data curation and access
- *Focus area 4 – Capacity building for open data sharing:* institutional policies, guidance, training and other support for open data practices
- *Focus area 5 – Providing services and enabling novel applications:* provision of data services (e.g. dataset registration, cross-archive search and other services), and support for further novel applications for archaeological research

The primary responsibility and agency with regard to the focus areas 2 – 4 is with research funding agencies, research institutions and associations, and mandated institutional and community data archives. ARIADNE can contribute to these goals, for example, by promoting open access to archaeological data, offering guides to good practice and supporting capacity building.

Many of the actions suggested for innovation through community-level data infrastructure and services in the focus areas 1 and 5 fall in the remit of ARIADNE. For example, without interoperability of (meta)data and knowledge organisation systems (e.g. terminology, ontologies), integrating e-infrastructures and data search and access across many archaeological archives in Europe will be hardly possible. Going beyond these services, ARIADNE expects to also enable novel applications based on enhanced data interoperability, e.g. creation of Linked Open Data applications by interested developer communities.

Towards the final version

The Final Innovation Agenda and Action Plan will focus on the 10-year horizon, innovation potentials and development paths of digital archaeology until 2025. Assuming that the 5-year objectives are met sufficiently until 2020, further progress in ICT-enabled archaeological research can be foreseen and targeted. Thus long-term scenarios will look into potential research-focused innovations, explore emerging new perspectives and capabilities, and suggest pathways towards innovative and potentially transformative "e-archaeology".

The focus of the final agenda and action plan will shift from the immediate need of fostering open access to re-usable and interoperable data, to required actions for enabling novel types of ICT-enabled archaeological research. The shift in horizon and focus (5 > 10 years, data infrastructure > e-research) necessitates the use of methods which allow road mapping for innovation, e.g. fore-sighting, scenario building and other methods.

The *preliminary* innovation agenda outlines the objectives, methods and selected topics of the road mapping. As the work on the 10-year horizon will be conducted until November 2016, it is foreseen interim results (e.g. scenarios) will be issued in the form of working documents, inviting suggestions and contributions by the wider ARIADNE stakeholder community.

The Final Innovation Agenda and Action Plan (D2.4, November 2016) will update the preliminary version, specifically it will:

- describe major new developments in the landscape of open access to archaeological data, digital archives and relevant e-infrastructures in Europe;
- adapt the recommendations for short- to medium-term actions in the 5-year horizon (2020) accordingly, taking account of responses from stakeholders to recommended actions for the initial innovation plan, and
- develop a good understanding of innovation potentials and development paths of ICT-enabled archaeological research until 2025, and suggest actions for advancing "e-archaeology" within this 10-year horizon.

A larger part of the final report will present the results of the work conducted until November 2016 for recommending approaches for the long-term horizon. Major elements that connect these approaches with the ARIADNE e-infrastructure may be a higher level of semantic interoperability for archaeological data and new tools which can exploit this interoperability, for example, within archaeological virtual research environments.

The final report will also include a chapter that presents the long-term sustainability model and plan for supporting the ARIADNE e-infrastructure and services beyond the funded period (Task 2.6, led by the project coordinator PIN). This period ends in January 2017. Around this time and onwards we may expect a growing demand for the core set of ARIADNE data services (e.g data registration, search and access) as well as for new research-focused applications.

Methods and sources

ARIADNE aims to provide solid evidence on the main topics of the innovation agenda and the suggested actions stakeholders may take to drive openness and innovation in archaeological research in a 5-10 year horizon. The agenda topics and suggested actions in the 5-year horizon mainly concern open research data, data archives and e-infrastructure services. The main intended users of these resources are the archaeological research and data curation communities in Europe.

The ARIADNE e-infrastructure and services are a significant element in the 5-year innovation horizon and beyond. Therefore the project has conducted extensive work to ascertain that the ARIADNE e-infrastructure and services meet the current needs and requirements of the intended user community. The e-infrastructure and services should meet the needs of large segments of the community in the 5-year horizon, as far as these needs can be supported with current technical means and data resources described by providers in the ARIADNE registry.

User needs surveys, interviews and background studies have been carried out and reported[1]. For example, the First Report on Users' Needs, among other work, presents results of an online survey with qualified input from over 500 respondents. The second report provides insights into what kind of services archaeological researchers expect from current and future online portals like the ARIADNE data portal. User needs and requirements research has an important role to play in the innovation agenda, and the available results (also of other studies) are frequently referenced in this report. Much further research will be required especially on the virtual research environments (VREs) archaeologists may wish to use, a topic which we see more in the 10-year horizon, at least concerning larger segments of the archaeological research community.

This report builds on the results of the ARIADNE user needs studies and provides further evidence for suggested actions in the 5/10-year innovation horizons. This includes background studies on core themes such as the current availability of different types of data resources, data repositories, re-use of data, and other special topics.

The report of course takes note of particularly relevant literature that has been published in the last 10-15 years, which is referenced in the thematic chapters of this report. In the literature on some themes archaeology is seldom addressed specifically, e-infrastructures and their impact on research practices, for instance. Archaeology is generally subsumed under the humanities. This often meant trying to discern significant differences between archaeology and other humanities, for example, taking into account archaeology's strong foothold in the applied sciences.

In general, however, the priorities in the 5-year horizon are the same for archaeology, the humanities in general as well as most other disciplines. They are not technological but concern growing an open culture of research. This is an immediate priority which, however, will require attention beyond the 5-year horizon. Other innovation topics clearly fall more in the 10-year horizon, in which we expect a stronger demand for ICT-enabled archaeological research ("e-archaeology").

A first overview of innovation priorities in the 5-year horizon has been published and presented in November 2014 on the occasion of the conference on "Research Infrastructures and e-Infrastructures for Cultural Heritage", 13-14 November 2014, at the National Central Library in Rome (Geser 2014a)[2]. Further suggestions by the ARIADNE project partners have been collected and discussed in the months thereafter. The initial priority topics and suggested actions have been extended also based on insights gained in additional background studies.

Interim results of the work on the 10-year horizon have been presented and discussed in the 2-day ARIADNE Expert Forum on Digital Futures of Archaeological Practice 2020-2025, 2-3 July 2015 in Athens[3]. The expert forum gathered ideas and discussed scenarios on how digital archaeological practices might evolve in the next 5-10 years. The forum has been organised by the Digital Curation

[1] The results of ARIADNE's work on user needs and requirements are documented in two deliverables, D2.1 (April 2014) and D2.2 (February 2015); the reports are available from the project website, http://www.ariadne-infrastructure.eu/Resources

[2] Geser, Guntram (2014a): From User Needs to the Innovation Plan, pp. 9-24, in: ARIADNE – The Way Forward to Digital Archaeology in Europe. ARIADNE publication, November 2014, http://www.ariadne-infrastructure.eu/content/download/4569/26666/version/2/file/Ariadne+Booklet.pdf

[3] ARIADNE Expert Forum: Digital Futures of Archaeological Practice 2020-2025 (Athens, 2-3 July 2015), http://summerschool.dcu.gr/?page_id=19#expert-forum

Unit of the IMIS-Athena Research Centre on behalf of the ARIADNE Special Interest Group on Archaeological Research Practices and Methods.

The expert forum brought together 22 senior and young researchers interested in the future of ICT-enabled archaeological research. The scenario work of the forum was co-moderated by the promoters of the Challenging Digital Archaeology Initiative, Jeremy Huggett, Gary Lock and Paul Reilly. The initiative has been launched in 2014 at the Computer Applications and Quantitative Methods in Archaeology (CAA) conference in Paris and continued at the CAA conference 2015 in Siena (Huggett *et al*. 2014 and 2015). The expert forum has been recorded and a report or paper, based on the summarised recording and further discussion, is expected before the end of 2015.

1.3 Overview of the focus areas and suggested actions

The objectives of the Innovation Agenda and Action Plan are to point out innovation needs and challenges of the archaeological research and data management community in Europe in the next 5-10 years, and to suggest actions for ARIADNE and other stakeholders on how to address the challenges. The aim of the suggested actions is to achieve favourable and potentially transformative advances in data sharing, access and (re-)use across institutional, national and disciplinary boundaries.

The overall objective of ARIADNE is to help overcome the existing fragmentation of archaeological data resources by making it easier to share, connect, access and (re-)use resources at the community-level. The main challenge for the research community is growing a culture of open data sharing and openness of research.

The preliminary innovation agenda emphasises two general, cross-cutting requirements for enabling e-infrastructures to promote innovation in research. These requirements are capacity building in open sharing of research data and coordination of stakeholders across the different levels of data infrastructures and services. The 5-year horizon also addresses some more specific requirements such as mobilization of data/content from all relevant sources and building sustainable community-level digital archives.

The 10-year horizon of the innovation agenda builds on and extends the short to medium-term actions. The overall focus of this horizon is innovative digital, ICT-enabled archaeological research and communication. To develop this part of the innovation roadmap a number of future-oriented topics have been investigated. Actions are suggested that stakeholders can begin to take now to promote progress in digital archaeology.

1.3.1 General, cross-cutting requirements

The innovation agenda highlights two general challenges of initiatives for community-level research data infrastructure and services.

Capacity building in open sharing of research data

Data infrastructure and services are being built to enable discovery, access and (re-)use of data. The investment makes little sense without open sharing of data through digital archives by the research community. Many studies on document and data repositories have shown that the “build it and they will come” approach does not work, the usage is often frustratingly small. One reason is that dedicated training and support of users is lagging behind the implementation of the technical infrastructure. There is also a need for skilled data managers at all levels, research projects, institutional and community repositories, and research e-infrastructure and services. The challenge of building and retaining a workforce of research data curators should not be underestimated.

Coordination across the different levels of e-infrastructure

Coordination of efforts is required across all levels of data management, from local efforts by university departments/institutes and research libraries, community-level national data archives, to major, European/international e-infrastructures. The primary challenges in building a rich, coherent and sustainable ecosystem of e-infrastructure and services do not concern technology, but coordination amongst the actors, building capacity and trust, managing legal aspects and, of course, costs. The involvement of very different stakeholders and resources (e.g. research institutes and laboratories, cultural heritage agencies, museums and archives), and generally tight funds, necessitate cost-effective solutions that allow benefits at the community-level. Proliferation of un-coordinated efforts that are very likely to be unsustainable should be prevented.

1.3.2 5-year innovation agenda and suggested actions

The 5-year innovation agenda comprises of five focus areas. Many of the actions suggested for innovation through community-level data infrastructure and services (focus areas 1 and 5) fall in the remit of ARIADNE. All data holders and service providers are addressed in actions suggested for the mobilisation of content/data. The research funding agencies, research institutions and associations, and institutional and community-level repositories bear the primary responsibility and agency for growing the culture and capability of open data sharing through accessible archives/repositories (focus areas 2–4). ARIADNE can contribute to these focus areas, for example by promoting open access to archaeological data, offering guides to good practice and supporting capacity building.

Focus area 1 – Research e-infrastructures and digital resources

This focus area addresses three fields of activities which aim to foster sustainable and coordinated e-infrastructures for heritage sciences, take account of their domain-specific and general requirements, and promote the mobilisation of high-quality data from all relevant sources.

E-infrastructure development and cooperation

E-infrastructures and services support scientific activities that have become increasingly collaborative, distributed and data-intensive. In the current drive to build new or upgrade existing infrastructures and resources for "e-science" there is a risk of limited funds being invested in many uncoordinated, possibly redundant and unsustainable initiatives. This risk is particularly critical in the heritage sciences, which are characterised by having many different stakeholders and data resources of research institutes and laboratories, cultural heritage agencies, and institutions such as museums and archives. ARIADNE aims to provide sustainable core e-infrastructure and services for archaeological research in Europe and to strengthen the cooperation with other e-infrastructures both within and beyond heritage sciences.

Suggested actions

- *Recognise the importance of e-infrastructure development and coordination*
- *Support the ARIADNE data infrastructure and portal and help mobilize data providers*
- *Ensure sustainability of the ARIADNE data infrastructure and portal*
- *Strengthen cooperation among e-infrastructures for heritage sciences*
- *Establish cooperation with other e-infrastructures and services that cover relevant data of related research domains*
- *Link up with providers of Distributed Computing Infrastructure and encourage use of their resources by archaeologists*

E-infrastructures and novel digital practices

Research e-infrastructures can play a significant role in enhancing established forms and ways of research as well as bringing about transformative innovation in practices and methods, in particular, in digital, web-based environments. Several general as well as domain-specific factors and requirements must be addressed so that e-infrastructures can promote innovation in archaeology and the humanities in general.

Suggested actions

- *Take account of disciplinary differences and foster the common ground*
- *Address the complex case of archaeology, which is a multi-disciplinary field of research*
- *Embed e-infrastructure and services in research practice*
- *Help overcome the current barriers to adoption and support*
- *Build and retain a skilled workforce of data curators*

Digital resources and services

The current landscape of content/data resources and services for archaeological and cultural heritage research is highly fragmented. Many resources are not easy to find and difficult to access or not accessible at all. A more targeted approach of resource development and access is required. The suggested actions for content/data holders and service providers are:

Register data archives, digital repositories and collections

- *Register data archives/repositories and other collections for archaeological research in the catalogue of the ARIADNE data portal*
- *The ARIADNE registry should become the prime location to document and discover available archaeological data resources*

Digitise cultural heritage content and make it discoverable

- *Step up the digitisation and documentation of archaeological material held by museums and other collections (e.g. unpublished excavations, grey literature, finds/objects)*
- *Implement mechanisms that allow harvesting and other programmatic access to collection metadata (e.g. OAI-based harvesting, SPARQL for Linked Data)*
- *Participate in the ARIADNE initiative for federated search and access specifically for archaeological content/data collections*

Develop online scientific reference collections

- *Take stock of existing reference collections for archaeological purposes and evaluate their relevance and current condition (e.g. actual demand, requirements for online access)*
- *Secure financial support and curatorial expertise for the development and maintenance of state-of-the-art web-based reference resources*

Mobilise laboratory facilities relevant for archaeometry

- *Investigate how major facilities as well as typical archaeometry laboratories could be mobilised to provide open access data (e.g. factors that impede open data sharing and how they might be removed)*
- *Improve the online documentation of archaeometry analyses on laboratory websites, and include pointers to publications and deposited analysis data*

- *Standardise the metadata of archaeometry documentation to promote consistent cataloguing and improve cross-laboratory discovery and access to publications and analysis data*

Virtual research environments and data processing services

- *Investigate if archaeologists need virtual research environments; e.g. what kind of research could archaeologists conduct online, what functionalities are necessary for such research, and how could they be provided in a VRE?*
- *Promote the development of relevant VREs with functionalities (tools, services) required by archaeologists to conduct research tasks online (e-research)*
- *Raise awareness of Distributed Computing Infrastructures (e.g. Grid/Cloud-based services of DCIs) amongst archaeologists who need data processing services*
- *Encourage use of DCIs by bringing together archaeological research groups with providers of DCI, especially with their developers of data processing software and services*

Focus area 2 – Culture of open sharing and re-use of data

There are many good arguments for open data such as preventing duplicative data collection, allowing replication of data analysis to scrutinize knowledge claims, and enabling new research questions to be addressed based on shared data. Data that is shared, curated, integrated, re-used and analysed with new methods gains in value rather than being only a cost factor. But there are many factors that work against open sharing of data, in particular the currently limited academic recognition and reward for making data available. All stakeholders should support open data policies/mandates and help ensure that data sharers are recognised and rewarded (e.g. data sharing could become a criterion for academic promotion and awarding of research grants). Specific measures are also required to foster openly licensed data resources, data re-use and consistent data citation. Archaeological data archives/repositories and data infrastructure will flourish only within a research culture that values open sharing of data.

Suggested actions

- *Promote a research culture of open sharing of data*
- *Support open data policies of funding bodies and institutions*
- *Suggest appropriate guidelines for open archaeological data*
- *Recommend use of open licenses for data, metadata and knowledge organization systems*
- *Foster consistent data citation so that data sharers can be recognized and rewarded*
- *Conduct studies of data re-use to better understand and support current and emerging practices*
- *Promote data re-use and highlight inspiring examples*

Focus area 3 – Data archives and curation of archaeological research data

Core services for the long-term curation and accessibility of data are provided by data archives which participate in initiatives for research e-infrastructures. State-of-the-art, certified and mandated community archives allow researchers to publish their data in a secure and trusted way. Domain-based archives are the most effective solution to overcome, or at least reduce, fragmentation and inaccessibility of archaeological data resources. The costs of post-project curation and online publication of archaeological data are only a small fraction of the total project costs at around 1-3%, depending on the type of investigation and data generated. Compared to the benefits expected from open and re-useable research data this investment seems well spent. But community archives should be stable in the long term and hence sustained support by the main funding bodies is required.

Suggested actions

- *Recognise that the costs of opening up archaeological research datasets are marginal and well spent*
- *Include the costs of open data sharing and digital archiving in project grants*
- *Recognise the advantages of domain-based community archives*
- *Ensure long-term sustainability of trustworthy data archives*
- *Encourage and support initiatives for data archives in countries where these are currently lacking for archaeologists*

Focus area 4 – Capacity building for open data sharing

To promote the open data agenda, universities, research institutes and other stakeholders should put in place policies, guidance and training. Institutional capacity building and support for researchers in the management of data is necessary so that open and re-useable data flows into data archives for long-term curation, access and (re-)use. The research community should also consider novel approaches to data description and review.

Suggested actions

- *Ensure that adequate institutional policies, guidance and other support are in place*
- *Step up capacity building and training for data management and sharing*
- *Provide support for managing data during project work*
- *Recognise high-quality metadata is required for data re-use*
- *Promote data papers for archaeological datasets*
- *Explore novel approaches to data peer review*

Focus area 5 – Providing services and enabling novel applications

Core and additional data services

ARIADNE has identified the core services which the archaeological research and data management community expects from the project. These services are:

- *A data portal that allows an overview of available but dispersed archaeological data resources*
- *Capability to search across different digital archives/repositories which hold such resources (i.e. data collections, databases, datasets of projects, etc.)*
- *Effective data discovery, browsing and filtering mechanisms, in particular based on geo-location (maps) and date-ranges/chronologies, but also other advanced options such as facetted search*
- *Data access methods according to the different access levels, types/products and interaction modes offered by the providers*

ARIADNE will provide the user community with an online facility to register and describe accessible data resources based on a common model that allows semantic integration of the information. It will offer portal services that provide the required set of data discovery and access functionalities. Strong demand has been perceived for some additional services such as visual media services for 3D models of objects, built structures and landscapes. ARIADNE has implemented such services in response to this demand.

Support for applications beyond ARIADNE

The ARIADNE data infrastructure and services will help overcome current limitations of discovery, access and (re-)use of available archaeological data. They can enable added value to be generated in the wider ecosystem of e-infrastructures and applications for the heritage sciences.

Suggested actions for ARIADNE

- *Deploy Linked Open Data (LOD) to integrate information within the portal and to link to external resources which follow LOD principles*
- *Provide interfaces that allow external applications exploit available data, metadata and conceptual knowledge*
- *Enable integrated access to data and publications (i.e. include metadata of document archives and publishers)*
- *Help enrich specialised community sites (thematic or focused on specific types of data) with relevant information, e.g. RSS feeds on newly available data*
- *Support data mobilisation through portal information (e.g. guides to good practice, available data archives)*

1.3.3 10-year innovation agenda and suggested actions

The innovation agenda includes a 10-year roadmap to go beyond the immediate innovation needs that concern archaeological research data, data archives, e-infrastructure and services, among others. Further progress in innovative digital, ICT-enabled archaeological research and communication can be foreseen, explored and targeted. For the preliminary view of the 10-year horizon a number of future-oriented topics have been investigated, further topics will be investigated as ARIADNE continues to develop the innovation agenda. The preliminary innovation agenda suggests actions which stakeholders could take today to bring about advances in digital, ICT-enabled archaeological research and communication.

Address the question of archaeology's societal relevance

- *Leverage archaeology's societal role and relevance, for example through connecting archaeological research and knowledge with current concerns and affairs such as climate change, environmental sustainability, urban agglomeration, globalization and geo-politics, regional conflicts, migration, and others*
- *Consider contributions which could allow archaeology a stronger voice in current debates*

Take account of the diversity of archaeological research practices and methods

- *Recognise that different archaeological schools of thought and research practices require different digital, ICT-based research environments and tools*
- *Focus on phases in the lifecycle of archaeological research in which significant progress in knowledge may be achieved. In the last decades data generation has seen enormous progress; in the future other phases may require more attention*
- *Recognise that issues of standardisation (e.g. data models) and cost-effectiveness are relevant for future research practices*

Target data integration for comparative and synthetic research

- *Foster the development of novel methods and tools that allow researchers to bring together and work with the variety of data required for cross-domain, interdisciplinary research*

- *Promote competence centres and programmes aimed at data integration for comparative and synthetic archaeological research*

Explore relevant future VREs for archaeological research

- *Look into VREs developed for other domains to conceive environments relevant for e-research in specific archaeological domains as well as in cross-domain collaboration*
- *Consider cases where researchers use data mediated by ARIADNE as well as data infrastructures and services of other disciplines (e.g. geo, environmental, biological data)*

Identify e-science practices based on data infrastructure and computing facilities

- *Promote collaborative way-finding for e-science approaches, methods and tools relevant to archaeological researchers*
- *Focus on e-science needs specific to archaeological research, which may differ from those of other humanities as well as the natural sciences*
- *Look for uses of low-level Grid/Cloud based services and emerging examples of archaeological applications of "big data" mining and other methods*

Propose grand challenges for the digital archaeology community

- *Seek grand challenges that inspire the research community to push the boundaries of digital archaeology*
- *Suggest challenges that promote mobilisation and integration of datasets for domain and cross-domain, interdisciplinary research*
- *Bring together domain experts and developers to create methods and tools for such research*

Evolve a Web of archaeological Linked Open Data for research

- *Promote publication of LOD datasets (collections, databases) by more archaeological and other cultural heritage institutions, especially based on mappings to the extended CIDOC-CRM*
- *Foster a community of LOD curators who ensure reliable availability and interlinking of LOD resources (datasets and vocabularies)*
- *Develop LOD-based applications that demonstrate advances in research capability, which may motivate a wider adoption of the LOD approach by research institutions and projects*

Promote new forms of scientific/scholarly publication

- *Promote novel forms of digital publication that could "work" for archaeological projects in terms of enhanced access to research outcomes as well as academic credit*
- *Start with moderately enriched familiar ways of publication (e.g. embedding explorable digital objects in online papers), and make new approaches as easy as possible*
- *Investigate fields of "data-driven" archaeological research and publication in which accessible datasets and executables (software, dynamic figures, etc.) could play an essential role*
- *Explore repositories as platforms for media/data-rich archaeological publications and value-added services for scholarly communication*

Foster participatory and reflective online public/community archaeology

- *Conceive and engage in participatory approaches based on online platforms (e.g. social media) in a highly reflective way*

- *Explore the concept and practicalities of "open research communities" that involve archaeologists and citizens in the production, dissemination and re-use of open data*
- *Instead of seeking "roots", focus on relevance of archaeology for societal, environmental and other issues, at regional as well as global scale*

The 10-year horizon goes beyond the formal lifecycle of the current ARIADNE project. The elaboration of a roadmap and action plan towards 2025 is an opportunity to stimulate a broader reflection and discussion among stakeholders of long-term needs in innovative digital, ICT-enabled research and communication practices.

2 Summary report of the preliminary agenda and suggested actions

The objectives of the Innovation Agenda and Action Plan are to point out the innovation needs and challenges for the archaeological research and data management community in Europe in the next 5-10 years, and to suggest actions for ARIADNE and other stakeholders on how to address the challenges. The aim of the suggested actions is to achieve favourable and potentially transformative advances in data sharing, access and (re-)use for further research across institutional, national and disciplinary boundaries.

The preliminary agenda and action plan for the 5-year horizon (2020) centres on challenges that are related to ARIADNE's focus on the e-infrastructure and services for data networking, discovery and access. These challenges include sharing of open and re-usable data, required community data archives, and interoperability of data and vocabularies (e.g. terminology, ontologies).

The 10-year horizon concerns potential innovations in ICT-enabled archaeological research ("e-archaeology"), which will depend on the progress in open and interoperable data in the next few years. For the 10-year horizon the preliminary innovation agenda presents results of first investigations, while the final version will cover this long-term horizon in greater detail.

We note that the 5/10-year horizons go beyond the formal duration of the ARIADNE project (January 2017). ARIADNE implements an e-infrastructure for archaeological data registration, integration, discovery and access. While the data infrastructure and services can support the archaeological community with regard to some of the challenges in the next 5-10 years, most challenges need to be addressed by stakeholders of the community, including academia, digital archives and funding bodies, among others. In particular these challenges concern growing a culture of open data sharing and openness of research in general.

2.1 General requirements in the 5-10 year horizons

Research e-infrastructures, digital archives and data services have become important pillars of the scientific enterprise that in recent decades has become ever more collaborative, distributed and data-intensive. There is an increasing need to enable resource sharing over e-infrastructures by pooling data resources, tools and services, aimed at supporting online team-based and cross-disciplinary collaboration. Moreover, there is the expectation that open access to re-usable and interoperable data will have an impact on the innovativeness of research, based on the exploitation of integrated datasets, for instance.

The primary objective of ARIADNE is to help overcome the existing fragmentation of archaeological data resources by making it easier to share, connect and provide access to resources at the community-level. However, the fact that today many data resources are not accessible or even discoverable is due to data-related attitudes and practices that are not favourable for open data sharing. These attitudes and practices should change so that data becomes open, can flow and be re-used for further research, progress in knowledge and innovation.

Therefore, as an introduction to the innovation agenda and suggested actions, we first address two general, cross-cutting challenges for research data sharing through community archives and e-infrastructure services. These challenges entail capacity building for open sharing of research data and coordination of stakeholders at and across the different levels of e-infrastructure and services. In addition there are discipline-specific requirements which must be fulfilled so that e-infrastructures can play a significant role in enhancing established research practices as well as bringing about transformative innovation in the humanities, and archaeology specifically.

2.1.1 Capacity building in open sharing of research data

Data infrastructure and services are being built to enable discovery, access and (re-)use of data. Investment in e-infrastructure makes little sense without open sharing of data through digital archives that store and curate the data, and make it available to overarching e-infrastructures and services for the research community.

The open data movement has proceeded quickly in some fields within the natural sciences, based on research culture, funder mandates, standardisation of data, and commonly available databases and archives. The situation in the humanities is more difficult and will require much capacity building. Leadership and support in the complex matters of open data sharing through accessible digital archives and other e-infrastructures is necessary.

Many studies on document and data repositories have shown that the "build it and they will come" approach does not work, the usage is often frustratingly small. One reason is that dedicated training and support of users is lagging behind the implementation of the technical infrastructure. Also, changes in curricula will be necessary for the transition to open science and data practices. The responsibility for open data readiness is clearly with university departments, research institutes, and academic and professional associations.[4]

The RECODE policy recommendations for different stakeholders for example urge research institutions

- *"develop an explicit institutional research data strategy with open access as the default position"*,
- *"develop educational and training programmes for researchers and staff to improve data management skills and to enhance data-intensive research"*, and
- *"include open access to high quality research data as a formal criterion for career progression"*.[5]

There is also a clear need for skilled data managers at all levels, research projects, institutional and community repositories, and research e-infrastructure and services. The challenge of building and retaining a workforce of research data curators should not be underestimated.

ARIADNE is not specifically tasked to offer training in the management of archaeological data. To support capacity building, the project however provides guides to good practice, trans-national access to partner facilities and expertise, and a limited training offer for individual researchers and institutions, e.g. on how to make legacy data accessible. Experiences from these activities will be collected and shared beyond the project partners.

[4] While open data readiness must be built at the local level, more national and European projects like FOSTER could help promote the case; FOSTER - Facilitate Open Science Training for European Research (EU, FP7, 02/2014-01/2016), http://www.fosteropenscience.eu; reportedly OpenAIRE2020 (started in January 2015) will support the Horizon 2020 Research Data Pilot *"through European-wide outreach for best research data management practices"* (OpenAIRE 2014).

[5] RECODE - Policy Recommendations for Open Access to Research Data in Europe project (EU, FP7-SiS, 02/2013-01/2015), http://policy.recodeproject.eu/research-institutions/recommendations/; see also their full report (RECODE 2015).

2.1.2 Coordination across the different levels of e-infrastructure

There is a need to coordinate efforts across all levels of data management, which include:

- *Local – university departments/institutes and research libraries:* provide support for individual researchers and projects (e.g. data management and access plans); strong, high-level promotion of open data is necessary at this level;
- *National – community-level data archives:* offer long-term storage, curation and access, and act as centres of expertise in data management, typically for certain domains (e.g. humanities or natural sciences); such data archives/centres are missing in many European countries;
- *European/international – major e-infrastructures:* enable transnational online access to data archives, services and tools, promote standards setting, leverage national efforts, and support international collaborative projects; such e-infrastructures should become more integrated to enable synergies and innovative, cross-disciplinary research.

The primary challenges in building a rich, coherent and sustainable ecosystem of e-infrastructure and services do not concern technology, but coordination amongst the actors, building capacity and trust, managing legal aspects and, of course, costs.

Archaeology and other heritage sciences address a multitude of complex research topics and objects. As noted in the Roadmap 2006 of the European Strategy Forum on Research Infrastructures (ESFRI), *"The complexity of the record of human cultures – a record that is multilingual, historically specific, geographically dispersed, and often highly ambiguous in meaning – makes digitisation difficult and expensive. (…) Data, information and knowledge are scattered in space and divided by language, cultural, economic, legal, and institutional barriers"* (ESFRI 2006: 20).

Building and maintaining e-infrastructure and services for research on the record of human cultures requires coordination among all stakeholders. Lack of co-ordination would mean that available digital resources remain difficult to access and integrate, and cost-efficient operation and sustainability of the e-infrastructure and services is unlikely.

The involvement of very different stakeholders and resources (e.g. research institutes and laboratories, cultural heritage agencies, museums and archives), and generally tight funds, necessitate cost-effective solutions that allow benefits at the community-level. For example, one state-of-the-art national community archive for archaeological data is preferable to investing effort in many local solutions which are difficult to integrate and therefore not helpful in overcoming data fragmentation. Proliferation of un-coordinated efforts that are very likely to be unsustainable should be prevented.

Integrating e-infrastructure initiatives at the national and European levels can exert a coordinative role, help prevent further fragmentation and provide means for integration of what proves to be sustainable. ARIADNE, as the European Union funded "Integrating Activity" project in the field of archaeology, can play a major role in this regard.

2.1.3 Overview of innovation actors and activities

The figure below gives an overview of the many actors and activities involved in enabling the ability to share, access and (re-)use research data through digital archives, e-infrastructures and virtual research environments:

	Data mobilisation and sharing	ARIADNE e-infrastructure and services	Collaborative data (re-)use
Data sharing, access and (re-)use	Open data policies & mandates Capacity building of institutions and researchers (training and other support) Data management & access plans Institutional repositories and community-level data archives (national, domain-based) Sharing of open, re-useable data (incl. standard metadata and common vocabularies)	Networking and interoperability of data archives / repositories Data registration, catalogue of data sets and collections Data portal services (cross-archive data discovery, access and other services) Data mobilisation support (e.g. guides to good practice) Collaboration with other other e-infrastructure and service providers	(Re-)use of shared data in new research projects (data from different institutions and countries) Virtual research communities / environments (VREs): collaborative e-research across organizations, countries, and disciplines Enhanced integrative research, comparative analysis and broad synthesis
Actors involved	Research funders Researchers and data managers of research projects Curators of domain archives and institutional repositories Research support services (e.g. university libraries) Academic/scientific publishers	Data archives / repositories connected to community e-infrastructure Providers and managers of data infrastructure and services, ARIADNE and others Providers/curators of special resources and services (e.g. vocabularies like gazetteers, thesauri, ontologies)	Research projects / collaborations Developers of open source e-research tools VRE builders and managers Distributed computing (Grid/Cloud) infrastructures and services

Figure 1: Overview of innovation actors and activities. ARIADNE, SRFG 2015

Figure 1 distinguishes three fields of activities: data mobilisation and sharing through accessible digital archives/repositories, integrated data access through community data infrastructures and services, and data (re-)use in further research, in particular collaborative research supported by e-research tools and environments. All innovation activities can be progressed in the 5-year horizon of the innovation agenda, while novel ICT-enabled collaborative e-research based on many shared data resources may see a strong take-off thereafter.

2.2 5-year innovation horizon

The objectives and recommended actions in the 5-year horizon (2020) of the Innovation Agenda and Action Plan centre on immediate challenges of open data, digital archives and e-infrastructures for data discovery and access. These innovation challenges are rather clear as they have been widely studied as well as discussed in the arenas of research policy and research infrastructure development.

The common understanding is that a research culture of open data sharing should be fostered through appropriate policies of research funders and institutions, availability of state-of-the-art data archives, and capacity building of research institutions, researchers and data managers to actually mobilise fit for re-use open data.

ARIADNE can contribute to these goals, for example, through promoting open access archaeological data, offering guides to good practice and supporting capacity development. But the primary responsibility and agency is with the research funding agencies, research institutions and associations, and mandated institutional and community data archives.

Concerning the need of an integrating e-infrastructure and data discovery and access services for the sector of archaeological research in Europe, many of the suggested actions for innovation fall within the remit of ARIADNE. The ARIADNE e-infrastructure and data portal will not replace existing data archives but provide integrating functionality and services on top of them.

The ARIADNE data portal will offer a core set of required services for the archaeological research and data management communities (e.g. dataset registration, search and access services). Required related innovation objectives pertain to enabling semantic interoperability of (meta)data and knowledge organisation systems (e.g. thesauri) based on novel methods and technologies.

The sections below summarise the preliminary ARIADNE innovation agenda and plan for the 5-year horizon. The plan comprises of five areas of actions which build on each other. The suggested actions should allow for providing enhanced access to a rich and growing stock of accessible and re-usable archaeological data.

2.2.1 Focus area 1 – Research e-infrastructures and digital resources

Background and situation

The ARIADNE project centers on the development of an e-infrastructure and services for archaeological and cultural heritage data resources (e.g. data registration, discovery, access and other services). This focus area therefore addresses the current situation and major issues of e-infrastructure development in Europe in general, and e-infrastructure and digital resources for archaeological and cultural heritage research in particular.

In the last about 20 years, e-infrastructures have become ever more important, indeed crucial for the conduct and progress of research in all branches of the scientific enterprise. There is an increasing need to build, share and integrate digital resources for research. This includes terminology and conceptual knowledge (e.g. thesauri, ontologies) which allow linking of and enhanced access to research publications and data.

Archaeological research is based on established practices and tools for structuring, analysing and presenting research results of individual projects. But the development of common e-infrastructures and services for data sharing and re-use is lagging behind, although necessary to allow for more effective, and possibly innovative research.

The European Archaeological Council (EAC) in the “Managing the sources of European history” theme of their Amersfoort Agenda emphasises *"the need to share, connect and provide access to archaeological information with the help of digital technologies. The key to this aspiration is to improve collaboration – we need to share rather than exchange. It is essential to encourage the development of European data-sharing networks and projects in the field of archaeology. The ARIADNE project is an excellent European initiative in this regard and participation in this project should be strongly encouraged"* (European Archaeological Council 2015: 21).

Archaeology, like other domains of research, indeed faces the challenge of supporting through common e-infrastructure and services the scientific enterprise that has become increasingly collaborative, distributed and data-intensive. But in the current drive to build new or upgrade existing infrastructures and resources for “e-science” in all research sectors there is a risk of investing limited funds in many uncoordinated, possibly redundant and unsustainable initiatives. This risk is particularly critical in the field of archaeology and cultural heritage, which is characterised by different stakeholders and data resources within research institutes and laboratories, cultural heritage agencies, and institutions such as museums and archives.

In the domain of archaeological research the risk of a "many-headed beast"[6] does not reside at the level of European-level data infrastructures, because no such infrastructure exists as yet. Rather, it will be up to ARIADNE to provide the archaeological sector with a common solution. But the solution will require sustained efforts to allow for maintenance and extension, e.g. mobilisation of data resources of all European Union Member States and integration with relevant other European data infrastructures.

Regular coordination activities at the European level are necessary to ensure interoperability of the e-infrastructures and digital resources as required by the heritage sciences, i.e. archaeological, cultural heritage and related other humanities research. Such coordination activities have already started in the PARTHENOS project[7]. In the medium to long term perspective also synergies with e-infrastructures within other related domains of research may be sought, environmental and life sciences, for instance. Notably, we do not see archaeology as belonging exclusively to the humanities, rather this field of research is multi-disciplinary, involving research across many disciplines.

Issues of coordination and integration at the level of common e-infrastructures may be solved relatively easily. But a critical situation exists at the level of archaeological data resources. The main issues here are fragmentation, disconnectedness, and lack of open access. A more targeted approach of resource development and access needs to be conceived and put into place. But there are also some critical general as well as domain-specific factors and requirements which need to be addressed so that e-infrastructure services and tools can help drive innovation in the humanities, and archaeology specifically.

Suggested actions: E-infrastructure development and cooperation

Recognise the importance of e-infrastructure development and coordination

Research e-infrastructures and services are important pillars and drivers of collaborative and data-intensive research. They provide access to distributed, but shared digital resources (data, services, tools) for advanced and innovative research across institutional and disciplinary boundaries. Archaeology is multi-disciplinary and therefore should devote particular attention to developing integrated e-infrastructure, data resources and services.

Considerable investments have already been made at the European and national levels in implementing research e-infrastructures for different disciplines. But the e-Infrastructure Reflection Group (e-IRG), the main advisory body for European e-infrastructures, notes insufficient coordination among the existing e-infrastructures, and asks all stakeholders to increase coherence and resource sharing. This is required to enable synergies, cost-effectiveness and sustainability in supporting innovative ICT-enhanced research within and across disciplines.

The sector of archaeological and other heritage research should prevent a proliferation of unsustainable e-infrastructure initiative and ensure that available funds are invested in a highly targeted and coordinated manner.

Support the ARIADNE data infrastructure and portal and help mobilise data providers

ARIADNE is the core EU-funded Integrating Activity project in the field of archaeology with the primary objective to help overcome the fragmentation of archaeological data resources. ARIADNE is

[6] The "many-headed beast" is one of the e-infrastructure scenarios of the eResearch2020 study, characterized by initiatives heading in different directions, incompatible technologies, duplicating efforts, etc. (cf. *Section 3.2.3*).

[7] PARTHENOS - Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies (EU, H2020, 05/2015-04/2019), http://www.parthenos-project.eu

an initiative to enable this field open sharing and re-use of data as needed for progress and innovation in archaeological research, and increasingly demanded by funding bodies. The immediate tasks of ARIADNE are to set up a state-of-the-art e-infrastructure and services (e.g. data registration, discovery, access and other services) and to support data providers in connecting their resources. All stakeholders in accessible archaeological data should support the ARIADNE data infrastructure and portal and help mobilise data providers.

Ensure sustainability of the ARIADNE data infrastructure and portal

ARIADNE will provide the archaeological sector with a common, European-level data infrastructure and portal. The solution will require sustained efforts to allow for maintenance and extension, e.g. mobilisation of data resources of all European Union Member States and integration with relevant other European data infrastructures.

Strengthen cooperation among e-infrastructures for heritage sciences

There is already a good level of information exchange between the existing European e-infrastructures for heritage sciences, e.g. ARIADNE, CLARIN, CENDARI, DARIAH, IPERION-CH, and others. But regular cooperation targeted at enabling a coordinated development of and synergies among the e-infrastructures is necessary. Coordination is required to ensure interoperability, common policies, sharing of resources, and integrated access of users to open data, services and tools as required by the heritage sciences. Also cooperation on training and professional development of e-infrastructure and data service managers can be recommended.

Establish cooperation with other e-infrastructures and services that cover relevant data of related research domains

Archaeology is a multi-disciplinary field of research in which researchers need information and datasets of different domains of research. Indeed, archaeological researchers address questions about past cultures and environments that require data in the micro to macro range, i.e. from micro-remains (e.g. molecular biology data) to landscapes and beyond (e.g. terrestrial and airborne or satellite remote-sensing data). Therefore cooperation on data access between e-infrastructures of different domains is required as no single e-infrastructure may be capable to cover all data resources needed by the multi-disciplinary archaeological research communities. Cooperation on data access will also create the setting for knowledge transfer, cross-fertilization and interdisciplinary approaches.

Link up with providers of Distributed Computing Infrastructure and encourage use of their resources by archaeologists

The European Research Area has many scientific Distributed Computing Infrastructures (DCIs) and the providers welcome research communities that are not yet avid users. Archaeological research could benefit from Grid/Cloud-based services of DCIs in some cases, for example, handling and processing large stocks of data objects ("big data"), detailed 3D objects (e.g. landscapes) or complex models of cultural change (e.g. adaption to gradual or disruptive environmental changes).

Use of DCI resources can be encouraged by bringing together archaeological research groups with providers of DCI, especially with their developers of data processing software and services. We assume that currently few archaeological research groups are familiar with DCI resources and methods like science gateways, virtual machines or workflow-based execution of data processing services. Therefore guidance and active knowledge transfer by experienced developers is required.

Suggested actions: E-infrastructures and novel digital practices

Research e-infrastructures can play a significant role in enhancing established forms and ways of carrying out research, as well as bringing about transformative innovation in practices and methods, in particular, in digital, web-based environments. But several general as well as domain-specific factors and requirements must be addressed so that e-infrastructures can become drivers of innovation in the humanities, and archaeology specifically.

Take account of disciplinary differences and foster the common ground

Instead of attempting to transfer what seemingly works in other domains (e.g. natural sciences), user-centred development of e-infrastructures, data resources and tools for archaeological research must be ensured. This requires close cooperation of researchers, data managers and technical developers in the development of innovative e-infrastructure, tools and services. The development requires fostering mutual understanding and shared perspectives. Such common ground cannot be taken for granted but must be built in close collaboration; involvement of the intended domain users will ensure that requirements are met.

Address the complex case of archaeology

Different areas on the multi-disciplinary map of archaeological research present distinctly different needs with regard to data, services or tools, classical studies versus environmental archaeology, for instance. Data infrastructures such as ARIADNE provide services required by all areas such as data registration, discovery and access for re-use. Additional specific services/tools or virtual research environments (VREs) can be built on top of data infrastructures. They are typically area-specific, e.g. support in building a scholarly edition of classical texts or analysis of aggregated environmental archaeology datasets (e.g. vegetation distribution maps).

Embed e-infrastructure and services in research practice

Research e-infrastructures are not primarily about technology but research practices supported by required services and tools. New technologies are only successful if they become embedded in the research culture and co-evolve with this culture. This requires taking account of the institutional, organisational, educational, legal and other non-technical aspects which determine the willingness and capability of the research community to adopt and use novel e-infrastructures, services and tools. A "technology-push" approach sidelining institutional, epistemic and methodological issues should be avoided, and conflicts due to vested interests in established research practices, technical systems and data resources must be taken seriously.

Help overcome the current barriers to adoption and support

There are several constraints that must be overcome to enable e-infrastructures to play a significant role in the humanities. These constraints pertain to lack of easy access to required data/content, effective and sustained tools, training offers and technical support. Furthermore, the priority of publications and low recognition of digital products keeps researchers disinclined to make data resources available or maintain research tools/software. Therefore the move towards open science practices will also necessitate a re-distribution of rewards, away from "high-impact papers" (or the scholarly monograph) to research databases and software. Stakeholders in open science should highlight the vital role of data(bases) and software in producing research results and ensure appropriate recognition and rewards for their sharing and maintenance.

Build and retain a skilled workforce of data curators

With regard to technical and other support we emphasise the need for skilled data managers/ curators at all levels, research projects, institutional and community repositories, and research e-infrastructure and services. Data curators are part of the human element of e-infrastructures, supporting established and emerging new forms of research and discipline-specific data practices. They are also increasingly important for the success of individual research projects, e.g. with regard to the design and implementation of effective data management plans. Stakeholders at all levels (e.g. research institutions, digital archives, common e-infrastructure) must recognise the challenge of building and retaining a workforce of research data curators. This challenge should not be underestimated. Among other requirements appropriate recognition and clear career paths for data curators are necessary.

Suggested actions: Digital resources and services

To support the elaboration of the ARIADNE Innovation Agenda and Action Plan, we looked in greater detail into the current landscape of digital content/data resources and services for archaeological and other cultural heritage research[8]. The study confirms the general perception of a high fragmentation and inaccessibility of relevant data resources. This perception, and a high dissatisfaction about the current situation, was expressed by many researchers who participated in the ARIADNE online survey on user needs and expectations from the ARIADNE e-infrastructure and services[9]. Many resources are indeed not easy to find (e.g. few are included in existing online registries) and difficult to access or not accessible at all. But a good understanding of the situation of different resources can allow for the conception and implementation of a more targeted approach of resource development and access.

The ARIADNE e-infrastructure, registry and portal may work toward this, but will for some time to come hardly cover all relevant data resources for archaeologists in Europe. This is due to factors on the side of the potential providers, such as lack of technical preparedness of data resources and access mechanisms. Dedicated efforts by all data/content holders will be necessary to prepare and register their resources so that they can be discovered and accessed, as well as interlinked with related information held by other institutions. Below we briefly address and suggest actions in areas of data resources and services covered in the background study. Furthermore a summary of perceived data mobilisation priorities is included.

Registration of digital repositories / collections

In the existing registries of digital repositories (e.g. OpenDOAR, ROAR, re3data) only few European repositories and collections of archaeological data can be found. They include the ARIADNE project partners Archaeology Data Service (UK), DANS-EASY (incl. the E-Depot for Dutch Archaeology), Swedish National Data Service (incl. archaeological deposits), and the ARACHNE IDAI.objects database (Germany). Furthermore there are some other, much smaller digital resources of interest to archaeology and classical studies. We think that many institutions have not yet considered registration as a way to make their collections discoverable.

Suggested actions

- *Register data archives/repositories and other collections for archaeological research in the catalogue of the ARIADNE data portal.*
- *The ARIADNE registry should become the prime location to document and discover available archaeological data resources.*

[8] This background study is included in *Section 9.1* and summarised in *Section 3.4*.

[9] ARIADNE First Report on Users' Needs (D2.1, April 2014; available on the project website).

Digitised cultural heritage content

The most comprehensive overview of cultural heritage content digitisation and access in Europe has been produced by the ENUMERATE surveys (2014) which cover the situation at 1179 institutions. The surveys give some indication concerning digital/digitised information of relevant museum collections and archaeological sites, monuments and landscapes. For example of the collections of 548 museums (including 180 “archaeology or history museums”) 24% are already digitised, 57% not yet digitised, and for 19% no need is perceived to digitise them. Collections of 239 institutions that hold mainly or significant amount of material concerning archaeological sites, monuments and landscapes show roughly the same pattern, but a smaller percentage (18%) have already digitised content. The ENUMERATE surveys also found that the institutions primarily provide digital access on their own website and show a low interest in allowing for programmatic access.

Suggested actions

- *Step up the digitisation and documentation of archaeological material held by museums and other collections (e.g. unpublished excavations, grey literature, finds/objects).*
- *Participate in the ARIADNE initiative for federated search and access specifically for archaeological content/data collections.*
- *Implement mechanisms that allow harvesting and other programmatic access to collection metadata (e.g. OAI-based harvesting, SPARQL for Linked Data).*

Online scientific reference collections

Scientific reference collections are difficult to maintain, re-vitalize and extend, due to shrinking budgets for their curation. In recent years these collections have been promoted as essential research infrastructures, but substantial investment will be necessary to turn many collections into state-of-the-art web-based resources. Some institutions have already digitised collections that are relevant for archaeological purposes but discovery and access is not well organised at the community level. Bottom-up initiatives of researchers aim to develop online community resources from individual contributions, but usually receive only few.

Suggested actions

- *Take stock of existing reference collections for archaeological purposes and evaluate their relevance and current condition (e.g. actual demand, requirements for online access).*
- *Secure financial support and curatorial expertise for the development and maintenance of state-of-the-art web-based reference resources.*

Major laboratory facilities relevant for archaeometry

In the MERIL inventory 17 laboratory facilities relevant for archaeometrical analyses are registered. These are large facilities that have special equipment such as synchrotron radiation and accelerator mass spectrometry instruments, typically not available at archaeometry laboratories of university institutes and major museums (of which there exist many more). An examination of the websites of the registered laboratories shows that documentation (metadata) of their analyses of archaeological and other cultural heritage material is hard to come by and access to data is offered nowhere. Better documentation could greatly improve the discovery of such analyses and finding related publications and data in repositories, which may not necessarily be at the laboratories.

Suggested actions

- *Investigate how major facilities as well as typical archaeometry laboratories could be mobilised to provide open access data (e.g. factors that impede open data sharing and how they might be removed).*

- *Improve the online documentation of archaeometry analyses on laboratory websites, and include pointers to publications and deposited analysis data.*
- *Standardise the metadata of archaeometry documentation to promote consistent cataloguing and improve cross-laboratory discovery and access to publications and analysis data*

Virtual research environments and data processing services

Virtual research environments (VREs) combine tools and services for domain-specific online research purposes (e-research). In the field of archaeological research few VREs are available or in development. Sometimes Web 2.0 applications are suggested, however, they mainly serve the professional networking and information exchange function, while other functions (e.g. data access and processing) require different systems. For data processing Grid/Cloud-based services of Distributed Computing Infrastructures allow user groups to share data resources and processing applications (virtual machines) and, thereby, form a virtual research community. ARIADNE does not foresee offering VREs, but VREs can be implemented on top of e-infrastructures for using existing as well as creating new data resources.

Suggested actions

- *Investigate if archaeologists need virtual research environments; e.g. what kind of research could archaeologists conduct online, what functionalities are necessary for such research, and how could they be provided in a VRE?*
- *Promote the development of relevant VREs with functionalities (tools, services) required by archaeologists to conduct research tasks online (e-research).*
- *Raise awareness of Distributed Computing Infrastructures (e.g. Grid/Cloud-based services of DCIs) amongst archaeologists who need data processing services.*
- *Encourage use of DCIs by bringing together archaeological research groups with providers of DCI, especially with their developers of data processing software and services.*

Summary of data mobilisation priorities

ARIADNE provides a facility to register data resources so that what is accessible can be documented, made visible and discoverable. The ARIADNE online survey found that gaining an overview of and ability to search across available resources is on top of what researchers expect from the project. The ARIADNE data registry and portal will allow this, for large and small datasets, given data holders make the effort to register the data.

E-infrastructure initiatives like ARIADNE of course want to have a critical mass not only with regard to the number but also volume of registered resources. Therefore incorporating available archaeological data from large community-level and institutional resources is clearly among the priorities of the initiative.

The overall tenor of the responses to the ARIADNE online survey (over 500 qualified responses) was that *all* types of data are important; exemplary comments were *"all sources are important and must be first-hand"*; *"all is important ...if it is available"*. This of course is related to the existing difficulty of finding and accessing online the many data resources researchers would like to consult and use for their research.

A comparison of what respondents considered as important for their research with their rating of the online accessibility of these resources gives some general indication of data mobilisation priorities: Data of excavations, prospection/field surveys, GIS, material/biological analyses and radiocarbon/

dendrochronology showed the largest gaps between importance and online accessibility.[10] Below we address data resources which may be priority targets and require specific actions by data holders:

Excavations, prospection/field surveys, including GIS

Some information about excavations, prospection/field surveys, including Geographic Information System/Services (GIS) used, is usually available online from project websites. But more could be done by research groups to make data available online during projects (which may extend over many years), and ensure that the full data/evidence of completed projects is deposited in digital archives. ARIADNE focuses on data deposited in accessible archives/repositories and available information about excavations, for example, the Fasti Online database of project partner AIAC[11].

Legacy data and grey literature

Much excavation and other archaeological work has been conducted which remained unpublished and poorly documented. National initiatives to identify and unearth important unpublished work could add greatly to the accessible archaeological record.

A lot of unpublished fieldwork reports ("grey literature") have been produced in the context of infrastructure development and other land use projects. While often the only record of the fieldwork undertaken, such reports are difficult to find and access in many European countries. Among the exceptions is the UK, where provisions have been put in place so that fieldwork reports are deposited with the Archaeology Data Service (at present over 34,000 reports)[12]. The extensive usage makes clear the need to make such material more accessible.

Laboratory analyses of archaeological and other heritage objects

Concerning material/biological analyses, our study confirms the researchers' perception of a difficult situation with regard to discovery and access; suggested actions are included in the brief summary above (more background is given in *Section 3.4.7* and *Section 9.1.3*). We did not look deeper into the situation concerning radiocarbon and dendrochronology data. Several databases are available[13], however, researchers may often need specific data which is not easy to identify or unavailable.

Non-priority data resources

The comparison of what the ARIADNE survey respondents considered as important for their research with their rating of the online accessibility of these resources shows two non-priority areas: One area is government site management data/information (e.g. sites and monuments records), although their online accessibility was perceived as low. Also not a mobilisation priority are large datasets or stocks of data ("big data") which could be used for data mining or other advanced computing methods. Reasons may be that respondents did not expect much from such approaches in archaeology and may lack expertise to apply the required technology/software and methods. The online availability of data resources for applying such methods was perceived as very low.

[10] Cf. ARIADNE D2.1 First Report on Users' Needs. April 2014, 76-84

[11] Fasti Online (International Association for Classical Archaeology), some 12,000 excavation reports and site summaries across the Mediterranean and other countries, http://www.fastionline.org/excavation/

[12] ADS: Library of Unpublished Fieldwork Reports, http://archaeologydataservice.ac.uk/archives/view/greylit/

[13] For example: Digital Collaboratory for Cultural Dendrochronology (DANS [NL] with national and European/ international partners); International Tree-Ring Data Bank (NOAA / World Data Center for Paleoclimatology); Archaeological Site Index to Radiocarbon Dates from Great Britain and Ireland (Council for British Archaeology), deposited with ADS; Radiocarbon Palaeolithic Database Europe (KU Leuven, Center for Archaeological Sciences); Radiocarbon Context Database (Univ. Cologne, Institute for Prehistory), and several others.

2.2.2 Focus area 2 – Culture of open sharing and re-use of data

Background and situation

ARIADNE is the core EU-funded "Integrating Activity" focused on archaeological data(sets). Through its e-infrastructure and services ARIADNE will enable enhanced discovery, access to, and (re-)use of archaeological data held by institutions across Europe (and beyond). But ARIADNE, like other data infrastructure initiatives, depends on underlying institutional repositories and community archives richly filled with open data shared by research projects.

Various benefits are expected from the ability to (re-)use open data for further research and other purposes. Research institutes, data archives and e-infrastructure providers, funding bodies and other stakeholders therefore cannot ignore obstacles which hinder researchers in sharing their data in an open manner. Rather they must support researchers in data sharing and help ensure that they receive appropriate credit for doing so.

Arguments for open data sharing

There are many good arguments for open data sharing such as preventing duplicative data collection, allowing replication of data analysis to scrutinize knowledge claims, and enabling new research questions to be addressed, based on combined datasets. Particularly strong are the arguments related to the re-use of data as re-use allows exploitation of previous investment. Data that is curated, integrated, re-used and analysed with new methods gains in value rather than being only a cost factor. The benefits expected from data re-use are among the core interests of research funders in open data.

Data sharing through accessible archives

Researchers often share data, but mainly with project collaborators and other trusted colleagues. Most data remains within a small circle of peers and is not available for other researchers or the public at large. The goal of open sharing is to have the data deposited in digital archives to allow for access and (re-)use in further research and other purposes. Instead of inaccessibility and potential loss, deposited data is curated and accessible for new, often not anticipated uses. However, strong barriers need to be addressed to move from a closed-circle to open sharing of data.

Barriers to open data sharing

Many factors work against open sharing of research data. In particular, researchers perceive a lack of incentives (e.g. little academic recognition and reward), fear that their data could be misinterpreted or misused, and preparing data for (re-)use by others means extra work (e.g. data description), while there are other priorities such as new projects and publications. The core requirements for open data sharing are not technical but institutional, especially the need for appropriate academic recognition and reward. But also other requirements must be met, such as available, trusted and sustainable data archives.

Strict open data mandates

A broad consensus is emerging that research data, in particular, data from publicly funded projects should become openly accessible for all. Ever more research funding bodies extend their open access mandates for publications to the data that underpins the research results. To ensure compliance, project funders also require grantees to implement a data management plan, with the objective of making the generated data publicly accessible through an appropriate repository. Experience shows that open access policies/mandates of research funders and institutions for publications and data need to be decisive, monitored and enforced. Also journals that demand research underlying reported results to be accessible have a significant impact on the availability of the data.

Data re-use and citation

Currently not a strong driver of open data sharing is evidence of individual benefits for researchers, which may be accrued through data (re-)use and citation by others. While a citation advantage for open access publications is well confirmed empirically, there is little evidence as yet of a similar advantage for open research data. Also solid empirical evidence of re-use of openly shared data is not available. But we may expect better insights into data re-use practices as more open research data becomes available. One important element in this regard is standardisation of data citation in the research literature and other publications (e.g. websites), including persistent identifiers of the data(sets).

The archaeological data sharing record

Archaeology was not addressed above, because the issues and required solutions are the same in all domains of research. But some are further ahead concerning the requirements, decisive open access policies and available and mandated data archives, for instance. In European countries where the requirements are fulfilled, archaeology is also part of the open data movement.

Elsewhere archaeological research is not at the forefront of open data sharing, despite factors that make an open data imperative particularly strong in this field: excavation of sites destroys the primary archaeological evidence, the work on archaeological heritage is done in the public interest, and there is little commercial interest in archaeological data (apart from the commercial archaeology sector).

However, many archaeologists are not yet well prepared or equipped for open data sharing. As the matter is complex, strong leadership with regard to policies/mandates, supportive institutional measures (e.g. capacity building, training of researchers), and state-of-the-art digital archives are required. Overall this is about growing a culture of open data sharing, based on common values and agreed principles. As a popular slogan of OpenAIRE puts it, *"Open Access is global – but implementation is local"*.

Openness should become embedded in archaeological research practices as *"the default modus operandi"* (e-InfraNet 2013) so that the advantages of accessible and re-usable data gain priority over the interest of the individual researcher to protect his/her "claim".

Suggested actions: Open data policies and practices

This focus area mainly concerns the challenge of growing an open culture of research of which open data is but one, although an important facet.

Promote a research culture of open sharing of data

Archaeological data archives/repositories and the ARIADNE data infrastructure will flourish only within a research culture that values open sharing of data. Therefore promoting this culture is of vital interest to data service providers which, however, have a supportive role. The main enablers are research groups, institutions, associations and funding bodies who opt for openness. There are many good arguments for open data such as preventing duplicative data collection, allowing replication of data analysis to scrutinize knowledge claims, and enabling new research questions to be addressed based on shared data. In particular, data that is shared, curated, integrated, re-used and analysed with new methods gains in value rather than being only a cost factor.

Support strict open data policies of funding bodies and institutions

Empirical studies make clear that only decisive open access policies for publications and data work effectively. Therefore it is recommended that funding bodies and research institutions issue strict mandates for open data sharing. Data archives and data infrastructure providers should give full support to such mandates.

Suggest appropriate guidelines for open archaeological data

A "one size fits all" approach may not work for archaeological data. In case open data guidelines do not take account of specificities of archaeological data (e.g. sensitive data such as findspots, human remains, knowledge of indigenous peoples), research institutions and associations should promote community consultation and suggestion of appropriate guidelines. The move towards open data offers an opportunity to discuss and get clear about disclosure or non-disclosure of certain information.

Recommend use of open licenses for data, metadata and knowledge organization systems

Research funders, institutes, archives and e-infrastructures should recommend (or mandate) that data, metadata and knowledge organization systems (e.g. thesauri) are shared under an appropriate open license. Data licenses should not impede effective re-use and further dissemination of the data, but ensure attribution of data sharers. As appropriate are considered Creative Commons (CC), Open Data Commons (ODC) and other licenses of type "BY" (Attribution), which ensure that data sharers (researchers, institutes and others) are recognised. For metadata the most open licenses are recommended, e.g. CC Zero (CC0), ODC-PDDL or other public domain dedication licenses. Licenses containing a No-Derivatives condition should generally be avoided, and there are strong arguments against Non-Commercial; applying instead ShareAlike will impact negatively on the ability of intermediation services to combine and utilize (meta)data effectively. Wide application of open licenses is necessary to remove barriers to legitimate and effective re-use, extension and integration of data(bases), metadata and knowledge organization systems (e.g. thesauri).

Help ensure that open data sharers are recognised and rewarded

Archaeology should be a leading example of open data sharing, because of factors that make an open data imperative particularly strong for archaeology: Excavation of sites destroys the primary archaeological evidence, archaeological work is conducted in the public interest, and there is little commercial interest in archaeological data. But there are many factors that work against open sharing of archaeological data. A particularly strong factor is little academic recognition and reward for making data available. The academic credit system values research publications while data sharing is considered much less, if at all. Therefore all stakeholders in open archaeological data should help ensure that data sharers are recognised and rewarded. Most impact in this regard will have institutions that make data sharing a criterion for academic promotion and funding agencies that take it into account when awarding new research grants.

Suggested actions: Data re-use, citation and altmetrics

Benefits associated with re-use of openly shared data arguably are the strongest drivers of the open data agenda. For example, re-use in further research (e.g. based on combined data) is expected to provide much return on investment. Re-use is also particularly important to individual researchers as it can bring recognition and rewards to data sharers based on data citations. Data archives and e-infrastructures could help mobilise more open research data more easily if convincing evidence for the expected individual benefits is available. However, empirical evidence for such benefits is still scarce. Existing archaeological data archives are accessed frequently, but little is known about actual data re-use and citation. Below we suggest how a better understanding of data re-use could be acquired, re-use promoted, and ensured that data sharer receive the credit they deserve.

Conduct studies of data re-use to better understand and support current and emerging practices

Surveys or a series of case studies could allow evaluation of current practices of data re-use in different areas of archaeology. Information about the main current and emerging forms, contexts and requirements of data re-use would be helpful to possibly better support these practices. In particular, data archives, e-infrastructure and tool developers could benefit from such information.

Promote data re-use and highlight inspiring examples

Research organisations, data archives/repositories and e-infrastructure service providers should promote data sharing and re-use wherever appropriate. Specific measures should be applied such as data re-use competitions and highlighting of inspiring examples. The introduction of a data re-use award by the Archaeology Data Service (UK) is a good example for this approach.

Foster consistent data citation so that data sharers can be recognised and rewarded

Researchers who (re-)use data from digital archives or databases should reference the data in a standardised way. The standardisation within and across disciplines is still in the making. We recommend that data archives/repositories, journal publishers and other service providers in the field of archaeology promote a common form of data citation in the research literature and other publications (e.g. websites). Among the requirements for consistent and accurate reference of (re-)used data we highlight persistent identifiers (e.g. DOIs or other). Standardised citation can greatly enhance the identification of data (re-)use and recognition of researchers and data archives that make data available. Eventually it will enable tracking of data (re-)use and support evaluation of the impact of shared data.

Capture and present data usage figures

Data archives/repositories and researchers will for some time to come lack information if accessed research material has been re-used and cited. Therefore other indicators of usage (e.g. views, downloads) should be applied and presented. Such indicators are called alternative metrics (altmetrics) and also include social media activities around research output.

Usage figures can allow data archives and other service providers to demonstrate relevance to the research community and funders, plan further development, and mobilise new data contributions. We recommend usage indicators that are directly related to the material that is being curated and made accessible (e.g. views, bookmarks, downloads); social media activities around the content/data are not seen as valid indicators by most researchers. It is also recommended to communicate the value curation work adds to data collections (e.g. indicators of improved discovery and access).

2.2.3 Focus area 3 – Data archives and curation of archaeological research data

Background and situation

Digital archives for deposit, long-term curation of and access to data provide core services underlying research e-infrastructures. Data archives richly filled by the research community are one key to the success of e-infrastructure initiatives. Because ultimately researchers want to discover and (re-)use available data for purposes such as comparison and further research. At the same time many researchers are reluctant to share their own data beyond project collaborators and other trusted colleagues. Therefore motivation and support of open data sharing is necessary.

State-of-the-art community data archives can foster trust in open data sharing as they provide a reliable environment for data publication, (re-)use and citation. We highlight national and international research community archives as they can reduce data fragmentation substantially. The "institutional" model, repositories of single universities or research institutes, provide a local solution for affiliated researchers but offer only limited potential to overcome the fragmentation of archaeological data.

Extending the productive lifecycle of data

Ensuring long-term curation of and access to archaeological datasets goes beyond the capability (and core business) of individual researchers, research projects and, arguably, most institutes. In any case, the existing difficulty of discovering and accessing archaeological datasets demonstrates that individual "local" solutions are not appropriate for the task. Therefore the data should be deposited in archives that ensure long-term curation and access.

Unfortunately, the lifespan of archaeological data often ends when researchers have published their results, but do not make the underlying data available. Extending the data lifecycle through measures of curation and dissemination allows use of the data in further research, for example, exploration of integrated datasets with novel methods and tools. A core role of digital archives and e-infrastructures therefore is mobilising and bringing together data resources so that they can be used in novel ways. The expectation is an increase in value of existing data through wider and new uses within and beyond the research community.

Towards an optimal solution

The fragmentation of archaeological data poses enormous difficulties to achieve aggregation and integration of data of many projects/institutes at the level of common e-infrastructure and services. As the most effective solution for tackling the fragmentation we see domain-based, community-level archives. The advantages of such archives include that they can:

- provide clear orientation for all stakeholders concerning data mobilisation, expected good practices, etc.
- act as centres of expertise required for the types of data and data-related issues common in the domain,
- foster trust in open data sharing through providing a reliable environment for data publication, (re-)use and citation,
- allow cost-effectiveness of data curation and access (e.g. economies of scale),
- promote common standards and act as hubs for data integration and access.

References for such digital archives exist, for example, the Archaeology Data Service (UK, established 1996) and the E-Depot for Dutch Archaeology of the Data Archiving and Networked Services (Netherlands, since 2005); both are mandated archives for depositing data of archaeological research in the respective country.

Many European countries lack a community archive for archaeological data. Not a good solution for overcoming data fragmentation in such cases, especially large countries, would be several research institutes each trying to establish their own data repository for archaeological data ending up in many general-purpose university-based repositories. Therefore the IANUS Research Data Centre for Archaeology and Classical Studies in Germany is a consequent initiative for a common solution. New initiatives in this direction have been inspired by ARIADNE in smaller countries, Austria, Ireland and Slovenia, for instance.

Mandates and certification

Research funders mandate domain-based archives that are acknowledged for their high standards of data curation or, if such an archive is not available, recommend use of a recognised and suitable other repository. Acknowledged state-of-the-art archives devote special attention to measures that promote trust and credibility. One measure is certification according to a standard of trustworthiness. Both the Archaeology Data Service (UK) and Data Archiving and Networked Services

(Netherlands), and several other digital archives, are certified according to the criteria of the Data Seal of Approval.

In the absence of a decisive deposit mandate, i.e. deposit is only recommended and several archives are considered as relevant, the certification of a data archive will be one factor in researchers' selection of an archive. According to an available study archaeologists appreciate most an archive's transparency with regard to the description and management of data collections, and indications of archive stability (e.g. sustained funding).

Benefits and costs

The benefits of a community-level, national/international data archive stem from its role as a central and sizeable hub of data in the research field it supports. Reliable, one-stop access to needed information and data resources makes research easier, faster and cheaper. In the case of the Archaeology Data Service (ADS) the increase in research efficiency of the users has been calculated to be worth at least 5 times the costs of operation. Including other advantages £ 1 investment in ADS yields up to £ 8.30 return. Like the ADS also other richly filled digital archives may have many non-academic users, including heritage management, educational, and private/general interest user groups (e.g. local/regional history).

Concerning costs, one important fact to bear in mind is that the costs of post-project data curation and online publication of archaeological projects are only a fraction of the total project costs, between 1-3%, depending on the type of investigation and data generated.

Another rule of thumb is that running an archaeological or other data archive of course costs considerably more than a typical institutional document repository. One major cost driver in archaeology is the variety and complexity of the data that needs to be ingested and curated. Data acquisition and ingest are the most costly curation activities, while archival storage and preservation activities are a much smaller segment of the overall costs, and likely to decline over time.

Keeping the operational costs stable while curating larger data collections allows economies of scales (lower per-unit cost). Enhancement of labour-intensive curation activities through streamlining and tool-support (e.g. easy submission of small deposits) can allow significant cost reduction.

Suggested actions: Data archives and curation of archaeological research data

Digital archives for long-term curation of and access to data provide core services underlying research e-infrastructures. Archives richly filled by the archaeological research community are one key to the success of the ARIADNE e-infrastructure initiative, which will provide the community with a solution for cross-archive search, access and (re-)use of available data resources.

Researchers are looking for relevant data, but are not necessarily willing to share their own data in an open manner. State-of-the-art, certified community archives allow researchers to publish their data in a secure and trusted way. The publication comes with a price tag, however compared to the total costs of archaeological investigations these costs are relatively small.

Substantial benefits are expected from the availability of ever more open research data. The chances that these benefits materialise appear to be higher if mandated community-level data archives are established. In any case, such archives will reduce the fragmentation of archaeological data and provide a more cost-effective solution than spreading funds across many data repositories of individual institutes.

Recognise that the costs of opening up archaeological research datasets are marginal and well spent

The costs of post-project data curation and online publication of archaeological projects are only a small fraction of the total project costs at around 1-3%, depending on the type of investigation and data generated. Compared to the many benefits expected from open and re-usable research data this investment seems well spent.

Include the costs of open data sharing and digital archiving in project grants

All surveys on open data sharing show that researchers consider the related effort as a significant barrier. Specifically, this concerns the effort required to prepare shareable data and detailed data description. Therefore research funders should allow inclusion of the costs of this work in project grants. A project data management plan, as increasingly requested by research funders, is the ideal place to present these costs as well as the expected archive charge for long-term data preservation and access.

Recognise the advantages of domain-based community archives

Building and mandating community data archives is the most effective strategy to overcome, or at least reduce, fragmentation and inaccessibility of archaeological data resources. Advantages of central, domain-based archives include clear orientation for all stakeholders, focused mobilisation of data deposits, economies of scale, among others. Substantial return on investment through research efficiency and use also by non-academic groups can be expected from domain archives that provide access to a wider range of relevant data resources. Funding one data centre is very likely a better solution than spreading funds across many hard to sustain data repository projects of individual institutes.

Ensure long-term sustainability of trustworthy data archives

A clear priority for open data initiatives is that archives are available where researchers can share their data in a secure and trusted way. Such archives should be certified and stable in the long-term in order to promote trust that the effort put into re-usable data is well spent. Commitment for sustained support (10+ years) by the main funding bodies would be helpful in this regard.

In the current economic climate it seems unlikely that additional funds for open data curation and access can be mobilised. This means that budgets may need to be shifted from already restrained research project funding to appropriate data curation and access solutions. Related concerns should be addressed by making clear the benefits of open research data.

Encourage and support initiatives for data archives in countries where these are currently lacking for archaeologists

The most effective solution for open archaeological data mobilisation, curation and access is a domain-based central archive (e.g. one place to go, community building, cost-effectiveness, and others). But such a data archive is missing in many European countries. Progress towards common solutions may be promoted through knowledge transfer between established data centres and initiatives for new archives in other countries. New entries may “leapfrog” to a state-of-the-art solution by learning from acknowledged benchmarks.

Devise interim archiving solutions for datasets that are inaccessible and at risk of loss

Archaeological research institutions should take stock of valuable old datasets as well as datasets not maintained by researchers actively using the data. Where a data archive for long-term curation and access is missing, devise an interim local or outsourced solution to prevent inaccessibility and potential loss of valuable datasets. A priority of course is excavation data that cannot be re-collected.

2.2.4 Focus area 4 – Capacity building for open data sharing

Background and situation

This focus area addresses the need of capacity building for open data sharing at the institutional level and issues concerning the required quality of data description and review. To support research project data management and sharing, universities and research institutes will have to put in place adequate institutional policies, guidance, training and other support. Strong institutional support of the open data agenda will allow more archaeological data (and metadata) flow into data archives for long-term curation, discovery and access, and (re-)use in further research.

Current situation of institutional data management support

At present the level of data management support by libraries/repositories of universities and other centres of research and education is rather low. The focus is mainly on institutional policy support and advice (e.g. issues of open access, data management plans, IPR/licensing, etc.), and organising training for research students. The main challenges for the research libraries, and institutional document repositories they manage, are limited resources (funds, personnel) and skills gaps with regard to research data curation. Moreover their potential role in managing or providing support in the management of research data is an on-going topic of discussion.

There are considerable doubts about attempts of individual universities to implement research data curation. Proper curation of research data requires specialisation, and leading examples of data archiving services indeed are specialising in research fields. Examples are the data centres funded and mandated by the UK Research Councils (e.g. Archaeology Data Service), or Data Archiving and Networked Services (Netherlands), which centre on social sciences and humanities (including archaeology). Specialisation is difficult to achieve for many different disciplines, e.g. all disciplines present at a university. Also a small or medium-size archaeology institute will hardly be in a position to maintain a data archive, and generally prefer to focus on research rather than data curation.

Support for managing data during project work

Rather than building research data archives, it is preferable that university departments and research institutes focus their efforts on training and support for researchers, so that shareable data emerge from the research process and are provided to appropriate subject/domain-based archives. More efforts could also be devoted to making legacy data, reference collections and other institutional assets available.

It is understood that proper research data management should start and be supported as early as possible, while researchers are working on their projects. Data management plans as requested by ever more research funders may provide a basis for this support. For the management of “active data” various methods are suggested, which range from plug-ins for software already used by researchers to automatic capture of information (metadata), which would reduce the effort required for regular documentation of research activities and results. At a minimum institutions or research networks could offer researchers a safe and controlled environment for storing and sharing data during collaborative project work.

High-quality metadata as required for data re-use

When researchers go about sharing data through an archive/repository the question of metadata comes up. All studies on data sharing through digital repositories (including the ARIADNE survey) found that researchers consider the effort to provide the required metadata as a barrier to open data sharing. While data repositories and users would benefit from high-quality metadata, data sharers face the burden and usually prefer not to invest much effort on providing metadata.

Asking for high-quality metadata is likely to result in fewer contributions, which is one reason why many repositories have rather shallow discovery metadata. Such metadata is insufficient to assist data re-use, which requires data description so that potential re-users can understand the data provenance/context, evaluate if the data is relevant for intended purposes, and use it properly to prevent incorrect conclusions. Requesting high-quality metadata arguably is possible only for domain-based archives that are mandated or recognised as the best place to share valuable data according to community standards. A further question addressed in this context is the effectiveness of metadata for cross-domain data re-use.

Data papers and data review

A novel approach that offers an incentive for researchers to provide rich data description is the peer-reviewed "data paper". Ever more journals are established that invite researchers to describe datasets they have deposited or databases they maintain. Moreover data review has become a topic of much recent discussion. It is anticipated that the open data policies of research funders and journals will bring about a wave of data in need of quality review. Therefore new models of peer-review need to be considered.

Suggested actions: Capacity building for open data sharing

To promote the open data agenda, universities, research institutes and other stakeholders should put in place policies, guidance and training. Institutional capacity building and support for researchers in the management of data is necessary so that open and re-usable data flows into data archives for long-term curation, access and (re-)use. The research community should also consider novel approaches to data description and review.

Ensure that adequate institutional policies, guidance and other support are in place

A growing number of funding bodies require research projects to implement a data management plan, with the objective that the generated data becomes openly available. Therefore researchers are looking for advice, guidance and support from their institution. Many institutions may not be prepared to meet this demand. Because the responsibility to manage and maintain data has traditionally been assigned to the researchers, based on the understanding that they own the data. We recommend that research institutions address issues of proper data management and sharing pro-actively and put adequate policies, guidance and other support in place.

Step up capacity building and training for data management and sharing

There is no lack of guidance material for good practice data management. For archaeological projects particularly the guides offered online by the Archaeology Data Service/Digital Antiquity and the ARCHES guide for data archiving (available in several European languages) merit highlighting. The challenge for publicly funded research projects and institutions now is to implement and support data management in view of open data sharing. This necessitates institutional capacity building, training offers and other measures aimed at ensuring data stewardship and accountability.

In practical terms, this means a range of assistance including, but not limited to, hands-on training in data management for PhD students and early-career researchers, help in drawing up a solid data management plan for discipline-specific data, expert advice concerning sensitive data and ethical issues, IPR/licensing (incl. rights clearance). Addressing "the data issue" therefore means more than pointing researchers to an appropriate data archive, institutional capacity building is necessary to provide the kind of assistance mentioned.

Below we give some specific suggestions on how research institutions and archives can foster proper data management and sharing:

- *Require that data management and subsequent open sharing of data are considered already in the project planning phase, e.g. data management and access plans.*

- *Promote the preparation of shareable data through dissemination of good practice guides and expert advice on specific matters (e.g. sensitive data, licensing of databases).*
- *Bring project data managers in contact with data archives/centres, not only in view of data deposit, but skills development (e.g. work on legacy data, use of specific standards).*
- *Emphasize the need for appropriate description of the methods used to collect, analyse and present the data, including technical and other requirements for data re-use (e.g. software).*
- *Suggest use of established open data formats, metadata standards as well as common terminology/vocabularies.*

Provide support for managing data during project work

Advocates of proper research data management suggest that it should start and be supported professionally as early as possible. This would make on-going research work more effective and data re-use easier. Suggested approaches include offering researchers lightweight curation tools and/or automated mechanisms integrated into researchers' normal workflows (so called "sheer curation"). Solutions ideally add immediate value to the creators and primary users of the data as well as prepare the ground for long-term data archiving and access.

This is still a field of research & development with some prototypic solutions. Also use of workflow management systems, for example, in archaeometry laboratories seems uncommon. Rather than looking for automation, we suggest that institutions or research networks offer researchers a safe and controlled environment for storing and sharing "active data" during project work. Moreover it would be beneficial if institutional data experts are available who support on-going projects, with regard to metadata generation and use of common vocabularies, for instance.

Recognise high-quality metadata is required for data re-use

Research datasets that are shared through repositories must be provided together with metadata, which are necessary for data discovery and access. But discovery metadata may not be sufficient for data re-use. Data description is required so that potential re-users can understand the data provenance/context, evaluate if the data is relevant for intended purposes, and use it properly to prevent incorrect conclusions. The effort to produce such data description is among the top barriers for open data sharing.

Metadata presents a conflict of interest between repositories, who would like to have good metadata, and data depositors, who may not be willing to invest much effort on the task. In practice this means that there are many repositories which have only shallow metadata, in particular general-purpose and university-based repositories. Both invite deposits of material from many disciplines hence cannot support specific domains like archaeology specifically. Domain-based, specialised and mandated archives can set high-quality metadata standards, which depositors will accept and follow, guided by archive curators, if necessary.

Promote data papers for archaeological datasets

A data paper is a peer-reviewed publication that describes a data resource (dataset, database or other), the methods and standards used to create it, its structure and size, and where and how it can be accessed. In particular, the paper should also describe the re-use potential of the data for further research or other purposes. Peer-reviewed and citable data papers offer an incentive for researchers to make their data available through repositories and provide rich description of the data intended to promote and enable re-use. One example of a dedicated, online and open access data journal is the *Journal of Open Archaeology Data*; the e-journal *Internet Archaeology* has initiated a series of data papers, and others may follow suit.

Explore novel approaches to data peer review

The open data policies of research funders and journals have raised the question of how the quality of the data should be evaluated. Peer reviewed data papers provide one solution, furthermore novel models recently developed for publications may be trialled. Such models are editor-mediated review, "crowd-sourced" review and post-publication review. For example, Data Archiving and Networked Services (DANS) collect ratings and other feedback from people who have downloaded data from their archive system, which includes the E-Depot for Dutch Archaeology.

2.2.5 Focus area 5 – Providing services and enabling novel applications

Background and situation

Archaeological research in Europe (and elsewhere) lacks common and integrated e-infrastructure and services for data curation, discovery, access and re-use. The consequence of this lack is a high fragmentation of archaeological data and limited capability for collaborative research across institutional and national as well as disciplinary boundaries.

Common and integrated e-infrastructure basically means accessible digital archives with structured, interoperable and re-usable data resources (e.g. collections, datasets of projects) at the bottom and discovery, access and other services on top. ARIADNE develops e-infrastructure that allows for interoperability of existing and newly built archaeological data archives and, based on this interoperability, cross-archive search, access and (re-)use of available data.

ARIADNE will not replace any of the underlying infrastructures (e.g. institutional repositories and community-level data archives), but provide integrating functionality and services on top of them. Thereby ARIADNE will help to make currently isolated archaeological data more accessible and useable for the research community as well as other groups such as heritage management agencies and citizens. The main intended user group of the ARIADNE e-infrastructure and services is the archaeological research community.

Identifying relevant services

Much care has been devoted to identifying the services end-users such as researchers and data managers expect from the ARIADNE data infrastructure and portal. This identification has been carried out through an online survey on user needs and a survey of existing data portals.

In brief, the online survey (with over 500 qualified responses) made clear that most researchers expect from ARIADNE to provide a data portal that allows an overview of existing archaeological data resources, and to search across the resources, using novel mechanisms for data discovery and access. Much less interest exists for typical features of Web 2.0 platforms such as content filtered based on tags or ratings of other users. However researchers appreciate effective mechanisms that save time in identifying relevant data (e.g. data preview mechanisms, clear licensing information).

Much further and more detailed insights for the development of the ARIADNE data portal have been acquired through a survey of existing portals and suggestions for the project's portal given by a panel of 23 archaeologists and data managers (reported in the project deliverable D2.2, February 2015). The 34 suggestions of the survey report have been evaluated by 28 experts of 21 partners in order to focus on the most relevant portal services. Hence, the project has a solid basis for developing a data portal that will serve the immediate and evolving needs of the archaeological research community.

In general, the service portfolio of the ARIADNE infrastructure and portal should meet core requirements of data overview, search and access. This includes data search based on geo-location (maps) and date-ranges/chronologies which the evaluators appreciated most. Not considered as a priority are personalized services and support of expert networking and discussion on the data portal. There is little scope to invest limited funds on specific services that are not appreciated, are provided by other portals, or may run ahead of the needs of broad user segments.

High relevance has been attached to deploying Linked Open Data to integrate information within the portal and to link to external resources. Also providing interfaces to allow external applications exploit available (meta)data was considered as very important. Clearly the ADRIADNE data infrastructure and portal should not be an "island" but enable added value in the wider information ecosystem of archaeology and beyond.

A "front-runner" category are services that support online research work (e-research), which is not an immediate concern, but may emerge in the 10 year horizon of the innovation agenda. We assume that the needs and requirements of the archaeological research community will evolve towards e-research capability when more open data becomes accessible through state-of-the-art community archives.

ARIADNE e-infrastructure architecture and services

This report includes an overview of the ARIADNE e-infrastructure architecture and service implementation. The architecture presents a layered approach with various components. A core role plays the data registry component. The component has been implemented based on the Data Catalog Vocabulary (DCAT) standard/recommendation of the World Wide Web Consortium (W3C), adapted for describing archaeological data resources. This adaptation is the ARIADNE Catalogue Data Model - ACDM.

The ARIADNE registry allows archaeological data providers, large and small, to describe their resources (collections, datasets), following a common model (the ACDM), which will allow for interoperability of the resources. This means that ARIADNE services can build on consistent information about how the resources are structured, what they contain, how they can be accessed, etc. The information will be available as Linked Open Data also to other developers to enable data interlinking and creation of additional, external services.

Data search, visualization and access on the ARIADNE data portal will be possible based on thematic, spatial and temporal information contained in the metadata of the data resources. Further integration of data can be achieved if providers map their databases to the extended CIDOC Conceptual Reference Model, as intended by several ARIADNE partners. Access to the data resources on the portal will be enabled in support of different study purposes, e.g. access to only one object (e.g. a 3D model of a building), an aggregated virtual reference collection of objects to allow for comparison, one or more datasets allowing for extraction of numeric data, etc.

For example, visual media services are already available that allow easy web-based publication, visualization and exploration of high-resolution images, reflection transformation images (RTI), and 3D models, including 3D landscapes. Other, more experimental stage services will exploit Linked Data (based on the CIDOC Conceptual Reference Model) or use data mining techniques to suggest patterns of relevance between data resources.

A major step towards integrated archaeological data and e-research

The overall objective of the ARIADNE e-infrastructure and data portal is building an environment for services that act as brokers between archaeological data providers and users. The creation of such an environment is a substantial step forward in the archaeological domain, in particular it provides a common platform where dispersed data resources can be uniformly described, discovered and accessed.

The ARIADNE e-infrastructure will allow overcoming some of the idiosyncrasies of the underlying different infrastructures of the data providers which currently prevent the collaborative exploitation of available data. At the same time, it represents an essential step towards the more ambitious goal of providing integrated services and tools capable to support web-based research aimed at creating new knowledge (e-archaeology). This may be achieved by future Virtual Research Environments (VREs) built on top of the ARIADNE e-infrastructure and services.

As noted above, support of online research work is not an immediate concern of ARIADNE, but may become an important topic in the medium to long-term as more open data and useful tools for data integration and exploitation become available.

Fostering data integration and added value services beyond ARIADNE

ARIADNE will promote data integration and added value services beyond the ARIADNE data portal. This can allow the initiative to play a significant role in the data service environment in and beyond the different domains of archaeological, cultural heritage and other humanities research. A core element of this scenario is Linked Open Data (LOD) which ARIADNE generates and employs for internal service provision, but can also be used for interlinking with external data resources as well as provided to developers for producing added value services. Such applications may combine data from different sources (ARIADNE-mediated data and others) and promote cross-domain use of data resources.

The ARIADNE e-infrastructure could also benefit from using services provided by other major e-infrastructures, including from domains such as earth, environmental, biological and other sciences, which are present on the multi-disciplinary map of archaeological research. Deep interlinking of the data from such services would, however, require integration of conceptual knowledge of the different domains (e.g. terminologies, ontologies). Such integration could be sought based on use cases with a clear added value for archaeological research communities. While this is not a priority for ARIADNE in the 5-year horizon, we assume that in the 10-year horizon such integration might be attempted.

The suggested actions in this focus area concern two main goals of the ARIADNE data infrastructure and portal: to provide the archaeological research community with core and high-value additional services, and to enable the e-infrastructure act as a node in the wider information ecosystem of archaeology and beyond. Furthermore it is recommended to use the ARIADNE data portal as a tool supporting mobilisation of open research data.

Suggested actions: Provide core and additional data services

ARIADNE has taken great care to identify the services which the archaeological research and data management community expects from the project. This includes results of a large online survey on user needs, evaluated service suggestions of a panel of archaeologists and data mangers, and regular consultation with project partners and other institutions. Therefore the project has a solid basis to provide data infrastructure and services according to current expectations, and take into account that user needs will evolve, for example, in view of the opportunities offered by ever more available open data resources.

In brief, the archaeological research and data management community expects from ARIADNE:

- A data portal that allows an overview of available but dispersed archaeological data resources,
- Capability to search across different digital archives/repositories which hold such resources (i.e. data collections, databases, datasets of projects, etc.),
- Effective data discovery, browsing and filtering mechanisms, in particular based on geo-location (maps) and date-ranges/chronologies, but also other advanced options such as facetted search,
- Data access methods according to the different data access levels, types/products and interaction modes offered by the providers.

These requirements will be met by providing the user community with an online facility to register and describe accessible data resources based on a common model, semantic integration of the information, and offering portal services that provide the required set of data discovery and access functionalities. We note that the data access levels of providers vary, so that the data portal can in general enable collection-level but in many cases not direct item-level access.

On the other hand, advanced online access and use of data items may be enabled with regard to products for which a strong demand exists, for example, visual media such as 3D models of objects, built structures and landscapes. The already available ARIADNE services in this area demonstrate that the project actively responds to perceived demand.

There is also clear evidence for services which the research community does not expect from the ARIADNE data portal. This concerns typical features of Web 2.0 platforms such as content filtered based on tags or ratings of other users, academic/professional profiles of users, and expert networking and discussion on the portal. Such services are provided by widely used portals such as Academia.edu and others.

Clearly the service portfolio of the ARIADNE portal should meet core requirements of data overview, search and access. There is little scope to invest limited funds on specific services that are not appreciated, are provided by other portals, or may run ahead of the needs of broad user segments.

Provision of the described services (registry, portal) will be a major step forward in the archaeological domain, allowing overcome current limitations of discovery, access and re-use of available data. At the same time, the ARIADNE e-infrastructure may open up future opportunities for online collaborative research in Virtual Research Environments (VREs), offered on top of the e-infrastructure and accessible data resources.

Suggested actions: Support for services beyond ARIADNE

The ARIADNE data infrastructure and portal should not be an "island" but serve as a node in the ecosystem of e-infrastructures and applications for the archaeological research community in Europe and beyond. To serve as such a node, support for, and interoperability with, external services is required. This concerns the question of how the ARIADNE data infrastructure can enable added value beyond the portal services.

In general this will be through the ARIADNE Linked Open Data (LOD). LOD are employed for internal service provision, but can also be used for interlinking with external data resources as well as provided to developers for producing added value services. The suggested actions below are evaluated recommendations of the ARIADNE portals survey.

Deploy Linked Open Data (LOD) to integrate information within the portal and to link to external resources which follow LOD principles

Using Linked Open Data (LOD) for information integration has been among the highest valued recommendations of the portals survey and is already followed in the development of the ARIADNE services. The development focuses on the information available in the data registry and employs widely used LOD vocabularies such as the Geonames gazetteer for geo-information, the Art & Architecture Thesaurus for subjects, and others. Use of such vocabularies will also allow interlinking with external resources which follow LOD principles. Rich interlinking with relevant resources within the domains of cultural heritage and humanities as well as of other disciplines will require dedicated efforts. To expand the initial web of ARIADNE LOD, related developer communities could be invited to create link-sets (i.e. sets of links between LOD resources).

Provide interfaces that allow external applications exploit available data, metadata and conceptual knowledge

This recommendation can be followed easily by allowing external application developers access and use the ARIADNE LOD (i.e. the SPARQL endpoint of the Linked Data [RDF] triple-store or a "dump" of the whole dataset). Available and well-documented interfaces can promote experimentation and generation of novel applications. Such applications may combine data from different sources, e.g. data “mash-ups” or assembled virtual collections of research objects. In the medium to long term we may also expect Virtual Research Environments (VREs) which use the ARIADNE LOD and others to support cross-domain collaborative research.

Enable integrated access to data and publications (i.e. include metadata of document archives and publishers)

ARIADNE focuses on data resources while aggregation of metadata of publications from document repositories (i.e. self-archived papers) and publishers is not considered yet. But the connection of research publications and underpinning data is of vital importance in the scientific enterprise. Therefore the project should investigate how interlinking of and integrated access to publications and data on the ARIADNE portal might be achieved. Established data citation standards (e.g. DataCite) will be instrumental in this regard.

Help enrich specialised community sites (thematic or focused on specific types of data) with relevant information, e.g. RSS feeds on newly available data

Such services were not appreciated by several project partners. There is a concern that this would add to the proliferation of specialised websites that are often not maintained and enriched with user contributions. However, established websites of research communities might benefit from ARIADNE information, e.g. notification about relevant new data. This would also allow enlarging ARIADNE’s footprint in and beyond the sector and may stimulate (re-)use of available data.

Suggested actions: Data mobilisation support

The ARIADNE data portal should also support data mobilisation through raising awareness of the importance of open research data and assisting capacity development. Thereby the ARIADNE initiative can align with open data policies, required data management and access planning, and requirements for high-value data in general.

The ARIADNE data portal, for example, could highlight and point to authoritative and supporting sources (e.g. the Horizon 2020 Open Research Data Pilot, OpenAIRE2020 and others), Guides to Good Practice in data management, open data licensing, available state-of-the art data archives, and inspiring examples of research projects that re-use archived data. Also promoted should of course be the option to register and share curated data resources through the ARIADNE Registry.

2.3 10-year innovation horizon

For the 10-year horizon of the Innovation Agenda and Action Plan this report outlines the objectives, methods, and selected topics which will be addressed in greater detail in the final agenda report. The overall goal is to create a roadmap towards potentially transformative innovations in archaeological research and communication, in particular digital archaeology (e-archaeology). With a roadmap at hand, new approaches to digital, ICT-enabled archaeological research and communication can be explored, challenges addressed, and feasible routes taken.

2.3.1 Introduction and overview

The current debate about the system of scholarly research and communication is all about “open”, including, among others, open access (publications), open data, linked open data, open science and, of course, open research infrastructures. As many proponents suggest, “openness” provides much potential for novel forms of research collaboration, including participation of citizens, and innovative generation and publication of new knowledge. Thus the request for “openness” (or “open science”, “science 2.0” and other labels) is closely tied to the expectation of transformative processes which could lead to favourable outcomes as envisioned by different proponents, advances in knowledge and relevance of science in/for society, for instance.

The ARIADNE Innovation Agenda and Action Plan are not about incremental innovations through small improvements of research practices and existing tools. The goal, in the long term, is transformative innovations, substantial changes in archaeological research practices, with a focus on digital resources and ICT-enabled research. A breakthrough in the next few years towards wide sharing and re-use of open data would already be a transformative innovation.

The impact of the request of open data is already felt in the ecosystem of research. Because compliance requires efforts such as negotiation of open data mandates, implementation of appropriate digital archives, solving intricate questions of IPR & licensing, and training of researchers (e.g. data management planning). The objective of open research data however are innovations in research practices that open new perspectives, new ways of doing research and, hopefully, new knowledge to address societal, environmental and other issues.

New technologies may trigger innovation in research practices, however, technology is but one factor in innovation and not necessarily the most important. Therefore the 10-year horizon of the ARIADNE innovation agenda is not primarily about technology, although some topics will require addressing questions of technology and data in greater detail.

Transformative innovation concerns the whole ecosystem of archaeological research and communication, including institutional, technical and socio-cultural dimensions. As an example: if we think of novel forms of digital publication, the core challenge is not technical but institutional. Because publication is tied to the system of scientific review, recognition and reward, and what does not fit with the established system will find it difficult to advance from prototypic solutions to wide adoption by the research community.

Approach and brief overview

The 10-year roadmap of the innovation agenda goes beyond the immediate innovation needs which concern open archaeological research data, data archives, e-infrastructure and services, among others. One major question in the long-term perspective is what new opportunities broader access to open research data will offer for innovative digital, ICT-enabled research in archaeology. Therefore the ARIADNE data infrastructure and portal services will be one element that connects the different, 5/10-year horizons. This does not mean, however, that all or even a large part of the innovation agenda for the 10-year horizon will relate to the ARIADNE data infrastructure and services. But considering their potential and further development will be part of the exploration of the 10-year horizon.

Next there is the question which innovation topics could become particularly relevant in the 10-year horizon. The preliminary innovation agenda addresses a number of topics which have been identified through a scanning of recent sources (literature, reports, project websites and other online sources), mostly of authors of the field of “digital archaeology” in its various forms. In the further work on the 10-year horizon some topics may appear as less relevant and others added.

Scanning and discussion of current topics is a first step in developing a roadmap towards innovative digital, ICT-enabled archaeology. Use of other methods such as scenario building and expert panels will be required to shape and evaluate the roadmap. But based on the discussion of the selected topics some actions can already be suggested which stakeholders could take today to bring about envisioned advances in digital archaeology in several respects.

Some of the topics for the 10-year horizon of the research agenda have been discussed in the ARIADNE Expert Forum on Digital Futures of Archaeological Practice 2020-2025, 2-3 July 2015 in Athens. The forum has been organised by the Digital Curation Unit of the IMIS-Athena Research Centre on behalf of the ARIADNE Special Interest Group (SIG) on Archaeological Research Practices and Methods[14]. But the content of this report does not result from this forum, hence, ideas and opinions expressed should not be understood as the view of participants, except the main authors of this report. The published results of the forum will of course feed into the work on the final ARIADNE Innovation Agenda and Action Plan, which will become available in November 2016.

The 10-year horizon goes beyond the formal lifecycle of the current ARIADNE project. The elaboration of a roadmap and action plan towards 2025 is an opportunity to stimulate a broader reflection and discussion among stakeholders of long-term needs in innovative digital, ICT-enabled research and communication practices.

2.3.2 Shift from data- to research-focused innovations

The development of the 10-year horizon (2025) of the innovation agenda builds on and extends the framework of the work on the 5-year horizon. The objectives and suggested actions for the first horizon (2020) primarily concern immediate innovation needs with regard to open archaeological research data, data archives, e-infrastructure and services. The 10-year horizon of the innovation agenda can focus more on innovations in digital, ICT-enabled practices of archaeological research.

We assume that without substantial achievements with regard to the objectives of the 5-year further progress in digital archaeology and collaboration across current boundaries of research domains, organisational and national settings is unlikely. Alternatively, if the immediate goals are met sufficiently, future-oriented perspectives will be strengthened or emerge in the 10-year horizon.

These perspectives will be less dominated by the need of growing and integrating the stock of open and interoperable research data. This will still require much attention and improvement. But instead of being blocked by seemingly unsurmountable barriers, further progress in digital, ICT-enhanced archaeological research (e-archaeology) can be foreseen, explored and targeted. Scenarios in the long-term horizon of 2025 can look more into potential research-focused innovations, explore emerging new perspectives and capabilities, and suggest pathways towards innovative and potentially transformative e-archaeology.

The overall focus of the 10-year horizon is digital, ICT-enabled research practices, with special emphasis on e-research that is conducted online. We assume that the needs and requirements of the archaeological research community will evolve towards e-research capability when more open data becomes accessible and re-usable through community archives, data infrastructure and services. We do not expect, however, that such novel research practices will be driven mainly by advances concerning data and technology, changes in research collaboration and methods will be just as or even more important.

The figure below illustrates the shift from the current focus of enabling open access to re-usable and interoperable data to enabling novel forms ICT-enabled archaeological research:

[14] ARIADNE Expert Forum: Digital Futures of Archaeological Practice 2020-2025 (Athens, 2-3 July 2015), http://summerschool.dcu.gr/?page_id=19#expert-forum; the expert forum has been recorded and a report or paper manuscript, based on the summarised recording and further discussion, is expected before the end of 2015.

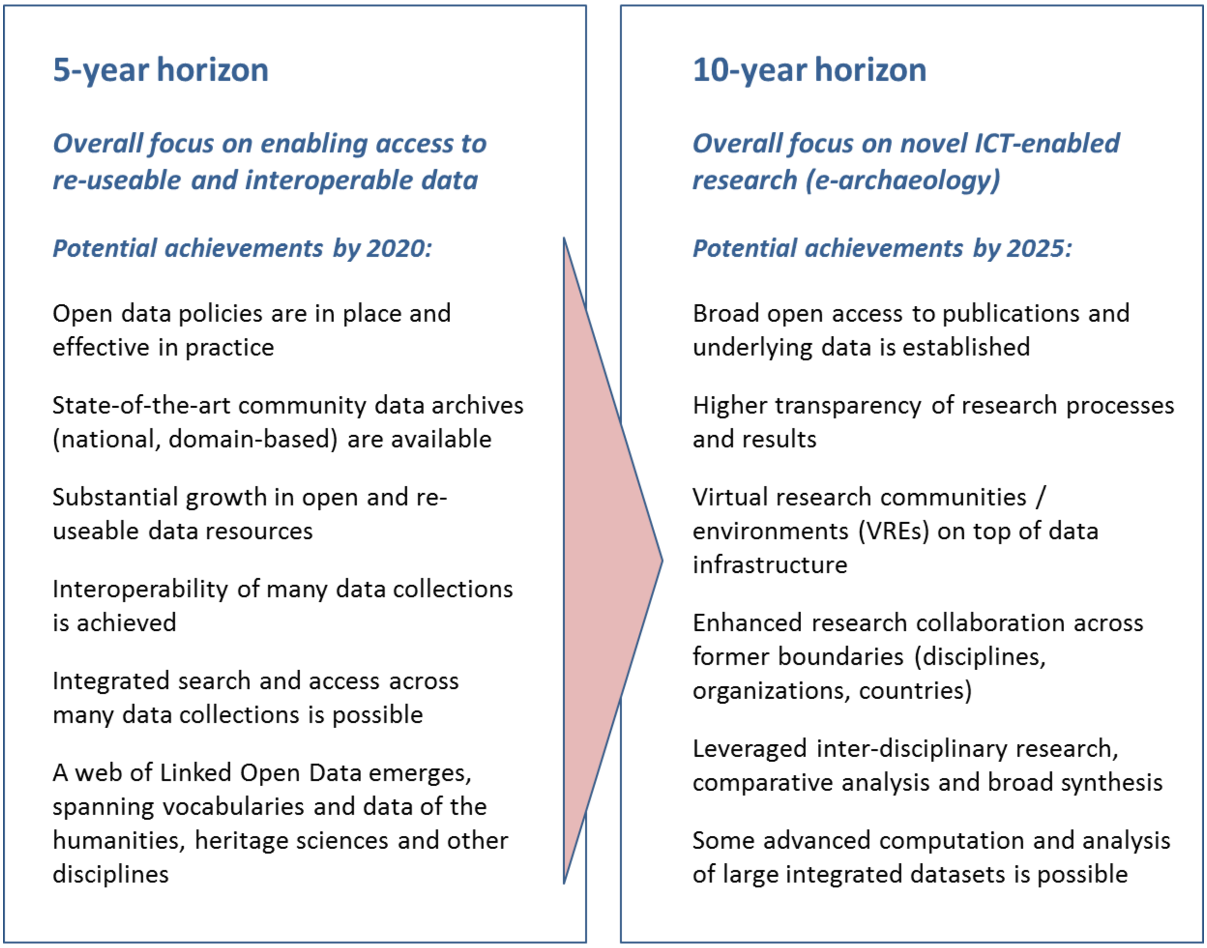


Figure 2: Towards open and transformative e-research. ARIADNE, SRFG 2015

The figure contains some potential achievements which may require some explanation:

The 5-year horizon includes community-level archaeological data archives. At present a common data archive is missing in many European countries, hence achievement of this goal will be rather demanding. Some large research institutes may be tempted to implement their own data repository, however a community-level solution will allow bringing together the research data in one place and will be more cost-effective than maintaining several repositories (cf. *Section 5.2.4*). Countries where the research community lacks a common data archive may "leapfrog" to a state-of-the-art solution by learning from acknowledged benchmarks such as the Archaeology Data Service in the UK.

Another difficult to achieve element in the 5-year horizon is the emergence of a web of Linked Open Data (LOD) that spans vocabularies and data of the humanities, heritage sciences and other disciplines. As a multi-disciplinary field of research archaeology could benefit greatly from such a LOD web as it allows discovery, browsing and retrieval of semantically related data and knowledge. But not many archaeological LOD resources have been published and interlinked as yet. The ARIADNE data catalogue will become available as LOD, which may promote some further LOD publication and interlinking. However, in the coming years the domain of archaeological and other heritage research would have to adopt the LOD approach much more so that a rich web of LOD emerges (cf. *Sections 7.3/7.4* and *Section 8.5.6*).

In the 10-year horizon we expect that virtual environments for archaeological research will become available, built on top of the ARIADNE and related data infrastructures and services. The archaeo-logical research communities currently lack such environments that are tailored to their specific requirements (cf. *Section 9.1.4*). In the figure some caution is expressed with regard to expectations of advanced computation and analysis of large integrated datasets. Archaeological data is difficult to integrate for such computation and "big data" mining technologies may not be applicable to many data resources (cf. *Section 8.5.4* and *Section 8.5.7*).

In the 10-year horizon there will still be a need to broaden the access to various data resources required for archaeological research. As a multi-disciplinary field of research, archaeology depends on a high-level of openness of all involved disciplines and research specialties. Furthermore required will be good access to data which are usually not produced by archaeologists or in an archaeological context, airborne or satellite remote sensing and imaging data, for instance.

2.3.3 Summary of topics and suggested actions

This section summarises the discussion of the selected roadmap topics and suggested actions. In the further work on the 10-year horizon topics may appear as less relevant and others added.

An explorative undertaking

In line with ARIADNE's focus on data infrastructure and services for the archaeological research community, the work on the 10-year horizon of the innovation and action plan focuses on innovative digital, ICT-enabled archaeological practices. Emerging new perspectives, innovation potentials and development paths towards "e-archaeology" in 2025 are being explored, and possible actions for enabling envisioned innovative practices, methods and tools considered.

In comparison to the rather clear objectives and requirements of the 5-year horizon, the 10-year horizon is more of an open field. It must count on considerable achievements in the 5-year horizon, with regard to open and re-usable data, for instance. What archaeological researchers, tool developers and others will do based on the achievements is much less clear.

The drive towards innovative e-research in archaeology will come less from research policies (as in the case of open data), rather visionary researchers, data scientists, tool developers and others will pave the way towards further achievements. However the ARIADNE innovation roadmap for the 10-year horizon may provide some orientation, point towards interesting avenues and, thereby, help in bringing about and shaping desirable futures.

As a first step for developing a roadmap towards innovative "e-archaeology" current future-oriented topics and discussion have been scanned, described, and some tentative suggestions added. The next step will be scenarios of innovative progress, involving consideration of required approaches and means (e.g. tools, data and others). Scenarios may also be provided by members of the ARIADNE expert community (e.g. the ARIADNE Archaeological Research Practices and Methods SIG). Based on an evaluation of the scenarios, actions will then be suggested on how the envisioned future digital, ICT-enabled practices and required means could be brought about.

As the work on the 10-year horizon will be conducted until November 2016, it is foreseen interim results (e.g. scenarios) will be issued in the form of working documents, inviting suggestions and contributions by the wider ARIADNE stakeholder community.

Suggested actions

The suggested actions below concern the development of scenarios of innovative progress in digital, ICT-enabled archaeology (e-archaeology) in the 10-year horizon of the ARIADNE innovation roadmap. They are, however, phrased as actions stakeholders could already take today.

Address the question of archaeology's societal relevance

The 10-year road mapping focuses on innovations in archaeological practices of research and communication. In this context questions of relevance are crucial. Such questions concern the role archaeology can and should play in society, with regard to societal and environmental challenges, for instance. What would be desirable future contributions of archaeological research to tackling such challenges? What kind of research – digital, ICT-enabled or not – might enable such contributions? What is required to allow for such research, for example in terms of researchers' perspectives, knowledge, methods, technologies and data?

Concepts of digital science will be futile in archaeology (and elsewhere) if not connected to vital current debates. It is felt widely that the archaeological research community has no strong role and voice in such debates. Gaining such a role is about demonstrating relevance for societal, environmental and other concerns. This may leverage citizens' appreciation, understanding and support of archaeology.

Suggested actions

- *Leverage archaeology's societal role and relevance, for example through connecting archaeological research and knowledge with current concerns and affairs such as climate change, environmental sustainability, urban agglomeration, globalisation and geo-politics, regional conflicts, migration, and others.*
- *Consider contributions which could allow archaeology a stronger voice in current debates.*

Take account of the diversity of archaeological research practices and methods

Archaeology is a multi-disciplinary field of research in which researchers in various domains address their research questions with different theories, methods, data and tools, e.g. classical studies versus environmental archaeology, for instance. The diversity of "schools" and research practices in archaeology should be considered when trying to identify and conceive pathways towards innovative digital, ICT-enabled research that makes a difference. One common feature is the research lifecycle from project idea to publication of results (incl. open sharing of the generated data). Also issues of standardisation (e.g. data models, terminology) and cost-effectiveness will remain common concerns.

Suggested actions

- *Recognise that different archaeological schools of thought and research practices require different digital, ICT-based research environments and tools.*
- *Focus on phases in the lifecycle of archaeological research in which significant progress in knowledge may be achieved. In the last decades data generation has seen enormous progress; in the future other phases may require more attention.*
- *Recognise that issues of standardisation (e.g. data models) and cost-effectiveness are relevant for future research practices.*

Target data integration for comparative and synthetic research

Open and re-usable data will allow easier combination of data for comparative and synthetic, cross-disciplinary research. Enlarged and integrated datasets may allow new insights through applying models of social systems and behaviours, adaptation to climate and environmental change, for instance. In recent years archaeological research has taken aboard an arsenal of data capture methods and produced growing volumes of field survey and excavation data for documenting individual sites and areas. On the other hand, data integration for comparative and synthetic research is lagging behind ever more.

There is a need of novel approaches and tools that allow researchers bringing together and working effectively with the variety of data required for cross-domain, interdisciplinary research, in fields such as historical ecology or urban archaeology, for instance. One solution that could drive progress in large-scale data integration could be setting up programmes and competence centres that support this task, for example, through facilitating collaboration between researchers with domain-specific knowledge and data science expertise.

Suggested actions

- *Foster the development of novel methods and tools that allow researchers to bring together and work with the variety of data required for cross-domain, interdisciplinary research.*
- *Promote competence centres and programmes aimed at data integration for comparative and synthetic archaeological research.*

Explore relevant future VREs for archaeological research

The e-research scenario of virtual research environments (VREs) build on top of e-infrastructure and underlying data repositories has not yet reached the archaeological research community. The ARIADNE data infrastructure and services may inspire VRE developers to create environments for archaeological researchers. Such environments can range from loosely coupled tools and services to tightly integrated virtual workbenches for specific research communities. Relevant VREs for archaeological research should be explored, taking account of the state-of-the-art in other disciplines and particular requirements of archaeological researchers in different domains as well as in cross-domain collaboration.

Suggested actions

- *Look into VREs developed for other domains to conceive environments relevant for e-research in specific archaeological domains as well as in cross-domain collaboration.*
- *Consider cases where researchers use data mediated by ARIADNE as well as data infrastructures and services of other disciplines (e.g. geo, environmental, biological data).*

Identify e-science practices based on data infrastructure and computing facilities

E-science has become an ever broader topic that now spans all forms of research activities conducted online, even involving “citizen scientists”. The classical, still vital variant of e-science is use of advanced and distributed computing by researchers in the natural sciences and other disciplines. Research in the heritage sciences, i.e. archaeology, cultural heritage and other humanities research seldom employs such computing. Available Grid/Cloud-based Distributed Computing Infrastructures did not find much use by researchers in the heritage sciences so far. The ARIADNE data infrastructure and services may promote e-science activities. In future scenarios this could involve distributed computing based on a seamless flow of data to computing infrastructures and *vice versa*.

There are expectations that mining of “big data” can allow also archaeologists novel insights, e.g. relevant patterns in data that suggest new research questions. However, many aggregated and integrated large archaeological datasets may not be available for quite some time. But archaeological researchers may increasingly use Cloud-based research support services for other purposes than computing, data transfer, temporary storage and access during research projects, for instance.

Suggested actions

- *Promote collaborative way-finding for e-science approaches, methods and tools relevant to archaeological researchers.*

- *Focus on e-science needs specific to archaeological research, which may differ from those of other humanities as well as the natural sciences.*
- *Look for uses of low-level Grid/Cloud based services and emerging examples of archaeological applications of "big data" mining and other methods.*

Propose grand challenges for the digital archaeology community

Experts of the archaeological informatics and computing community suggest seeking "grand challenges" which are research-directed and contribute to the development of new theories and methods. Thereby the community could play a stimulating and transformative role rather than simply support well-established research practices. Grand challenges for digital archaeology should go beyond what seems feasible in the short to medium term through applied research and engineering. The challenges would inspire the research community to push boundaries and explore new avenues of research with potential revolutionary impact, i.e. a shift in established paradigms, theories and methods of archaeological research. Meanwhile also grand challenges for archaeology have been identified, most of which require large-scale and integrated datasets. It appears that the sought for advances require bringing together domain experts (e.g. theories, methods, data) and developers (software, computing) to create novel research tools, and a strong focus on cross-domain, interdisciplinary research.

Suggested actions

- *Seek grand challenges that inspire the research community to push the boundaries of digital archaeology.*
- *Suggest challenges that promote mobilisation and integration of datasets for domain and cross-domain, interdisciplinary research.*
- *Bring together domain experts and developers to create methods and tools for such research.*

Evolve a Web of archaeological Linked Open Data for research

The last 10 years have seen much progress in Linked Open Data (LOD) know-how required to produce, publish and interlink LOD of archaeological and cultural heritage collections/databases. In practice, however, not many LOD datasets have been produced and interlinked so far. As a multi-disciplinary field of research archaeology could benefit greatly from a web of LOD that spans vocabularies and data of the humanities, heritage sciences and other disciplines. This would allow discovery and retrieval of semantically related data and knowledge of different domains of research within and beyond archaeology.

The ARIADNE data catalogue will become available as LOD, which may promote some further LOD publication and interlinking. However, a much wider uptake of the LOD approach for semantic interoperability in the archaeological and other domains is necessary so that a rich web of LOD can evolve. Two core requirements must be met: effective interlinking of LOD requires use of common or mapped vocabularies (thesauri, ontologies), and the LOD resources need to be curated to ensure reliable interlinking and access. A central role in the LOD scenario plays the CIDOC Conceptual Reference Model, which recently has been extended for scientific observations and argumentation, and domain-specific modelling (e.g. archaeological excavations). It is expected that mapping of archaeological data collections/databases to the extended CIDOC-CRM will enable enhanced search as well as research-focused applications.

Suggested actions

- *Promote publication of LOD datasets (collections, databases) by more archaeological and other cultural heritage institutions, especially based on mappings to the extended CIDOC-CRM.*
- *Foster a community of LOD curators who ensure reliable availability and interlinking of LOD resources (datasets and vocabularies).*
- *Develop LOD-based applications that demonstrate advances in research capability, which may motivate a wider adoption of the LOD approach by research institutions and projects.*

Promote new forms of scientific/scholarly publication

The next 10 years will very likely see some advances with regard to new forms and ways archaeological researchers could publish project outcomes. However, publications are tied to the scholarly review and credit system. Without some adjustments in this system new forms of online publication may find it difficult to progress from prototypes to wide adoption by publishers and authors. Uptake will be higher for moderately enhanced familiar forms (e.g. embedding explorable objects such as 3D models in online papers).

Advanced forms will be driven by a process-view of research rather than static documents, for example, figures that auto-update as new data becomes available. This will require deep interlinking of publications and datasets. Also the distinction between journal and repository may become increasingly blurred if digital repositories become publishing platforms and value-added services for scholarly communication.

Archaeologists are aware that the finally "published excavation" in book form is not the optimal solution. Novel repository-based forms of publication may allow significant steps towards media/ data-rich reports of investigations. We may also envision the "digital record" of an on-going excavation as a stream of data from the field and laboratories, continuously made available, analysed and discussed by subject experts. In such a setting "publications" would be snapshots of the state of knowledge at a certain time, instead of the annual excavation report and some papers by researchers in need of taking care for their academic record.

Suggested actions

- *Promote novel forms of digital publication that could "work" for archaeological projects in terms of enhanced access to research outcomes as well as academic credit.*
- *Start with moderately enriched familiar ways of publication (e.g. embedding explorable digital objects in online papers), and make new approaches as easy as possible.*
- *Investigate fields of "data-driven" archaeological research and publication in which accessible datasets and executables (software, dynamic figures, etc.) could play an essential role.*
- *Explore repositories as platforms for media/data-rich archaeological publications and value-added services for scholarly communication.*

Foster participatory and reflective online public/community archaeology

Many community/public archaeology projects have found it difficult to engage the communities they claimed as stakeholders in the ways described in theory, i.e. non-hierarchical, participatory or, even, "rooted" in the community. There appears to be a large gap between what is expected from involving citizens in the archaeological research process and what is actually possible in such involvement. The shift of engagement activities into the digital realm aggravates conceptual issues of public/community archaeology, for example, the (sociological) concept of community becomes illusive.

Online environments specifically built for “crowd sourcing” contributions by non-experts often present a one-way participation approach. Concerning expectations from open research data, archaeological documentation and data is not something many non-experts may easily understand and use for own research work (“citizen science”). Mediation by archaeologists might again reproduce hierarchical and expert-directed involvement. Archaeological institutions and projects already use social media to increase visibility and disseminate information. Conceiving novel, participatory approaches based on such media will require a highly reflective usage.

Suggested actions

- *Conceive and engage in participatory approaches based on online platforms (e.g. social media) in a highly reflective way.*
- *Explore the concept and practicalities of “open research communities” that involve archaeologists and citizens in the production, dissemination and re-use of open data.*
- *Instead of seeking “roots”, focus on relevance of archaeology for societal, environmental and other issues, at regional as well as global scale.*

3 Focus area 1 – Research e-infrastructures and digital resources

3.1 Introduction and overview

The ARIADNE project centers on the development of an e-infrastructure and services for archaeological and cultural heritage data resources (e.g. data registration, discovery, and access and other services). This focus area therefore addresses the current situation and major issues of e-infrastructure development in Europe in general, as well as of e-infrastructure and digital resources for archaeological and cultural heritage research in particular.

In the last about 20 years, e-infrastructures have become ever more important, indeed crucial for the conduct and progress of research in all branches of the scientific enterprise. There is an increasing need to build, share and integrate digital resources for research. This includes terminology and conceptual knowledge (e.g. thesauri, ontologies) which allow linking of and enhanced access to research publications and data.

Archaeological research is based on established practices and tools for structuring, analysing and presenting research results of individual projects. But the development of common e-infrastructures and services for data sharing is understood to be lagging behind, although necessary to allow for more effective, and possibly innovative research.

The European Archaeological Council (EAC)[15] in the "Managing the sources of European history" theme of their Amersfoort Agenda emphasises *"the need to share, connect and provide access to archaeological information with the help of digital technologies. The key to this aspiration is to improve collaboration – we need to share rather than exchange. It is essential to encourage the development of European data-sharing networks and projects in the field of archaeology. The ARIADNE project is an excellent European initiative in this regard and participation in this project should be strongly encouraged"* (European Archaeological Council 2015: 21).

Like other disciplines archaeology indeed faces the challenge of supporting through common e-infrastructure and services a scientific enterprise that has become increasingly collaborative, distributed and data-intensive. But in the current drive to build new or upgrade existing infra-structures and resources for "e-science" there is a risk of limited funds being invested in many uncoordinated, possibly redundant and unsustainable initiatives. This risk is particularly critical in the heritage sciences, which are characterised by having many different stakeholders and data resources of research institutes and laboratories, cultural heritage agencies, and institutions such as museums and archives

In the domain of archaeological research the risk of a "many-headed beast"[16] does not reside at the level of European-level data infrastructures, because no such infrastructure exists as yet. Rather it will be up to ARIADNE to provide the archaeological sector with a first common solution. But the solution will require sustained efforts to allow for maintenance and extension, e.g. mobilisation of data resources of all European Union Member States, integration with relevant other European data

---

[15] The European Archaeological Council (EAC), according to their mission statement, *"is a democratic network of heads of national services responsible under law for the management of the archaeological heritage in the Council of Europe member states. The EAC therefore represents the managers of the historic environ-ment and the associated cultural heritage"*, http://european-archaeological-council.org

[16] The "many-headed beast" is one of the e-infrastructures scenarios of the eResearch2020 study, character-ized by initiatives heading in different directions, incompatible technologies, duplicating efforts, etc. (cf. *Section 3.2.3*).

infrastructures, among others. Regular coordination activities at the European level is necessary to ensure interoperability of the e-infrastructures and digital resources as required by the heritage sciences. Such coordination activities have already started in the PARTHENOS project[17].

A critical situation however exists at the level of digital resources and environments for archaeological and other heritage research. The main issues here include, but are not limited to, fragmentation, disconnectedness, and lack of open access. At this level a more targeted approach of resource development, access and usage needs to be conceived and put into place.

3.2 Growing importance but need of coordination and synergies between e-infrastructures

This section addresses the current situation and major issues of e-infrastructure development in Europe in general and e-infrastructure for cultural heritage, archaeology and other humanities research in particular. It highlights the need for coordination of the e-infrastructure ecology, which in the multi-disciplinary field of archaeology is of critical importance. Moreover cooperation with providers of relevant other e-infrastructures, services and data resources is seen as beneficial.

3.2.1 Importance of e-infrastructures

The term Research Infrastructures (RIs) refers to facilities, resources and related services used by the scientific community to conduct research in their respective fields. RIs typically are of national or international/European importance, however it is now common to describe also local arrangements like a campus-wide information system or laboratory facilities of a natural sciences department as research infrastructures.

Major RIs, both physical (e.g. laboratories, telescopes, research vessels) and multi-tier *e*-infrastructures (e.g. digital archives, data federation, access and computing) require large investments on the national and international levels. In Europe, the European Strategy Forum on Research Infrastructures (ESFRI) was launched in 2002 to support a coherent approach to policy-making on major infrastructures, which are understood to be one of the pillars of the European Research Area. In the first ESFRI Roadmap, Research Infrastructures (RIs) have been defined as follows:

"This definition of Research Infrastructures, including the associated human resources, covers major equipment or sets of instruments, as well as knowledge-containing resources such as collections, archives and databases. Research Infrastructures may be 'single-site', 'distributed', or 'virtual' (the service being provided electronically). They often require structured information systems related to data management, enabling information and communication. These include technology-based infrastructures such as grid, computing, software and middleware" (ESFRI 2006: 16).

This definition is often used in studies of RIs as it distinguishes different types of RIs and highlights the need of "virtual" infrastructure, e.g. databases, digital collections and archives, data federation and portal services. Indeed, since 2006 e-infrastructure has been recognised as crucial for the operation of any single-sited RI (e.g. a research facility like CERN) or distributed RI (e.g. a formalised, long-term collaboration of similar or complementary research facilities in different countries). The

[17] PARTHENOS - Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies (EU, H2020, 05/2015-04/2019), http://www.parthenos-project.eu

fact that most research depends on the capability to generate, store, share, access, process and analyse data blurs the distinction between "virtual" and other research infrastructures[18].

The increasing role of ICT-based *e*-infrastructures for research or "cyberinfrastructure" (the term used in the USA) corresponds to the overall trends in scientific research, which are that the research has become increasingly collaborative, distributed and data-intensive (cf. Wuchty & Uzzi 2007; Frenken *et al*. 2010; Riding the Wave 2010). Indeed, *e*-infrastructures are understood as important pillars and drivers of innovative scientific research. There is a growing need to enable resource sharing over e-infrastructures by pooling data resources, tools and services, and to support team-based, cross-disciplinary collaboration through virtual research environments. Moreover there is the expectation that with large datasets ("big data"), e-infrastructure and advanced computing new scientific questions can be tackled (NSF 2007; IWGDD 2009; Riding the Wave 2010).

3.2.2 Lack of coordination in the e-infrastructure ecology

In recent years many research *e*-infrastructure projects have been funded in Europe on the national level as well as through the European Union's 6th and 7th Framework Programmes and other programmes like eContent(plus). The landscape of e-infrastructures and digital resources that emerged from these projects presents a patchy, highly fragmented picture.

The e-Infrastructure Reflection Group (e-IRG), since 2003 the main advisory body for European e-infrastructures, warns of negative effects of this fragmentation. The e-IRG asks for more co-ordination among all stakeholders and suggests establishing a single "e-Infrastructure Commons" for knowledge, science and innovation (e-IRG 2013). Lack of co-ordination would mean that Europe misses opportunities for creating coherent, cost-efficient and sustainable e-infrastructures. In turn, this would impede opportunities for both disciplinary and cross-disciplinary research and innovation (cf. Holmgren 2014).

Among several related issues, the e-IRG observes insufficient integration of existing e-infrastructures and services, high awareness by users of borders between technologies/interfaces of individual components, lack of coherence of most user communities, disparate legal frameworks in different countries, and lack of sustainable funding of many e-infrastructures (e-IRG 2013: 11).

The European Science Foundation with regard to e-infrastructures for the humanities notes: *"Digital infrastructures are developing rapidly but unevenly, and there is an urgent need for coordination, standardisation and sharing of experience to prevent unnecessary duplication and the atomisation of good initiatives"* (ESF 2011).

The EPIRIA evaluation of Research Infrastructures activities under the Seventh Framework Programme for Research of the European Union identified a particularly critical situation of research infrastructures (RIs) for the Social Sciences and Humanities (SSH). The EPIRIA report (2014) notes that the RI policy does not meet their specificity, and that there is too little funding, which will make it difficult to sustain, least extend and enhance their services. It is worth quoting the main observations and conclusions of the report with regard to SSH RIs:

"It is obvious that transnational access to RIs in the whole SSH domain is a crucial point. The aim of many of the RIs in the SSH domain is to offer open access to a comprehensive dataset or collections supported by advanced (ICT) tools. In order to serve the SSH research community and other users, the RIs should offer not only transnational access, but also offer multilingual and multimodal approaches. In this respect a (more) international approach is needed. Only a common exploitation of the RIs can take into account the European diversity and improve open access.

[18] The e-Infrastructure Reflection Group (e-IRG) in February 2015 issued guidelines on "Best Practices for the Use of E-Infrastructures by Large-scale Research Infrastructures", http://e-irg.eu/guidelines; see also Vandenbroucke (2015).

The lack of a sustainable funding stream for RIs in both the Arts & Humanities and the Social Sciences is considered to be the main problem. RIs in the Humanities and Social Science are underfinanced, both on the national and European level. In order to foster the competitive position of European research in the SSH domain, a substantial effort on the funding side should be undertaken in the next years and in a mid-term perspective, with clear milestones concerning funding strategies. Since there is a need of overarching RIs, special attention should be given to funding at a European level." (EPIRIA 2014: 78)

It appears that in the European Research Area (ERA) the importance of dedicated "soft sciences" RIs is underestimated. Indeed, we think that such RIs (especially *e*-infrastructures) should be extended, enhanced and stronger inter-connected, including with RIs of other domains. But more investment should come with more coordination with regard to creating synergies and collaborative innovation at different levels of e-infrastructure and services. Coordination will ideally also create the setting for knowledge transfer, cross-fertilization and interdisciplinary approaches.

3.2.3 How e-infrastructure development can go wrong

The study eResearch2020 - The Role of e-Infrastructures in the Creation of Global Virtual Research Communities (eResearch2020, 2010a/b) provides a useful account of the expectations from research e-infrastructures and how their development can go wrong. The study has been carried out by a consortium led by Empirica (Germany) for the European Commission.

The main goal of eResearch2020 was to understand better the organisational, collaborative and technological developments in e-infrastructures which are effective in supporting virtual research organisations in different fields of research. The study results are relevant for innovation agendas of different research communities, including archaeology.

eResearch2020 looked into current trends in e-infrastructures, conducted an international online survey of virtual research communities, and produced case studies of e-infrastructure providers in Europe. The study elaborated scenarios of e-infrastructure development towards 2020 and provided a set of recommendations for policy makers and research communities.

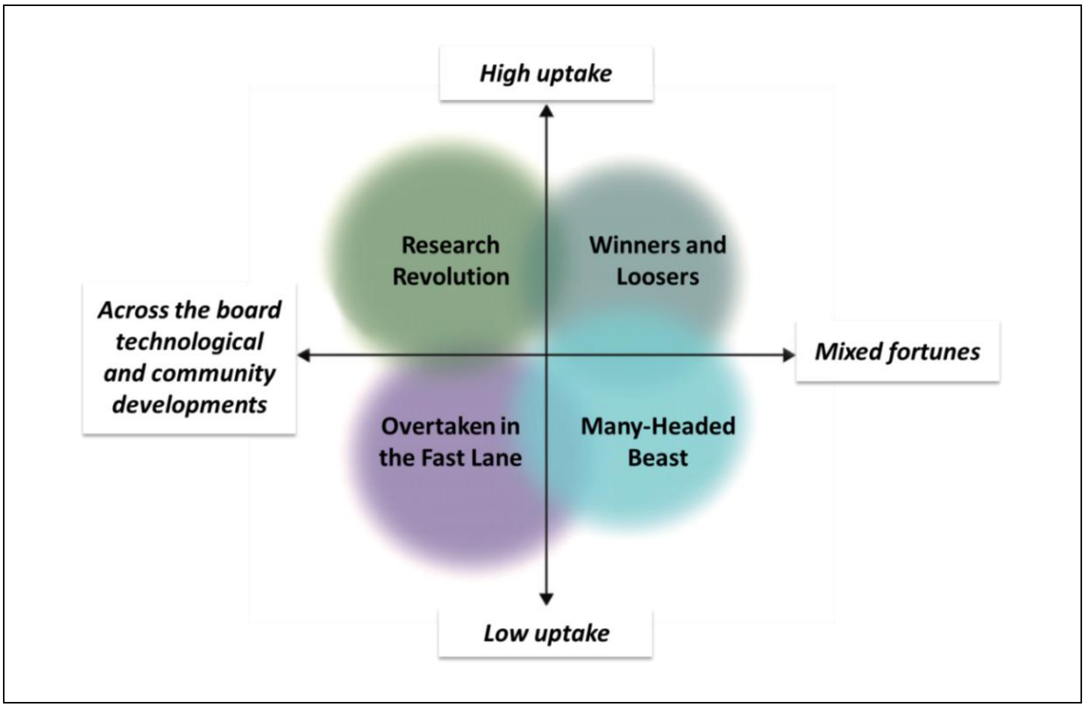


Figure 3: eResearch2020 scenario matrix. Source: eResearch2020, 2010b: 7

Our primary interest here is in the eResearch2020 scenarios, which the project developed as an e-infrastructures road mapping tool (eResearch2020 2010a: ix-xiv). The four scenarios identify different states of the e-infrastructure landscape in Europe 2020 depending on the consistency, level and domain-coverage of e-infrastructure adoption (see figure above).

Achievement of the scenario "Research Revolution" is understood as the best case. The other three involve a failure to reach goals of this scenario, which is described as follows:

- *"Large-scale collaboration, data- and tool- intensive,*
- *The nature of research is fundamentally transformed and carried out in distributed mode,*
- *Change takes place across all disciplines and there is cross-disciplinary fertilization,*
- *Change takes place on all levels of research (infrastructures, applications, daily practices) and at all levels, including in schools,*
- *Industry joins up with the research community and there are links to e-Government, e-Health and the public,*
- *Public funding is complemented by private funding, and an 'open science' ethos prevails."*

The authors emphasise the transformative and beneficial character of this scenario. However, they also note that it requires the largest amount of funding and targeted efforts of the research communities, and that the benefits might materialise only after a considerable time.

Keywords of this scenario for archaeological research are: transformed practices, data- and tool-intensive research, collaboration at a larger scale and in distributed mode, cross-disciplinary fertilization, and "open science" ethos. The scenario might also include closer ties with the commercial sector (e.g. contract archaeology) and a stronger involvement of citizens (public archaeology).

The eResearch2020 scenario "Overtaken in the Fast Lane" can be summarised as a decline of the Grid-focused paradigm of European research e-infrastructures and a slow uptake of alternative Cloud-based services. In the US and Asia such services would be offered as a commodity by commercial providers and find also wide usage in the research sector. The potential failure on the route to the best case scenario in Europe is seen as a lack of Cloud-based services for the research communities, for example, offered by the academic/public sector or public-private partnerships. In the eResearch2020 online survey a large majority (80%) of respondents found it likely or very likely *"that new resource delivery models such as Software as a Service, Cloud Computing or Utility Computing will spread and have a significant impact in science in the next five years"* (eResearch2020 2010b: 5). The main message for archaeology here is to look for Cloud-based services that are appropriate for research purposes.

The two other, "mixed fortunes" scenarios are most relevant to ARIADNE in its role as an Integrating Activity by means of e-infrastructure and services for the archaeological research and data management communities:

"Winners and Losers" foresees that some research communities are moving towards the best case scenario while others do not get beyond planning and are left behind. The "losers" are characterised by weakness in terms of ICT-enhanced research (e-research), especially in collaborative and cross-disciplinary modes, lack of new avenues of research, and retreat into disciplinary silos. The scenario authors note that winners and loser could be found also within disciplines, and that some fields of research might gain through data- and resource-sharing whereas others remain unable to benefit.

The "Many-Headed Beast" scenario depicts an e-infrastructures landscape characterised by a mixture of incompatible technologies being used, strong and weak research identities, funding concentrated in pockets (with little for the humanities), e-research efforts ignored in some areas and heading in different directions in others, lack of possibilities to collaborate, some areas duplicating efforts, and so forth.

The core message for the archaeological research and data management community here is the need of **coordination** at all levels, i.e. digital archives, e-infrastructure and services, virtual research environments. The objective is to prevent a *"many-headed beast"* landscape of digital resources and services. Proliferation of many and very likely not sustainable solutions would reduce the capability to provide shared data resources and services.

The situation is critical as the level of funding for the humanities and cultural heritage research, despite their recognised importance, is rather low compared to the natural and life sciences. Therefore available funds must be invested in a highly targeted and coordinated manner. Regular "spending reviews" are required to allow smart investments and optimize the impact of allocated funds.[19] Indicators of quality, effectiveness, and success with the user communities are necessary for targeted funding of sustainable resources and planning of new resources the research community may need.

3.2.4 Cooperation and synergies between e-infrastructures

In order to build a coherent and sustainable ecosystem of e-infrastructures coordination and synergies need to be sought between major initiatives. Through the ARIADNE e-infrastructure and services the archaeological sector in Europe can acquire a seat at the table of such initiatives. Indeed, one goal of ARIADNE is to make the sector ready for participation in programmes and concerted actions on the level of e-infrastructure and service providers in Europe and beyond.

The e-Infrastructure Reflection Group (e-IRG), the main advisory body for European e-infrastructures, asks for more coordination among all stakeholders to increase coherence and resource sharing. This is required to enable synergies, cost-effectiveness and sustainability in supporting innovative ICT-enhanced research within and across disciplines.

The immediate tasks of ARIADNE are to set up an e-infrastructure and services (e.g. dataset catalogue, data discovery and access) and support data providers in connecting their archaeological data resources. In the first place this is about building common, community-level resources, ensuring proper "housekeeping" with regard to data standards as well as organisational aspects of e-infrastructure governance.

Established governance then can allow ARIADNE to engage in

- Coordination with relevant other e-infrastructure initiatives to build mutual understanding with regard to objectives and development priorities, remove gaps in common knowledge, etc.;
- Cooperation on defined joint activities such as tackling current barriers or constraints to resource sharing, combination and integration.

Compared to such direct and targeted activities among e-infrastructure, data and service providers we see little value in a proliferation of interest and working groups around all matters of open data sharing and access. For example, many of the groups accepted by the Research Data Alliance[20] appear to be not connected to data archives and e-infrastructures that are being put in place to allow for data sharing and access.

[19] Cf. PARTHENOS (2015): It's a start. Voices and impressions from the PARTHENOS kick-off meeting. PARTHENOS website, http://www.parthenos-project.eu/kick-off-article-tbu/

[20] Research Data Alliance (RDA), Working and Interest Groups, https://www.rd-alliance.org/groups; at present archaeology is not present in the RDA; there is only one active humanities interest group: Digital Practices in History and Ethnography, https://rd-alliance.org/node/508

With regard to e-infrastructures, data resources and service provision cooperation will be beneficial in three areas:

Strengthen cooperation among e-infrastructures for heritage sciences

In the first place cooperation targeted at enabling a coordinated development, interoperability and synergies among e-infrastructures for the heritage sciences is required. The heritage sciences span all research on cultural heritage as conducted in the humanities and research involving the use of methods of other disciplines, e.g. for material analysis and conservation. The heritage sciences include of course archaeology and archaeometry.

Cooperation among the relevant e-infrastructures is required to enable interoperability, sharing of resources, and integrated access of users to data, services and tools. Also digital archives ask for a closer collaboration of e-infrastructures as they want to see the content/data they curate being used in new, ideally cross-disciplinary research. As one statement of an archive participating in ARIADNE: "*To promote cross-disciplinary research, create collaborations between different infrastructures and create synergies through this, e.g. by promoting use of tools developed or offered within CLARIN or other infrastructures"*.

Addressing the issue of lack of e-infrastructure interoperability, Lossau (2012) in an article for the European research libraries community (LIBER) stresses: *"Creating seamless access to publiccations, research data and cultural heritage material or indeed supporting interdisciplinary research will never become a reality if interoperability at various levels is not addressed properly. How do the DRIVER-OpenAIRE guidelines relate to the data models of the ESFRI infrastructures, or to those of the Europeana Data Model (EDM)? How can researchers not only search and access data but also automatically export large quantities of data from RIs to their own domain or theme-specific virtual research environment? What ICT-services are required to support this usage scenario and what should the terms and conditions be for those services?"* (Lossau 2012).

Coordination of activities among European infrastructures for heritage sciences has already started with the PARTHENOS project[21] that aims at sharing resources (services, tools and data) as required in domain and cross-domain research. PARTHENOS involves members of the humanities infrastructures CLARIN[22], DARIAH[23] and several research organisations. A regular exchange between the relevant European infrastructure projects exists through dedicated conferences.[24]

Forging closer ties may be necessary in some areas, relevant natural and life sciences centres and facilities, for instance. One ARIADNE partner suggested: *"Be in contact with other research infrastructures like DARIAH and CENDARI in the sector of Arts and Humanities, and facilities belonging to the so called 'hard sciences' that offer innovative diagnostic laboratories tools and manage huge databases of scientific data, e.g. ELETTRA Sincrotone di Trieste*[25]*, CHNet of INFN*[26] *or the LAMS Laboratoire d'archéologie moléculaire et struc*tural[27]".

[21] PARTHENOS - Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies (EU, H2020, 05/2015-04/2019), http://www.parthenos-project.eu; PARTHENOS and ARIADNE have the same coordinator, professor Franco Niccolucci (PIN Vast-Lab, Italy).

[22] CLARIN - Common Language Resources and Technology Initiative / Infrastructure, established as a legal entity (ERIC), http://www.clarin.eu

[23] DARIAH - Digital Research Infrastructure for the Arts and Humanities, established as a legal entity (ERIC), https://www.dariah.eu

[24] For example, Research Infrastructures and e-Infrastructures for Cultural Heritage, Rome, 13-14 November 2014; European Research Infrastructures for Heritage Science, Florence, June 30 to July 3, 2015.

[25] ELETTRA Sincrotone di Trieste, https://www.elettra.trieste.it

[26] Istituto Nazionale di Fisica Nucleare (INFN), CHNet, http://infnbeniculturali.net

[27] LAMS - Laboratoire d'archéologie moléculaire et structural, http://www.umr-lams.fr/?lang=fr

The Association of European-level Research Infrastructure Facilities (ERF-AISBL)[28] unites 15 major research infrastructures / facilities in this field, including the Central European Research Infrastructure Consortium (CERIC-ERIC)[29] with facilities in nine countries. Common policies for example concerning how to approach the open data challenge are a major concern of the association (cf. ERF-AISBL 2014). The IPERION-CH project[30] develops a distributed European research infrastructure for heritage science, restoration and conservation.

Establish cooperation with other e-infrastructures and services that cover relevant data of related research domains

Archaeology is a multi-disciplinary field of research in which researchers need information and datasets of different domains of research. Indeed, archaeological researchers address questions about past cultures and environments that require data in the micro to macro range, i.e. from micro-remains (e.g. molecular biology data) to landscapes and beyond (e.g. terrestrial and airborne or satellite remote-sensing data).

Therefore one cooperation objective for ARIADNE is enabling discovery and integrated access also of resources that are aggregated and made accessible by other domain e-infrastructures. Relevant data resources of (sub-)domains of the environmental, biological and social sciences might be considered. To provide but one example, the earth & environmental sciences avail of major digital archives and services like PANGAEA or the NASA Global Change Master Directory (GCMD), which include relevant data.[31]

Cooperation on data access and interlinking is required as no single e-infrastructure may be capable to cover all information/data resources needed by the multi-disciplinary archaeological research communities. However, a major requirement for the ARIADNE e-infrastructure in this context is capability to handle a variety of metadata and semantics of scientific datasets.

Link up with providers of Distributed Computing Infrastructure (i.e. Grid/Cloud) resources

The European Research Area has many scientific Distributed Computing Infrastructures (DCIs) and the providers welcome research communities that are not yet avid users. The EGI Foundation (EGI.eu) coordinates a pan-European platform of distributed computing infrastructures (incl. the EGI Federated Cloud) and their users. The EGI-Engage project (2015-2017) promotes open science commons and supports research communities with technical and data solutions.[32]

Archaeological research could benefit from DCI services in some cases, e.g. when processing large 3D objects (e.g. landscapes), stocks of images, or complex models of cultural change, e.g. adaption to gradual or disruptive environmental or other changes. One ARIADNE partner suggested linking up with DCI providers as they *"make available to researchers and scholars – today, especially in the so-called hard sciences – simple and controlled online access to services, resources, and collaboration tools. It harnesses the power of ICT for computing, connectivity, storage, long-term preservation, and instrumentation. Indeed, a basic assumption is that the existence of e-infrastructure for research and academia is also an efficient channel for providing efficient services for the archaeology and more in general for the cultural heritage field".*

[28] Association of European-level Research Infrastructure Facilities (ERF-AISBL), http://www.erf-aisbl.eu

[29] Central European Research Infrastructure Consortium (CERIC-ERIC), http://www.ceric-eric.eu

[30] IPERION-CH - Integrated Platform for the European Research Infrastructure on Cultural Heritage (H2020, 2015-2019, http://www.iperionch.eu

[31] PANGAEA - Data Publisher for Earth & Environmental Science, http://www.pangaea.de, includes about 3100 datasets related to "archaeology"; the Global Change Master Directory (GCMD), http://gcmd.nasa.gov, covers more than 34,000 earth & environmental data set and service descriptions including, for example, World Data Center for Paleoclimatology and other relevant datasets.

[32] EGI Foundation, http://www.egi.eu

Use of Distributed Computing Infrastructure services can be encouraged by bringing together archaeological research groups with providers of DCI, especially with their developers of data processing software and services. We assume that currently only few archaeological research groups can use DCI resources like Science Gateways, virtual machines or workflow-based execution of data processing services "out of the box". Therefore guidance and active knowledge transfer by experienced developers is required. Only few projects have so far introduced cultural heritage institutions and researchers to DCI offerings, one example is Digital Cultural Heritage Roadmap for Preservation (DCH-RP).[33]

3.2.5 Suggested actions

In the area of e-infrastructure development and cooperation the following actions are suggested:

Recognise the importance of e-infrastructure development and coordination

Research e-infrastructures and services are important pillars and drivers of collaborative and data-intensive research. They provide access to distributed, but shared digital resources (data, services, tools) for advanced and innovative research across institutional and disciplinary boundaries. Archaeology is multi-disciplinary and therefore should devote particular attention to developing integrated e-infrastructure, data resources and services.

Considerable investments have already been made at the European and national levels in implementing research e-infrastructures for different disciplines. But the e-Infrastructure Reflection Group (e-IRG), the main advisory body for European e-infrastructures, notes insufficient coordination among the existing e-infrastructures, and asks all stakeholders to increase coherence and resource sharing. This is required to enable synergies, cost-effectiveness and sustainability in supporting innovative ICT-enhanced research within and across disciplines.

The sector of archaeological and other heritage research should prevent a proliferation of unsustainable e-infrastructure initiative and ensure that available funds are invested in a highly targeted and coordinated manner.

Support the ARIADNE data infrastructure and portal and help mobilise many data providers

ARIADNE is the core EU-funded Integrating Activity project in the field of archaeology with the primary objective to help overcome the fragmentation of archaeological data resources. ARIADNE is an initiative to enable this field open sharing and re-use of data as needed for progress and innovation in archaeological research, and increasingly demanded by funding bodies. The immediate tasks of ARIADNE are to set up a state-of-the-art e-infrastructure and services (e.g. data registration, discovery, access and other services) and to support data providers in connecting their resources. All stakeholders in accessible archaeological data should support ARIADNE data infrastructure and portal and help mobilise data providers.

Ensure sustainability of the ARIADNE data infrastructure and portal

ARIADNE will provide the archaeological sector with a common, European-level data infrastructure and portal. The solution will require sustained efforts to allow for maintenance and extension, e.g. mobilisation of data resources of all European Union Member States, integration with relevant other European data infrastructures, among others.

[33] DCH-RP - Digital Cultural Heritage Roadmap for Preservation - Open Science Infrastructure for Digital Cultural Heritage in 2020 (EU FP7 project, 10/2012-09/2014, coordinated by ARIADNE partner MiBACT-ICCU), http://www.dch-rp.eu

Strengthen cooperation among e-infrastructures for heritage sciences

There is already a good level of information exchange between the existing European e-infrastructures for heritage sciences, e.g. ARIADNE, CLARIN, CENDARI, DARIAH, IPERION-CH, and others. But regular cooperation targeted at enabling a coordinated development of and synergies among the e-infrastructures is necessary. Coordination is required to ensure interoperability, common policies, sharing of resources, and integrated access of users to open data, services and tools as required by the heritage sciences. Also preventing duplication of efforts with regard to the mobilisation of specific data resources may be necessary. Furthermore cooperation on training and professional development of e-infrastructure and data service managers can be recommended.

Establish cooperation with other e-infrastructures and services that cover relevant data of related research domains

Archaeology is a multi-disciplinary field of research in which researchers need information and datasets of different domains of research. Indeed, archaeological researchers address questions about past cultures and environments that require data in the micro to macro range, i.e. from micro-remains (e.g. molecular biology data) to landscapes and beyond (e.g. terrestrial and airborne or satellite remote-sensing data). Cooperation on data access is required as no single e-infrastructure may be capable to cover all information/data resources needed by the multi-disciplinary archaeological research communities. One cooperation objective for the ARIADNE initiative here is enabling discovery and access also of resources that are aggregated and made accessible by other domain e-infrastructures. Cooperation on data access will also create the setting for knowledge transfer, cross-fertilization and interdisciplinary approaches.

Link up with providers of Distributed Computing Infrastructure and encourage use of their resources by archaeologists

The European Research Area has many scientific Distributed Computing Infrastructures (DCIs) and the providers welcome research communities that are not yet avid users. The EGI Foundation (EGI.eu) coordinates a pan-European platform of distributed computing infrastructures (incl. the EGI Federated Cloud) and their users. The EGI-Engage project promotes open science commons and supports research communities with technical and data solutions.

Archaeological research could benefit from Grid/Cloud-based services of DCIs in some cases, for example, handling and processing large stocks of data objects (“big data”), detailed 3D objects (e.g. landscapes) or complex models of cultural change (e.g. adaption to gradual or disruptive environmental changes). Use of DCI resources can be encouraged by bringing together archaeological research groups with providers of DCI, especially with their developers of data processing software and services. We assume that currently few archaeological research groups are familiar with DCI resources and methods like science gateways, virtual machines or workflow-based execution of data processing services. Therefore guidance and active knowledge transfer by experienced developers is required.

3.3 Requirements for transformative innovation through research e-infrastructures

Research e-infrastructures are understood as one factor for potential innovation in research practices. It is expected that the ARIADNE e-infrastructure and services can help overcome the current fragmentation of archaeological resources as well as promote transformative innovations in archaeological research. However such innovations depend on many factors and requirements. Below we discuss perceived requirements for innovations through e-infrastructure in the humanities, and archaeology specifically.

3.3.1 Taking account of disciplinary differences

Researchers in the humanities perceived the initial visions of research e-infrastructures as overly technology-centred ("technology-push" model). These visions were mainly informed by examples of advanced scientific computing as used in (some) natural sciences and did not resonate well with the situation in the humanities. For instance, the first report of the US National Science Foundation's Advisory Panel on Cyberinfrastructure (NSF 2003, known as the Atkins report) and subsequent ventures were received with strong reservations by the humanities sector, e.g. the "Our Cultural Commonwealth" report of the Commission on Cyberinfrastructure for the Humanities & Social Sciences of the American Council of Learned Societies (ACLS 2006).

According to the Atkins report, cyberinfrastructure encompass grids of computational centers, comprehensive libraries of digital objects, well-curated collections of scientific data, online instruments and vast sensor arrays, convenient software toolkits, and more. The ACLS responded: *"Humanities scholars and social scientists will require similar facilities but, obviously, not exactly the same ones: 'grids of computational centers' are needed in the humanities and social sciences, but they will have to be staffed with different kinds of subject-area experts; comprehensive and well-curated libraries of digital objects will certainly be needed, but the objects themselves will be different from those used in the sciences; software toolkits for projects involving data-mining and data-visualization could be shared across the sciences, humanities, and social sciences, but only up to the point where the nature of the data begins to shape the nature of the tools"* (ACLS 2006: 8; cf. Wouters 2007, who recognises the ACLS report as *"a very convincing delineation of what digital scholarship is about"*).

The widely referenced *"Riding the wave"* (2010) report of the European High level Expert Group on Scientific Data mentions only one humanities example, the CLARIN e-infrastructure for language resources and tools, *"already tackling proper data management as a key dimension of the system for the scholarly community"* (ibid. 15). Elsewhere, under the concept "virtual lab", the report notes: *"Researchers with widely different backgrounds – from the humanities and social sciences to the physical, biological and engineering sciences – can collaborate on the same set of data from different perspectives".* Next the report goes into data-driven research: *"Indeed, we begin to see what some have called a "fourth paradigm" of science – beyond observation, theory and simulation, and into a new realm of exploration driven by mining new insights from vast, diverse data sets."* (ibid. 8; the highly influential book "Fourth Paradigm" has been edited by Hey *et al*. 2009).

Also in other reports major differences between the natural sciences and the humanities as well as social sciences are not recognised, instead they seem to play the same game as showcase "big data" disciplines like astronomy and genetics, only not yet in the same league. Concerning the social sciences the most prominent example is mining online social media whereas for the humanities convincing examples are missing (some first examples are mentioned below).

The ACLS report emphasises *"the intractability, the rich ambiguity, and the magnificent complexity that is the human experience"* (ACLS 2006: 8). This summarises well major differences between the humanities and the natural sciences with regard to research subjects and content. These differences include, but are not limited to, natural vs. socio-cultural systems, i.e. behaviour dictated by immutable natural laws vs. intentional and changing human goals, values and norms; determined manifestations of natural phenomena vs. creative, meaningful but ambiguous human agency and expression, as present in the archaeological record, for instance.

These essential differences suggest that the humanities and social sciences may need specific research e-infrastructures and tools which can support their different methods, models and forms of reasoning. Hermeneutic and other interpretive methodologies play a major role in these disciplines, but there is also a mass of research based on systematic comparative and statistical methods present in these disciplines.

Furthermore we think that an overemphasis on subjective interpretation can be misguiding as it is too much focused on individual scholarship and, consequently, digital tools that might somehow assist, but not constrain, individual acts of interpretation. Rather, what is needed appears to be a re-evaluation of scholarship taking account of the opportunities offered by increasing volumes of accessible content/data and ICT-enabled research.

Wouters (2007) praises the *"Our Cultural Commonwealth"* (ACLS 2006) report for having put on the agenda a major challenge, *"the building of tools that support new ways of working in the humanities",* especially with regard to computer-assisted interpretation: "*Interpretation is at the heart of the humanities, and we are still far from the situation that computer-assisted interpretation is common in the humanities or qualitative social sciences. We need new tools for finding patterns in large data sets and collections (heuristics) as well as tools that support the interpretation (hermeneutics). The ACLS report is, I think, a very convincing delineation of what digital scholarship is about: it is not only tool building, or making infrastructures available and then hope that scholars will indeed make use of them, but is the reconfiguration of the very essence of knowledge making"* (Wouters 2007). The question than appears to be how welcome a reconfiguration of current modes of interpretive knowledge making is by humanities scholars.

3.3.2 "Digital humanities" versus ICT-supported archaeology

There are substantial differences between ICT-supported archaeological and other humanities with regard to the tools required for making sense of certain "data".[34] Scholars in the so called "digital humanities" analyse mainly cultural content produced by others, e.g. literary texts, paintings, photographs, music, films, media, etc., as mostly held in collections of libraries, media archives and museums. Without online access to this cultural heritage in digital format and tools for rich annotation and interlinking of various content (and pieces thereof) the current "digital humanities" boom would not have been possible. It allows creating scholarly works based on hyperlinks, a web of sophisticated references and knowledge rather than linear textual discourse.

In contrast, archaeologists create most of their data themselves, a multitude of different data produced in surveys, excavations, laboratory analyses of physical and biological finds, etc. The array of data also includes data from specialised laboratories which serve archaeologists among other clients (e.g. synchrotron facilities or sequencing labs with regard to ancient DNA). Furthermore, data not produced but used by archaeologists (if affordable) are airborne or satellite remote sensing and imaging data.

Most "digital humanities" research is informed by a paradigm of textual studies, whereas archaeologists work with various data in the range of micro-remains (e.g. molecular biology data) to landscapes and beyond (e.g. terrestrial and airborne or satellite remote-sensing data). The data represent material remains and traces of cultures in the environment, ranging from micro-finds and parts of artefacts, to architecture and other features of individual sites, and on to cultural landscapes and beyond. Accordingly the paradigms and tools for "reading" these data are very different from textual analysis or other "digital humanities" studies of cultural content.

Material objects and traces usually are *"not among the standard resources of academic research in the humanities and social sciences. Historians are used to dealing with literary sources, and all our sophisticated methodology focuses on the investigation of these. Meanwhile, we observe the march of images into the historian's study, but physical objects are still mostly kept at arm's length"* (Sibum in Auslander *et al*. 2009: 1384). But there are some hybrids where researchers work with content (e.g. historical records) and analysis of material remains, for instance some of the Classics (Babeu 2011), Medieval History/Archaeology (Graham-Campbell & Valor 2007; Carver & Klápště 2011), and

[34] Cf. ARIADNE First Report on Users' Needs (D2.1, April 2014), pp. 34-36.

Historical Archaeology (Hall & Silliman 2006; Mehler 2013). In such cases the researchers often use the same models and tools.

Arguably more relevant are fields of research where the use of different models and tools may trigger transformative innovation. Examples are networks analyses as initially developed in the social sciences (cf. the review in Brughmans 2013; Knappett 2013; Collar *et al*. 2015)[35]; models of the bio-sciences (e.g. cultural evolution theories based on phylogenetic approaches), and human-environment interactions, also called coupled human and natural systems. The latter is seen as a particularly important field where archaeologists can contribute their long-term perspective and connect to issues in societal and environmental sustainability (cf. *Section 8.5.10*).

3.3.3 Need for different research e-infrastructure and tools

There has been much debate about the question if research e-infrastructures initially developed for the natural sciences and engineering disciplines can be adjusted to the needs of the humanities. The majority appears to hold that scholarship in the humanities affords different tools which typically do not require “heavy” e-infrastructure for massive Grid or Cloud based data processing (Svensson 2010; Wouters & Beaulieu 2006; van Zundert 2012[36]). The limited use of such data processing by humanities researchers, except by research groups that avail of large corpora of texts, supports this assumption. However, there are examples that demonstrate potential for wider usage, e.g. in the “Digging into Data Challenge” program[37]. There is also an increasing recognition that the opposition of qualitative vs. quantitative analysis, close reading of individual vs. processing of masses of texts or objects, individual stories vs. impersonal statistics graphs is unproductive (cf. Anderson & Blanke 2012).

Nevertheless, instead of processing large volumes of texts, images or numeric data, most “digital humanities” most humanities scholars will need effective tools to search & retrieve, handle, identify, describe, decipher, translate, annotate, link, compare, etc. digitised cultural objects. The variety of such scholarly digital practices is described in available schemes of so called “scholarly primitives” (e.g. Palmer *et al*. 2009; NEMO - NeDiMAH ICT methods ontology 2014; TaDiRAH - Taxonomy of Digital Research Activities in the Humanities 2014), and investigated in studies on the actual conduct and level of use in different fields of humanities research (e.g. RIN 2011a; Svensson 2010; Rutner & Schonfeld 2012).

Scholars in the “digital humanities” analyse cultural expressions (e.g. engravings, literary texts, paintings, photographs, music, films, theatre, dance, etc.) produced or documented in different media and kept in libraries, archives and museums. E-research on this cultural heritage requires access to the digitised or born-digital content and effective tools that support established as well as potentially new forms of humanities studies. The amount of “data items” of such studies and their outcome (e.g. a digital scholarly edition) is typically not “big data”, but can range from one richly

[35] A rich online resource on all related topics is https://archaeologicalnetworks.wordpress.com

[36] Most outspoken in this regard is van Zundert (2012: 1-2): *“In the middle of a small-scale-focused, multi-faceted, patched-together, interconnected, very slow but ever developing technological humanities landscape these tall big bulky structures will be waiting for a horde of uniformly behaving humanities scholars that will never come. These infrastructures will be like the disastrously wrongly planned and developed highways that connect nothing to nowhere.”* Van Zundert argues that innovation in digital humanities is better served by small scale, research-focused development.

[37] Digging into Data Challenge, http://diggingintodata.org; Williford & Henry (2012) provide an extensive analysis of the first round of funded projects (2009). Among the winners of the third round (2013) is DADAISM - Digging into Archaeological Data and Image Search Metadata (UK, AHRC/ESRC, 06/2014-12/2015), http://dadaism-did.org, aims to develop effective search and analysis procedures for large volumes of images and grey literature.

annotated cultural work to thousands of "hand-collected" objects (e.g. a scholarly edition of epigraphical material like the Epigraphic Database Heidelberg[38]).

In case of a large amount of high-resolution images also advanced e-infrastructure and tools may be required, not only for data processing but, in the first place, to store, manage, remotely retrieve, load, render, juxtaposed display etc. Then comparison, transcription, annotation and other scholarly work may start (cf. Ainsworth & Meredith 2009). Archaeologists may use the same or similar tools, and ideally the same e-infrastructure, if they archival records relevant for an excavation can be assumed to be available and can be discovered and accessed online. Also tools for comparing new finds with digitised specimen of scientific reference collections will be helpful but, unfortunately, such collections are often not available online for inclusion in research e-infrastructures (cf. *Section 9.1.2*). But some areas in the multi-disciplinary field of archaeology present distinctly different needs with regard to data, tools and services, classical studies versus environmental archaeology, for instance.

3.3.4 Archaeology as a multi-disciplinary field of research

The classical "tree of knowledge" model splits the scientific and scholarly knowledge into different branches (i.e. natural sciences, social sciences, arts & humanities), major disciplines within these branches, and the disciplines into sub-disciplines and specialties. While it is necessary to take account of differences between disciplines when building research infrastructure, archaeology poses a challenge because it is a multi-disciplinary field of research, in which theoretical concepts, models and methods from different branches of knowledge are being used.

In a volume of papers on "Archaeology 2.0", Eric C. Kansa notes: *"Archaeology is an inherently multidisciplinary enterprise, with one foot in the humanities and interpretive social sciences and another in the natural sciences. As such, case studies in digital archaeology can help illuminate changing patterns in scholarly communications across a wide array of disciplinary contexts."* (Kansa 2011: 2) In other words, archaeology comprises of different disciplinary contexts in that some fields of research present mainly characteristics of the humanities (e.g. history of arts, classical studies), while others lean heavily towards the physical sciences and employ methods of archaeometry, some build on knowledge and methods of biology and other life sciences, others relate to the earth & environmental sciences, and still others use models and methods of the social sciences, models of social structure and change or ethnological methods, for instance.

Thus different areas of archaeology address their research questions with concepts, models, methods and/or data from a wide spectrum of disciplines, with regard to the natural sciences indeed ranging from genetics to astrophysics. To illustrate the case, the figure below is an extract of a larger network of 1554 journals that were co-cited in the Web of Science Web Core Collection index more than 35 times in 10,639 archaeology-related publications between 2004 and 2013 (Sinclair 2014). Beside "archaeological" clusters (e.g. Method and theory, Early Prehistory, Mediterranean and Near Eastern, African, Pacific and Australasian archaeology, Archaeological prospection and remote sensing), the author identified related clusters of publications in Geoarchaeology and earth sciences, Scientific analysis of materials, Life sciences, medical sciences and forensics, and Astrophysics.

[38] Epigraphic Database Heidelberg (begun 1986, launched online 1997, currently at about 24,000 find spots, 68,700 inscriptions, 34,000 photographs [up from about 20,000 in 2009]), http://www.epigraphische-datenbank-heidelberg.de; Cayless *et al*. (2009) note that *"the majority of new epigraphic editions are still published solely in print, but by 2017 we believe this circumstance will change drastically".* Porter (2013) presents survey results on the use of scholarly digital editions by medievalists.

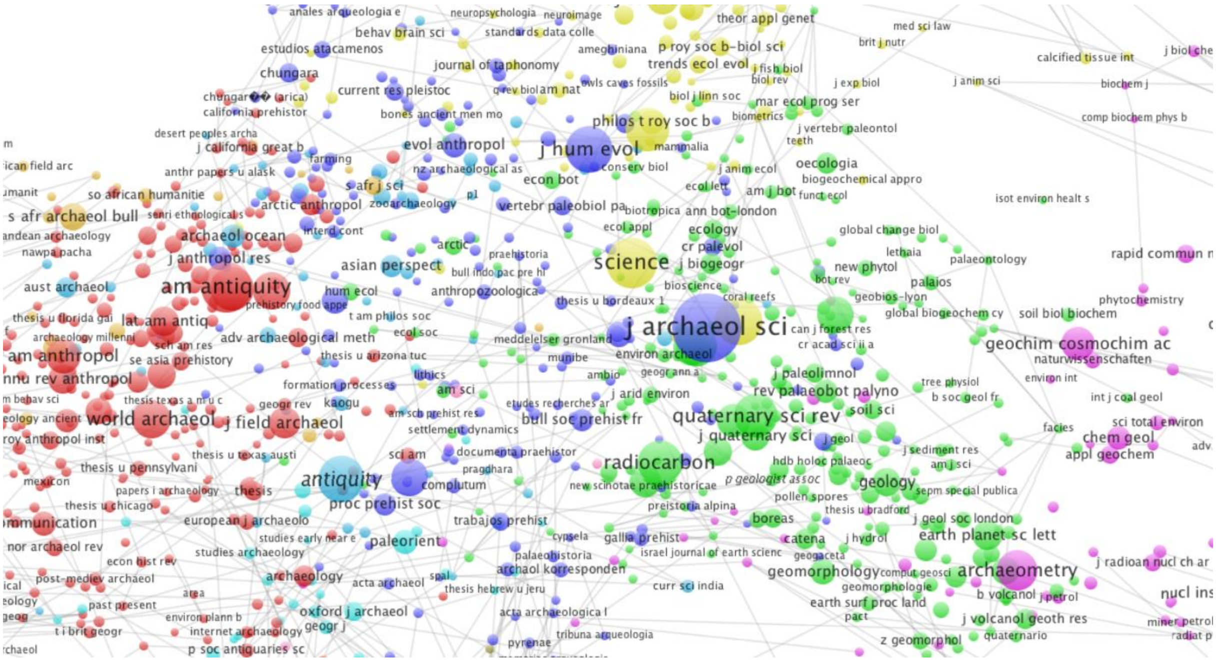


Figure 4: Presents an extract of a larger figure produced by Sinclair (2014), based on often co-cited journals in over 10,000 archaeology-related publications between 2004 and 2013.

The main anchor and integrating force of archaeology arguably is the primacy of the cultural interpretation of "the archaeological record". While this primacy seems clear, there has been an increasing impact of the use of natural sciences methods, the "radiocarbon revolution" for instance (Manning 2015; cf. Harding 2008 [interview with Colin Renfrew]). Some struggles were necessary for accommodating natural sciences views and results with the overall cultural paradigm of archaeology (Pollard & Bray 2007), with arguably most difficulty concerning the bio-sciences. A current example may be to accommodate properly results of ancient human DNA analyses (see for example Di Cosmo 2011; Hofmann 2015).

The website of the ArchLab Consortium of Swedish Archaeological Research Laboratories excellently describes how in archaeological studies different, multi-disciplinary fields and methods contribute to solving research questions: *"Archaeology studies human activities, driven by cultural and natural causes, environment, climate and the feedbacks between these systems. Natural science methods are fundamental to the analysis of archaeological source materials, deposits and related contexts. Archaeobotany, for example, requires elements of biology (taxonomy, identification) and ecology (habitat, environment), alongside ethnobotany (plant use) and traditional archaeological methods. In turn, archaeology requires geology (chronology, stratigraphy, evolution of landforms), geography (landscape context) and behavioural and social sciences (explaining human actions). Radiocarbon dating owes its theory and method to physics and chemistry, but an understanding of archaeological contexts is essential for sampling and interpreting results. Empirical data and theoretical frameworks are intertwined in the quest for understanding the past, both as an object of study in itself, but also as an aid to understanding the human condition and predicting our possible futures"* (ArchLab website, http://archlab.se).

Degryse & Shortland (2013) highlight that "*there are very few sciences that have no relevance to archaeology"*, and see an increasing influx of scientific methods in archaeological research. Once primarily used in research on prehistoric sites and finds, the application of such methods has become much more widespread, if not standard practice. This has also impacted on the traditional, hierarchical organisation of archaeological work, interpretation and publication. As David Killick elaborates: *"The explosive growth of new scientific methods for archaeology over the last twenty years is changing archaeological practice. The old model of site excavation resembled a wagon wheel,* *with the principal archaeologist (site director) at the hub, a ring of specialists around the rim, and the spokes of the wheel representing the separate two-way communication between the site director and each specialist. Today the specialists often need to communicate directly with each other to do their job – for example, the archaeometallurgist needs the geologist to identify ore samples, the botanist to identify charcoals, the lead isotope specialist to distinguish local from exotic metals. Synthesis and interpretation today tends to be made by teams rather than by the site director. Trends in publication reflect this tendency; scientific studies now tend to be published separately rather than included as dry appendices in the site monograph, as used to be the case”* (Killick 2015: 170).

Turning back to Sinclair’s clusters of co-citing journals above, these clusters do not present autonomous, self-contained areas of archaeological research. Rather, they show that scientific studies are published in, and more frequently co-cited by other papers, in relevant journals of the respective fields (e.g. bio-archaeology or geo-archaeology). But the research will often have been conducted on the same site or area on which other researchers, using other methods, have published in other journal clusters.

This raises several questions: One is how this distributed knowledge could be integrated and synthesized not in “canned” site monographs, but open, digital, rich-media and interactive environments, including cross-site comparative analysis and synthesis. With regard to research e-infrastructures and services this relates to questions about the support they can give to integrative research in the multi-disciplinary field of archaeology. In a discussion about the question if “digital archaeology” is a self-contained sub-discipline of archaeology, Bernardou (2015) suggests, *“Digital research infrastructures should look at Archaeology - Digital or non - as a set of numerous hardly self-contained sub-disciplines; Archaeological research sub-communities are not autonomous or self-sufficient. Rather, they are co-dependent, interdisciplinary and largely inhomogeneous. Only those digital research infrastructures which address this granularity in sub-communities of practice and their respective user needs are to be meaningful to researchers”*.

We generally agree to this view, because different areas of the multi-disciplinary map of archaeological research will present distinctly different needs with regard to data, services or tools, classical studies versus environmental archaeology, for instance. Data infrastructures such as ARIADNE provide services required by all areas such as data registration, discovery and access for re-use. Additional specific services/tools or virtual research environments (VREs) can be built on top of data infrastructures. They typically will be area-specific, e.g. support in building a scholarly edition of classical texts or, in the case of environmental archaeology, for example allow aggregation, visualization and analysis of pollen data. The further question then is if there is any need of the research community for integrative virtual research environments.

3.3.5 Conflicts about proposed e-infrastructure solutions

Multi-disciplinary areas of research like Archaeology must accommodate different disciplinary paradigms of research, different perspectives on research objects, different types of data etc. Novel e-science environments may allow researchers tackling new or unsolved questions in novel ways. Yet the current push for building such environments of integrated tools and data resources is technology-driven whereas epistemic and methodological issues are side-lined. Contrary to this “technology-push” we advocate “research-pull”, actual demand stimulated by scientific controversy over questions that are critical for advancing cross-disciplinary fields of research, setting a common research agenda and design the required research environment.

The “technology-push” approach typically offers an array of generic tools and services to be “customized” for the research community. But the research questions are primary and the design of the e-research environments, development of tools, acquisition of data, etc. secondary, although this may be challenging as well. It is not about adapting research questions to tools, but challenge developers of scientific tools to come up with tools for questions that are vital for advancing knowledge and potential breakthrough insights.

Establishing an innovative research agenda, i.e. defining the questions to address, collaboration to forge, resources to mobilise, etc. can mean conflicts which arise from different views of the direction in which a field of research should go as well as vested interests in established theories, ways of working, methods and tools in use. As Paul Wouters suggests, *"conflicts about innovation in humanities methods and ways of working are in themselves very interesting and strategic research sites and moments. The same holds for non-use of tools and infrastructure or reluctant use. Studying them as moments in which new scholarly identities are being forged is I think a promising research line"* (Wouters 2007).

E-infrastructure initiatives should therefore be accompanied by studying any conflicts about innovation that may arise:

- Very likely conflicts will first surface at the level of existing data which should be mobilised and shared. "Data" or, more precisely, the anxieties and tensions the request for data sharing occasions, represent *"the front line of cyberinfrastructure development"*. Tensions and conflicts in research e-infrastructure *"are very often played out and resolved (or not) at the level of data"* (cf. Edwards *et al*. 2007, 32-34)[39].
- Second, concerning semantics, the common terminology and conceptual / ontological commitments which should allow for integration and reasoning over of interlinked data (where the linking represents a formalized discourse of statements about conceptual relations and facts). For instance, Limp (2011: 278) argues that *"the reward structure in archaeological scholarship provides a powerful disincentive for participation in the development of semantic interoperability and, instead, privileges the individual to develop and defend individual terms/structures and categories".*
- Third, many conflicts can also be expected due to vested interests in technical systems that are already in use and support established research practices. Fit or non-fit with these interests, systems and practices will determine if stakeholders/users are willing and capable to adopt novel e-infrastructure components. Non-fit will certainly face resistance and little chance that proponents of assumed better technology can prove their case.

Thus there is much scope in research e-infrastructure development for socio-technical systems research with a focus on conflicts and how they are resolved (or not).

3.3.6 Fostering collaboration and embedding e-infrastructure in research practice

Particularly important in the development of e-infrastructure, tools and services is common ground of technical developers, researchers and data managers. Successful development and implementation of research e-infrastructures requires their mutual understanding and close cooperation. This requirement has been emphasised especially by scholars in the humanities (e.g. Anderson *et al*. 2010; Borgman 2009; Svensson 2011; Wouters & Beaulieu 2006). Common interest, perspective and language of technology developers and users cannot be assumed, but must be built in close cooperation. This has been observed not only in the development of e-infrastructures for the humanities, but also in very different disciplines such as astronomy (Gray & Szalay 2004) or marine

[39] As the authors note: *"From one view at least, cyberinfrastructure is principally about data: how to get it, how to share it, how to store it, and how to leverage it into the major downstream products (knowledge, discoveries, learning, applications, etc.) we want our sciences to produce. It should come as little surprise then that the flat-sounding word 'data' stands at the center of some of the most vexing tensions confronting the development of cyberinfrastructure."* (Edwards *et al*. 2007: 32-34)

sciences (Lee *et al*. 2010). The requirement must be considered also in domains where researchers are more likely to have the technical background and language to collaborate with technical specialists.

Research e-infrastructures are not primarily about technology but research practices supported by required information and communication systems and tools. The use of the e-infrastructure and services must become embedded in the research culture to ensure relevance and wide uptake in the day-to-day work of the researchers. This requires taking into account the institutional, organisational, educational, legal and other non-technical aspects which determine the willingness and capability to adopt and use novel tools and services. (Borgman 2007; Lee *et al*. 2006; Procter *et al*. 2013) The adoption will be a process of "co-evolution" in which potential technical benefits and disciplinary requirements are successively evaluated and appropriated by the research community; therefore evolutionary rather than radical, disruptive change can be expected (cf. Voss *et al*. 2009 and 2010).

3.3.7 Overcoming current barriers to adoption and support

The "Our Cultural Commonwealth" report notes several constraints that must be overcome to enable "cyberinfrastructure" to play a significant role: *"insufficient training, outdated policies, unsatisfactory tools, incomplete resources, and inadequate access. These constraints are not primarily technological but, instead, cultural, economic, legal, and institutional"* (ACLS 2006: 3; cf. Green & Roy 2008). Observers of the use (and non-use) of ICT-supported research indeed perceive that humanities researchers show resistance to adopt e-research methods not because of "conservative" attitudes, but due to lack of easy access to digital content in their fields of study, effective and sustained tools, and technical support (cf. Anderson *et al*. 2010; Borgman 2009; Wouters 2007)

Furthermore in the humanities the recognition of digital products (e.g. research tools/software, database building, digital scholarly editions) compared to academic papers or the scholarly monograph is perceived as low (specifically with regard to archaeology see Harley *et al*. 2010b). If digital products are not acknowledged as equally valuable scholarly contributions (e.g. in decision on tenure and promotion) there is little hope for more scholars engaging in their creation.

The move towards making "open data" a first-class citizen in the research enterprise will also necessitate a re-distribution of rewards, away from "high-impact papers" (or the "scholarly monograph") to research databases and software which enable the expected new insights generated through "data-driven" science. It must also be considered that skilled data managers/curators are needed at all levels, in research projects, institutional and community repositories, and research e-infrastructure and services.

This requires capacity building and training as well as career paths for such professionals. While the novel category of "data scientists" may fit into academic/scientific frames, the much needed data managers/curators will be in a difficult position if appropriate recognition and clear paths of career advancement are not offered. As one interviewee in the evaluation of the UK e-infrastructure for research and innovation strategy noted: *"The career structure for those people with expertise is miserable, because the number of places they can work at is not large, and the universities don't treat them particularly well as key staff"* (RIN 2010b: 21).

There is an increasing demand for a skilled workforce of data curators (Palmer *et al*. 2013; Swan & Brown 2008; Weber *et al*. 2012b), hence more capacity building and training needs to be provided. A training needs survey conducted by the DigCurV project[40] in July/August 2011 received 454 valid responses from 44 countries, mostly from Europe. The survey included lists of expected curation skills of which almost all were regarded as relevant. But among the general skills beside an affinity for

[40] DigCurV - Digital Curator Vocational Education project (EU, Leonardo da Vinci, 01/2011-06/2013), http://www.digcur-education.org

working with technology the ability to communicate to and collaboration with others were the most prized in digital curation staff (Gow & Molloy 2014).

Data curators are needed as active agents in the e-infrastructure ecology based on understanding the interdependencies involved in curating and providing access to research data, i.e. the interrelatedness of stakeholders, organisational arrangements, technologies, and data standards and resources. Data curators are part of the human element of e-infrastructures, supporting established and newly emerging forms of scholarly work, discipline-specific data practices, including domain vocabularies, and are increasingly important for the success of individual research projects, e.g. designing and executing effective data management plans. At present, however, data curators are in short supply for example in university-based research libraries and repositories aspiring to take on responsibilities in research data management (cf. *Section 6.2.3*).

3.3.8 Suggested actions

Research e-infrastructures are expected to play a significant role in enhancing established research practices as well as bringing about transformative innovation in practices. Such enhancements and innovations will impact on the advancement of knowledge and its relevance for addressing societal, environmental and other matters. But several general as well as domain-specific factors and requirements must be considered so that e-infrastructures can become drivers of innovation in the humanities, and archaeology specifically. Below we highlight factors and requirements that are likely to foster innovation through e-infrastructures for the heritage sciences, i.e. archaeological, cultural heritage and related other humanities research. Several specific actions are suggested.

Take account of disciplinary differences

Researchers in the humanities perceived the initial visions of research e-infrastructures as overly technology-centred ("technology-push" model). These visions were informed by examples of advanced scientific computing as used in (some) natural sciences and did not recognise major differences between those sciences and the humanities. But e-infrastructure for ICT-enabled scholarship in the humanities has to take account of affordances which are different because of the differences in research subjects, methods, content and required tools, for instance. Therefore research e-infrastructures developed for natural sciences (e.g. Grid-based distributed computing) may not simply be adjusted to the needs of researchers in the humanities.

Specific suggestions: Instead of attempting to transfer what seemingly works in other domains, user-centred development of e-infrastructures, data resources and tools for archaeological research must be ensured. Involvement of the intended domain users will more likely ensure that requirements according to research subjects, methods, tools and services are met.

Address the complex case of archaeology

Archaeology poses a challenge because it is a multi-disciplinary field of research, in which theoretical concepts, models and methods from different branches of knowledge, the humanities, natural and life sciences are being used. One major factor that distinguishes archaeology from other humanities is that archaeologists create most of their data themselves, in surveys, excavations, laboratory analyses of physical and biological finds, etc. Other humanities depend much more on digitised collections of libraries, archives and museums. They apply methods such as textual and visual studies whereas the analysis of material remains of past cultures often requires natural sciences methods (e.g. archaeometry). There are of course hybrids where researchers study historical content and material remains (e.g. medieval archaeology/history).

Specific suggestions: Different areas of the multi-disciplinary map of archaeological research present distinctly different needs with regard to data, services or tools, classical studies versus environmental archaeology, for instance. Data infrastructures such as ARIADNE provide services required by all areas such as data registration, discovery and access for re-use. Additional specific services/tools or virtual research environments (VREs) can be built on top of data infrastructures. They are typically area-specific, e.g. support in building a scholarly edition of classical texts or analysis of aggregated environmental archaeology datasets (e.g. vegetation distribution maps).

Foster common ground and embed e-infrastructure in research practice

The development of e-infrastructure, tools and services requires fostering mutual understanding and shared perspectives of technical developers, researchers and data managers. Such common ground cannot be taken for granted but must be built in close collaboration. Research e-infrastructures are not primarily about technology but research practices supported by required services and tools. New technologies are only successful if they become embedded in the research culture and co-evolve with this culture. This requires taking account of the institutional, organisational, educational, legal and other non-technical aspects which determine the willingness and capability of the research community to adopt and use novel e-infrastructure, services and tools.

Specific suggestions: Require close cooperation of researchers, data managers and technical developers in the development of innovative e-infrastructure, tools and services. While this is more likely to produce useful results, consider and address any non-technical aspects which may impede their actual adoption and use.

Help overcome current barriers to adoption and support

There are several constraints that must be overcome to enable e-infrastructures to play a significant role in the humanities. These constraints pertain to lack of easy access to required data/content, effective and sustained tools, training offers and technical support. Furthermore, the priority of publications and low recognition of digital products keeps researchers disinclined to make data resources available or maintain research tools/software. Therefore the move towards open science practices will also necessitate a re-distribution of rewards, away from “high-impact papers” (or the scholarly monograph) to research databases and software.

Specific suggestions: Remove conditions that impede effective e-infrastructures, in particular, lack of data sharing for open access. Highlight the vital role of data(bases) and software in producing research results and ensure appropriate recognition and rewards for their sharing and maintenance.

Build and retain a skilled workforce of data curators

With regard to technical and other support we emphasise the need for skilled data managers/ curators at all levels, research projects, institutional and community repositories, and research e-infrastructure and services. Data curators are part of the human element of e-infrastructures, supporting established and emerging new forms of research and discipline-specific data practices. They are also increasingly important for the success of individual research projects, e.g. with regard to the design and implementation of effective data management and access plans.

Specific suggestions: Recognise the challenge of building and retaining a workforce of research data curators, especially at the local level. This challenge should not be underestimated. Among other requirements appropriate recognition and clear career paths for data curators are necessary.

3.4 Fragmentation of digital resources and services

To support the elaboration of the ARIADNE Innovation Agenda and Action Plan, we looked in greater detail into the current landscape of digital content/data and other resources and services for archaeological and other cultural heritage research. The detailed results are included in the Annex (*Section 9.1*). The study confirms the general perception of a high fragmentation and inaccessibility of relevant data resources, as expressed by many researchers who participated in the online survey on user needs the ARIADNE e-infrastructure and services might help to address[41].

Many resources are indeed not easy to find (e.g. few are documented in existing online registries) and difficult to access or not accessible at all. But a good understanding of the situation of different resources can allow for the conception and implementation of a more targeted approach of resource development and access. The paragraphs below summarise the survey results and suggest actions for the different areas covered in the background study.

3.4.1 Common e-infrastructure and services

Concerning European-level infrastructures, there are two infrastructures for the humanities: CLARIN and DARIAH. Both have been on the European Strategy Forum on Research Infrastructures (ESFRI) Roadmap from 2006 onwards and are now legal entities (ERICs) and sustained through funding commitments of European Union Member States.

CLARIN[42] provides networked access for scholars in the humanities and social sciences to language data and tools of several research centres. The CLARIN central portal and registry of members' language resources offers several search, access and other services. DARIAH[43] addresses any kind of cultural content, historic as well as contemporary, hence none specifically. DARIAH did not implement a research e-infrastructure like CLARIN but is organised as a network of national branches and centres where most activities take place. Both research infrastructures do not serve substantially the needs of the archaeological sector. At the European level of data infrastructure and services it will be up to ARIADNE to provide the archaeological domain with a common and, hopefully, sustainable solution.

It is worth to address also Europeana[44], the main gateway to digital cultural heritage content of European libraries, archives and museums. The gateway provides search functionality on over 30 million items, which are a large part of the estimated 10% of European cultural heritage digitised as yet (ENUMERATE 2014). Europeana has not been built as a research e-infrastructure, but a recent policy document "Europeana for Research" (June 2014), issued by Member States' policy makers and sector experts, suggested *"to accelerate, subject to funding, the implementation of tools and services that will serve the needs of researchers, provide good practices of the use of the platform for research, and increase its collaboration with relevant initiatives and e-infrastructures at the European level"* (Europeana 2014).

Europeana has taken up the challenge[45], and related projects (e.g. Europeana Cloud[46]) explore requirements, analyse the available information, and aim to provide relevant tools. One major issue is that the Europeana information often is too limited for research purposes with regard to provenance / contextual information. It is also difficult to envision how text or image based analysis

[41] ARIADNE First Report on Users' Needs (D2.1, April 2014; available on the project website).

[42] CLARIN - Common Language Resources and Technology Initiative / Infrastructure, http://www.clarin.eu

[43] DARIAH - Digital Research Infrastructure for the Arts and Humanities, https://www.dariah.eu

[44] Europeana, http://www.europeana.eu

[45] Europeana Research, http://research.europeana.eu

[46] Europeana Cloud project (EU, ICT-PSP, 02/2013-01/2016), http://pro.europeana.eu/web/europeana-cloud

of a large number of the Europeana content items might be conducted effectively. Material of archaeological documentation, which some institutions have provided to Europeana, should anyway be studied based on the full contextual information as available at the institutions.

Another question is how Europeana might be used by the archaeological sector in fulfilling its public mission, e.g. ensure the presence of archaeological themes and information in the public sphere. In this regard collaboration with Europeana might be worth to explore.

Suggested actions:

- *Support the ARIADNE data infrastructure and portal, e.g. help mobilise data providers and ensure sustainability.*
- *Collaborate with Europeana with regard to making archaeological themes and information present on their platform; this concerns providers of information to Europeana but also sector institutions in general.*

3.4.2 Digital archives & repositories

In the main registries of digital repositories (e.g. OpenDOAR, ROAR, re3data) only few European archives for archaeological data can be discovered. We found three, national-level digital archives, namely the ARIADNE project partners Archaeology Data Service (UK), DANS-EASY (incl. the E-Depot for Dutch Archaeology), Swedish National Data Service (incl. archaeological deposits), and the ARACHNE IDAI.objects database (Germany).

Also other European digital resources of interest to archaeology and classical studies are registered, however only about 10 and of different types and sizes. We think that many institutions with relevant collections, which may comply with the requirements of existing registries (e.g. with regard to accessibility), did not yet consider registration as a way to make their collections discoverable.

Some examples of registered digital collections are Human Origins (University of Southampton, UK), Propylaeum-DOK (Heidelberg University Library, Germany), the document repository of the Norwegian Directorate for Cultural Heritage (about 900 items), and a few small collections such as the Acropolis Educational Resources Repository (270 items) and DIGIMOM (Maison de l'Orient et de la Méditerranée, France, 130 digitised books).

Beyond this there are many, primarily university-based document repositories registered in OpenDOAR and ROAR. Very few of them hold any data, but some contain (self-)archived publications on archaeological subjects. A search of the most comprehensive BASE - Bielefeld Academic Search Engine (over 78 million documents of 3725 repositories and other sources) on the subject heading "archaeology" yields about 387,000 records worldwide.

This illustrates the current gap between (self-)archived archaeology papers, which reside in document repositories, and research data, which should go into proper data archives (but are yet to be built in many countries). In any case, if in the future papers and research data are not being deposited in the same archive/repository, stable identifiers and linking between the paper and data records should be a prime concern.

Suggested actions

- *Register data archives/repositories and other collections for archaeological research in the catalogue of the ARIADNE data portal.*
- *The ARIADNE registry should become the prime location to document and discover available archaeological data resources.*

3.4.3 Digitised cultural heritage content

There are numerous cultural heritage collections of archives, libraries and museums across Europe that provide access to digitised content. A segment of these digital collections is included in the Europeana search catalogue. The most comprehensive survey of digitised collections has been conducted by the ENUMERATE project (2014). The survey gives a good overview of the situation at 1179 institutions. Some relevant findings concerning archaeological content are that of the collections of 548 museums (including 180 "archaeology or history museums") 24% are already digitised, 57% not yet digitised, and for 19% no need is perceived to digitise them. The collections of 239 institutions that hold mainly or a significant amount of material of/about archaeological sites, monuments and landscapes show roughly the same pattern, but a smaller percentage (18%) have already digitised content.

The ENMERATE survey also provides figures for the accessibility of the digital content of 905 institutions, however, not broken down according to different categories. Given that the situation at the archaeology and/or history museums and other institutions mentioned above roughly corresponds to the average across all institutions, we may assume: that these institutions primarily provide digital access on their own website and in-house (e.g. at computer stations). Some also participate in the Europeana initiative. There is considerably less interest in allowing service developers direct access to (meta-)data through an Application Programming Interface (API). Also there is little, but somewhat growing interest in making information available through Wikipedia or social media channels.

Suggested actions

- *Step up the digitisation and documentation of archaeological material held by archives and museums (e.g. unpublished excavations, grey literature, finds/objects).*
- *Participate in the ARIADNE initiative for federated search and access specifically for archaeological content/data collections.*
- *Implement mechanisms that allow harvesting and other programmatic access to collection metadata (e.g. OAI-based harvesting, SPARQL for Linked Data).*

3.4.4 Online scientific reference collections

The development of online scientific reference collections for archaeological purposes is not well organised at the community level. The result of this lack of coordination and collaboration is that there are many collection websites of individual institutions which are difficult to find and need to be browsed one by one to possibly find some relevant information. Some bottom-up initiatives address the issue and aim to develop an online community resource, but usually receive only few contributions.

It must be noted that since decades scientific collections are difficult to maintain, re-vitalize and extend due to shrinking budgets for their curation. In recent years reference collections have been promoted as essential research infrastructures, but major targeted investment is necessary to turn them into state-of-the-art web-based resources.

Suggested actions

- *Take stock of existing reference collections for archaeological purposes and evaluate their relevance and current condition (e.g. actual demand, requirements for online access).*
- *Secure financial support and curatorial expertise for the development and maintenance of state-of-the-art web-based reference resources.*

3.4.5 Laboratory analyses of archaeological material

In the MERIL inventory, that boosts itself as the most comprehensive documentation of European research infrastructures (RIs), we found 84 RIs for humanities and heritage sciences. A closer inspection showed that 31 RIs might provide at least some content/data relevant to archaeological research. The majority of these RIs (17) are laboratories that conduct physical, chemical and biological analyses of objects and materials for archaeologists and other cultural heritage researchers. It seems that these laboratories have been included in the MERIL inventory because they are large ones and avail of special facilities such as synchrotron radiation and accelerator mass spectrometry instruments. Such instruments are not available at typical archaeometry laboratories of university institutes and major museums of which there exist many more.

We examined the websites of the MERIL sample of laboratories to see if they provide access to documentation (metadata) and data of their archaeometrical and other analyses (i.e. may have a data repository). We found that online documentation is hard to come by, access to data is offered nowhere, and even some illustrative webpages of conducted analysis are missing in most cases. Lists of papers are typically presented, but few entries link to an open access journal paper or self-archived, pre-print copy. We assume that the situation with regard to open data at the smaller laboratories, of which there are many more, is not much better. A survey on the actual situation would be helpful to understand better factors that impede open access and how they might be removed.

The main reason for the lack of study documentation and access at the laboratory facilities arguably is the understanding that a study is available if it is published in a paper. The number of publications which we could identify on the laboratory websites ranged from 0 up to 40 per year, some with supplementary material. A large number of listed publications can be found if the laboratory analyses have been conducted for researchers of related institutes.

We think that minimum information standards for analyses conducted at laboratory facilities could greatly improve the cataloguing and discovery of data of such analyses. This would allow researchers to request data from repositories where they are deposited, which may not necessarily be at the laboratories. With regard to establishing minimum information standards the MIBBI initiative in the bio-sciences might be a good reference example.[47]

Suggested actions

- *Investigate how major facilities as well as typical archaeometry laboratories, could be mobilised to provide open access data (e.g. factors that impede open data sharing and how they might be removed).*
- *Improve the online documentation of archaeometry analyses on laboratory websites, and include pointers to publications and deposited analysis data.*
- *Standardise the metadata of archaeometry documentation to promote consistent cataloguing and improve cross-laboratory discovery and access to publications and analysis data*

3.4.6 Virtual research environments and data processing services

Virtual Research Environments (VREs) is an umbrella category of online environments that support functions such as professional networking and information exchange, collaborative digital collection formation and use, and tools for data processing and analysis. Ideally research communities could flexibly select and combine solutions for these functions according to their specific needs. In general

[47] MIBBI - Minimum Information for Biological and Biomedical Investigations https://www.biosharing.org/standards/

the availability and types of VREs indicates if and how far digital research tasks are already conducted online and in a collaborative manner.

The results of our study on VREs in the field of archaeology suggest that at present there are only few advanced, multi-functional VREs in use. Arguably the main virtual research environments of archaeologists are Geographic Information Systems. GIS allow them to bring together data of individual excavations as well as regional analyses (e.g. settlement patterns).

In the literature the most often referenced example of an archaeological VRE is the VERA - Virtual Research Environment for Archaeology that was developed for the large-scale excavations of the Silchester Town Life Project (UK). Based on the Integrated Archaeological Database (IADB) the project achieved an improvement in data organisation and access of the excavators and subject experts involved in the project. Alison Babeu (2011) presents many digital environments developed for Classical Studies of which most are focused on digital scholarly editions and some also allow advanced collaborative work.

Other archaeological VREs described in the literature hardly qualify as such. For instance articles in the “Archaeology 2.0” volume (Kansa *et al*. 2011) beside VERA present examples like Open Context (data publishing platform) or the UCLA Encyclopedia of Egyptology. “Web 2.0” applications are perceived as bottom-up, flexible and arguably more inclusive than other VREs. However they mainly serve the professional networking and information exchange function, while the building and curation of digital collections and data processing require other systems.

Many researchers use “Web 2.0”-type professional networking platforms such as Academia.edu, Mendeley and ResearchGate; especially Academia.edu attracts many users who are active or interested in specific fields of archaeological research. Such a networking platform is not intended as part of the ARIADNE data portal; it would duplicate available solutions which seem to satisfy very well the needs of archaeologists in this regard.

Arguably the largest distance of most archaeological research projects exists to Distributed Computing Infrastructures (DCIs). DCIs provide a so called Science Gateway to Grid/Cloud infrastructure and software applications for data processing, storage and transfer. User groups can share data resources and computing applications (virtual machines) and, thereby, also form a virtual research community. Although providers of DCIs have sought to expand their mainly natural sciences user base to other disciplines, uptake in the humanities, including archaeology, is very limited as yet. Lack of large datasets and of a solid base of skills for advanced computing can limit the use of DCIs. The major impediments for DCI usage of archaeologists arguably are the diverse and complex types of data, lack of consistent data structures, isolated and often not openly available data sources.

Building of virtual research environments or providing facilities for data-intensive research are not within the remit of ARIADNE, but not outside the scope of interest of ARIADNE partners. For example, one partner considered, *“It could be interesting to be engaged in creating virtual research environments (VREs) devoted to archaeology that can meet the needs of modern forms of scientific collaboration. VRE for archaeology could be conceived to support research activities that require access to great quantities of data, the intensive use of computing resources, and collaborative efforts involving research communities in related areas of the ‘hard sciences’.”*

In our study we found no advanced VREs for archaeology and use of Distributed Computing Infrastructures appears to be low; we may have overlooked a few, but not many existing examples. Users of VREs, DCIs or other research environments are also potential users of data resources provided through the ARIADNE data infrastructure and services. Therefore we give the suggestions below.

Suggested actions

- *Investigate if archaeologists need virtual research environments; e.g. what kind of research could archaeologists conduct online, what functionalities are necessary for such research, and how could they be provided in a VRE?*
- *Promote the development of relevant VREs with functionalities (tools, services) required by archaeologists to conduct research tasks online (e-research).*
- *Raise awareness of Distributed Computing Infrastructures (e.g. Grid/Cloud-based services of DCIs) amongst archaeologists who need data processing services.*
- *Encourage use of DCIs by bringing together archaeological research groups with providers of DCI, especially with their developers of data processing software and services.*

3.4.7 Data mobilisation strategies and specific actions

The lack of access to many relevant data resources in the field of archaeology and other heritage sciences raises the question if there are any priority resources which might be mobilised with specific actions. This is a difficult question because in this field researchers need access to a wide range of information for preparing, carrying out, and interpreting results of projects. Below we first answer the question from the perspective of community-level research e-infrastructure and services and then briefly address some resources which require specific actions by data holders.

Registration of available data resources

From the perspective of community-level research e-infrastructure and services like ARIADNE the answer is rather clear: ARIADNE provides a facility to register data resources so that what is accessible can be documented, made visible and discoverable. The ARIADNE online survey found that gaining an overview of and ability to search across available data resources is on top of what researchers expect from the project. The ARIADNE registry will allow this, for large and small datasets, given data holders make the effort to register the data. Importantly, this is not only about registration, there are also requirements with regard to the readiness for access to the data in terms of access mechanisms and conditions (e.g. appropriate licensing).

E-infrastructure initiatives like ARIADNE of course want to have a critical mass with regard to the number as well as volume of registered resources. Incorporating available archaeological data from large community-level and institutional resources is clearly among the priorities of the initiative. The focus of ARIADNE is data resources, not publications such as research papers, proceedings, book chapters, etc. which are available from other services (e.g. publishers' websites or self-archived papers in institutional document repositories). The relevant resources are archaeology and other heritage science data; also data resources that are not produced by domain researchers but relevant for their research may be considered (e.g. satellite remote sensing data). Next we may ask about priorities for specific data resources.

Some data mobilisation priorities

The overall tenor of the responses to the ARIADNE online survey (over 500 qualified responses) was that *all* types of data are important. This of course is related to the existing difficulty of finding and accessing online the many data resources researchers would like to consult and use for their research. A comparison of what respondents considered as important for their research with their rating of the online accessibility of these resources gives some general indication of data mobilisation priorities: Data of excavations, prospection/field surveys, GIS, material/biological analyses and radiocarbon/dendrochronology showed the largest gaps between importance and online accessibility[48].

The comparison also shows that government site management data are not a priority, although their online accessibility was perceived as low. Also not a mobilisation priority are large datasets or stocks of data ("big data") which could be used for data mining or other advanced computing methods. Reasons may be that respondents did not expect much from such approaches in archaeology and may lack expertise to apply the required technology/software and methods. The online availability of data resources for applying such methods was perceived as very low, and exemplary comments were *"Datamining is currently undervalued due to a lack of repositories"* or *"Not much available in the way of data mining - it may be important in the future"*.

Below we address data resources which may be priority targets and require specific actions by data holders. We note, however, that survey respondents were not particularly fond of distinguishing between relevant and less or not relevant resources (e.g. *"all sources are important and must be first-hand"*; *"all is important ...if it is available"*). However, resources important for one user segment may not be relevant to others. For example, some may need access to legacy data of excavations while others would like to use, but cannot afford, satellite or airborne remote sensing data.

Excavations, prospection/field surveys, including GIS

Some information about excavations, prospection/field surveys, including Geographic Information System/Services (GIS) used, is usually available online from project websites. But more could be done by research groups to make data available online during projects (which may extend over many years), and ensure that the full data/evidence of completed projects is deposited in digital archives. ARIADNE focuses on data deposited in accessible archives/repositories and available information about excavations, for example, the Fasti Online database of project partner International Association for Classical Archaeology (AIAC)[49].

Legacy data

Much excavation and other archaeological work has been conducted which remained unpublished and poorly documented. There is a growing interest to identify, study, document, digitise and make accessible information about and from such work. One exemplary initiative in this field is the Odyssee programme in the Netherlands, a large number of projects funded by the Netherlands Organisation for Scientific Research (NWO)[50]. Within this programme DANS 2009/2010 made an inventory of 1800 old, unfinished projects. Fifty were described in more detail (available documentation, material, etc.) and researchers/students invited to work on them. DANS also helped digitising old material (maps, drawings etc.) and stored it[51].

More initiatives in different countries to identify and unearth important unpublished work could add greatly to the accessible archaeological record[52]. Such initiatives will need effective processes for the

[48] Cf. ARIADNE D2.1 First Report on Users' Needs. April 2014, 76-84

[49] Fasti Online, some 12,000 excavation reports and site summaries across the Mediterranean and other countries, http://www.fastionline.org/excavation/

[50] NWO: Odyssee programme (in total 32 projects have been funded, the first projects started in 2009 and some ran until 2013), http://www.nwo.nl/onderzoek-en-resultaten/programmas/odyssee

[51] DANS: Thematische collectie: Odyssee onderzoeksprojecten, https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:34359

[52] The Heritage Council of Ireland some years ago addressed the issue of unpublished excavations straight-forward by a survey (2001) of such excavations. The survey categorised 81 sites/excavations as being of national importance and a further 340 as being of regional significance but also worthy of full publication. Some sites were published thereafter while others in 2007 were still noted as in the process of excavation and due for publication in the near future: Significant Unpublished Irish Archaeological Excavations 1930-1997, website (2007), http://heritagecouncil.ie/unpublished_excavations/section1.html

digitalization, documentation and analysis of the material. In order not to re-invent the wheel, existing knowledge and tools should be collected, evaluated and put to work. As one project in this area we note the RetroDig project at the University of Heidelberg which develops an environment and workflows for retro-digitalisation, annotation and analysis of excavation records, exemplified with archival material of an unpublished excavation of the 1970s held at a regional museum (Decker & Volkmann 2015).

Grey literature

In archaeological terms, "grey literature" means unpublished fieldwork reports, a lot of which have been produced in the context of infrastructure development and other land use projects. While often the only record of the fieldwork undertaken, such reports are difficult to find and access in many European countries. This is especially the case where contract archaeologists conducted the work and there was no obligation for the involved parties to make the reports publicly available.

Among the exceptions is the UK, where provisions have been put in place so that fieldwork reports are deposited with the Archaeology Data Service (ADS)[53]. At present the digital library contains over 34,000 reports and downloads in some months range in the thousands. The extensive usage makes clear the need to make such "grey literature" more accessible. In the UK, it has been estimated that much archaeological knowledge was 10 years out of date because results of fieldwork did not find their way into the academic research. Now the fieldwork reports are made available from ADS, alongside the academic work, which acknowledges the importance of the fieldwork undertaken by contract archaeologists.[54]

Also other institutions hold larger stocks of "grey" material, including documents from unpublished excavations (as addressed above) and other relevant content for which typically little or no metadata is available. A special line of activity within ARIADNE therefore is to enable better access to such documents[55]. Natural language processing techniques and controlled vocabulary (e.g. thesauri) are employed for tagging and extracting terms in the documents. Consequently these terms (keywords) can be indexed and used in search services to discover relevant documents in otherwise difficult to explore document collections. It is expected that many archives will benefit from improved methods and tools in this field.

Laboratory analyses of archaeological and other heritage objects

In the ARIADNE online survey material/biological analyses and radiocarbon/dendrochronology data showed a large gap between importance and online accessibility as perceived by the researchers. We did not look deeper into the situation of radiocarbon and dendrochronology data. Several databases have been developed[56], however, researchers may often need specific data which is not easy to identify or unavailable.

---

[53] ADS: Library of Unpublished Fieldwork Reports, http://archaeologydataservice.ac.uk/archives/view/greylit/

[54] ARIADNE: Grey Literature Special Interest Group, http://www.ariadne-infrastructure.eu/Community/Special-Interest-Groups/Grey-Literature

[55] This experimental work is conducted by Universiteit Leiden, University of South Wales and ADS. See their report ARIADNE D16.2 First Report on Natural Language Processing, May 2015, http://www.ariadne-infrastructure.eu/Resources

[56] Some examples: DCCD - Digital Collaboratory for Cultural Dendrochronology (developed by DANS [NL] and national partners since 2006, 2010-2013 with other European partners, and since 2013 some further work is conducted within ARIADNE; measurement series of different wood species derived from over 5000 objects and sites dating between 6000 BC and the present, ca. 50% of the DCCD collection is derived from archaeological sites and structures, including maritime archaeological sites), http://dendro.dans.knaw.nl; International Tree-Ring Data Bank (NOAA / NCEI Paleoclimatology Program & World Data Center for Paleoclimatology, raw ring width, wood density, and isotope measurements of material from 4000 sites on six continents), http://www.ncdc.noaa.gov/paleo/treering.html. Radiocarbon Palaeolithic Database Europe (Katholieke Universiteit Leuven, Center for Archaeological Sciences; about 13,000 dates),

Concerning material/biological analyses our study (cf. *Section 9.1.3*) confirms the researchers' perception: At large, well-equipped laboratories documentation of archaeometrical and other analyses is hard to come by online, access to data is offered nowhere, and even some illustrative webpages of conducted analysis are missing in most cases. Lists of papers are typically presented, but few entries link to an open access journal paper or self-archived, pre-print copy. The main reason for the lack of study documentation and access at the laboratory facilities arguably is the understanding that a study is available if it is published in a paper.

Among the factors that impede proper documentation of analyses may be a lack of common metadata standards, which would be required to create a laboratory-based or common data catalogue. One ARIADNE partner that operates a laboratory mentioned that metadata for archaeometrical analyses are not standardised or such standards are not in wide use. Arguably because of the many types and variants of analyses and that they are applied according to particular research questions. But recommended description and terminology for research objects are generally followed in research papers and reports, for examples, the *Guidelines of the Prehistoric Ceramic Research Group* (Hamilton *et al*. 1992; Prehistoric Ceramics Research Group 2011) and *International Code for Phytolith Nomenclature 1.0* (Madella *et al*. 2005). It seems that these codes are not available in a machine-readable format.

In the ARIADNE portals survey one partner explained that databases of scientific analyses (archaeometry) are not integrated because of a lack of common description standards; "*if one wants to upload scientific data there should be minimum requirements of that data; e.g. minimum requirements of ceramic thin section description etc. This will allow better cross country/regional comparison of results".* There could be a need for minimum information standards for archaeometrical analyses (e.g. data types, methods and procedures used in the data generation). In recent years such standards have been developed in the bio-sciences by the MIBBI initiative for different biological investigations (Taylor *et al*. 2008). Some 40 Minimum Information reporting guidelines for different analyses have been produced and are included in the BioSharing catalogue and portal[57]. MIBBI could be a exemplary reference for the development of description standards; they also provide a suite of tools that facilitate standards compliant description of laboratory analyses[58].

http://ees.kuleuven.be/geography/projects/14c-palaeolithic/; Radiocarbon Context Database (University of Cologne, Institute for Prehistory; 1274 Near Eastern sites, 2874 dates), http://context-database.uni-koeln.de; RADON - Radiocarbon Dates Online 2014 (University of Kiel, Germany, about 4300 dates for the Neolithic of Central Europe and Southern Scandinavia), http://radon.ufg.uni-kiel.de; C14.sk (Comenius University in Bratislava, Department of Archaeology, over 2400 dates from Neolithic to Iron Age samples from Slovakia, Czech Republic, Austria, Hungary and Poland), http://www.c14.sk; ADS: Archaeological Site Index to Radiocarbon Dates from Great Britain and Ireland. Council for British Archaeology, 2000 (updated 2012), about 22,000 dates, http://archaeologydataservice.ac.uk/archives/view/c14_cba/overview.cfm

[57] MIBBI - Minimum Information for Biological and Biomedical Investigations https://www.biosharing.org/standards/

[58] ISA (Investigation/Study/Assay) tools, managed by the Oxford e-Research Centre, http://www.isa-tools.org

4 Focus area 2 – Culture of open sharing and re-use of data

4.1 Introduction and overview

ARIADNE is the core EU-funded "Integrating Activity" focused on archaeological data(sets). Through its e-infrastructure and services ARIADNE will enable enhanced discovery, access to, and (re-)use of archaeological data held by institutions across Europe (and beyond). But ARIADNE, like other data infrastructure initiatives, depends on underlying institutional repositories and community archives richly filled with open data shared by research projects.

Various benefits are expected from the ability to (re-)use open data for further research and other purposes. Research institutes, data archives and e-infrastructure providers, funding bodies and other stakeholders therefore cannot ignore obstacles which hinder researchers in sharing their data in an open manner. Rather they must support researchers in data sharing and help ensure that they receive appropriate credit for doing so.

Arguments for open data sharing

There are many good arguments for open data sharing such as preventing duplicative data collection, allowing replication of data analysis to scrutinize knowledge claims, and enabling new research questions to be addressed, based on combined datasets. Particularly strong are the arguments related to the re-use of data as re-use allows exploitation of previous investment. Data that is curated, integrated, re-used and analysed with new methods gains in value rather than being only a cost factor. The benefits expected from data re-use are among the core interests of research funders in open data.

Data sharing through accessible archives

Researchers often share data, but mainly with project collaborators and other trusted colleagues. Most data remains within a small circle of peers and is not available for other researchers or the public at large. The goal of open sharing is to have the data deposited in digital archives to allow for access and (re-)use in further research and other purposes. Instead of inaccessibility and potential loss, deposited data is curated and accessible for new, often not anticipated uses. However, strong barriers need to be addressed to move from a closed-circle to open sharing of data.

Barriers to open data sharing

Many factors work against open sharing of research data. In particular, researchers perceive a lack of incentives (e.g. little academic recognition and reward), fear that their data could be misinterpreted or misused, and preparing data for (re-)use by others means extra work (e.g. data description), while there are other priorities such as new projects and publications. The core requirements for open data sharing are not technical but institutional, especially the need for appropriate academic recognition and reward. But also other requirements must be met, such as available, trusted and sustainable data archives.

Strict open data mandates

A broad consensus is emerging that research data, in particular, data from publicly funded projects should become openly accessible for all. Ever more research funding bodies extend their open access mandates for publications to the data that underpins the research results. To ensure compliance, project funders also require grantees to implement a data management plan, with the objective of making the generated data publicly accessible through an appropriate repository. Experience shows that open access policies/mandates of research funders and institutions for publications and data need to be decisive, monitored and enforced. Also journals that demand research underlying reported results to be accessible have a significant impact on the availability of the data.

Data re-use and citation

Currently not a strong driver of open data sharing is evidence of individual benefits for researchers, which may be accrued through data (re-)use and citation by others. While a citation advantage for open access publications is well confirmed empirically, there is little evidence as yet of a similar advantage of open research data. Also solid empirical evidence of re-use of openly shared data is not available. But we may expect better insights into data re-use practices as more open research data becomes available. One important element in this regard is standardisation of data citation in the research literature and other publications (e.g. websites), including persistent identifiers of the data(sets).

The archaeological data sharing record

Archaeology was not addressed above, because the issues and required solutions are the same in all domains of research. But some are further ahead concerning the requirements, decisive open access policies and available and mandated data archives, for instance. In European countries where the requirements are fulfilled, archaeology is also part of the open data movement.

Elsewhere archaeological research is not at the forefront of open data sharing, despite factors that make an open data imperative particularly strong in this field: excavation of sites destroys the primary archaeological evidence, the work on archaeological heritage is done in the public interest, and there is little commercial interest in archaeological data (apart from the commercial archaeology sector).

However, many archaeologists are not yet well prepared or equipped for open data sharing. As the matter is complex, strong leadership with regard to policies/mandates, supportive institutional measures (e.g. capacity building, training of researchers), and state-of-the-art digital archives are required. Overall this is about growing a culture of open data sharing, based on common values and agreed principles. As a popular slogan of OpenAIRE puts it, *"Open Access is global – but implementation is local"*.

Openness should become embedded in archaeological research practices as *"the default modus operandi"* (e-InfraNet 2013) so that the advantages of accessible and re-usable data gain priority over the interest of the individual researcher to protect his/her "claim".

4.2 Open Data and Science

A brief definition of "open data" is that anyone can freely access, use, modify, and share the data for any purpose.[59] With regard to data licensing there should be no or only minimal restrictions on re-use (e.g. attribution of the producers).

In the area of research data the request for open data in particular pertains to data from publicly funded research. Such data should be open to the greatest extent and with the fewest constraints possible while, at the same time, respecting concerns with regard to issues of privacy, security, and legitimate commercial interests of private research partners.

Specific requirements concern the description of the data and technical aspects: There should be metadata so that users can discover, understand the provenance, and evaluate the trustworthiness and quality of the data. For effective re-use the data should be machine-readable and in open, non-proprietary formats. In some research areas it is also important to make available the specific

[59] Cf. the "Open Definition" of the Open Knowledge Foundation, http://opendefinition.org

research software (e.g. programme code for computational analysis), as this may be necessary to reproduce and validate research results and repurpose the data.

In this report we often refer to open, re-usable and interoperable data, putting particular emphasis on three aspects:

- Ability to re-use data in further research or for other purposes – getting more from shared data. This requires that re-use is not restricted technically (e.g. difficult to process formats such as PDF) or by licensing conditions;
- Data and metadata is also (re-)used for providing services, which require interoperability, i.e. ability of systems to exchange, read and interpret the data, based on the codification of the data and vocabulary;
- Moreover, we note re-use and interoperability principles of Linked Open Data (LOD): providers of LOD should, wherever possible, re-use existing vocabularies (instead of creating new ones) and link their data to data of others; re-use of concepts/terms allows linking and integrating of datasets from different providers, i.e. create an enriched web of related information.

The Royal Society's *Science as an Open Enterprise* report urges researchers to strive for "intelligent openness" for which data should be "accessible, useable, assessable and intelligible" (The Royal Society 2012a: 12 and 14). The report emphasizes that research data *"must provide an account of the results of scientific work that is intelligible to those wishing to understand or scrutinise them"*. Recipients should be able to understand and judge what is communicated, e.g. the nature of the claims that are made, the reliability of the source and evidence provided, etc. The report also considers that users of open data could range from the highly expert in the field to the non-specialist. Hence data communication needed to be differentiated, but based on a deeper understanding of what the non-scientific wider public may need, in order to prioritize relevant data and justify the additional effort.

A report of the European ERA-net project e-InfraNet discusses the concept of openness in various contexts (e.g. research, content/data, software, infrastructure, standards, innovation). The report suggests openness as the *"default modus operandi"* for all publicly funded research and educational resources, with "open" as the preferable approach *"not as an end in itself or as an ideology"* (e-InfraNet 2013: 10).

Open data is understood as a key element of "open science" or "science 2.0". Fecher & Friesike (2014) distinguish five "open science" schools of thought and literatures that centre on different goals: making knowledge freely available for everyone ("democratic school"), making science accessible for citizens and enable participation ("public school"), creating openly available platforms, tools and services for research ("infrastructure school"), making the process of knowledge creation more efficient through collaboration ("pragmatic school"), and developing new metrics for relevant impacts of scientific works ("measurement school").

The overall vision may be summarised as making science and knowledge as accessible, transparent and effective as possible, including a deeper involvement of citizens in science. The need and challenges of open data and research practices are of course also discussed in the archaeological community (e.g. Beck & Neylon 2012; Beck 2013; Costa *et al*. 2012; Kansa 2012; Lake 2012; Wilson & Edwards 2015).

4.3 Open data policies and practices

A broad consensus is emerging that research data, especially data generated in publicly funded research projects, should be openly accessible for all. Beside the "tax-payer argument" there are many other arguments for open sharing of research data and, hence, that such sharing should be standard practice for researchers. However the actual sharing of data through accessible digital archives is rather low in most disciplines, including archaeology. Ever more research funders therefore extend their open access mandates for publications to the data that underpins the research results. To ensure compliance with the open data mandate, the funders also require grantees to implement a data management plan with the ultimate objective of making the data generated in the funded project openly accessible.

4.3.1 Arguments for open sharing of research data

There are many good arguments for open sharing of research data and benefits open data are expected for the scientific enterprise as well as society at large (e.g. Beagrie 2011; Borgman 2012; OECD 2007; The Royal Society 2012a). The advantages of open access to well-managed research data include that it

- prevents the duplication of data collection as well as potential loss of valuable data (especially observation data which cannot be reproduced);
- allows to replicate data analysis and scrutinize knowledge claims, i.e. bolster scientific rigour and guard against data fabrication;
- enables that new research questions – not envisioned by the initial investigators – can be asked of existing data, thereby feeding and potentially speeding up the research process;
- allows to combine open data resources for building larger and richer databases, which may be exploited with new research tools (e.g. data mining, pattern detection, meta-analysis);
- enables a higher return on public investment in the generation of research data, ideally economic benefits through innovations based on advances in scientific knowledge;
- promotes a higher relevance of science in/for society, e.g. through transparency of and potential engagement of citizens in the scientific process.

The set of expected benefits of open sharing of research data thus includes elements of the scientific method and new approaches to research (e.g. “data science”) as well as societal aspects. Indeed, the quest for open data is part of an envisaged transformation of the research system towards “open science” (cf. *Section 8.4*).

Several of the arguments also apply to the open data agenda concerning governmental and other public sector information, albeit within different frameworks: transparency and democratic control of governmental procedures and decisions (Huijboom & Van den Broek 2011; Open Data Barometer 2013). Furthermore there is the expectation that public sector information (PSI)[60] – readily available and for free or at marginal costs – enables the generation of commercial added value services (Manyika *et al*. 2013). Much of the governmental and PSI information cannot be considered “scientific”, but quite some of it can of course also be used for research purposes (census data and other social statistics, for instance).

It is worth noting the emphasis of many open data proponents on fundamental scientific principles, i.e. that the scientific method is based on evidence, which should be readily available in case of

[60] The core reference for Public Sector Information (PSI) is the Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information. Official Journal of the European Union, L 175/1, 27.6.2013, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32013L0037. The PSI Directive does not prescribe “open data” but transparent, accessible and non-discriminatory terms of use. On legal aspects of PSI see the results of the LAPSI projects, http://ec.europa.eu/digital-agenda/en/news/legal-aspects-public-sector-information-lapsi-thematic-network-outputs

doubt (e.g. misreporting, lack of validity or outright fraud). Indeed, the increase in the number of papers that are withdrawn from scientific journals (Van Noorden 2011), detected errors in many scientific papers (Bakker & Wicherts 2011; Wicherts *et al*. 2011; The Economist 2013) and cases of outright fraud have added to the open data agenda in the research community. The expectation that the data on which published knowledge claims are based should be publicly available has been greatly reinforced.

Particularly strong are the arguments related to the re-use of data, e.g. combining and asking new questions of shared data, as this allows exploitation of previous investment in data resources. Data that is curated, re-used, enriched, combined and analysed with new or enhanced methods gains in value rather than being only a cost factor.

The core interest of research funders is expected benefits of data re-use. They are eager to see re-use happen und confirm that the required investment in data sharing and preservation pays off. As Clifford Lynch notes: *"Funders, and particularly public funders, are under great pressure to show how their funding contributes to broad economic growth, how it addresses the needs of society, and to demonstrate that the requirements that they impose on the work they fund makes discovery ever more rapid, extensive, and cost-effective. From this perspective, they are not interested in data preservation or even data sharing other than as a necessary precondition to data reuse; they are interested in conformance to their data management and sharing policies because it is the only way they can create the preconditions for data reuse. They are hungry for examples of how data reuse has improved the processes of scholarship and discovery, or contributed to economic growth, job creation, control of health care costs, or public policy."* (Lynch 2013: 397)

Above we only addressed expected benefits of open sharing of data, while related issues and requirements, and evidence for data re-use are addressed in other sections of this chapter.

4.3.2 Open data policies of research funders and institutions

Referring to the many good arguments for open data sharing, ever more governments, research councils and other funding agencies are extending their existing policies for open access to research publications to the data underlying the research. These policies arguably are the strongest driver motivating research institutions and researchers to take appropriate actions (e.g. data management plans for funded projects). Other important actors are publishers and editorial boards of journals who request that the research data underlying a paper has to be made accessible upon submission of the publication for review.

Policies of research funders

Open access to research data first has been strongly promoted by the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003)[61] and the OECD Declaration on Access to Research Data from Public Funding (2004; Principles and Guidelines 2007) – but arguably with limited impact for some years. For instance, in September 2009 an editorial of *Nature* noted "Data's shameful neglect" and stressed: *"All but a handful of disciplines still lack the technical, institutional and cultural frameworks required to support such open data access – leading to a scandalous shortfall in the sharing of data by researchers. This deficiency urgently needs to be addressed by funders, universities and the researchers themselves. Research funding agencies need to recognize*

[61] The Berlin Declaration under the foreseen open access contributions includes *"original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material"*. Since 2003 the Declaration has been signed by representatives of over 500 universities, research institutes, academies and professional associations, cultural heritage institutions, and funding agencies, http://openaccess.mpg.de/Berlin-Declaration

that preservation of and access to digital data are central to their mission, and need to be supported accordingly." (*Nature*, editors 2009)

Today the demand for open data is voiced and supported by many governments (e.g. G8 Open Data Charter 2013; Executive Office of the President 2013) and research funders, including the European Commission, European Science Foundation, Research Councils UK, among others. The European Union's Horizon 2020 program includes an Open Research Data pilot which requires data management plans aimed at accessible data of projects that are funded under specified areas.[62]

One leading example of open data policy is the Research Councils UK. The Common Principles on Data Policy of the seven councils state: *"Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property".* [63] Since April 2013 RCUK-funded researchers are required to include a statement in published papers setting out how and on what terms supporting materials, including data may be accessed. With regard to data retention most of the seven councils require 10 years. The Research Councils have also implemented state-of-the-art data centers, for example, the Natural Environment Research Council (NERC) has several such data centers.[64] The mandated repository for NERC-funded research in science-based archaeology is the Archaeology Data Service; also the Arts and Humanities Research Council requires grant holders to deposit the funded project data with the ADS.

Concerning the landscape of open data policies across Europe, some research funders already attach data management conditions to grants and name preferred repositories where the data should be deposited and shared openly. Other funders are still hesitant to become a regulatory force with regard to research data, arguably because of perceived reservations of the research community and the required additional investments (e.g. implementation of data archives, costs of data deposits, etc.).

A survey of 18 research funders in the European area conducted in the second quarter of 2012 (SURF Foundation 2013) found that 50% had a funding policy that covered data management; of the other funders 10% thought it would take more than 2 years and 35% did not intend to introduce such a policy. Concerning data management plans, 25% of the research funders required such a plan as part of the grant agreement, 10% recommended it while 65% did not require one.

Science-Metrix (2014e: 20-21) analysed the Open Access policies of 48 research funders within European Research Area (ERA) countries, Brazil, Canada, and the United States listed in ROARMAP: 23% of the funders explicitly excluded data from Open Access requirements and 38% did not mention data at all. Conversely, 10% encouraged open data archiving without a strong mandate and 29% mandated it.

The SHERPA/JULIET[65] service gives an overview of research funders worldwide that have a policy (mandates) for open access publishing and/or archiving of publications and data. In May 2015 only only 17 funders in Europe required data archiving: Twelve were research funders in the UK that is clearly the leading country in this regard (Jones 2012). Only five research funders in other European countries had a data archiving mandate: Austrian Science Fund, Fondazione Cariplo (Italy), Hungarian Scientific Research Fund, Higher Education Authority (Ireland), and Swedish Research Council.

[62] European Commission (2013-12): Commission launches pilot to open up publicly funded research data, Brussels, 16 December 2013, http://europa.eu/rapid/press-release_IP-13-1257_en.htm; current Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 and Guidelines on Data Management in Horizon 2020; versions 2.0, 30 October 2015 (European Commission, DG for Research & Innovation 2015a/b).

[63] Research Councils UK: Common Principles on Data Policy, http://www.rcuk.ac.uk/research/datapolicy/

[64] NERC Data Centres, http://www.nerc.ac.uk/research/sites/data/

[65] SHERPA/JULIET, www.sherpa.ac.uk/juliet/

Policies of research institutions

The SURF Foundation surveyed 94 higher education and research institutions in Europe with regard to research data management policy (SURF Foundation 2013): 37% of the institutions indicated that they had such a policy, mostly established in the last two years, but only few already prescribed a data management plan. The most important drivers for requesting such plans were compliance with requirements of funders, a national or institutional code of conduct, or the influence of major reports and guidelines issued by relevant institutions. 25% of the institutions indicated that they provide incentives to make available data (e.g. free data publishing services, translation of metadata into English, etc.). 42% of the institutions that did not have a data policy in place intended to define one in the next year (2013) while 30% had no intention to introduce such a policy.

Science-Metrix (2014c: 12-13) surveyed 162 head librarians at universities and higher education institutions of which 73% agreed that providing Open Access (OA) scholarly publications is a priority in their institution, whereas only 45% saw open data as a priority; 42% of respondents said that the institution has an OA policy for publications, but only 11% indicated that their institution has an OA policy for data. The head librarians were also asked if their institution maintains a repository for open access scientific data. 78 (47%) said no, while 59 (35%) indicated that such a repository is available either centralised (28%) or decentralized (12%). The study authors therefore conclude that the institutional infrastructures for scientific data are already more developed than the policy frameworks.

However, it is unlikely that so many universities and higher learning institutions have repositories that are specialised for data management and access. Research data, if deposited, is mainly kept in repositories that have been set up for managing documents (i.e. self-archived papers, proceedings, theses, etc.). In the Directory of Open Access Repositories (OpenDOAR) at present 136 of 2404 (5.7%) institutional, mostly university-based repositories report that their content includes datasets among other items, while 5 repositories (0.2%) manage datasets only[66]. The low percentage of the category "datasets" in OpenDOAR compared to the result of the Science-Metrix survey of head librarians suggests that OpenDOAR is not a good information source with regard to data repositories of universities and higher education institutions, which may have been developed in the last few years (cf. Science-Metrix 2014c: 12-13; see also our study of repositories in *Section 9.1.1*).

Journals' data policies

In some scientific areas it is already standard practice that authors make available their research data together with published papers, typically as supplemental material, but increasingly also by depositing the data in an accessible repository. Indeed, editorial boards of journals that request that data underlying papers has to be archived and accessible, ideally upon submission of a paper for review, are strong forces in the move towards open research data. A survey with 1317 American scientists in 43 domains of research found that regulative pressure by journals had a significant positive relationship with data sharing behaviours (Kim & Stanton 2013).

Alsheikh-Ali *et al*. (2011) studied papers published 2009 in 50 of the highest-impact journals in biomedicine: 22 journals required public sharing of specific raw data as a condition of publication, a further 22 encouraged data sharing without binding instruction, while six of the 50 journals had no published policy for data sharing. The study reviewed the first ten papers published in each journal (500 in all) and found that of 351 papers covered by some data-sharing policy, only 143 (41%)

[66] Fluorophores.org (Technical University Graz [Austria]; chemistry, 838 data items); Imperial College London - Chemistry (185,071 items); University of Edinburgh's DataShare (multi-disciplinary, 211 items); Harvard Dataverse (multi-disciplinary, 59,467 items); Odum Institute Data Archive at University of North Carolina (social sciences, 690,355 items)

adhered to that policy; of all 500 papers only 47 (9.4%) had the full raw data corresponding to the reported results publicly available online.

Vines *et al*. (2013) examined the effectiveness of different data archiving policies of journals on papers in the field of population genetics – no stated archiving policy, recommending (but not requiring) archiving, and data deposition mandates with or without a required data availability statement. Compared to having no policy, mandates with a required data availability statement increased the odds of finding the data online almost 1000-fold; without such a statement the likelihood was only 17 times higher, and "recommended" archiving did not show a significant effect.

A required data availability statement in published papers (and control of compliance) thus can greatly enforce data archiving policies of journals, research funders and institutions. For instance since 2011 many bio-sciences journals (e.g. *Evolution*, *Heredity*, *Molecular Ecology*, *Systematic Biology* among others) have adopted the Joint Data Archiving Policy (JDAP) which is promoted by the NSF-funded DRYAD data repository[67]. But the JDAP data availability statement can be adopted or used as a model by any journal and is of course not tied to depositing with DRYAD.[68]

If open data mandates of research funders become more widespread, journals will experience an increasing demand from authors to support depositing and/or linking of research data, beyond current practices of managing supplementary material (Schwarzman 2010). Many have already recognised this challenge and are looking for solutions. In principle, publishers are willing to support open sharing of research data, for instance, the over 50 publishers, including, among others, Blackwell, Elsevier, Macmillan, Sage, Springer, Taylor & Francis, Wiley, who endorsed the STM "Brussels Declaration". Principle 7 of the declaration reads: *"Raw research data should be made freely available to all researchers. Publishers encourage the public posting of the raw data outputs of research. Sets or sub-sets of data that are submitted with a paper to a journal should wherever possible be made freely accessible to other scholars"* (STM Association 2007).

But publishers' business is producing and marketing of publications, which includes managing information about authors, books, papers, references, etc. Most publishers do not see data management and preservation as part of their core business. Rather they understand it as a responsibility of the authors, the author's institutes or the research community as a whole (PARSE.Insight 2009: 68-70). Still many publishers accept and manage a growing amount of supplementary material submitted with papers. However, a survey in 2011 found that some journals started to limit the size of such material or do not accept it any longer (Opportunities for Data Exchange 2011: 45-46).

Ever more journals now suggest depositing the data in appropriate domain archives, also because research funders wish to see re-usable data made available, which supplementary material typically is not. Such material often presents data in a highly processed state (e.g. summary tables, graphs or charts which do not fit into the paper) and contained in a PDF document, which does not allow data re-use, e.g. inclusion in other datasets. As the urge to ensure access to and re-usability of data underlying research papers becomes ever more pressing the main approach of publishers will be cooperation with state-of-the-art digital archives.

[67] DRYAD (http://datadryad.org) is a repository for bio-sciences data made accessible under journal data policy; it has supported by National Science Foundation (NSF, USA) grants since September 2008.

[68] The JDAP statement (http://datadryad.org/pages/jdap) consists of the following text: *"[Journal] requires, as a condition for publication, that data supporting the results in the paper should be archived in an appropriate public archive, such as [list of approved archives here]. Data are important products of the scientific enterprise, and they should be preserved and usable for decades in the future. Authors may elect to have the data publicly available at time of publication, or, if the technology of the archive allows, may opt to embargo access to the data for a period up to a year after publication. Exceptions may be granted at the discretion of the editor, especially for sensitive information such as human subject data or the location of endangered species."*

This approach will also help to establish credit mechanisms for deposited research data. For instance, Elsevier cooperates with PANGAEA in the field of earth system research. Data sets deposited with PANGAEA are automatically linked to the corresponding articles in Elsevier journals on their ScienceDirect platform and *vice versa*. The citable data sets are geo-referenced and displayed on a Google map indicating their geographical coverage. The reciprocal linking and geo-location feature were launched in February 2010 but have been implemented also for earlier papers and deposits.[69]

Lin & Strasser (2014) suggest many ways in which journal publishers and editors together with repositories can promote and contribute to increasing access to data. The recommendations and examples address mandatory data availability policies, data citation standards, indicators for data recognition and usage (e.g., views, bookmarks, downloads, references, etc.), incentives such as awards for authors whose data are frequently re-used by other researchers. But in the first place publishers and editors should listen to the researchers they serve in order to help establish community standards for data sharing they accept, follow and re-enforce.

Need of strict mandates

Experience shows that open data mandates need to be decisive, monitored and enforced. Many studies found that without strict requirements scholars avoid self-archiving of publications – which is much easier than for research data (cf. the overview of studies in Xia *et al*. 2012). The latest evidence comes from the PASTEUR4OA project: The project updated the ROARMAP[70] database of institutional and other repositories worldwide that work on the basis of a mandatory or less strict open access policy, in total 663 policies of which 481 set by universities or research institutions. Then the project analysed the effectiveness of the policies by comparing the article output of the institutions to the number of articles deposited in the repositories. Across all institutions more than three-quarters of published articles were not deposited, 12% were open access deposits, 3% restricted access, and for 8% only metadata was available. Open access deposit was over four times higher at institutions with a mandatory policy than at those without. The study found that significantly better results are achieved if depositing is mandatory, cannot be waived, and is linked with research evaluation measures. The project recommends that OA policies should include these three elements. (PASTEUR4OA 2015)

While the effectiveness of mandated archiving of publications is low, the case of research data deposits is much more difficult. To gain traction with many researchers who will take a "wait and see" position and ignore requirements, data repositories and curators will need decisive mandates and a strong backing of research funders as well as their own institution.

Somewhat surprisingly therefore a recent report issued by the EU-funded Research Data Alliance Europe project contains a recommendation that runs against decisive open data policies & mandates. *"DON'T regulate what we don't yet understand. One of the greatest risks in government is for politicians to act too soon. A global data library sounds like a big deal – and it is. But we don't yet know which technologies will work best, which policies will safeguard our privacy and security, and who will benefit most and how. There will be a great temptation – especially in the post-Snowden era – to regulate. We urge forbearance – at least until the world's scientific community has more experience with the costs and benefits. Issues such as privacy and ethics should be handled in consultation with the data and scientific communities, as well as with society at large."* (RDA Europe 2014: 36)

In the light of the massive evidence for the need of decisive open data policies such a position is counter-productive, although the mentioned requirements (e.g. protection of sensitive personal data, costs & benefits assessment, optimal technical solutions, etc.) of course need to be considered

---

[69] See for example: http://doi.pangaea.de/10.1594/PANGAEA.704574

[70] ROARMAP - Registry of Open Access Repositories Mandatory Archiving Policies, http://roarmap.eprints.org

thoroughly. The key point arguably is that a "one size fits all" approach will not work, hence consultation with and taking account of the specificities of the research communities is necessary (cf. Pryor 2009).

Ensuring appropriate guidelines for open archaeological data

Planned or already issued open data guidelines of research funders or publishers may not necessarily take account of particular aspects of archaeological data. In such cases research institutions and associations should promote community consultation and suggestion of appropriate guidelines (i.e. on the national level, where specific conditions may apply).

One issue regularly raised is that archaeological information frequently includes spatial data which may risk facilitating looting[71] (Bevan 2012: 7-8). Hence it may be argued that spatial data should be degraded and full resolution data made available only to "approved" users.

McKeague *et al*. (2012: 47) note that greater access to information *"does pose a threat to the cultural heritage resource through theft and, for the archaeological resource, illegal metal detecting or 'night-hawking' in particular. Some information is considered sensitive and is withheld for this reason, but in a mature society it is in the interests of the historic environment that information is visible and consulted alongside other considerations rather than shielded from the public, who, by and large, have funded most of its collection. Restricting information runs counter to novel ways of encouraging the public to appreciate their heritage through participatory approaches to both national and local inventories".*

Similarly, one ARIADNE partner notes "a *seemingly sound argument that archaeological data should not be published in complexity because such a publication may bring a threat to archaeological sites"*, and suggests: "*This should be discussed. Our opinion is that in the long-term the public knowledge about sites in the surrounding landscape can be an important aspect of their protection. Means should be found how to educate the public to understand that archaeological heritage belongs to the public and no one else can better protect it than the public itself".*

Beck (2010) notes that the argumentation against opening up archaeological information *"is based on the premise that 'accessible knowledge will inevitably be abused' and eschews any of the benefits that data sharing can provide"*. The argumentation tends to fortify a default position of doing nothing or accept only the lowest common denominator of accessibility. Is the belief that publication of detailed data might lead to greater levels of archaeological looting supported by solid evidence? Costa *et al*. (2012: 453) think that it may be *"in dire need of some formal demonstration"*.[72]

We assume that the move towards open data offers an opportunity to discuss and get clear about issues concerning the disclosure or non-disclosure of certain information that is understood as sensitive. Such information concerns find spots, human remains, and knowledge of indigenous communities, for instance. Also personal information such as contained in field notes (site diaries) or oral history records require special attention. The "Guidance on the Deposition of Sensitive Digital Data"[73] of the Archaeology Data Service and other professional guidance may be consulted for reference. Concerning knowledge and intellectual property rights of traditional cultures see the report of Torsen & Anderson (2010) for the World Intellectual Property Organization (WIPO), and the

---

[71] Colin Renfrew defines looting as *"the illicit, unrecorded and unpublished excavation of ancient sites to provide antiquities for commercial profit"* (Renfrew 2001: 15).

[72] This is of course not meant to say that there are no critical issues, including in European countries. For example see the papers in Lagerlöf (2013) concerning metal detectorists and other "treasure hunters".

[73] Archaeology Data Service ADS: Guidance on the Deposition of Sensitive Digital Data, http://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy

work that has been conducted in the Intellectual Property Issues in Cultural Heritage (IPinCH) project[74].

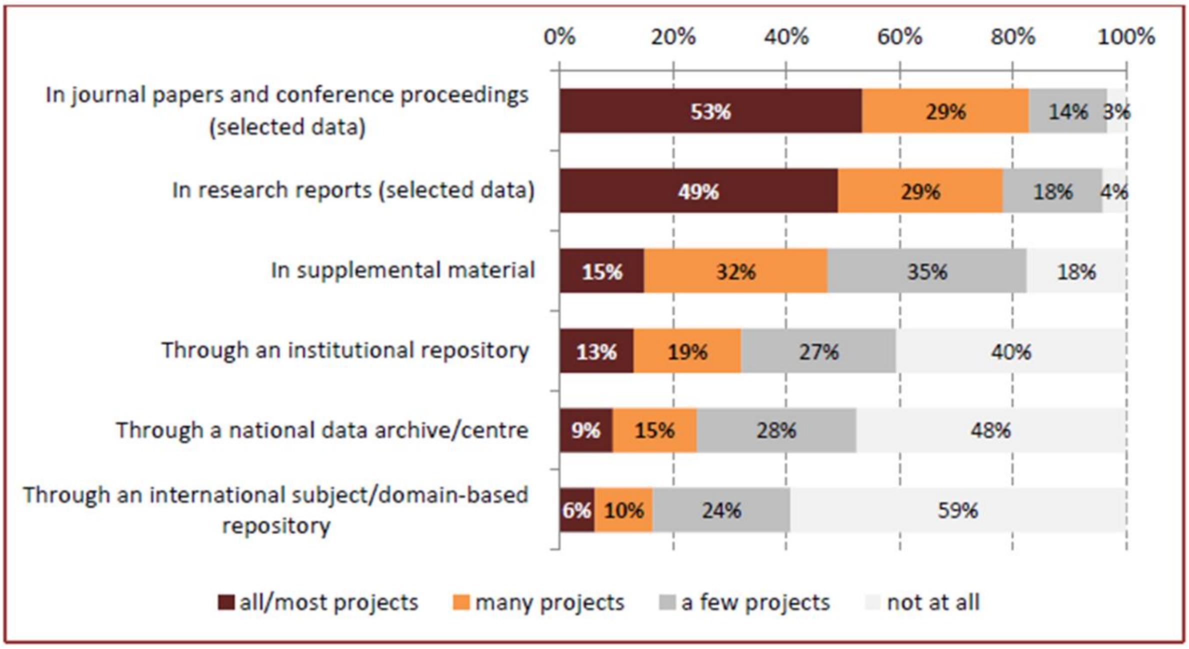
4.3.3 Sharing of research data in archaeology

Current practices

The current data practices of researchers run against what advocates of proper data management and open data sharing would like them to do. Representative surveys across all disciplines show that after the completion of research projects most data remains locked away, resides on PCs, storage devices, and restricted access servers, out of reach of other researchers. Only some 6-8% of researchers sometimes share data through a community repository (PARSE.Insight 2009; Science 2011)[75]. Hence, proper management to prevent data loss and the value of the data for re-use by the research community are often not considered.

The ARIADNE online survey generally confirms the insights of the mentioned surveys with regard to data sharing, but with somewhat better results. Many of the over 500 respondents claimed to publish the data of "all/most projects" in ways so that they are available to others beyond the researchers own institute: 6% through an international subject/domain-based repository, 9% through a national data archive, and 13% through an institutional repository. If the percentages for "many projects" are included, the figures even are 16%, 24% and 32%, respectively.

Figure 6.2-20: Question C.6 – "To what extent and in what way is data which your research group is producing typically being published (i.e. made available to a certain community beyond your own institute)?"



N = 517-547 per item (depending on number of respondents without answer)

The comparatively good results can be explained in part by the participation of many respondents from the UK (79) and the Netherlands (35). In these countries, archaeological data archives are available and mandated or recommended by research funders. In the UK, the Archaeology Data

[74] Intellectual Property Issues in Cultural Heritage (IPinCH), http://www.sfu.ca/ipinch/

[75] Not representative but with similar result for example Mowers *et al*. (2013); Andreoli-Versbach & Mueller-Langer (2013) examined the academic webpages of 488 economists and found that about 90% neither have a data section nor indicate whether and where their data is available, about 9% share some of the their data, and 2% fully share.

Service (ADS)[76] is the mandated archive for data of projects funded by the Arts and Humanities Research Council and the Natural Environment Research Council as well as the archive recommended by the British Academy, Council for British Archaeology, English Heritage and the Society of Antiquaries. In the Netherlands archaeologists since 2007 are formally obligated to deposit their data with the Data Archiving and Networked Services (DANS, established 2005), according to the Quality Standard for Dutch Archaeology (Kwaliteitsnorm Archeologie). The DANS-EASY system includes the E-Depot Dutch Archaeology (EDNA); over 80% of the data deposited in EDNA are publicly accessible.[77]

Where a mandated community archive is not available the situation is very likely the same as in Germany about which a questionnaire survey of the IANUS Research Data Centre initiative in 2013 provides insights (IANUS 2014). The 243 respondents were from the field of "Altertumswissen-schaften" (Prehistory, Archaeology of the Ancient World, Classical Archaeology and other specific fields) and held positions at universities (45%), other research institutions (32%), museums (8%), cultural heritage departments (7%), and various others (8%).

21 of the respondents said that they allow download of their data from an online portal. Only documentation of data was provided by 19 on an openly accessible portal and by 29 on a restricted access portal. Most respondents kept data of completed projects at arm's length on carrier media (136), their computer (98) and/or local network/central server (82). Only few deposited data in a professional archive (11) or provided it to a data center (32). The data was mostly maintained either by the researcher (96) or institute (81). About half of the respondents were obliged to maintain the data, while others were not (15%) or did not know (34%). Besides obligations formulated by research funders, heritage management bodies, own institution or a professional standard, many respondents were motivated personally to maintain the data. However the majority (161) declared that they would make data available to third parties only on individual request.

Barriers to open data sharing

Surveys show that the majority of researchers in different disciplines have a positive attitude to sharing of research data (e.g. Tenopir *et al*. 2011), only a minority however shares data in an open manner such as depositing it in an accessible repository. Researchers still perceive more obstacles than incentives for open data sharing (Borgman 2010; LeClere 2010; Pryor 2009; RIN 2008; RIN & NESTA 2010). Among the main obstacles are

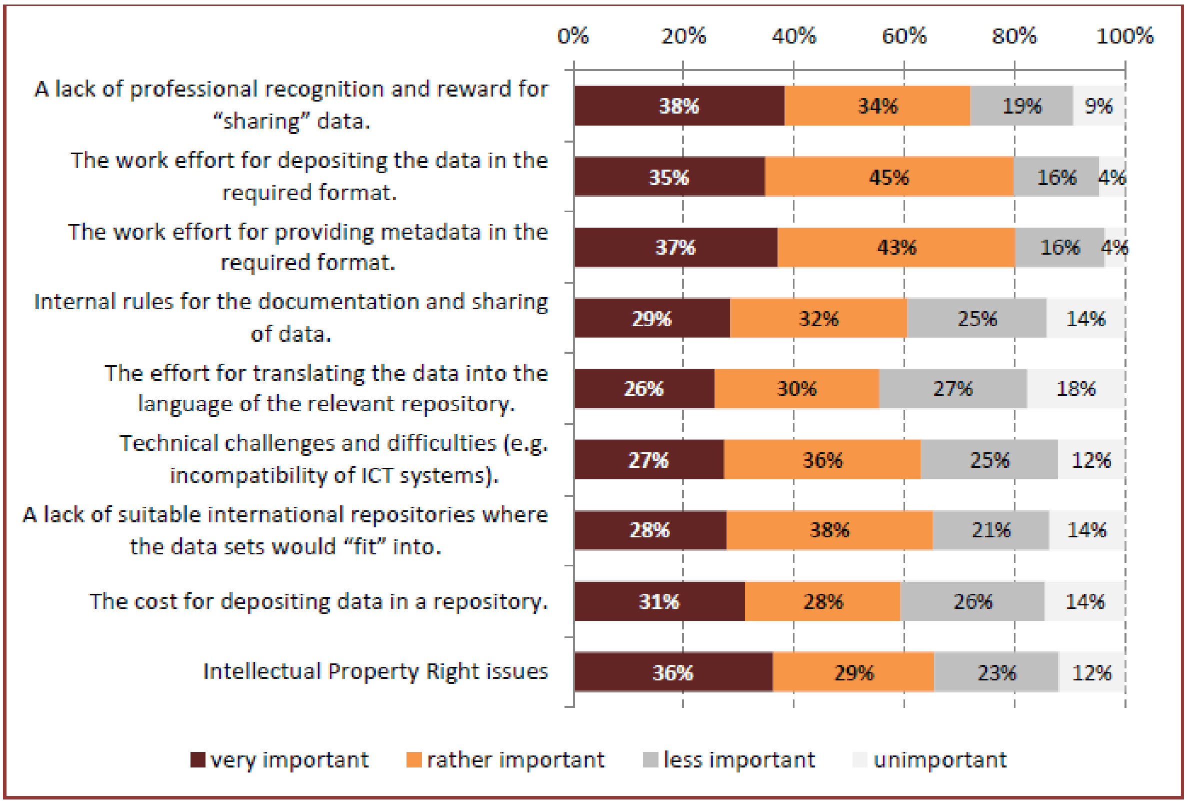
- Little academic recognition and reward for data curation and sharing (priority of publications),
- Required additional effort for providing shareable data, including formatting, metadata creation, licensing, among others,
- Missing or not appropriate data repositories, including lack of confidence in the sustainability of newly established repositories,
- Unclear data ownership and rights of use, copyright issues, confidential and sensitive data,
- Concerns that the data might be misused or misinterpreted.

In the ARIADNE survey the required additional effort for the preparation of shareable data and lack of professional recognition and reward for data sharing were perceived as the strongest barriers. But also other factors like IPR issues and lack of appropriate international archives (e.g. subject-based data repositories) show high scores.

[76] Archaeology Data Service (ADS), established 1996 (until 2008 Arts and Humanities Data Service – Archaeology), http://archaeologydataservice.ac.uk

[77] DANS: E-Depot Dutch Archaeology (EDNA), http://www.edna.nl

Figure 6.2-22: Question C.9 – "The following table describes potential barriers for researchers to deposit their research data in digital repositories and sharing them with colleagues. How important are these barriers in your view?"



N = 488-512 per item (depending on number of respondents without answer)

The results of the ARIADNE survey are consistent with many other studies in which lack of academic reward and additional effort for sharing data appear to be the strongest obstacles to open sharing of re-usable data. Making the case for this sharing even more difficult, many researchers fear a loss of control if their data it is accessible to others, e.g. potential misinterpretation and professional vulnerability if shortcomings of data are detected (Rouder 2015: 13-14). The benefits of open data sharing are not apparent, rather the personal return on investment may be negative because of a competitive disadvantage, for example, if full exploitation of the data through publications is not possible.

Consequently attempts at making researchers more willing to share their data must demonstrate that sharing produces measurable personal benefits that outweigh the additional effort and potential risks. In this difficult equation, it can be assumed that scholars are striving for academic reputation, expressed e.g. in recognition and many citations of their works, and also expect this if they share data. For instance, Fecher *et al*. (2015b) surveyed 1564 academic researchers of different disciplines on their motives for sharing or withholding research data. By far the strongest enablers of data sharing were *"if I had enough time beforehand, to publish on the basis of my data"* and *"if I were cited in publications using my data"*.

Also other studies make clear that researchers must be granted the time to exploit their data appropriately, i.e. sufficient time until data of funded research has to be archived or embargos on deposited data. Furthermore it is understood that open data will not become the norm if the costs of preparing and making available such data are not factored in (e.g. in research grants, funding of research community archives), and if the academic credit system does not reward data sharing.

A survey conducted by the FAIMS project in Australia illustrates the case for archaeology. The survey on the use of information technology included one question about sharing primary datasets

(Sobotkova 2013). Of the 79 respondents 41% were academic and 37% consulting archaeologists; others such as government or museum employees were less present, 4% and 3%, respectively. 46% of the respondents were willing to share data after they had finished their own publication, 24% before publication but only with selected persons or groups, 20% without restriction (even before publication); 5% said that they are prohibited from sharing by their employer, and 5% were "not at all" willing to share their data. When asked *"Would you agree that peer reviewed publication/sharing of data online should be given research credit or professional acknowledgement as a publication?"*, 86% of survey respondents agreed with the idea of allocating credit for the publication of substantial, peer-reviewed primary datasets; 14% opposed this idea.

The special case of archaeological research data

There are at least three factors that make an open data imperative particularly strong for archaeology: Excavation of sites destroys the primary archaeological evidence, archaeological work is conducted in the public interest, and there is little commercial interest in archaeological data.

Excavation is destruction

In the first place, there is the fact that excavation destructs the primary archaeological evidence, e.g. associations and stratigraphic relationships between various remains. As archaeological sites are limited and non-renewable resources of evidence it is argued that excavation should be avoided (Scarre 2014) or, if interventions are necessary, the disturbance of archaeological deposits should kept as minimal as possible (e.g. in preventive archaeology).

The Standard and Guide to Best Practice in Archaeological Archiving in Europe notes the destruction of primary evidence as a unique aspect of archaeology: *"By its very nature, archaeological field research destroys its own evidence by removing objects from their context by excavation. This makes archaeology unique compared to other scientific disciplines. As observations in the field can never be repeated, the process of excavation must be carefully registered and documented."* (ARCHES 2014)[78]

The fact that an excavation cannot be replicated entails an ethical and scientific responsibility of archaeologists to document the excavation as comprehensively as possible, and make the documentation available so that others can validate the empirical findings and use the data for further research (e.g. comparison between sites). The same may be argued for data acquired in scientific analyses of objects from excavations. Such analyses can be replicated, but the physical objects are not easily available to other researchers and reproducing analyses would not make sense, at least if there are no doubts about the presented results.

There are of course many forms of data acquisition in archaeological field work which are non-destructive and can be repeated rather easily, for instance geo-physical prospection, airborne remote sensing etc. for which an open data imperative is less obvious.

Archaeological work is conducted in the public interest

Archaeological fieldwork as well as other research is conducted in the public interest in archaeological heritage and knowledge. For example the European Convention on the Protection of the Archaeological Heritage (1969, revised 1992) aims *"to protect the archaeological heritage as a source of the European collective memory and as an instrument for historical and scientific study"*,

[78] The Standard and Guide to Best Practice in Archaeological Archiving in Europe (2014), http://archaeologydataservice.ac.uk/arches/Wiki.jsp?page=INTRODUCTION; in this context also the difficult situation of many physical archives of excavation material must be noted (Childs 2006; Edwards 2013; The Archaeologist 90/2013).

and considers that the preservation and study of the archaeological heritage *"help to retrace the history of mankind and its relation with the natural environment"*.[79]

The primacy of public interest in the protection, study and communication of archaeological heritage exists irrespectively if the work is paid from the public purse, by a charity, or by a private for-profit company (e.g. a developer). Arguably the majority of archaeologists in Europe undertake work that relates to protecting or rescuing archaeological heritage in the context of land-use where change ("development") can significantly affect the heritage, and the Environmental Impact Assessment (EIA) Directive applies.[80] It should be clear that the results of archaeological mitigation and rescue actions ("preservation by record"), but also other fieldwork and examinations, need to be documented, curated and accessible to other researchers.

No commercial value of the data

There is little scope for commercial exploitation of archaeological data, which is an aspect archaeology shares with some other domains, for example astronomy and oceanography. This does not mean that there are no commercial interests in archaeological research. For instance, there is the issue of who is allowed to conduct rescue excavations in the context of development work (only public institutions or also private, for profit companies); there are companies selling data capture equipment to archaeologists, etc. But the produced data as such has little or no commercial value. Therefore "hoarding" and "withholding" archaeological data must have other reasons, in which the economics of data sharing play an essential role (cf. *Section 4.3.4*).

High responsibility but low sharing of data

Given that archaeological research data presents a case with particularly high responsibility that the data becomes publicly available, one may wonder why archaeology is not one, if not the, flagship discipline of open data sharing. Indeed, many excavations did and still do not comply with the special responsibility, the results of many are not even published at all[81].

The Archaeology Data Service describes the situation as follows: *"Archaeology is in a special position in that much of the creation of its data results from destruction of primary evidence, making access to data all the more critical in order to test, assess, and subsequently reanalyse and reinterpret both data and the hypotheses arising from them. Over the years, archaeologists have amassed a vast collection of fieldwork data archives, a significant proportion of which remain unpublished. Access to data, even those which are published, is often difficult or inconvenient at best".* [82]

[79] Council of Europe: European Convention on the Protection of the Archaeological Heritage (Revised), Valetta, 16.I.1992, http://conventions.coe.int/Treaty/en/Treaties/Html/143.htm; other relevant Council of Europe conventions are the Convention for the Protection of the Architectural Heritage of Europe, Granada, 3.X.1985, the European Landscape Convention, Florence, 20.X.2000, and the Framework Convention on the Value of Cultural Heritage for Society, Faro, 27.X.2005. While the "Valetta Convention" (cited above) focuses very much on the evidential value of archaeological heritage, the more recent conventions, especially the "Faro Convention" have a wider scope, involving communal values and participation.

[80] The European Environmental Impact Assessment Directive (http://ec.europa.eu/environment/eia/eia-legalcontext.htm) addresses "cultural heritage" in Article 3 and mentions "architectural and archaeological heritage" in the listing of aspects of the environment that may be significantly affected by a proposed development project (Annex IV). The EIA Directive of 1985 and its three amendments (1997, 2003, 2009) have been codified by the Directive 2011/92/EU (13 December 2011) and recently amended by the Directive 2014/52/EU (16 April 2014).

[81] Called "archaeology's dirty secret" by Brian Fagan (1995) and often quoted in the archaeological ethics literature.

[82] Archaeology Data Service: What is the ADS, and what do we do? Accessed 2015-04-26 at: http://archaeologydataservice.ac.uk/about/background

Eric Kansa of the data publishing service Open Context (Alexandria Archive Institute) writes, *"One would naively think that data sharing should be an uncontroversial 'no-brainer' in archaeology. After all, archaeological research methods, particularly excavation, are often destructive. Primary field data documenting excavations represent the only way excavated (i.e. destroyed) areas can ever be understood. One would think this would make the dissemination and archiving of primary field data a high priority, particularly for a discipline that emphasizes preservation ethics and cultural heritage stewardship"* (Kansa 2014b: 36)

However, Kansa observes that the priority of proper data archiving and sharing is often not followed. As the main reasons he notes traditional modes of conduct and publishing practices that remain highly resistant to change, despite the fact that the archaeological recording of excavations and other field work has become digital. Curation and publication of the data is largely avoided as it does not fit into the institutional regimes of recognition and reward.

It does not come as a surprise that the research community so far has been remarkably silent about openness of archaeological data and research practices. But there are some voices arguing for "open archaeology" (e.g. Beck & Neylon 2012; Costa *et al*. 2012; Kansa 2012; Lake 2012; Wilson & Edwards 2015), new journals for addressing issues and making other researchers aware of open data have been established recently (e.g. *Open Archaeology*[83], *Journal of Open Archaeology Data*[84]), and there are some groups pushing for open data (e.g. Open Pompei[85] in Italy). Much attention also received the ARIADNE Open Access Session at the Annual Meeting of the European Association of Archaeologists (EAA) on the 3rd of September 2014 in Istanbul (Geser 2014b). But it will still be a long way until the discussion about open archaeology becomes centre stage.

Concerning current practices of data management and access of archaeological projects, it may be argued that a non-sharing mindset already impacts negatively on project-internal data organisation. For instance, if in an excavation the data flow is centralized around the site director, who may wish to control the data but is not versed in state-of-the-art data management. One of the ARIADNE Special Interest Groups conducted a survey of excavation data management systems and concludes that the traditional, hierarchical structure of excavations impedes the adoption of innovative solutions in terms of technology and openness. The group assumes that team-driven research, a stronger role of data managers and technical assistants, and use of open source software could allow data to become more accessible (ARIADNE Excavation and Monument Data SIG, 2015).

4.3.4 Some basic economics of data sharing

Data from publicly funded research is typically not sold and bought, hence by economics we mean frameworks and considerations of the data producers which determine if they do or do not share data. The concept of *data sharing* (used throughout this report) covers very well the characteristics of the "data market" in the field of academic research.

Researchers often share data directly with colleagues, however this means that the data remains within a small circle of peers and is not available for other researches or the public at large. The goal of open sharing is to have the data deposited in accessible archives to allow for re-use in further research and other purposes. We use the concept of open data sharing to emphasise requirements that need to be fulfilled to move from non- and closed-circle sharing to open sharing of data.

[83] Open Archaeology (since 2014), http://www.degruyter.com/view/j/opar

[84] Journal of Open Archaeology Data (Ubiquity Press, since 2012), http://openarchaeologydata.metajnl.com

[85] Open Pompei, http://www.openpompei.it; aims to ensure transparency and open data in the context of the intended large-scale investment on the Pompei site (over 100 million EUR), but also in the profession overall, e.g. through the Archaeological Open Data Manifesto, open data training schools and hackathons.

The related concept of *data publication* (Kratz & Strasser 2014) fits as well, as it allows to frame sharing of data through accessible archives as publication, and relate to the familiar notion of a product that "counts" in the academic system of recognition and reward. There are of course considerable differences between traditional forms of publication and sharing/publication of data which, however, may gradually disappear in the digital environment with novel forms of so called "enhanced publications".

Contrary to Parsons & Fox (2013) we do not question if data publication is the right metaphor, rather we think that at this stage it is still the best analogy. As Eric Kansa of the open access data dissemination platform Open Context notes, *"The phrase 'data sharing as publication' helps to encapsulate and communicate the investment and skills needed to make data easier to reuse. It conveys the idea that data dissemination can be a collaborative undertaking, where data 'authors' and specialized 'editors' work together contributing different elements of expertise and taking on different responsibilities"* (Kansa 2014a). Similar to conventional publishing of papers or books data sharing through an accessible archive involves effort and expertise of both the researchers, who know best their data, and professional data curators, who know what is required to describe and make data available online for re-use.

Publications as the main academic currency

It is clear that researchers see their data as an asset to be exploited as far as possible for publications. Publications are the basic currency in the system of academic recognition and rewards (e.g. tenure and promotion) as well as necessary to underpin applications for research grants. If one's career largely depends on prestigious publications (ideally scholarly monographs or papers in high-impact journals) and activities in academic and professional circles, requests to invest effort on curating and sharing data must appear as irrelevant, counter-productive, if not damaging. As noted by one scientist, *"Although things appear to be changing, the idea of sharing data seems counterintuitive. Grant money is harder than ever to come by, and competition for space in professional journals is fierce. Factor in the desire for personal advancement and prestige, and the allure of possibly striking it rich, and it would seem that sharing data, although a lofty ideal, hasn't got a prayer if it remains voluntary"* (Samson 2008: 13).

Mainly direct sharing of data between colleagues

In the established academic system individual benefits of open data sharing are not apparent, rather the return on investment may even be negative, e.g. competitive disadvantage if full exploitation of the data through publications is not possible, loss of control of data once it is accessible to others (e.g. misinterpretation of the data), etc. Therefore data is primarily shared directly between trusted colleagues of the research community. In a survey of 1564 academic researchers of different disciplines 58% said that they have shared data with other researchers they know personally, 49% with colleagues at their institute, and 40% with scholars that work on similar topics, while only 13% shared data publicly (Fecher *et al*. 2015b).

Hence researchers *do* share data, but based on personal relationships and collaboration. This so called "gift culture" of scholarship is a disincentive for depositing data in a repository for anyone to use, because the researchers lose the ability to barter data privately. Data sharing between colleagues allows a direct and reciprocal relation, and ensures that the data are understood and used correctly. Importantly, it does not require the data sharers to produce metadata, because the colleagues are familiar with the types of data and procedures employed in their creation, and unclear aspects may be clarified easily.

As in all social relations, trust is a core factor also in data sharing: *"Questions of trust loom large here, and run both ways. Can I trust those I agree to share my data with to make reasonable and appropriate use of it, and on a timeline which doesn't impede my own requirements re: publication,* *credit, and priority? On the receiving end, can I trust the data I'm getting, particularly as collaborative webs lengthen and my first-hand knowledge of the data and its producers recedes?"* (Edwards *et al*. 2007: 32-33). Therefore researchers are mainly willing to share data directly within their research group, with close colleagues of other institutes, and in collaborative projects. They are much less inclined to make data openly available (to anybody, including commercial actors) or re-use data of providers of research fields they are not familiar with. Also data of "citizen scientists" are unlikely candidates of re-use, if the data collection is not controlled based on scientific standard.

Automated vs. hand-collected data

In the data sharing debate special attention deserve the conditions under which data is acquired. It appears that practices of sharing (or not sharing) are strongly determined by the different ways and contexts of data generation in different subject areas. Factors of the respective economics are access to required resources (e.g. research material, facilities, technical support), the degree of automation of research work, etc. One factor that makes researchers more inclined to share research data is if the data can be largely produce automatically (Pritchard *et al*. 2005: 9-11). Automated data capture reduces the perception of a high personal effort invested in data acquisition and, thereby, the sense of data ownership.

Astronomy is a forerunner in open data because the data is captured automatically and data sharing is expected, because of limited availability of the costly instruments for the research groups. The allocation of usage time is usually decided by an international committee. The data must be released after a grace time of one year or less during which the research group can be the first to publish findings based on the newly acquired data. The archived data is often used by other researchers, for instance the amount of papers written based on data in the Hubble telescope archive is higher than the number of papers resulting from new observations (Hubblesite 2011).

Another example is the field of genetics in which DNA sequencing is conducted automatically and with massively increasing output. The reference model for open data here is the Human Genome Project in which data has been produced by different laboratories and made available prior to publications (Birney *et al*. 2009). While this is seen as the "gold standard" in the field, it has been enabled by large-scale funding and setting clear rules participants had to follow. Genetic data sharing however has increased significantly because funders and journals require data archiving. A study explored the availability of datasets underlying 11,603 papers on gene expression microarrays published between 2000 and 2009 (Piwowar 2011). Available data was identified in best-practice repositories for on average 25% of the papers, increasing from less than 5% in 2001 to 30%–35% in 2007–2009. The study found that *"authors were most likely to share data if they had prior experience sharing or reusing data, if their study was published in an open access journal or a journal with a relatively strong data sharing policy, or if the study was funded by a large number of NIH [US National Institutes of Health] grants. Authors of studies on cancer and human subjects were least likely to make their datasets available."*

It should be noted that DNA sequence or protein structure data independent of a particular experimental context is very basic data. The closer the context gets to commercial exploitation, the less data sharing takes place. As explained by a researcher involved in both genomics and brain cancer research: *"Genomic data we share freely because there's so much data and so many ideas that you just don't care – it's wide open. But in our private research we try not to share anything until a patent is awarded - period. (…) With private research and commercial trials data, the rule is you don't talk about it – not on cell phones or even land lines – there's a lot of paranoia, but it may be justified"* (cited in Samson 2008: A14).

Closer home to archaeology, the sharing of data of ancient human DNA analyses (mitochondrial DNA, Y chromosomal, autosomal) has been shown to be "close to 100%", much higher than in evolutionary, forensic and medical genetics (Anagnostou *et al*. 2014). Of 207 analyses reported in

162 papers published between October 1988 and December 2013 only the data of 5 analyses were withheld (i.e. only data-derived statistics were presented). The study authors counted all forms of making data available, in the body text of the paper, as supplementary material, in an open online database or repository. In-text presentation of the data was most common (on average 57.7%), supplementary material came next (23.8%), an online database was used for about 20% of all analyses for which such a database was available (e.g. GenBank), and an online repository was used only by one study. The authors stress that it is important to consider what sharing options for different types of data are actually available and if the best one in terms of preservation, access and re-use are employed. They suggest ancient human DNA analyses as a "flagship research field" of data sharing, but again under the condition of no commercial interest, mostly publicly funded research, and increasingly database supported data sharing (since about 2006).

A different case, involving "hand-collected" data, is crop QTL [Quantitative Trait Loci] mapping. Zamir (2013) notes that for over 5000 plant QTL mapping papers in the Web of Knowledge database the raw data of less than 1% is publicly available. QTL mapping is conducted to identify which chromosomal regions are associated with plant traits that are of agricultural importance, e.g. disease or drought resistance. The phenotypic crop data of the published studies, which is "hand-collected" in multiple locations and field conditions, is not made available.

In archaeology a lot of data, especially excavation data, is "hand-collected", e.g. sieving and counting of small biological remains, documentation of archaeological finds with recording templates, "double entry" of data, i.e. first on paper templates then into a database. Attempts at introducing digital documentation in the trench with hand-held devices have not been as successful as envisioned, because of advantages of paper and pen in the recording process as well as work patterns of on-site archaeological teams (cf. Warwick *et al*. 2009; Terras *et al*. 2010, based on observations in the Silchester Town Life Project in 2007/2008).

In recent years the use of "paperless" recording applications in field survey and excavation projects, mostly with Apple iPads (available since 2010) or other tablet devices, has become more widespread. Some examples since 2011/12 in Europe are the Pompeii Archaeological Research Project: Porta Stabia (PARP:PS, Italy)[86], Sangro Valley Project (Tornareccio, Abruzzo, Italy)[87], Zagora Archaeological Project (Andros island, Greece)[88], Athienou Archaeological Project (Cyprus)[89] and Pyla-Koutsopetria Archaeological Project (Cyprus)[90], Eastern Vani Survey (Georgia)[91].

Users in general report improvements in data collection, processing and availability to team members, but a re-organisation of workflows (especially in excavations) and (re-)training of staff is required[92]. Among the advanced mobile recording systems is the digital environment developed by PARP:PS and the system offered by the FAIMS - Federated Archaeological Information System

[86] Pompeii Archaeological Research Project: Porta Stabia, http://classics.uc.edu/pompeii/

[87] Sangro Valley Project, http://www.sangro.org

[88] Zagora Archaeological Project, http://www.powerhousemuseum.com/zagora/

[89] Athienou Archaeological Project, http://sites.davidson.edu/aap/

[90] Pyla-Koutsopetria Archaeological Project, http://pylakoutsopetria.wordpress.com

[91] Eastern Vani Survey, http://www.lsa.umich.edu/kelsey/fieldwork/currentfieldwork/vanirepublicofgeorgia

[92] cf. Averett *et al*. 2014: 53-55 [Athienou Project]; Wallrodt 2013 [PARP:PS project]; Motz & Carrier 2012 [Sangro Valley Project], Caraher 2013 [Pyla-Koutsopetria Project]; Hughes 2012 [Eastern Vani Survey], see also Goodale *et al*. 2013).

project[93] (trialled for example in surveys of the Zagora Project: Havlicek & Wilson 2012; about FAIMS see Ross *et al*. 2013).[94]

The question in our context is if the advantages of mobile data collection will have any effect with regard to open sharing of the data. For the mentioned projects we did not find evidence for such an impact during the project, e.g. data made available on project websites. But the Pyla-Koutsopetria Archaeological Project (Cyprus) of American archaeologists made about 9000 data items accessible through the Open Content data publication platform; the data is archived with the California Digital Library.[95] The data has been published online in 2013, which is about 10 years after the start of data collection.

4.3.5 Data licensing

The definition of open data includes that the data should be licensed with no or only minimal restrictions on re-use. This is required to enable legitimate and effective re-use of data/content which include also metadata as well as terminology and vocabularies[96]; in what follows "data" are meant to cover all of those.

As IPR/copyright laws set as the default "all rights reserved" rights holders must state clearly what uses of the data they allow[97]. Data producers may feel comfortable with a situation of "all rights reserved", which requires that others who wish to re-use some data must ask them for permission. But such a situation is hardly acceptable for archaeological research data which has been publicly funded and mostly concerns public property. One ARIADNE partner, who aims to follow "*the principle of free access to data, providing data to everybody who needs it"*, points out major problems to fulfil this objective in their country:

"[there is] a collision, or at least ambiguity in the law concerning authors´ rights which might be understood in such way that for using any part of the archaeological field documentation you have to ask its author for permission. This is, of course, not possible and it is generally not respected. So far, these matters never got to court, so there is no precedent how such a problem would be legally decided. Anyway, it would be very useful to stimulate a legal change or a European instruction about this matter. We understand all archaeological field documentation to be part of the project (paid from public means and concerning archaeological finds which are public property) and, therefore, should be generally available. In case anyone who uses any information should quote the source (author), but the presumed duty to ask the author before seems ridiculous in the present world".

[93] FAIMS Mobile Platform, https://www.fedarch.org/wordpress/faims-mobile-platform/

[94] The recent workshop "Mobilizing the Past for a Digital Future: the Potential of Digital Archaeology", Wentworth Institute of Technology, Boston, 27-28 February 2015 has brought together several presenters of projects which use mobile recording tools (Averett *et al*. 2015 [programme]; Wallrodt 2015 [videos].

[95] Open Context: Pyla-Koutsopetria Archaeological Project, http://opencontext.org/projects/3F6DCD13-A476-488E-ED10-47D25513FCB2

[96] Vocabularies or knowledge organization systems (KOSs), e.g. thesauri or ontologies, are generally included under Open Data and Linked Open Data (LOD). For example, the Getty Research Institute publish their thesauri as LOD, http://www.getty.edu/research/tools/vocabularies/lod/ (license: Open Data Commons Attribution License - ODC-By 1.0); the UK vocabularies for cultural heritage made available by the SENESCHAL project as LOD are licensed under Creative Commons (CC BY) or under the Open Government License for public sector information (http://www.heritagedata.org/blog/vocabularies-provided/).

[97] As noted, we here use "data" for all kinds of work while "facts" such as data expressed as numeric values are not protected by IPR/copyright law, but the manner in which they are selected, arranged and presented can be protected. Carroll (2015) provides an excellent "primer" on IPR/copyright law and sharing of research data.

This situation obviously is due to a lack of proper data licensing which should be addressed by the authorities responsible for the preservation and research on archaeological heritage resources.

Most referenced as appropriate open data licensing solutions are some of the licenses offered by Creative Commons (CC)[98] and Open Data Commons (ODC)[99] (cf. Ball 2014). The licenses are formalized (but explained jargon-free, CC in many languages), can be embedded in any digital content, and are machine-readable. The ODC licenses have been conceived particularly for data/databases, however CC licenses at least from version 4.0 can serve the purpose to make them openly available just as well[100].

Korn & Oppenheim (2011) note that the CC and ODC licences do not guarantee to provide any information about third party data a licensed resource may contain, and of course do not include any indemnities for the user in the case that they do. Data providers therefore are strongly adviced to seek permission of third-party holders of rights to distribute it under an open license.

Public Domain

No restrictions on data re-use means that it is placed in the Public Domain. The licenses available from Creative Commons and Open Data Commons are CC Zero (CC0) and ODC-PDDL (Public Domain Dedication and License), respectively. The licenses facilitate the release of data and databases into the Public Domain. The copyright owner waives all rights, including the European database rights[101] and the right to be identified as the creator. If data is already in the Public Domain, it can be marked using the CC Public Domain Mark.

A Public Domain dedication fits well for the metadata of content collections or databases, for example, Europeana publishes metadata received from its data providers based on the Europeana Data Exchange Agreement under the terms of the CC0 1.0 (Universal Public Domain Dedication).

BY (Attribution)

As particularly appropriate for data are considered licenses of type “BY” (Attribution) which allow that data sharers are recognised and receive the credit they deserve. Such licenses are CC-BY and ODC-by which only require proper attribution.

Share-Alike

The condition Share-Alike (e.g. CC BY-SA or ODC-ODbL) allows data users who modify, transform or build upon the data(base) to distribute the resulting work only under “the same, similar or a compatible license”. This can impact negatively on the ability of intermediation services to combine and utilize (meta)data effectively; e.g. (meta)data licensed under CC-BY and CC-BY-SA cannot be combined and distributed. Data federation and integration initiatives therefore recommend not to

[98] Creative Commons (CC) licenses, https://creativecommons.org/licenses/?lang=en

[99] Open Data Commons (ODC) licenses, http://opendatacommons.org/licenses/

[100] Creative Commons: Data, https://wiki.creativecommons.org/wiki/Data

[101] Directive 96/9/EC of the European Parliament and of the Council on the Legal Protection of Databases, 11 March 1996, http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31996L0009; the EU law on database rights was introduced to recognise the substantial investment made in compiling databases, and to prevent unauthorized copying or re-use of their content. Database rights are established automatically and cover both substantial extraction and copying of the database, and also piecemeal copying of data and subsequent reassembly. In principle, non-substantial or “fair use” is possible, but any substantial copying of relevant data requires obtaining permission and agreeing terms of use with the database owner. In addition to the database rights, the arrangement, selection and presentation of the data may also be protected by copyright. Guibault & Wiebe *et al*. (2013) provide extensive discussion of the database right, *“by far the most important IP right in the context of e-infrastructures”* in the context of OpenAIRE and other e-infrastructures.

apply licenses including the Share-Alike condition. Such initiatives should generally avoid the use of different, incompatible open licenses of (meta)data providers.

No-Derivatives

In the context of open sharing of research data it should be clear that a license including a No-Derivatives condition is inappropriate as it does not permit to distribute modified, derived or re-purposed data(base). Creative Commons does not recommend use of its NoDerivatives (ND) as well as NonCommercial (NC) licenses on data(bases) intended for scholarly or scientific use. None of the Open Data Commons licenses includes a ND or NC condition.

Non-Commercial

Licenses which include Non-Commercial (e.g. CC BY-NC or CC BY-NC-SA) allow use of the data for non-commercial purposes only. The Non-Commercial (NC) condition has been very popular among users of CC licenses, however poses many difficulties. They include that NC licensed data cannot be combined and distributed with other CC licensed material as well as other, often unwanted consequences by licensors, e.g. if a not-for-profit organisation needs to generate some income to sustain its work (Möller 2007; Hagedorn *et al*. 2011; Klimpel 2012)[102]. Many authors suggest using ShareAlike instead of Non-Commercial because it allows commercial use but requires that works that re-used content, including some “added value”, are shared under the same license.

In the context of publicly funded research it seems inappropriate to apply the non-commercial condition to the data generated, because the public purse is filled by tax-paying commercial enterprises as well as citizens. Some research funders take an active stance against prohibiting businesses benefit from outcomes of publicly funded research. Most prominently, the European Commission aims to ensure *“that the results of Europe’s publicly-funded research are fully accessible for researchers, businesses and citizens”* (European Commission 2012: 10).

Goals of this endeavour include acceleration of innovation, fostering economic growth, addressing societal challenges, and improving the transparency of the scientific process. Furthermore the European Commission in this context highlights the important role of e-infrastructures such as OpenAIRE (for publications and datasets resulting from EU-funded projects) and thematic data infrastructures for storing and providing access to data.

Finally it seems appropriate to mention the (preliminary) recommendations for the ARIADNE data sharing policy framework, which have been derived from studying best practices, a survey of data sharing and access practices within the consortium, and consultation with project partners[103]. The framework is meant mainly for data made available through community-level archives and institutional repositories. It is recommended that data sharing partners be requested to provide:

- A collection description (of the whole collection and sub-sets within the collection) published under a CC BY licence for each dataset ingested into the ARIADNE infrastructure. Collection description is a useful way of capturing the provenance and contextual information about data collections, and can be used to underpin data citation.
- Item level metadata records should be published under a CC0 licence – to enable integration of multiple datasets within the metadata repository, support resource discovery and enable Linked Open Data.

[102] The development of the Creative Commons 4.0 license suite included an extensive consultation on the NonCommercial condition (Creative Commons 2013). The final decision was not to modify or rename it, but improve the communication of its meaning and implications. One major reason for no changes was that a study on Internet users’ understanding of “non-commercial” highlighted the need for caution *“so that expectations of those using NC licenses are preserved, not broken”* (Creative Commons 2009).

[103] ARIADNE D3.3 Report on Data Sharing Policies, January 2014 (p.37), http://www.ariadne-infrastructure.eu/Resources

- The CC licences suite (version 4.0 is preferred) should also be used for content (databases, document archives, images, 3D models, etc.) provided to ARIADNE by partners under licence permissions agreed with the content owner. CC BY is recommended for open access; other, less open CC licenses may also be applicable.

Moreover the recommendations for the data sharing policy framework include that the allocation of persistent unique identifiers (e.g. DOIs) to datasets ingested to the ARIADNE infrastructure should be investigated. The system used should be capable of identifying sub-sets within collections. Persistent identification of datasets is important in underpinning data sharing and data citation.

Furthermore a common method of data citation should be established for adoption by partners and promotion by ARIADNE to the archaeological research community. This relates to the fact that academic recognition (formally expressed by citation) is an important motivation for encouraging researchers to share access to their datasets.

4.3.6 Suggested actions

This focus area mainly concerns the challenge of growing an open culture of research of which open data is but one, although an important facet.

Promote a research culture of open sharing of data

Archaeological data archives/repositories and the ARIADNE data infrastructure will flourish only within a research culture that values open sharing of data. Therefore promoting this culture is of vital interest to data service providers which, however, have a supportive role. The main enablers are research groups, institutions, associations and funding bodies who opt for openness. There are many good arguments for open data such as preventing duplicative data collection, allowing replication of data analysis to scrutinize knowledge claims, and enabling new research questions to be addressed based on shared data. In particular, data that is shared, curated, integrated, re-used and analysed with new methods gains in value rather than being only a cost factor.

Support strict open data policies of funding bodies and institutions

Empirical studies make clear that only decisive open access policies for publications and data work effectively. Therefore it is recommended that funding bodies and research institutions issue strict mandates for open data sharing. Data archives and data infrastructure providers should give full support to such mandates.

Suggest appropriate guidelines for open archaeological data

A "one size fits all" approach may not work for archaeological data. In case open data guidelines do not take account of specificities of archaeological data (e.g. sensitive data such as findspots, human remains, knowledge of indigenous peoples), research institutions and associations should promote community consultation and suggestion of appropriate guidelines. The move towards open data offers an opportunity to discuss and get clear about disclosure or non-disclosure of certain information.

Recommend use of open licenses for data, metadata and knowledge organization systems

Research funders, institutes, archives and e-infrastructures should recommend (or mandate) that data, metadata and knowledge organization systems (e.g. thesauri) are shared under an appropriate open license. Data licenses should not impede effective re-use and further dissemination of the data, but ensure attribution of data sharers. As appropriate are considered Creative Commons (CC), Open Data Commons (ODC) and other licenses of type "BY" (Attribution), which ensure that data sharers (researchers, institutes and others) are recognised. For metadata the most open licenses are recommended, e.g. CC Zero (CC0), ODC-PDDL or other public domain dedication licenses. Licenses containing a No-Derivatives condition should generally be avoided, and there are strong arguments against Non-Commercial; applying instead ShareAlike will impact negatively on the ability of intermediation services to combine and utilize (meta)data effectively. Wide application of open licenses is necessary to remove barriers to legitimate and effective re-use, extension and integration of data(bases), metadata and knowledge organization systems (e.g. thesauri).

Help ensure that open data sharers are recognised and rewarded

Archaeology should be a leading example of open data sharing, because of factors that make an open data imperative particularly strong for archaeology: Excavation of sites destroys the primary archaeological evidence, archaeological work is conducted in the public interest, and there is little commercial interest in archaeological data. But there are many factors that work against open sharing of archaeological data. A particularly strong factor is little academic recognition and reward for making data available. The academic credit system values research publications while data sharing is considered much less, if at all. Therefore all stakeholders in open archaeological data should help ensure that data sharers are recognised and rewarded. Most impact in this regard will have institutions that make data sharing a criterion for academic promotion and funding agencies that take it into account when awarding new research grants.

4.4 Data re-use, citation, and altmetrics

Open data sharing through digital archives means enabling (re-)use of the data by others, researchers as well as other users. Because of the substantial effort this takes it is important to look into the use side and how data sharers can receive the credit they deserve. This section explores the topic of data re-use, on which only few studies are available as yet, but including some work on data re-use in archaeology.

The quest for open research data and rewarding data sharers also necessitates standardising data citation and integrating data in research information services. While open access publications did not pose much of a problem for these services, data does, also with regard to metrics and studies of "impact", which is a prime concern of research performance assessments. We describe the on-going efforts in this direction as well as summarise available studies on citation of open access papers and data as well as not citation-focused altmetrics (alternative metrics). One finding of this background study is that specificities of the humanities deserve special attention. We also suggest that data archives should take a special effort to present usage data, in particular, data which demonstrate the added-value of their data curation and dissemination work.

4.4.1 In search of evidence

There are many good arguments for open access data (cf. *Section 4.3.1*), but also many obstacles, in particular, lack of clear evidence of benefits for researchers who share their data. Many researchers appear to be unwilling to share their data openly through a data repository due to the additional effort and potential disadvantages, among other factors. Sharing may be good for science and society, but for the researchers, who in surveys and interviews often express that they would like to share it presents a dilemma (Scheliga & Friesike 2014).

Advocates of open data argue that such data will often be re-used and cited, bringing recognition and rewards to data sharers. But there is little empirical evidence as yet for such benefits, which makes open data sharing practices depending on the researchers' altruism or "sticks" such as strict open data mandates. This has a negative impact on data archives and e-infrastructures that are being implemented to curate and provide access to data resources and, thereby, support their re-use. Data archives and e-infrastructures could help mobilise more open data more easily if better evidence for benefits, especially benefits for individual researchers and research groups, are available.

There are voices that are very dismissive concerning claims that data re-use could bring tangible benefits for researchers who share data, at least not within a few years. For example, one American doyen of library and information sciences and practices considers:

"The vast majority of faculty will, at least in the near term, see little real benefit from making their data available for sharing. Despite work on data citation practices and on changing evaluation criteria for researchers, it will take a long time for faculty contributions of data for potential community reuse to make a compelling and widespread difference in tenure and promotion cases; the inertia and conservatism in this system is enormous. So developing and subsequently implementing data management plans will most often be viewed as just one more burden imposed by the funding agencies; faculty will want to satisfy these new requirements in the most time-efficient and easiest fashion. Some faculty (we don't know how many, or in what disciplines) will be very creative in exploiting the growing amounts of data available for reuse and will find their own scholarly work advanced. There will, of course, be some high-profile cases where faculty who obtain important new results through data reuse gain important recognition (keep in mind that the funders are eager to identify, encourage, and recognize these scholars). Even researchers who provide data that is subsequently reused to significant effect may find their contributions honoured – but there's a sizeable luck factor here, as it is not so much that they make data available for possible reuse as it is that they were lucky enough to have someone actually reuse it and then make an important discovery" (Lynch 2014: 397-8).

Important questions indeed are if open data will often be re-used and cited in a discipline, and if the data sharing and citation record of a researcher, research group or institute will be considered appropriately in the evaluation of research performance (e.g. in decisions on research grants, tenure and promotion, etc.). In general, we think that cautions as expressed in the above quote under-estimate the capacity of the research community to adapt properly to new demands of research funders and take advantage of the rules and means implemented for this purpose. Furthermore, that empirical evidence of tangible benefits of open data sharers could of course help a lot in driving the agenda of open research data.

4.4.2 Data value

The value of research data depends on various factors such as quality, uniqueness, risk of loss, repeatability, production costs, and potential for re-use. The archive community advises holders of research data (e.g. research projects and institutes) on general evaluation and selection criteria for valuable data which should be curated and made accessible to the wider research community. But, as Tjalsma & Rombouts note, *"Only researchers working in the academic discipline / subdiscipline itself can judge the academic value of 'their' data on the basis of content, whether it is data from experiments, simulations, observations or ancient but now digitised texts"* (Tjalsma & Rombouts 2010: 18-19).

Researchers indeed have notions about the value of their data. Within the established system of recognition and rewards they see it as an asset to be exploited as far as possible for publications. The possible re-use of data by the original producers for publications arguably is limited, and an approach of "salami publication" (Šupak-Smolčić 2013) based on the same data is not really rewarding or appreciated by the research community.

After some reasonable exploitation by the original producers, when the data is unlikely to allow for new insights, a change of research focus, etc. the data loses its value for the researchers. Then the data becomes obsolete, remains on PCs, carrier media, restricted access servers, and is eventually discarded or lost otherwise. On the other hand, data that is made accessible, re-used and enriched by other researchers (e.g. combined with their data) may gain in value and become a well-curated asset of the research community.

We assume that an increase in the value of research data can be achieved if the data is shared openly and re-used. Palmer *et al*. (2011) support this assumption and highlight the analytical potential of scientific data as a core factor of re-use value. Weber *et al*. (2012a) add that data increases in value through exposure to diverse contexts of use.

There is little evidence as yet for a citation advantage of publications that are shared with the data (cf. *Section 9.2.3*), or the unlikely case that researchers make data available without any related research publication or a “data paper” (described in *Section 6.6*).

Another valuable incentive may be that data sharers are invited as co-authors of publications which build on the data. But in a survey of 1564 academic researchers of different disciplines only 34% respondents considered this as a motivation for data sharing, while 44% rejected it (Fecher *et al*. 2015b: 9). Arguably co-authorship is more welcome if researchers not only provide the data but are involved in projects that re-use their data and can contribute to the study design and interpretation.

It appears that the quest for open research data must for now depend more on sticks than carrots, sticks of funders, publishers and peers. For example, journals which do not accept papers if the underlying research data is not accessible, funders that refuse to give grants to researchers who did not share the data of previous research, or peers who do not accept withholding of data that are vital for the research community. More favourable however is rewarding those researchers, research groups and institute who share and re-use data.

4.4.3 Data re-use

Barriers to data sharing have been studied extensively, while a solid base of knowledge about data re-use does not exist as yet. Therefore we collected and looked into available information on the topic which, fortunately, also includes some work on data re-use in archaeology.

A neglected topic of open data studies

Issues in the re-use of research data have so far received much less attention than those related to data sharing. As Wallis *et al.* (2013) note, *“Underlying the arguments for data sharing are assumptions that available data will be used or reused by others. However, surprisingly few studies have addressed how and when researchers reuse data they obtain from other researchers”.*

The lack of empirical knowledge about data re-use is surprising because the benefits associated with re-use are among the strongest arguments for open sharing of data. However studies on data re-use are not easy to undertake, and data archives often do not have information if downloaded data actually has been re-used. We expect however that data re-use will become an important research topic because of the increasing interest of all stakeholders, researchers, digital archives and research funders in the impact of open data.

Among the first things to consider and study for well-informed decision making on matters of open research data indeed is if researchers in certain (sub-)disciplines actually do re-use data of others, or would do so, if more re-usable data becomes available. “Just in case” certainly is not a good basis. If re-use is unlikely it would make little sense to invest in building digital archives and ask researchers to invest the effort to prepare their data and metadata for this purpose. The Royal Society’s report *Science as an Open Enterprise* notes, *“Sharing research data can be complex and costly and needs to be tempered by realistic estimates of demand for those data”* (The Royal Society 2012a: 60)

Wallis *et al*. (2013) emphasise the need to investigate which data are worth the effort for long-term curation, access and potential re-use as follows: *“Data sharing is perhaps better understood as the problem of making best use of research resources. Researchers produce large amounts of data, some* *of which may be useful to others. Making those data useful to others requires a substantial investment in documentation, and often in interpersonal negotiation, above and beyond the conduct of the research per se. It is not possible to justify making that level of investment in all data just in case someone, somewhere, at some future time, might wish to use them. The originating investigator bears the cost of data preparation. Other entities such as data repositories, universities, libraries, and funding agencies are likely to bear the cost of curating those data for sustainable access. Unknown – and often non-existent – reusers reap the benefits. This equation is not viable in economic or social terms. Thus, the better question to ask is which data are worth the investment for reuse?"*

Also Palmer *et al*. (2011) stress that there is a need to determine *"if and how much potential there is for re-use and how much investment should be put into curation to fill the knowledge base gap and improve the fit for new purpose"*.

If substantial re-use of data from digital archives can be assumed in a field of research, we may start asking questions about how to enable easy re-use of data once disconnected from their producers. For example, what provenance and context information should be documented to allow also re-use by researchers who are not familiar with the data generation practices in the particular field of research. Studies of research areas with frequent data re-use by different user groups may suggest best practices with regard to standards and services.

Data re-use ideally could be tracked and studied based on data citations in the research literature, however the infrastructure for such analysis is not available yet. Therefore studies often use the much wider concept of data usage which includes views and downloads of data as well as other so called altmetrics. Such metrics however remain at the surface of the matter. Detailed analyses of data re-use, for example, with regard to different contexts and forms, are necessary to understand practices of re-use (and lack of re-use) in different areas of research, including different areas of archaeology.

Purposes, forms and models of data re-use

Data can be used for many purposes, for example, to evaluate knowledge claims of other researchers, conceive hypotheses for own studies or to support conclusions of study results. A survey conducted by the Making Data Count project[104] in November/December 2014 asked researchers how frequently they use data from public sources for different purposes (Kratz & Strasser 2015b). The question was answered by 247 respondents: Roughly 70% of the researchers said that they used publicly available data "often" or "occasionally" to reproduce or confirm a published analysis, 90% to generate ideas/hypotheses before collecting own data, 90% to support the main conclusions of a paper, 70% to reach the main conclusions of a paper. More than half of the respondents (53%) were biologists, 17% researchers in environmental sciences, and 10% social scientists.

In our study we are mainly interested in data re-use for producing new knowledge, for instance based on combining own and available data from other sources. Among the purposes of using data addressed in the survey above mainly "to reach the main conclusions of a paper" indicates data re-use. Notably the figures for this practice are about 30% "often", 40% "occasionally, and 30% "never".

Data re-use can be defined as *"the use of data collected for one purpose to study a new problem"* (Zimmerman 2008). This definition makes clear that data re-use is not about "replication" or "reproduction" but use for addressing new research questions or shedding new light on established ones. Data is of course often re-used by the data creators based on their knowledge of how the data was collected and prepared for analysis. But the term is associated primarily with re-use by other researchers who have not been involved in the production of the data.

[104] Making Data Count, http://mdc.lagotto.io

In the business world data has been dubbed the "new oil" which, refined and analysed, can drive profitable activities. The metaphor has sprung over to research policy makers who expect that available open data – ideally in massive volumes ("big data") – can be easily used to produce new knowledge. But the idea of data as raw oil of research is of course flawed. Scientific data are heavily informed by the questions, designs, methods, tools, etc. the researchers apply, and re-use requires taking full account of these conditions. Re-use therefore is not about refining a raw resource such as oil, but requires understanding the conditions under which the data have been produced (they are already products) and making them work for new research.

The conditions are described in "contextual information", i.e. detailed information about the production and subsequent work on the data. Thus the data are separated from the framework and apparatus of their production, and reproduction (e.g. for verifying the data) requires to put in place again the latter. Re-use for new research however will in many – if not most – cases be within a different framework.

Many different forms of data re-use can be distinguished, for example, comparing older with new data (e.g. longitudinal series of data), processing and analysis of data in ways not considered by or not possible for the data producers (e.g. data mining), re-analysing data from a new research perspective, bringing together data from multiple studies for meta-analysis. Re-use therefore in many, if not most, cases means re-contextualising of the data.

There are of course quite some differences concerning the re-use of natural sciences and social sciences data, and particular challenges in re-using qualitative social sciences data (Heaton 2004; Moore 2007). Attention to research contexts and careful, reflexive re-use is important in all disciplines but particularly crucial when qualitative social data and personal information is involved.

Some authors have proposed a research agenda, framework or model for data re-use. Faniel & Zimmerman (2010) suggest to go beyond the perception of a "data deluge" (i.e. ever more data produced and managed with different technologies) by investigating factors that increase participation in the production, sharing and re-use of data, the role of different intermediaries such as digital archives and virtual organisations (e.g. data networks/Grids), and opportunities as well as challenges posed by the creation of new types of digital products that include data or integrate / interlink data resources.

With regard to more practical aspects of data re-use, Gonçalves-Curty & Qin (2014) summarised interviews with social scientists in a model of six categories of factors which may determine if data is re-used or not. The categories are perceived benefits, perceived risks, perceived effort, reusability assessment (difficulty to evaluate others' data for re-use), enabling factors (including conditions and infrastructure that facilitate re-use), and social factors which influence scientists' intention to re-use or not re-use data.

Faniel & Jacobsen (2010a/b) propose a framework specifically for data reusability assessment to guide research on how researchers in different disciplines evaluate if potentially re-usable data is relevant, understandable and trustworthy. According to the framework data resources (including documentation) are *relevant* if they meet the requirements of the research problem at hand; *understandable* if the researchers can comprehend the intent, design and other aspects of the data creation; and *trustworthy* if the data are credible, reliable and valid. Credibility concerns attributes of data producers (e.g. reputation), while reliability and validity requires direct evaluation of the data and available documentation.

Examples of (re-)use of research data from data archives

While a prime source for re-usable data, archives often do not have or chose not to present information about data usage. But we could identify some examples in the natural and life sciences as well as include one archaeological example.

Astronomy: Data of the Hubble telescope is re-used often so that the number of papers written based on Hubble archival data is higher than the number of papers resulting from new observations (Hubblesite 2011).

Oceanography: A citation analysis of three databases curated by the US National Oceanographic Data Center (Belter 2014) showed that they are highly cited in the research literature. If the databases were counted as journal articles in Web of Science (WoS), two would have citation counts higher than 99% of all articles in Oceanography in the WoS Science Citation Index from any single publication year from 1995 to the present.

Genetics (microarray studies): A data citation survey analysed 10,557 microarray studies published in the years 2001-2009 without or with datasets archived in the major repositories ArrayExpress and Gene Expression Omnibus (Piwowar & Vision 2013). The survey found that a substantial fraction of the archived data was re-used by third-party investigators. The re-use was distributed across a broad base of datasets and the intensity of re-use had steadily increased since 2003. In a conservative estimate, 20% of the datasets archived between 2003 and 2007 were re-used at least once by third parties.

Biodiversity: The Global Biodiversity Information Facility (GBIF) is the core integrating database of worldwide biodiversity data, holding 579 million species observations (occurrences) of over 1.6 million species from 15,234 datasets of 770 data publishers. Nevertheless in recent years GBIF felt the need to demonstrate their relevance to secure continued support from stakeholders (e.g. data publishers) and long-term funding of their operation. Therefore GBIF developed a data usage index, based e.g. on server log data and web-analytics (Ingwersen & Chavan 2011), and since 2008 documents research publications that use GBIF-mediated data; until April 2015 a total of 1409 publications have been identified (Schigel 2015: 15).

Archaeology: The Archaeology Data Service (UK) has a broad user group which goes well beyond academia: 38% of users are conducting academic research, 19% private research, and 8% commercial research; 11% are heritage management users; 17% use ADS for general interest enquiries, 6% to support educational activities, and 1% for family history research (Beagrie & Houghton (2013a: 33-34). During the 23 month period from February 2013 to December 2014, the ADS website had 693,622 visitors who carried out a total of 4,620,070 actions including 4,244,893 page views and 324,492 downloads. During this period ADS modified the website's direct access filter functionality, allowing the website metrics to take account of users finding PDFs and images directly via Google. This has resulted in a dramatic difference between the average monthly users in the first eight months of the 23 month period (February - September 2013) and the figure for later months, 14,000 versus 39,000.[105] The ADS presents detailed weekly access statistics, for example, between the 5th and 11th of October 2015 the website had 12,659 visits, over 50,000 page views, and about 3500 downloads[106]. Moreover the website shows archive-level usage figures for each of the deposited project archives, theses, journals & series and bibliographies (but not "grey literature", i.e. fieldwork reports). This includes visits to the archive, page views within the archive, file downloads, and totals. The figures cover the period since July 2011 or since the online release of the project archive thereafter.[107]

[105] ARIADNE D5.2 Initial Report on the Assessment of Online Access, January 2015, p.19.

[106] ADS Access Statistics, http://archaeologydataservice.ac.uk/about/accessStatistics

[107] For example see: Roman Amphorae: a digital resource, University of Southampton, 2005 (updated 2014), http://archaeologydataservice.ac.uk/archives/view/amphora_ahrb_2005/stats.cfm

Qualitative studies of data re-use

At present qualitative studies of data re-use in different research communities are still rare. But available examples can illustrate the potential scope of such studies.

A case of little data re-use

One important finding is that there can be fields of research in which data is re-used seldom. One example is the community of ecological researchers who employ environmental sensors studied by Wallis *et al*. (2013; based on 43 interviews). The researchers in this field use external information primarily from observatories, to calibrate sensors and provide ecological context (e.g. flora and fauna of an area), for instance. As reasons for not using others data Wallis *et al*. (2013) mention difficulties in discovering and interpreting other researchers' data, but refer to the literature rather than their interviews. Their results however suggest that the researchers studied simply do not need others' data but focus on capturing and analysing their own data.

The consequence is "dark data", data that is not made available and difficult to discover. Such data has been associated with "hand-collected" data which is common in small-scale ecological research (Heidorn 2008). It is estimated that less than 1% of the ecological data ever collected is readily discoverable and accessible (Reichman *et al*. 2011). However also automated data recording with in-field sensors appears to have little impact on data availability, at least in areas or research in which researchers are narrowly focused on their own data.

How researchers evaluate the reusability of colleagues' data

Faniel & Jacobsen (2010a/b) studied earthquake simulation researchers with a focus on how these scientists evaluate the reusability of colleagues' data for model validation. They conducted 14 semi-structured interviews and an online survey (117 respondents) to identify how the researchers evaluate if potentially re-usable data is relevant, understandable and trustworthy. As such evaluation is relevant to re-use in all research areas we briefly summarise the results.

Relevant data: The researchers apply problem-specific criteria to evaluate if the data meet the problem at hand, which in this case is validation of models. Certain data structures, parameters, etc. may be required and for details colleagues are often asked directly.

Understandable data: With regard to understandability the researchers mainly use available data documentation, benefitting from the fact that in their research field such documentation typically spans hundreds of pages describing equipment, configuration, data acquisition, processing, analysis, etc. In other research areas assistance in understanding the data under consideration may be required even if the researchers have prior experience in the research field.

Trustworthy data: Trustworthiness concerns whether the data are credible, reliable, and valid. With regard to credibility typically attributes of the data producers are considered. Such attributes (e.g. reputation) were less important than the other two criteria, which usually require direct evaluation of the data. In the case study however the available detailed documentation was also very useful for data evaluation.

The role of tacit knowledge in data re-use

The role of tacit knowledge in understanding other researchers' data is emphasised in many studies. Tacit knowledge means knowledge that is acquired through practice in a particular field of research and is hard to formalise and communicate in publications.

Carlson & Anderson (2007) conducted interviews with participants of four projects which produced different kinds of data collections, including astronomy data, socio-economic survey data, digitised cultural artefacts, and anthropological materials collected in several countries over one researcher's academic career. Not surprisingly the study found that the projects presented different contexts in technical and social terms with considerable implications concerning data organisation and access.

Carlson & Anderson highlight the role of tacit knowledge in all projects but greater difficulty in creating shareable data where data derived from people, craft-like research methods, and personal research collections are involved (on the latter see also Kaye *et al*. 2006).

Most often referenced in the literature is the work of Ann Zimmerman on tacit knowledge in ecological research, starting with her PhD thesis and subsequent studies on related matters (Zimmerman 2003, 2007, 2008). Her ethnological research shows, on the one hand, that ecologists aim to use "standard methods" to create their complex data and refer to notions of scientific practice that emphasize objectivity. On the other hand, information about the practicalities of research in specific local contexts is often left behind when results are published. Hence when ecologists consider re-using data of other researchers, they rely on formal and tacit knowledge gained through training and own data collection experiences. Without the tacit knowledge it would be difficult to understand and assess the data and re-use it appropriately. Therefore Zimmermann suggests that research e-infrastructures *"must find ways to support the sharing of tacit knowledge and to encourage interactions around data between users of data and those with knowledge of them. These tools should be part of digital libraries and other places where knowledge creation occurs"* (Zimmerman 2007: 23).

Data re-use in archaeology

Ten years ago Naylor & Richards (2005) argued that data re-use has become an important theme in archaeological research because of a number of factors, which include: reduced opportunities and increased costs of excavations, pressures for a "quick win" rather than long-term field research (e.g. doctoral researchers encouraged to look for existing data sets rather than create their own), the development of online data resources, and available data analysis techniques. At the same time, Naylor & Richards noted an abundance of books and manuals on primary data collection but little literature on how to re-use of existing third-party data. They saw the situation as "worrying" because new generations of archaeologists needed training in difficult matters of data re-use, including from different sources, often with inadequate metadata, varied and ill-defined terminologies, lack of contextual information, difficulties of data integration, etc. Ten years on, the situation arguably is worse, because still there is little guidance on but increased expectation of data re-use by research funders. Moreover, there is a lack of studies on actual data re-use in archaeology as well as other disciplines.

One exception is the DIPIR project[108] which investigated requirements for data re-use of three research communities, archaeologists, quantitative social scientists, and zoologists. Faniel *et al*. (2013) report results of semi-structured interviews with 22 archaeologists about their data re-use experiences. Re-use was mainly determined by the availability of sufficient contextual information, but also other criteria such as the background of the data creators and the perceived trustworthiness of the data archive were considered.

To re-use data created by others the archaeologists needed contextual information, especially about the research design and procedures of data collection (e.g. recovery, sampling and identification methods). Required contextual information could mean the map system or type of GPS used, the distance between field walkers in surveys, the size of the mesh screen used to sift dirt when collecting objects, etc.

Lack of sufficient contextual information was noted specifically for museum objects. Also the data collection procedures of older excavations were perceived as often not well-documented, but respondents found ways to re-use their data based on information identified in publications, field reports and other material, as well as talking to people who participated in an excavation.

[108] DIPIR - Dissemination Information Packages for Information Reuse project (USA, funded by the Institute of Museum and Library Services), http://dipir.org

While the amount and quality of documented contextual information improved over time respondents still experienced considerable difficulty re-using data. In part this was due to lack of readily available documentation, especially documents created in the field (e.g. maps, drawings, field notes), use of unclear and different terminology (concerning e.g. chronology or object classification), and differences between data recording methods due to particular research settings and objectives. Often the researchers had to be contacting directly to better understand the rationale and details of the field work (including decisions taken in situ) in order to evaluate and re-use the data accordingly.

The archaeologists also looked into available documents created in the field to assess how diligently the field work and organisation of data was carried out (e.g. the quality of the maps of excavated areas, the labelling of stratigraphic drawings, etc.). Furthermore the wording and ordering of the narrative was considered as an indicator for trustworthy data (e.g. a report first addressing some major finds instead of talking about the excavation matrix, soil, etc. would be perceived as less reliable).

One important additional criterion for data re-use was the background of the data creators, i.e. the reputation of the archaeologists for quality and reliability of their work. This did not mean their current institutional affiliation, but their doctoral degree granting institution and advisor, and where and with whom they trained and worked in the field. Furthermore knowledge about the archive was considered as a factor for trusting the deposited data, e.g. data submission requirements such as comprehensive metadata and transparency of data curation processes[109].

Based on their study results Faniel *et al*. (2013) briefly address the need to improve documentation standards in order to support data re-use. They note the recent introduction of comprehensive and interoperable vocabularies which, however, may help to address only some of the identified issues, and have yet to see wide adoption in the discipline. To support data re-use the research community and archives would have to address the issue of missing contextual information, in particular detailed information about the procedures of data collection and analysis which potential data re-users need to understand, evaluate and trust the data.

The quality of the contextual information is a decisive factor for the possibility and likelihood of data re-use, particularly if the archaeologists who produced the data cannot be contacted, i.e. a legacy dataset with little accompanying context information. In such cases researchers would be in the same situation as the archaeologists who participated in an experiment organised by the Alexandria Archive Institute (Atici *et al*. 2012). A dataset of 30,000 animal bone specimens from excavations during the 1960s at Chogha Mish, Iran was analysed by three researchers, each with over fifteen years of experience working with Near Eastern zooarchaeological assemblages. The zoo-archaeologists analysed and interpreted the data independently and arrived at markedly different interpretive conclusions. The conclusions were based on objective measures of the bones as well as necessary conjectures about the original data collection, because important contextual information was not available.

While the study by Faniel *et al*. (2013) makes clear the paramount importance of detailed contextual information for data re-use, it will be hardly possible to consider and provide the information for any questions potential re-users may have concerning their specific purposes. Therefore a feasible approach for the research community could be to define baseline standards of contextual information for data sets and more specific standards for specific types of data collection, processing and analysis as common in the respective community of practice.

The study by Faniel *et al*. (2013) also confirms that archaeologists sometimes re-use data, albeit on a small basis of interviews with 22 archaeologists. Also a small questionnaire survey conducted in Ireland provides evidence of data re-use (Sands 2009: 54-59): The question *"Do you routinely use*

---

[109] The DIPIR publication Yakel *et al*. (2013) reports on factors that influence trust in digital repositories of archaeologists and quantitative social scientists (summarised in *Section 5.4.2*).

data created by others?" was answered by 37 respondents of which 30 said "yes". 29 respondents received data directly from colleagues (e-mailed, on carrier media, etc.), 25 extracted it from written reports, 17 downloaded it from dedicated websites (e.g. Excavations.ie or the National Roads Authority website), but only 4 used and contributed data to an online facility. The question *"When you reuse data how much restructuring is required?"* was answered by 29 respondents of which 9 had to do extensive and 16 light re-working, while 4 could use it directly.

An online survey undertaken by the Archaeology Data Service on the creation, use and preservation of "big data" also provides evidence of data sharing and re-use (Austin & Mitcham 2007: 36). The survey addressed data from "big data" technologies such as airborne LiDAR (Light Detection and Ranging), terrestrial 3D laser scanning, maritime survey techniques and others. Responses were received from 48 respondents, not only archaeologists, but also from the earth sciences, for instance.

70% of respondents had somehow re-used data at least once a year (others "very infrequently"). Over 80% stated that they had received large datasets from other researchers or organisations, and nearly 80% noted that they would allow access by others to their data. All respondents stated that they consider using existing datasets for a new project, for example, to avoid duplication of costly data collection or conceive new surveys. The report notes, *"Clearly there is both a strong desire to, and sound reasoning for, reuse of data"*.

It is important to note that not all of the data addressed by the ADS survey are produced by archaeologists. For example, maritime surveys are conducted in marine science and oceanography. In this field willingness to share data and conduct research collaboratively is high, even by commercial actors (cf. the VENUS[110] and SPLASHCOS[111] projects). Archaeologists increasingly employ 3D laser scanning, while LiDAR data tend to be commercial, but some may be available because they have been commissioned by public agencies or acquired in publicly funded research programmes. For example, the UK Environment Agency in 2013 made some of their LiDAR data freely available for non-commercial use and since 2014 is committed to publish as much as possible as Open Data; from September 2015 all their LIDAR data will become Open Data (Matthew 2015).

In summary

Major criteria of archaeological researchers to re-use available data collected by others are if the data is relevant, understandable and trustworthy. The quality of data documentation, especially the contextual information, plays a major role in the decision-making. Thoroughly documented research data with the required contextual information is necessary to allow researchers evaluate if the data is relevant for intended uses such as comparative analysis or inclusion in a larger dataset.

Representative surveys are necessary to evaluate the current level of data re-use in archaeology (ideally in different fields of research), and if an increase could be expected if more open data becomes available. Such surveys should also collect information on the main forms of re-use (e.g. extraction of specific data points, combination of data, comparative analysis or other) in order to possibly better support these practices. In particular, archaeological data archives and tool developers may benefit from such surveys. Also conducting a series of case studies on data re-use in different fields of archaeological research could yield relevant insights.

[110] VENUS - Virtual Exploration of Underwater Sites project (EU, FP6-IST, 07/2006-06/2009), http://archaeologydataservice.ac.uk/research/venus; see Alcala *et al*. (2008)

[111] SPLASHCOS - Submerged Prehistoric Archaeology and Landscapes of the Continental Shelf (EU, COST Action, 2009-2013), http://www.splashcos.org; see European Marine Board (2014).

4.4.4 Data citation

Rationale and background

It is assumed that data which is openly shared through digital archives will often be re-used and cited, bringing academic recognition and rewards to data sharers like other publications. One guiding idea behind this rationale is citation metrics that complement the established, publications-based evaluation of research performance. The activity of leading actors in the field of research evaluation and the discussion of data citation in the scientific community show a substantial buy-in for this idea.

The scenario is that data citation indicates and acknowledges providers of valuable data, promotes further data sharing and re-use, and enables the use and impact of the data to be tracked and measured. Most importantly, it would drive the emergence of an academic credit system that appropriately recognises data creation, curation and sharing as scholarly activities which deserve to be rewarded. In turn, this would raise the status of data-related scholarly activity and motivate more scholars to invest the effort required for sharing their data through accessible digital archives.

So far the metrics of scientific achievement are based on publications (e.g. number and types of publications, journal impact factor, h-index, etc.), understood as the main record of scholars' contribution to their field of research. Publication of research data currently has no relevance because solid metrics for measuring and comparing the impact of shared data are missing. Due to this lack committees that decide on career progression (e.g. tenure & promotion), research funders who decide about grants and other stakeholders will find it difficult to reward data sharers. It is worth to note that this also applies to the sharing, curation and re-use of scientific software (cf. *Section 9.2.8*).

The objective to standardise and facilitate data citation can count on a broad consensus of the research community. As shown in a survey of scientists' data sharing practices (Tenopir *et al*. 2011), researchers see a formal citation of shared data as appropriate credit. 95% of 1234 researchers of different disciplines agreed that formal citation of the data providers in all disseminated work making use of the data would be a fair condition of data re-use.

There are of course many requirements for consistent data citation and appropriate academic recognition, including standardisation, discovery, aggregation, analysis, and reasonable metrics (e.g. taking account of data-related differences between fields of research). At present basic requirements are not fulfilled, agree upon and consistently applied data citation standards across the research community, for instance. Without consistent data citation and easy discovery etc. reliable metrics will not be possible.

Integration of data in the information ecosystem

Data citation standards to promote data sharing and ensure appropriate credit have been suggested since the 1980s (cf. the examples in Mooney & Newton 2012). But such standards have gained substantial traction only in recent years with the building of ever more data repositories and, overall, the quest for open sharing of research data. Altman & Crosas (2013) distinguish four phases of data archiving and the role assigned to citation principles. The current phase, since 2009, centres on the integration of data in the research and publication ecosystem (e.g. data catalogues, tool chains, scholarly attribution systems), building on previous work that was more interested in data description, access and verification.

Publishers, data archives and information services are working busily on this integration, e.g. DataCite[112] (founded in December 2009), Thomson Reuters with the Data Citation Index[113] (launched

[112] DataCite: http://www.datacite.org

[113] Data Citation Index, http://wokinfo.com/products_tools/multidisciplinary/dci/

in October 2012), and data archives that integrate data deposition with publishers' workflows, e.g. DRYAD, Figshare and others. Lately also the research community has started discussing data citation as a cornerstone of leveraging research data to the status they deserve in the scientific enterprise, for example the multi-facetted *For Attribution* report (Uhlir 2012), which presents the discussion of an international consultation facilitated by the US National Academies. Other substantial contributions are CODATA's *Out of Cite, Out of Mind* (CODATA - ICSTI Task Group 2013), and a report of the Opportunities for Data Exchange project (2012), which addresses the role and responsibility of different stakeholders concerning the citability of data in scholarly publication and communication.

A study commissioned by JISC (UK) also evaluated the issue that publication and data citation information is mainly provided by commercial services, hence, such information does not comply with the objective of openness of all research-related information (Curtis *et al*. 2013). The study authors conclude that a radical intervention in the citation information market would be difficult to justify, because of the difficulty and expected costs of creating alternative open services as well as the general, although not total, satisfaction of the research community with the current systems. Rather research policy should focus on encouraging open access to research resources and support the development of appropriate citation metrics to reduce concerns of the research community regarding data citation metrics.

Much attention and support has received the Joint Declaration of Data Citation Principles of FORCE11 (2014). FORCE11[114] is an open community aimed to shape the future of research communication and e-scholarship that, among other activities, since 2011 hosts the annual "Beyond the PDF" conference. The FORCE11 principles are a combination and synthesis of principles suggested by the community, CODATA, DataCite and Digital Curation Center (UK). In 2014 they were endorsed by 93 organisations and 209 individuals. The eight principles address the importance of citable data, attribution and credit enabled by citation, evidence provided by data, unique identification of data, access, persistence, specificity and verifiability, and interoperability and flexibility. The Preamble to the principles stresses that *"data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse."*

One important element in the ecosystem of publication and data services is cross-linking, persistent links between publications and data, and *vice versa*. This linking is crucial to prevent that data becomes detached from the published research design, methods and conclusions of the data sharers (Hoogerwerf *et al*. 2013; Wessels *et al*. 2014). Successive publications (papers and data) represent the knowledge base and discourse of a research community, which should exist in an integrated, accessible and persistent form. One initiative to ensure the interoperability of publication and data information services is the cooperation of CrossRef and DataCite that was announced in November 2014 (Ruiz 2014).

Another noteworthy integrative activity is the Horizon2020 THOR - Technical and Human infrastructure for Open Research project[115], started in June 2015 with the aim to leverage the persistent identifier systems of DataCite[116] and ORCID (unique identifiers for researchers)[117], build enhanced services for researchers and institutions concerning research objects of different disciplines (including the humanities and social sciences), and help build expertise of the research community in related matters by running a training programme. The international consortium of THOR comprises

[114] FORCE11 - Future of Research Communication and e-Scholarship, https://www.force11.org

[115] THOR - Technical and Human infrastructure for Open Research project (EU, H2020, 06/2015-12/2017), http://project-thor.eu

[116] DataCite, https://www.datacite.org

[117] ORCID, http://orcid.org

of the Australian National Data Service, British Library, CERN, DataCite, DRYAD, Elsevier Labs, EMBL-EBI, ORCID, Pangaea and Public Library of Science (PLOS).

Current data citation practices

The standardisation of data citation is on-going so that the guidance for citation by different institutions (journals, repositories and others) and actual practice of researchers are varied. Data citations tend to differ according to the community of practice and the recommended citation styles of the repositories that host data. There is also often a lack of guidance on how data should be referenced.

A study by Enriquez *et al*. (2010) found that good practice citation guidance was given only by 8 of 26 repositories (31%) surveyed, only by 16 of 307 (6%) journals, and only by 1 of 53 of research funders. The study team also manually reviewed 500 papers published between 2000 and 2010 across six environmental sciences journals and identified 198 papers that re-used datasets. But only 14% gave a unique dataset identifier in their attribution and a partially-overlapping 12% mentioned the names of the data author and repository. Few data citations were made in the article references section. The attribution patterns were steady over time and more correlated with repository than journal or type of data.

Mooney & Newton (2012) reviewed 65 science, social science and humanities papers that re-used data and the author instructions of the 44 journals in which they were published. They found that data citation was poorly practiced across the papers surveyed. Reference was mainly given in the text through mention of the dataset title; also only few data retrieval locations were provided. The journals' author instructions were largely silent on the issue of data citation. The authors note that not much has changed since a study of dataset citation in 1995 (Sieber & Trumbo 1995) and reckon that it might take another 15+ years until widespread good data citation practice may be expected.

Wide adoption of proposed data citation standards across many stakeholders and, particularly, that researchers follow suggested good practice, will certainly not be easy to accomplish. However consistent citation and capability to trace the use of datasets are necessary to enable the expected benefits of open sharing of research data.

Standardisation of data citation practices

The standardisation of data citation practices is still in the making, although there are growing volumes of standardised data citation, especially in the user community of DataCite DOIs (Digital Object Identifier) and metadata schema, including the ARIADNE partners Archaeology Data Service and Swedish National Data Service, for instance. Also practical guidance for good practice in data citation is available (e.g. Ball & Duke 2015) and should be promoted.

There is agreement among the different standardisation actors that a data citation should include a unique and persistent identifier (PID) for the data[118]. The PID may be a Digital Object Identifier (DOI)[119], Archival Resource Key (ARK)[120], Universal Numeric Fingerprint (UNF)[121] or other identifier that fulfils the requirements of a unique and persistent identifier; there must of course also be a service that resolves the identifier and allows access to the cited data. Furthermore there is wide agreement that a data citation should include: Author(s), Year, Title, Version, Data Repository/Archive (or other Publisher/Distributor), and PID (always at the end of the citation). Thus

[118] On the importance of identifiers and related information service requirements see: Open Data Institute & Thomson Reuters (2014): Creating Value with Identifiers in an Open Data World, October 2014, http://site.thomsonreuters.com/site/data-identifiers/

[119] Digital Object Identifier (DOI), http://www.doi.org

[120] Archival Resource Key (ARK), https://wiki.ucop.edu/display/Curation/ARK

[121] Universal Numeric Fingerprint (UNF), http://guides.dataverse.org/en/latest/developers/unf/index.html

there is a core set of what should be included in a data citation and additional ones may be included if necessary or beneficial (Kratz 2013). This could for instance be the date of access, subset used (if referencable), or type of resource.

As important as a standardised data citation format is the way in which data (re-)users actually cite the data, i.e. where they put the citation. This should be the references section of the publication (or a similar section for data citations) to allow easy identification and extraction, e.g. by citation aggregation services. If a subset of a dataset has been used, but is not citable, details about the data used can be given in another section of the publication.

Open issues, which need to be solved according to the research community's needs and digital archive capability, are the granularity at which the data can and should be cited, and how to deal with dynamic datasets (e.g. versioning). The granularity issue, i.e. possible "deep citation" of subsets and components of a dataset, is not only a technical one, but requires also a solution concerning data crediting and impact evaluation. Because there is a trade-off of exactness and crediting: For research review or reproducibility one would like to know exactly what data has been used. But for the evaluation of accumulated credit diverting citation count to subsets (e.g. by minting identifiers for them) would better be avoided (cf. Ball 2013; Kratz 2015).

A further issue that deserves attention is how the often many contributors to a data product, and their different types of contribution, can be credited (Hames 2012). Contributors for example can be study designers, field team leaders, instrument builders, algorithm developers, data editors and others. This can be best solved through metadata produced by the data providers and associated with the data product.

The CRediT - Contributor Roles Taxonomy project[122] addresses the need for transparency in contributions to published research in order to enable improved systems of attribution, credit, and accountability. The taxonomy distinguishes 14 types of contributorship including methodology, data curation, software, formal analysis, validation, review & editing, among others. The project involves a prominent group of developers and supporters, for example, the Wellcome Trust and CASRAI (Consortia Advancing Standards in Research Administration). The taxonomy is included in the CASRAI Dictionary of research administration information.[123]

4.4.5 Scholarly publication culture, citations and altmetrics

This section summarises a background study in which we collected information and empirical results on citation-related topics and issues (*Section 9.2*), such as the citation advantage for open access publications, alternative metrics (altmetrics), and the importance of usage data for data archives. Special attention is also devoted to specificities of the humanities publication culture, which should be taken account of in the development of comparative metrics of data sharing, re-use and citation.

Specificities of the humanities publication culture

Bibliometric and other studies point out specificities that set the publication culture of the humanities apart from other disciplines, in particular the natural sciences, but also in some respects the social sciences. They include that the humanities scholarly literature

- is under-representation in the main citation indexes Web of Science (Thomson Reuters) and Scopus (Elsevier), and that the indexes are biased towards English-language publications,

[122] CRediT – Contributor Roles Taxonomy project, http://credit.casrai.org

[123] CASRAI Dictionary, http://dictionary.casrai.org/Contributor_Roles

- presents a substantially different composition than other disciplines in that monographs, book chapters, conference papers, reports and non-scientific literature play a greater role than journal publications,
- usually cites few journal papers (in 2007 over 90% of the cited literature were not papers), and more often much older literature than other disciplines,
- has a more context-specific character than natural sciences publications, i.e. are more often published in national journals, addressing the readership in the native language.

The specificities of the humanities publication culture may also play a role in the context of data sharing, re-use and citation, and need to be taken account of in metrics which are being developed for these practices. For example, important drivers of open data are journals which request that data underpinning papers is made publicly available. As journal papers are not the main form of publication in the humanities this development will have a much lower impact in this sector. But archaeology is a multi-disciplinary field of research, involving branches of the applied natural sciences, so that some impact can be expected via journals of these branches. Another specificity of the humanities which should be considered is their stronger national focus, which in archaeology will mean the preference of national archives where researchers can deposit their data, with metadata in the native language.

Suggestions

- *Take account of the stronger context-specific character of the humanities and social sciences, with consequences such as preference of publication in the local language and national / native-language digital archives.*
- *Expect a lower impact of data policies of journals in the humanities because research papers are not the main form of scholarly publication in this; but in the field of archaeology some impact may occur as papers are also published in journals of the applied natural, life and other sciences.*
- *Consider specificities of the humanities publication culture in the development of comparative metrics of data sharing, re-use and citation.*

Citation advantage for open access papers

The citation advantage for open access, freely available papers over papers published in subscription-based journals is very well supported by many studies. The SPARC Europe[124] documents 70 studies (2001 to present) across many disciplines of which 46 found a citation advantage, 17 did not confirm the advantage, and 7 were inconclusive, found non-significant data or measured other things than citation advantage for articles. For publications in the humanities the few available studies do not show an advantage. No studies on archaeological publications are available as yet.

It can be expected that open access publishing will become ever more widespread, and that high-ranked OA journals in terms of impact will trump low-ranked ones. Thus there will be increasing competition between OA journals for readers’ attention, authors and references. Many of the OA journals which have been launched in recent years, including in the field of archaeology, classics and other domains of ancient world studies, will be difficult to maintain – if they do not occupy a well-defined niche in terms of sponsors (e.g. a professional association), authors and readers.

Suggestions

- *Continue promoting open access to research publications; the strong evidence of a citation advantage can help in establishing OA publication in archaeology. If appropriate, non-OA publications should be deposited in accessible repositories.*

[124] SPARC Europe: Open Access Citation Advantage Service, http://sparceurope.org/oaca/

- *Expect an increasing competition between OA journals for readers' attention, authors and references; many OA journals (incl. in archaeology, ancient world studies, classics) will be difficult to maintain.*

Citation advantage for papers with underlying data

Papers published with the data underlying the research are expected to receive more citations than others for which the data is not accessible. The main assumption that this will be the case is that such publications are more transparent and reliable and therefore more often consulted, used and cited. Furthermore, the data of the author/s as such may attract interest, for example as a source for further research, if it can be re-used easily.

However there is little evidence as yet for such advantages. This is due to several factors, for instance, low open data sharing in many disciplines as well as that empirical evidence for a citation advantage for papers with underlying data is difficult to establish. Indeed, evidence is only available for biomedical research, specifically microarray clinical trial studies, with on average 9% more citations and higher percentages of more recent studies (Piwowar & Vision 2013).

Areas of research where a citation advantage may currently be identified are those which avail of state-of-the-art data repositories and depositing of data is mandatory. A study of the Research Trends journal (Huggett 2014) identified about 180,000 publications of the period 1996-2014 that referenced a data repository, most of which addressed health-related topics. The publications received a rather high average of 105 citations.

Suggestions

- *Take note that there is little evidence as yet for a citation advantage for papers that are published with the data underlying the reported research results.*
- *Areas of research where a citation advantage may currently be identified are those which avail of state-of-the-art data repositories and depositing of data is mandatory.*

Citation of data

Citations of both papers and data are generally rather low. For example, article-level metrics for 63,771 papers in Public Library of Science (PLOS) open access journals in November 2012 show that the papers received about 124 million HTML page views, 27.7 million PDFs of papers were downloaded (22% of the views), but only 375,000 CrossRef citations of the papers were found, 0.3% of the views or 1.35% of the downloads (Pattinson 2012: 4).

Peters *et al*. (2015) found that 85% of the about 4 million data records of the period 1960-2014 in Thomson Reuters' Data Citation Index (DCI) were not cited; but the authors report a growing trend in citing data published since 2007. Of the 10,934 data items that received two or more citations 21 items were deposited with the Archaeology Data Service and together had 75 citations in the DCI.

While leading digital archives such as the Archaeology Data Service display usage figures down to individual data deposits (e.g. visits, page views, downloads), actual re-use and citation of the data in the literature are difficult to identify. Special citation analysis using sources such as Thomson Reuters' Web of Science need to be done or references collected via different information sources; for example, the Global Biodiversity Information Facility (GBIF) collected 1409 publications in the period 2008 to April 2015 which used GBIF-mediated data (Schigel 2015: 15).

Data citation will for many years remain an area where reliable figures will be difficult to establish, in particular because it requires standardised citation by researchers in the literature. Moreover, the citation context would need to be analysed, i.e. what kind of use the citations stands for (e.g. consultation or actual re-use).

Suggestions

- *Recognise that at present data citations are difficult to identify, due to lack of standardised citation practices and other factors.*
- *Consider that a lot of deposited data(sets) may be consulted and (re-)used but not referenced due to many reasons, e.g. citation of related papers rather than the data/sets.*

Altmetrics (alternative metrics)

Many papers and data items are viewed, downloaded and consulted, but not cited. Therefore altmetrics have been developed to evaluate better the dissemination, reach, use and impact of scholarly work. Such metrics are based on online activities and interactions, including communication on the scholarly and social Web (e.g. Facebook, Mendeley, Slideshare, Twitter, etc.).

The fuel that is driving the adoption of altmetrics is the increasing pressure for recognition and impact at all levels. Research funders wish to see impact of their grants, university and research institutes worry about their ranking in national and internal rankings, journal publishers and editors are competing for authors and readership, and scholars are of course keen to demonstrate that their work is recognised.

Altmetrics are not alternatives but provide additional and complementary measures to citation-based metrics of scholarly impact. Therefore it can be argued, despite many claims to the contrary, that altmetrics may have little effect on established frameworks of research evaluation. Adoption by the research community will also be low as most scholars see citations as the core currency of prestige and regard social media with suspicion.

It should also be noted that the main sources of altmetrics data as well as altmetrics service providers, are private, for profit enterprises. The leading providers are PlumX/Plum Analytics (owned by EBSCO Information Services) and Altmetrics.com (own by Macmillan Publishers via their Digital Science brand).

The background study also presents some results for humanities and archaeology. The most comprehensive study by Costas *et al*. (2014) analysed altmetrics provided by Altmetric.com for about 50,000 publications of different disciplines covered in Web of Science. The study found a relatively low but growing presence and density of social media activity around the publications. The biomedical & health sciences and social sciences & humanities publications showed the highest presence of at least one altmetrics scores, 22.83% and 22.50% respectively, followed by life & earth sciences (15.94%) and natural sciences & engineering (8.98%). Only the social sciences & humanities had a (slightly) higher density of altmetrics per publication than citations (1.04 versus 0.9).

Suggestions

- *Take note that the fuel driving the adoption of altmetrics is the increasing pressure for recognition and impact at all levels. While a legitimate goal, indicators of recognition and impact need to be developed and applied carefully. This concerns classical citation as well alternative metrics (altmetrics).*
- *Give priority to indicators/metrics based on actual usage of curated material; social media activities around the content/data are not seen as valid indicators by most researchers.*

(Alt)metrics for data archives

Altmetrics can allow data archives to demonstrate their relevance to the research community and funders, plan archive development, and mobilise new data contributions. However many digital archives currently do not present usage metrics (Costas *et al*. 2013: 23).

We suggest that digital archives should in the first place focus on indicators that are directly related to the research output they curate and make accessible (e.g. searches, views, downloads). Social media altmetrics are currently relevant mainly as illustrative evidence of usage and may be convincing only for large and heavily used archives.

Systems log-analysis and web analytics can allow correlating user activities with the effectiveness of the information system and relevance of the data being served. Particularly important are indicators that can expose the value that curation work adds to datasets such as organising the data in ways that ease discovery, access and re-use.

The background study includes examples of data archives of different disciplines that have developed data usage indexes or deposit-level metrics to prove their case, including the Archaeology Data Service.

Suggestions

- *Capture and present usage figures that are directly related to the research material that is being curated and made accessible (e.g. views, bookmarks, downloads).*
- *Use such figures/metrics to demonstrate relevance to the research community and funders, plan archive development, and mobilise new data contributions.*
- *Communicate the value curation work adds to data collections (e.g. indicators of improved discovery and access).*
- *Support the development of reasonable and comparable metrics / usage statistics for data archives as well as single data resources (e.g. data of an archaeological project). Make sure that differences between domains of research are taken account of in such endeavors.*

Not all suggestions of this section are included in the main suggested actions in this focus area, which are focused on data usage. However it is recommended to continue promoting open access also to research publications (i.e. open access publication and self-archiving of other publications in accessible repositories). Moreover specificities of the humanities should be considered in the development of any open access metrics, publications as well as data.

4.4.6 Suggested actions

Benefits associated with re-use of openly shared data arguably are the strongest drivers of the open data agenda. For example, re-use in further research (e.g. based on combined data) is expected to provide much return on investment. Re-use is also particularly important to individual researchers as it can bring recognition and rewards to data sharers based on data citations. Data archives and e-infrastructures could help mobilise more open research data more easily if convincing evidence for the expected individual benefits is available. However, empirical evidence for such benefits is still scarce. Existing archaeological data archives are accessed frequently, but little is known about actual data re-use and citation. Below we suggest how a better understanding of data re-use could be acquired, re-use promoted, and ensured that data sharers receive the credit they deserve. Furthermore some specific suggestions from the background study on citation-related topics and issues are included.

Conduct studies of data re-use to better understand and support current and emerging practices

Surveys or a series of case studies could allow evaluation of current practices of data re-use in different areas of archaeology. Information about the main current and emerging forms, contexts and requirements of data re-use would be helpful to possibly better support these practices. In particular, data archives, e-infrastructure and tool developers could benefit from such information.

Promote data re-use and highlight inspiring examples

Research organisations, data archives/repositories and e-infrastructure service providers should promote data sharing and re-use wherever appropriate. Specific measures should be applied such as data re-use competitions and highlighting of inspiring examples. The introduction of a data re-use award by the Archaeology Data Service (UK) is a good example for this approach.

Foster consistent data citation so that data sharers can be recognised and rewarded

Researchers who (re-)use data from digital archives or databases should reference the data in a standardised way. The standardisation within and across disciplines is still in the making. We recommend that data archives/repositories, journal publishers and other service providers in the field of archaeology promote a common form of data citation in the research literature and other publications (e.g. websites). Among the requirements for consistent and accurate reference of (re-) used data we highlight persistent identifiers (e.g. DOIs or other). Standardised citation can greatly enhance the identification of data (re-)use and recognition of researchers and data archives that make data available. Eventually it will enable tracking of data (re-)use and support evaluation of the impact of shared data.

Capture and present data usage figures

Data archives/repositories and researchers will for some time to come lack information if accessed research material has been re-used and cited. Therefore other indicators of usage (e.g. views, downloads) should be applied and presented. Such indicators are called alternative metrics (altmetrics) and also include social media activities around research output.

Usage figures can allow data archives and other service providers to demonstrate relevance to the research community and funders, plan further development, and mobilise new data contributions. We recommend usage indicators that are directly related to the material that is being curated and made accessible (e.g. views, bookmarks, downloads); social media activities around the content/data are not seen as valid indicators by most researchers. It is also recommended to communicate the value curation work adds to data collections (e.g. indicators of improved discovery and access).

5 Focus area 3 – Data archives and curation of archaeological data

5.1 Introduction and overview

Digital archives for deposit, long-term curation of and access to data provide core services underlying research e-infrastructures. Data archives richly filled by the research community are one key to the success of e-infrastructure initiatives, because researchers want to discover and (re-)use available data for purposes such as comparison and further research. At the same time many researchers are reluctant to share their own data beyond project collaborators and other trusted colleagues. Therefore motivation and support of open data sharing is necessary.

State-of-the-art community data archives can foster trust in open data sharing as they provide a reliable environment for data publication, (re-)use and citation. We highlight national and international research community archives as they can reduce data fragmentation substantially. The "institutional" model, repositories of single universities or research institutes, provide a local solution for affiliated researchers but offer only limited potential to overcome the fragmentation of archaeological data.

Extending the productive lifecycle of data

Ensuring long-term curation of and access to archaeological datasets goes beyond the capability (and core business) of individual researchers, research projects and, arguably, most institutes. In any case, the existing difficulty of discovering and accessing archaeological datasets demonstrates that individual "local" solutions are not appropriate for the task. Therefore the data should be deposited in archives that ensure long-term curation and access.

Unfortunately, the lifespan of archaeological data often ends when researchers have published their results, but do not make the underlying data available. Extending the data lifecycle through measures of curation and dissemination allows use of the data in further research, for example, exploration of integrated datasets with novel methods and tools. A core role of digital archives and e-infrastructures therefore is mobilising and bringing together data resources so that they can be used in novel ways. The expectation is an increase in value of existing data through wider and new uses within and beyond the research community.

Towards an optimal solution

The fragmentation of archaeological data poses enormous difficulties to achieve aggregation and integration of data of many projects/institutes at the level of common e-infrastructure and services. As the most effective solution for tackling the fragmentation we see domain-based, community-level archives. The advantages of such archives include that they can:

- provide clear orientation for all stakeholders concerning data mobilisation, expected good practices, etc.
- act as centres of expertise required for the types of data and data-related issues common in the domain,
- foster trust in open data sharing through providing a reliable environment for data publication, (re-)use and citation,
- allow cost-effectiveness of data curation and access (e.g. economies of scale),
- promote common standards and act as hubs for data integration and access.

References for such digital archives exist, for example, the Archaeology Data Service (UK, established 1996) and the E-Depot for Dutch Archaeology of the Data Archiving and Networked Services (Netherlands, since 2005); both are mandated archives for depositing data of archaeological research in the respective country.

Many European countries lack a community archive for archaeological data. Not a good solution for overcoming data fragmentation in such cases, especially large countries, would be several research institutes each trying to establish their own data repository or archaeological data ending up in many general-purpose university-based repositories. Therefore the IANUS Research Data Centre for Archaeology and Classical Studies in Germany is a consequent initiative for a common solution. New initiatives in this direction have been inspired by ARIADNE in smaller countries, Austria, Ireland and Slovenia, for instance.

Mandates and certification

Research funders mandate domain-based archives that are acknowledged for their high standards of data curation or, if such an archive is not available, recommend use of a recognised and suitable other repository. Acknowledged state-of-the-art archives devote special attention to measures that promote trust and credibility. One measure is certification according to a standard of trust-worthiness. Both the Archaeology Data Service (UK) and Data Archiving and Networked Services (Netherlands), and several other digital archives, are certified according to the criteria of the Data Seal of Approval.

In the absence of a decisive deposit mandate, i.e. deposit is only recommended and several archives are considered as relevant, the certification of a data archive will be one factor in researchers' selection of an archive. According to an available study archaeologists appreciate most an archive's transparency with regard to the description and management of data collections, and indications of archive stability (e.g. sustained funding).

Benefits and costs

The benefits of a community-level, national/international data archive stem from its role as a central and sizeable hub of data in the research field it supports. Reliable, one-stop access to needed information and data resources makes research easier, faster and cheaper. In the case of the Archaeology Data Service (ADS) the increase in research efficiency of the users has been calculated to be worth at least 5 times the costs of operation. Including other advantages, £ 1 investment in ADS yields up to £ 8.30 return. Like the ADS also other richly filled digital archives may have many non-academic users, including heritage management, educational and private/general interest user groups (e.g. local/regional history).

Concerning costs, one important fact to bear in mind is that the costs of post-project data curation and online publication of archaeological projects are only a fraction of the total project costs, between 1-3%, depending on the type of investigation and data generated.

Another rule of thumb is that running an archaeological or other data archive of course costs considerably more than a typical institutional document repository. One major cost driver in archaeology is the variety and complexity of the data that needs to be ingested and curated. Data acquisition and ingest are the most costly curation activities, while archival storage and preservation activities are a much smaller segment of the overall costs, and likely to decline over time.

Keeping the operational costs stable, while curating larger data collections, allows economies of scales (lower per-unit cost). Enhancement of labour-intensive curation activities through streamlining and tool-support (e.g. easy submission of small deposits) can allow significant cost reduction.

5.2 Curation of archaeological research data

Our survey of the landscape of archaeological and other cultural heritage data resources shows that this landscape is diverse and fragmented (cf. *Section 9.1*). In archaeological research diversity and heterogeneity of resources (content, vocabularies, information technologies) is an issue but a matter of fact. As stated in a document of the European Commission's e-Infrastructures unit (2013), *"diversity is likely to remain a dominant feature of research data – diversity of formats, types, vocabularies, and computational requirements – but also of the people and communities that generate and use the data"*. This is certainly the case also in archaeology. But it should be considered how the management of archaeological research data can be improved so that open and re-usable data flows into data archives. The existing fragmentation of archaeological data resources must not be accepted, but might be overcome at least to some degree. Further fragmentation should of course be prevented.

5.2.1 Levels of data curation and access

Data curation means maintaining and adding value to a collection of data or a database for current and future use. Curation thus is a broader concept than preservation, emphasizing opportunities for added value through systematic collection management, extension and integration of data resources, and linking to related information (e.g. similar or complementary data resources).

Below we distinguish four levels of data curation and access and summarise some study results concerning the field of archaeological research:

Level 1: Project-focused data curation of research groups or individual researchers

Such data curation spans the researchers' own management of data during and after completion of projects. Practices and means of data management during a project are of course focused on doing the research. Access to the data is usually restricted to the project collaborators. Requirements for curation in the long-term and sharing the data beyond the collaborators may not be considered before the final stage of the project.

If the researchers make data available online on a project website it often becomes inaccessible rather quickly. Many websites of archaeological projects present a situation of insufficient content longevity and re-usability. Researchers of the Alexandria Archive Institute reviewed over 60 archaeology-themed websites. Only one quarter of the websites had content in a format that made it easy to re-use such as clear licensing information, open formats, stable URIs, and machine-readable data (Kansa & Whitcher Kansa 2011).

Exceptions with regard to longevity and accessibility are international long-term excavation projects with a strong impetus of openness (e.g. the excavations at Çatalhöyük[125]) or initiatives to bring together online the documentation of several projects (e.g. the "Portal to the Past" of the Canadian Institute in Greece[126]). One example of an open science approach to data curation and access in archaeology is the DART - Detection of Archaeological residues using Remote sensing Techniques project (Beck & 2014). DART made their extensive data openly accessible already during the project work through a purpose-build data repository based on the open source CKAN platform; the data is still available throught the DARTPortal[127].

[125] Çatalhöyük - Excavations of a Neolithic Anatolian Höyük, http://www.catalhoyuk.com

[126] Portal to the Past (online archive of 18 archaeological projects undertaken under the aegis of the Canadian Institute in Greece), http://www.portal.cig-icg.gr

[127] DART - Detection of Archaeological residues using Remote sensing Techniques (UK, Science and Heritage programme, AHRC and EPSRC, 10/2010-09/2013), DARTPortal, https://dartportal.leeds.ac.uk; as a reminder of insustainability of project websites, the information-rich DART project website disappeared; it has been

Level 2: Data curation at the level of individual universities and research institutes

At this level the current situation of archaeological data curation corresponds to The Royal Society's description of the situation at individual universities and research institutes in general: "*Their approaches vary greatly. Although many have policies for research data, this tends to take the form of broad advice to researchers rather than stronger, targeted support or active curation based on comprehensive oversight of the full range of data generated by the research efforts that they house"* (The Royal Society 2012a: 61).

At universities the implementation of repositories for research data is a relatively recent development. In general such repositories do not support one discipline but all or a group of disciplines that are present at the university. What many universities have is a document repository that invites self-archiving of material, which often is much below expectations even where an institutional mandate exists. Few of these repositories contain datasets because researchers perceive them as inadequate or are unwilling to deposit their data. Newly implemented data archives such as Heidelberg University's HeiDATA[128] (launched in June 2014) contain only few datasets so far.

Archaeological research institutes, at least small or medium sized institutes, will hardly have the resources to implement and maintain a data repository that brings together field survey, excavation and laboratory data of their researchers. At smaller institutes even support for project data management tends to be minimal. The ARIADNE online survey found that the responsibility for maintaining the data after project completion is mainly with the researchers (54% with the project manager, 27% an appointed member of the research team, 19% "other").[129] Major reasons arguably are the understanding that the data is owned by the researchers, including the difficult case of collaborative projects, and lack of data curation expertise and resources at the institutional level.

Institutions in some European countries may also struggle because of a slow transition from analogue to digital curation of archaeological and cultural heritage information. In the consultation with ARIADNE partners on the innovation agenda one partner explained the situation in their country; *"(…) even though the standards and procedures are imposing an unique model of registration and management of data, these are not respected because of (a) the quantity and quality of the data managed by each institution; (b) the digitalization degree of each institution (many of these institutions have not managed to transform the old inventory lists, field charts, registries, etc., from the paper format to digital format; and (c) the degree of data gathering and administration in digital format, given the lack of computerized resources at some institutions".*

Level 3: National community data archives

National data archives curate data of research communities in different domains. Such archives are often managed by institutions on behalf of research funding agencies (e.g. research councils) and/or mandated or recommended as preferred archive. They avail of the special expertise as required for the types of data and data-related issues common in the domain, promote common standards, and provide guidance and support (e.g. guides to good practice, training). Examples in the field of archaeology are the ARIADNE partners Archaeology Data Service (UK) and DANS - Data Archiving and Networked Services (Netherlands), which includes the E-depot for Dutch Archaeology. In the United States, Digital Antiquity aspires to make The Digital Archaeological Record (tDAR) archive a data home for a substantial fraction of North-American archaeological research projects (Kintigh &

preserved by the Web Archive but not with the full functionality and content (Web Archive, https://web.archive.org/web/20140707074632/http:/dartproject.info/WPBlog).

[128] HeiDATA, https://heidata.uni-heidelberg.de/dvn/

[129] ARIADNE First Report on Users' Needs (D2.1, April 2014), p.100

Altschul 2010)[130]. In Europe coverage of the archaeological data might be expected by national-level data archives, but such archives are missing in many countries.

Level 4: Major international data archives and databases

Such data archives and databases have been initiated by leading national institutes (and adopted by the international community) or established by international initiatives aimed at bringing together research data either centrally or through a network of distributed databases. Typically they centre on a specific (sub-)domain such as biomedical, social or earth sciences.

The networked variant aggregates and provides access to (meta-)data from a larger number of trusted sources. Data curation thus takes place at the level of the providing sources and aggregating central database. This can take the form of a common catalogue of dataset metadata or regular aggregation of data into one database. Consequently such networks operate based on conformance to well-established protocols, metadata and exchange standards, often co-developed by the participating research and archive organisations. Budgets for maintaining the archive or database typically come from several committed sponsors.

Examples of international data archives where researchers can deposit and provide access to data are PANGAEA[131] (earth & environment data) and GenBank[132] (genetic sequence database), which is part of the International Nucleotide Sequence Database Collaboration. The NASA Global Change Master Directory (GCMD)[133] is an example of the catalogue variant, and the Global Biodiversity Information Facility (GBIF)[134] of the data aggregation approach.

In the field of archaeology there are no international data archives and no comprehensive directory of data resources (but many lists of various resources). ARIADNE will provide a catalogue of European collections of data sets and databases, with capability to aggregate and integrate their metadata for cross-archive data discovery and access. A high integration of datasets/databases would require mappings to the (extended) CIDOC Conceptual Reference Model[135], which is intended by several ARIADNE partners.

The distinction of different levels of data curation above is inspired by the schemes of Berman (2008) and The Royal Society (2012a: 60-69), which both present a data curation pyramid, with project-level curation at the bottom and major curated international resources on top. The rationale of the pyramid is that data should ideally go up into, and be curated by, digital archives or databases at a higher level. Project-level long-term data curation is generally understood as not viable. Most discussed is the role of institutional repositories, while curated resources at the national and international level are appreciated.

---

130 Digital Antiquity, http://www.digitalantiquity.org; tDAR - The Digital Archaeological Record (hosted at Arizona State University, USA), http://tdar.org

131 PANGAEA - Data Publisher for Earth & Environmental Science (hosted by the Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Germany), http://www.pangaea.de, includes about 3100 datasets related to "archaeology".

132 GenBank, http://www.ncbi.nlm.nih.gov/genbank/, includes about 10,000 records of ancient DNA of humans, animals, plants, bacteria, etc.

133 GCMD, http://gcmd.nasa.gov, covers more than 34,000 earth & environmental data set and service descriptions including, for example, World Data Center for Paleoclimatology and other relevant datasets.

134 GBIF, http://www.gbif.org, 579 million species observations (occurrences) of over 1.6 million species from 15,234 datasets of 770 data publishers.

135 CIDOC Conceptual Reference Model (CIDOC-CRM), http://www.cidoc-crm.org; there are several recent extensions of the CIDOC-CRM, e.g. for archaeological excavations (CRMarchaeo), scientific observations and argumentation (CRMsci, CRMinf), http://www.ics.forth.gr/isl/CRMext/

Rusbridge (2008) mentions more actors in data curation, including individual researchers, research groups, institutes and departments, communities of institutions (as consortia or informal), scientific disciplines, national services (e.g. national libraries or archives), publishers and other 3rd party services. He does not point out who may be best placed to curate research data, but it seems clear that a situation where all mentioned actors engage in managing research data will hardly be an optimal solution.

5.2.2 Factors that produce data fragmentation in archaeology

Archaeological data is fragmented and often difficult to access. This situation is caused by many factors which include, but are not limited to, the fact that such data is created and/or managed by different actors and that the data types are varied.

The different actors and contexts include academic research institutes and projects, cultural heritage agencies, contract excavators, scientific laboratories, museums, and others. Accordingly the data may be embedded in, or attached to, documentation of excavations or field surveys, scientific laboratory analyses, museum collections, monuments records, and others.

The data types are more varied and complex than in other fields of research (e.g. social sciences) and include, for example, textual descriptions, drawings, photographs, photogrammetry, laser scans, laboratory data analysed with different archaeometry instruments. Another important characteristic is that a lot of archaeological information is contained in “grey literature”, i.e. unpublished reports of contract excavation work.

Perspective and forms of data organisation

Based on the insights gained from interviews and online surveys for the ARIADNE First Report on Users’ Needs[136] we see major additional factors that cause fragmentation of archaeological research data. The factors are the general perspective and forms of data organisation in archaeological research. The pre-dominant perspective appears to be project-centred. In this perspective the research institute mainly serves as an administrative umbrella and home-base of several researcher-driven projects.

A stronger institutional perspective would more likely motivate data curation and integration across projects. But the project data management seem to be conducted with limited supervision, coordination and support by many institutes. Some basic support from the university’s central IT services or technical staff of the institute (or a technology-savvy colleague) is usually available, to set up a project database, for instance. Typically the data resides on the researchers’ computers, carrier media and, in the case of a collaborative project, on a file server with access restricted to the research group.

In terms of data organisation, the limited institutional curation means isolated “project archives” (one per excavation site), efforts of small groups or even one researcher (e.g. PhD candidate or post-doc) to create, fill and maintain a research database, and other project-centred products. There are of course also some projects aimed at integrated results, for example, a database that documents field surveys or finds of several excavations in a region. Access to such databases may be restricted to members of the research institute or open also for external consultation (sometimes with less detail).

In addition, there is the website of the research institute where results of completed and on-going project work are described and illustrated (e.g. images of field surveys or laboratory results), lists of publications presented, etc.

[136] ARIADNE First Report on Users’ Needs (D2.1, April 2014; available on the project website).

We think that this description summarises the situation at most small and medium-size archaeological research institutes, while larger organisations may be able to invest more effort in institution-level management of the data that is produced by their research groups.

The fragmented situation of data organisation poses enormous difficulties to achieve aggregation of (meta)data of many projects/institutes at the level of common e-infrastructure and services. A more integrated approach at the basis, the research institutes, would require a change in the research framework and mind-set (i.e. priority of the institutional perspective instead of individual projects), and implementation of a common data organisation and curation system.

In general, however, data curation is not a priority of most researchers and research institutes, and would require resources (funds, technical support personnel) that are not available at many research institutes. The question then is how the data curation and access might be organised, and the answer, at least since some years, is: a community data archive or an institutional repository (in the case of larger research institutes or universities).

Use of repositories

Results of the ARIADNE online survey illustrate the current situation in this regard.[137] The following figures are based on well over 500 responses by researchers: 40% said that their organisation (e.g. university / department, research centre, or other) has an institutional repository that is managed by dedicated staff. However, 67% of respondents said that they had shared data/content through an institutional repository only in a few projects or not at all. A national data archive or international data repository was not an alternative solution for most of the archaeologists. 76% did not or only in a few projects share data through a national data archive, 83% in the case of subject/domain-based international repositories.

40% of the respondents (569) said that their organisation has an institutional repository. The survey did not ask if it is a document repository or a dedicated data repository. It is well known that institutional repositories are mainly document repositories which contain material such as self-archived papers, research reports, theses, course material, etc. Datasets are rare in such repositories: of 2504 institutional repositories registered in OpenDOAR only 101 (4%) include datasets among their content categories.

But some repositories seem to fare better in this regard. For example, among the 186 UK-based institutional repositories registered in OpenDOAR 14 (7.5%) include datasets. The Repositories Support Project survey (2013) of university-based repositories in the UK found that in their sample of 89 repositories 33 (37.1%) include datasets. But not many may hold archaeological datasets because in the UK the Archaeology Data Service is the mandated archive for data of projects funded by the Arts and Humanities Research Council and the Natural Environment Research Council as well as the archive recommended by the British Academy, Council for British Archaeology, English Heritage and the Society of Antiquaries.

Indeed, we do not argue that universities should include archaeological data in their institutional repository or that archaeological research institutes should implement a data repository. Many such local solutions would not help overcoming the fragmentation of archaeological data overall. Therefore we emphasise national archives as arguably the most effective approach to reduce the fragmentation in a cost-effective way as well as to make the data available to higher-level, European/international e-infrastructure and services.

We note, however, that at present such archives are available only in a few European countries. Examples of existing archives, all ARIADNE consortium partners, are the Archaeology Data Service (UK, established 1996), DANS-EASY (NL, since 2005) with the E-depot for Dutch Archaeology, and the

[137] ARIADNE First Report on Users' Needs (D2.1, April 2014), p.99 and p.102.

Swedish National Data Service, which archives archaeological data since 2011. In Germany the community repository IANUS is being built (coordinated by ARIADNE partner German Archaeological Institute), and new initiatives have been inspired by ARIADNE, for instance in Austria, Ireland and Slovenia.

5.2.3 Research data management by libraries?

Several studies have shown that the majority of institutional repository managers belong to the library community (e.g. Markey *et al*. 2007; Kennan & Kingsley 2009; Wickham 2010; Repositories Support Project 2013). This does not come as a surprise as most university-based repositories – which are the bulk of existing repositories – have been developed by university libraries, often in cooperation with the university's central IT services. These repositories are conceived to cover content of all subjects and mainly document-type academic and educational content such as articles, theses, course material, lecture notes, etc.

Standard tasks of such repositories include support of authors in self-archiving of their publications and other content, metadata creation, indexing of resources, provision of access on a local search website, and metadata exposure for harvesting in a content federation network. Since several years now the question if university-based research libraries should engage in data management and e-science support has been discussed (e.g. Antell *et al*. 2013; Gold 2007a/b; Joint Task Force on Library Support for E-Science 2007; Kruse & Thestrup 2014; Lewis 2010; LIBER 2012; Lyon 2012; Salo 2010)

Most contributions emphasise favourable aspects, pointing to relevant existing expertise of research libraries as well as the opportunity to revitalize and strengthen the libraries' relation with their user community. At the same time many difficulties of research libraries to engage effectively in research data management are noted, so that they might *"run the risk of embarrassing missteps as they come to grips with the data challenge"* (Salo 2010).

The main challenges for the research libraries, and institutional document repositories they operate, are limited resources (funds, personnel) and skills gaps with regard to research data management. Consequently, at present their engagement in this area is rather low (cf. *Section 6.2.3*).

Indeed, how could librarians accustomed to standardised publication formats and bibliographic metadata, journal subscription services of publishers, and maybe some experience in the digitisation of special collections dare to manage research data? How could the document-centric institutional repositories be adapted for the requirements posed by curating data of research projects?[138] How could the librarians advice researchers of different disciplines in appropriate project-based data management?

And how to do this with little dedicated staff time? The Repositories Support Project survey (2013), based on responses from 86 UK universities, found that the average university repository employed a total 2.52 WTE (whole time equivalents) – combining 1.29 administrative, 0.94 managerial and 0.29 technical WTEs.

Studies on document repositories of universities found a largely disengaged faculty, unwilling to self-archive papers and other scholarly output. The Supporting Research project (2010-2011) identified a high dissatisfaction of researchers with such repositories. The scholars perceive them as more in the service of the institution rather than the research community and, particularly, see no personal benefits of sharing material through institutional repositories (MacColl & Jubb 2011: 3-4). One major issue of university-based repositories is that they have to cater to all subjects und therefore find it

[138] Some can: For example institutions that have implemented EPrints are working hard on making it capable to manage data or at least provide a catalogue of data deposited elsewhere. Others implement a data repository system alongside an existing document repository. Lewis (2014a/b) gives an overview of on-going work in this area at universities in the UK.

difficult to provide special services for any of them. There is little reason to assume that this would not also be the case with regard to research data management.

5.2.4 National and other community data archives as a solution

Based on our review of the current situation of archaeological research data sharing and access we assume that building and mandating national and other community archives is the most effective strategy in several respects. This includes clear orientation of all stakeholders, formation of centres of expertise, focused mobilisation of data deposits, economies of scale and cost-effectiveness, among others. The community archives need not necessarily be dedicated to archaeology only like in the case of the Archaeology Data Service (UK). Also solutions with archaeology as one among several subjects covered by a data centre are possible, e.g. the E-depot for Dutch Archaeology as part of DANS-EASY in the Netherlands.

Preventing a proliferation of repositories managing archaeological data

The existing examples of successful digital archives for archaeology are at the national / community level. Such digital archives are not yet available in many European countries, but arguably are the most effective solution. The alternative is a "many-headed beast" scenario in which many universities and other research institutes invest in data repositories, with a lot of duplication of efforts, re-invention of the wheel, competition for data deposits, etc.

The issue appears to be most critical in countries with many universities that may feel obliged to offer their researchers a data repository, in addition to or adapting existing document repositories to also manage properly datasets of different disciplines. The costs of maintaining a university-based repository for research data have been estimated to be *"an order of magnitude greater than that for a typical institutional repository focused on e-publications"* (Beagrie *et al*. 2008: 70).

In Germany a survey identified 141 university-based as well as some other open access (document) repositories (Vierkant 2013). If we assume that only the larger ones, those hosting 5000-50,000 content items, might aspire to manage also research data this would be 37 of the repositories (26%). Surveys on the demand of research data management support and pilot projects have been conducted already at several German universities.

One institutional repository for research data is already implemented at the Heidelberg University, HeiDATA[139], managed by a competence centre established by the university's library and computing centre. Only 50 studies (data sets) have been deposited since June 2014, including one dataset in the subject category History of the Ancient World (deposited on July 30th, 2014 and downloaded 101 times as of October 27th, 2015). If many such institutional repositories are being build and used, the result would be dispersed archaeological data sets. Therefore the IANUS Research Data Centre for Archaeology and Classical Studies[140] is a consequent initiative for a common solution in Germany.

In the UK the Archaeology Data Service is the acknowledged central archive for archaeological data, which makes efforts to manage such data by institutional repositories unnecessary. For example, the REWARD project[141] asked the British Library for DataCite DOIs for archaeological research data to be deposited in the institutional repository of the University College London[142]. But the British Library, the DataCite DOIs Allocation Agent in the UK, was not willing to provide DOIs due to their concerns about the suitability of institutional repositories for curating such data (cf. Hole 2013). A survey on

[139] HeiDATA, https://heidata.uni-heidelberg.de/dvn/

[140] IANUS, http://www.ianus-fdz.de

[141] REWARD - Researchers using Existing Workflows to Archive Research Data project (UK, 10/2011-03/2012), http://www.ucl.ac.uk/reward

[142] UCL Discovery, http://discovery.ucl.ac.uk

the value of data centres in the UK, including the Archaeology Data Service, found that their widely-agreed benefits relate to research efficiency and *"for the most part are a result of the data centre's status as a central and sizeable hub within its field"* (RIN 2011b: 7).

We think that the whole domain of archaeological research in Europe will be better off if universities and other research institutions can outsource the long-term curation and accessibility of archaeological research data to national or other community-level data archiving centres. In turn, the university departments and institutes could invest more effort in making researchers willing and ready to share their research data.

Such collaboration will be beneficial for both parties. As colleagues of one of ARIADNE's data archive partners suggest, service providers such as university libraries "*have an understanding of Open Access processes in general and might be good partners in promoting Open Access to research data. As service providers, they also have good relations to (their local) researchers. This kind of cooperation can thus create synergies – draw on other's knowledge and networks".*

Some local support with regard to data management planning, issues of IPR/licensing, funding of data deposit, etc. will be helpful and could be provided by dedicated staff of a research library (Green & Gutmann 2007). For institutions that wish to have information about deposited data, the central community archive could provide an institutional view/index of all research data provided by their researchers. Instead of maintaining an own data archive, more local efforts could, for example, be devoted to making legacy data and reference collections accessible (cf. *Section 9.1.2*).

Addressing the current lack of community-level data archives

While there are doubts about the suitability and cost-effectiveness of a potentially large number of institutional repositories, at the same time a central, national-level solution for archaeological data is missing in many countries. According to a DASISH[143] report the availability of digital archives for the humanities and social sciences in Europe is characterised by considerable regional differences; furthermore the report warns that many *"have insufficient resources to live up to future requirements"* (Marker 2014: 210).

In the area of social sciences research data the Consortium of European Social Science Data Archives (CESSDA)[144] has a large footprint of data archives and services. In the SERSCIDA project[145] members of CESSDA helped Bosnia-Herzegovina, Croatia and Serbia assess their readiness to establish an archive and take first steps in this direction. The project raised awareness and support for the importance of data archives, open access and other good practices among government institutions, funding bodies, universities and other data producers. Such support would also be necessary to support the building of archaeological data archives in many European countries.

National research funders that extend their open access mandate to data, require data management plans, etc. will indeed also have to consider if appropriate archives are available. In one ARIADNE partner country the main agency funding scientific research has extended their open access mandate to data and points to the Databib[146] list of data repositories. The list includes the archaeological repositories ADS (UK), DANS-EASY (NL), Open Context (USA) and tDAR (USA). A dedicated repository in the partner's country, where most of the funded archaeological research is conducted, is not present and does not exist as yet. But the researchers' would prefer to deposit their data in a

[143] DASISH - Data Service Infrastructure for the Social Sciences and Humanities (EU, FP7, 01/ 2012-12/2014), http://dasish.eu

[144] CESSDA - Consortium of European Social Science Data Archives, http://cessda.net

[145] SERSCIDA - Support for Establishment of National/Regional Social Sciences Data Archives (EU, FP7, Support action, 01/2012-06/2014), http://www.serscida.eu

[146] Databib, http://databib.org; now incorporated in the re3data registry, http://www.re3data.org

national archive for archaeology, together with data of other national research groups, and in the national language, rather than elsewhere in Europe or overseas.

It appears that archaeology is a case for which research funders will have to help implement an appropriate national solution. In Germany such a discipline-oriented solution is being prepared, the IANUS - Research Data Centre for Archaeology and Ancient Studies[147], funded by the German Research Foundation. The IANUS initiative is coordinated by the ARIADNE partner German Archaeological Institute. Furthermore, as reported by Corns *et al*. (2014), participation in ARIADNE has already promoted initiatives in other countries: In Ireland, the Discovery Programme together with archaeological institutes aims to develop a framework for archaeological research data that might be deposited in the recently formed Digital Repository of Ireland. Inspired by the example of the DANS E-Depot for Dutch Archaeology, the Society of Slovenian Archaeologists, including researchers of ARIADNE partner ZRC SAZU - Institute of Archaeology, promotes an initiative for a national archaeological data repository under the Slovenian Ministry of Culture.

ARIADNE partners understand the importance of community-level national or other archives (e.g. international subject/discipline-based archives). In the consultation for the innovation agenda colleagues (archaeologists) of two different countries suggested:

"Promote the idea of creating National or European Digital Repositories in order to: a) support a generally accepted standards for creating and documenting the digital archaeological data, and b) avoid the creation of multiple small repositories that cannot be easily tracked and searched in the world wide web";

"National(!) repositories: To archive archaeological data, get support in data management / guidance throughout the project, have a central access point to get information about all archaeological data from a country".

Community-level solutions are more likely to foster common standards and act as hubs for data integration and access. The institutional model, repositories of single universities or research institutes, provides a local solution but does not help much in setting common standards and overcoming the fragmentation of archaeological data overall.

5.3 Extra costs of open data sharing

There are of course not only benefits but also costs of open data sharing through digital archives. While open data advocates in the research community emphasise the benefits, sceptical voices note that little funding for data curation has been invested since long, and question if the allocation of resources for open data initiatives will be up to the challenge. In the current economic climate it seems unlikely that extra money for open research data initiatives can be mobilised, which means that it would need to be taken from already restrained research funds.

A clear priority is that digital archives are available where researchers can deposit and provide access to their data in a trustful and open manner. Archives set up for open data curation and access should be stable in the long-term in order to promote trust that the extra effort researchers have to put into sharable, fit for re-use data is well spent. Uncertainties around the sustainability of mandated digital archives may reduce the enthusiasm of the proponents of open data and increase the unwillingness of the yet-to-be-engaged. Long-term commitment for sustained support (10+ years) by the main funding bodies would be helpful in this regard.

But research funders that issue an open data mandate, and will have to support related cost factors, need solid information about expectable costs. The total costs of open access data, which include the researchers' effort for preparing shareable data (e.g. data description) and the costs of the

[147] IANUS, http://www.ianus-fdz.de

organisations that curate the data for long-term access, are not fully clear. Rather well known are the costs of preservation solutions (Rosenthal 2014), but these costs are only a part of digital curation, which has several cost elements, including those for adding value to available data (e.g. adding or linking to related information).

5.3.1 Cost models and cost/benefit analysis of digital curation

Cost models for digital curation services have been developed and/or evaluated by many projects, more recently for example by LIFE3[148], APARSEN[149] and Collaboration to Clarify the Cost of Curation (Costs4C)[150]. Individual organisations have developed their own models, e.g. Danish National Archives and Royal Library (Bøgvad-Kejser *et al*. 2011)[151], Data Archiving and Networked Services - DANS (Palaiologk *et al*. 2012), or the University of California Curation Center (UC3)[152].

A common feature of the cost models is that they cover the different phases/activities of the curation lifecycle of digital objects, i.e. data acquisition, ingest, management, preservation (incl. migration), and of course providing regular access to the preserved data. Most models relate cost components to the Open Archival Information System (OAIS) Reference Model and some to auditing schemes for trusted digital repositories, TRAC or ISO 16363:2012 (which are addressed in *Section 5.4.1*).

One highly elaborated model has been developed by the Data Archiving and Networked Services (DANS) institute for their digital archive, which focuses on research data of arts & humanities and social sciences (Palaiologk *et al*. 2012). The activity-based costing (ABC) model has been tested on empirical cost data from activities performed by all DANS employees. This resulted in a detailed insight into the labour intensity of activities and a “euros per dataset” unit of cost measurement. The DANS model also applies a Balanced Scorecard approach to connect derived information to institutional goals and support decision-making on data archive services (e.g. improvement, extension, etc.).

Wheatley (2012) addresses the issue of too many and too complex cost models, and suggests the modelling community should converge towards one curation lifecycle based model that can be employed easily by most institutions[153]. The APARSEN project surveyed research libraries (members of LIBER) on their awareness and use of curation cost models. The results for “On what basis would you select a cost model” (81 respondents) were: 63% would chose a model that has been validated by similar organisations in their sector, followed by model scope (55%) and easy to use and

[148] LIFE - Life Cycle Information for E-Literature (UK, JISC, three phases, 2006-2010), http://www.life.ac.uk; see Hole *et al*. (2010).

[149] APARSEN - Alliance Permanent Access to the Records of Science in Europe Network (EU, FP7, 01/2011-12/2014), http://www.alliancepermanentaccess.org/index.php/aparsen/; APARSEN (2013b/c) provide an evaluation of a number of cost models and results of model testing.

[150] Costs4C - Collaboration to Clarify the Cost of Curation (EU, FP7, 02/2013-01/2015), http://4cproject.eu; see in particular Costs4C (2014a/b). The project has also established the Curation Costs Exchange (CCEx), which provides rich information on cost concepts, cost models and drivers, and sustainability planning; also a cost comparison tool is provided, http://www.curationexchange.org

[151] Danish National Archives / Danish Royal Library, http://www.costmodelfordigitalpreservation.dk

[152] University of California Curation Center (UC3): TCP - Total Cost of Preservation (2011-), https://wiki.ucop.edu/display/Curation/Cost+Modeling; UC3 provides data services for the Open Context archaeological data publishing platform; at about 1 million objects the Open Context collection is the largest by number of objects in the Merritt repository of the California Digital Library (Willett 2015); UC3 - Merrit, http://www.cdlib.org/uc3/merritt/

[153] Wheatley (2013) also provides a list of modelling projects; summaries of several models are available from Costs4C, http://4cproject.eu/summary-of-cost-models

adaptable (44%); other criteria such as payment for model use or available support were seen as less important (APARSEN 2013b: 27).

The purposes of modelling, collecting and analysing cost figures of curation activities are to evaluate current and predict/control future costs, establish and adapt pricing schemes, and establish transparency of curation costs for stakeholders (e.g. researchers, research councils, funding bodies). Being clear about the costs can contribute greatly to the sustainability of digital archives, but this is not enough, also – and arguably even more so – the benefits of data curation and access need to be established and communicated.

Cost/benefit analysis is difficult and hence a missing component of most projects. Fry *et al*. (2008) conceived a framework for "making a business case" and estimated cost-benefit ratios of curating and open sharing of research data produced by UK higher education and research institutes. However, this framework is based on hypothetical cost figures and assumptions. Particularly noteworthy for individual institutions are the Keeping Research Data Safe (KRDS) Benefits Framework and Analysis tools[154] for assessing the benefits of digital curation of research data. The framework includes about 30 examples of benefits for researchers, institutions, society. Some examples relevant to researchers and institutes are scholarly communication/access to data, verification of research integrity, no re-creation of data, re-use/-purposing of well-curated data, increased visibility/citation, motivating/input for new research, stimulating new networks/collaborations.

5.3.2 Archive curation costs – some rules of thumb and examples

The Keeping Research Data Safe (KRDS) projects have collected and analysed curation costs of several digital archives, both institutional as well as domain-based archives (Beagrie *et al*. 2008; Beagrie *et al*. 2010). Briefly summarised the following "rules of thumb" may apply for different archives (cf. Beagrie *et al*. 2010: 79-80; Beagrie & JISC 2011):

- Staff is the major cost component, c. 70% of the overall costs (other components include offices, hardware & software, external services, among others).
- Data acquisition and ingest costs most, while archival storage and preservation activities are a significantly smaller part of the overall costs; providing access ranges in between. For example, the staff costs for these activities of the Archaeology Data Service (2005-2009) on average were: c. 55% outreach/acquisition/ingest, c. 15% archival storage and preservation, c. 31% access (cf. Beagrie *et al*. 2010: 31-34; see also Richards *et al*. 2010).
- Access services is an area where costs can vary considerably (depending on the kind of access services, level of user support), while archival storage and preservation costs are likely to decline significantly over time.
- A certain level of skilled staff (different categories) and IT is needed to provide reliable services, representing the regular "fixed costs" of an archive. Keeping these costs stable while curating larger data collections allows economies of scales (lower per-unit cost).
- Improvement of labour-intensive curation activities, particularly ingest and access, through streamlining and tool-support can allow significant cost reduction.

The actual costs depend on many factors which include, but are not limited to, the type of data repository/archive (institutional, national domain-based or other) and the types of research data/ objects.

[154] Charles Beagrie Ltd.: Keeping Research Data Safe (KRDS), http://beagrie.com/krds/; the UK JISC funded KRDS projects ran from 2007 to 2011, led by Charles Beagrie Ltd.; the Archaeology Data Service has been one of the project partners.

The operation of a research data repository costs more than running a document repository, indeed the costs are *"an order of magnitude greater"* (Beagrie *et al*. 2008: 70). At the level of institutional repositories Keeping Research Data Safe case studies indicate that a document repository may do with 1 FTE (full-time equivalent) staff and equipment costs of £ 1300 p.a. [SHERPA project], whereas a data repository 2.5 FTE and equipment costs of £ 27,500 p.a. [King's College London case study] or 4 FTE and £ 58,800 p.a. equipment cost [Cambridge case study] (Beagrie *et al*. 2008: 67-71). The *Science as an Open Enterprise* report of The Royal Society (2012a [Appendix 3]) provides more or less detailed figures of cost components of seven other document and data repositories[155].

Also the types of research data/objects that are curated can make a large difference in terms of costs as well as in other respects. For example, DANS in the development of their cost model (addressed above) found that the curation of archaeological data sets costs considerably more than other humanities and social sciences data sets. This is due to higher variety and complexity of data formats (databases, images, CAD, geodata/GIS and others)[156]. DANS curates the archaeological data as a separate collection, the E-depot Netherlands Archaeology, within their archiving system (EASY). The reported curation cost breakdown for the archaeological data is 36% Ingest, 44% Data Management, 12% Archival storage, 1% Archival administration, 1% Preservation, 6% Access (Palaiologk *et al*. 2010; Palaiologk 2013).

Kintigh & Altschul (2010) discuss the cost model of Digital Antiquity's tDAR online repository for archaeological data and documents in the United States[157]. They note that the Archaeology Data Service (UK) had projected *"that proper digital curation alone costs between one and three percent of total project costs. In the US total curation costs for both physical and digital materials on many projects do not reach three percent".*

As a major difference between ADS and Digital Antiquity they consider that the ADS curation model *"requires relatively little of the user but charges a relatively large fee to cover its staff's curation efforts. This model is successful both from the standpoint of the quality of the curation and from the standpoint of financial sustainability. However, a by-product of this high staff investment, high charge model is that the volume of projects processed is relatively low – on the order of a few hundred annually. In contrast, Digital Antiquity's design anticipates eventually being able to process a substantial fraction of the projects conducted each year in the US".*

Kintigh & Altschul (2010) assumed that Digital Antiquity can bring down the tDAR fee through online tools and guidance so that the depositor executes more of the curation process, particularly related to metadata, with no or only little assistance by a curator. The same approach has been taken by the Archaeology Data Service with ADS-easy[158] for small-sized project archives of files of a common type; the first ADS-easy submitted archive was released in June 2014 (Moore *et al*. 2012; Moore 2014). Kansa (2015) describes the labour intensive data preparation and publishing process of Open Context, which allows a highly granular access and citation of excavation and other items.

---

[155] Document repositories: Arxiv.org (international), Eprints Soton and Dspace@MIT (institutional); data repositories: Oxford University Research Archive and DataBank (institutional), DRYAD (bio-sciences research community), UK Data Archive (social and economic research data, national archive), Worldwide Protein Data Bank (international).

[156] Costs4C project: Summary of DANS' Cost Model for Digital Archiving, http://4cproject.eu/summary-of-cost-models/16-community-resources/outputs-and-deliverables/107-cost-model-for-digital-archiving-cmda

[157] Digital Antiquity, http://www.digitalantiquity.org; tDAR - The Digital Archaeological Record (hosted at Arizona State University, USA), http://tdar.org

[158] ADS-easy: http://archaeologydataservice.ac.uk/easy/

5.3.3 Research data archives costs and funding

Research funders demanding open access research data will also have to consider how to fund appropriate curation solutions. In the current economic climate it seems unlikely that extra money for data curation can be mobilised; a shift of funding budget from research projects to appropriate and cost-effective data curation and access solutions seems more likely.

In this regard it would be interesting to know the current percentage of data curation costs of the total publicly funded research in different country, ideally per discipline, and curation levels, i.e. institutional repositories, domain-based national archives, and contributions to international archives and databases. We could not find such figures, but expect that in the current push for open access research data the interest in such figures will increase rapidly.

At the national level domain-centred data archives are considered the most appropriate solutions for research data. Because they establish common, community-level standards and resources, and avail of the special expertise as required for the types of data and data-related issues common in the domain. Such archives will very likely receive the largest share of funding for research data curation, directly (basic funding, to maintain data curation and access) and indirectly (from research grants, for new deposits).

The *Science as an Open Enterprise* report of The Royal Society (2012a) considers that large scale digital curation at the level of national domain-centred data archives *"is typically between 1-10% of the cost of research"* (The Royal Society 2012a: 66). Examples that suggest this range are that the British Geological Survey has an annual budget of about £30 million of which it spends £350,000 on the National Geosciences Data Centre (1.2%). The UK National Centre for Atmospheric Science receives £9 million funding a year of which about £1 million (11%) is spent on curating 228 datasets through the British Atmospheric Data Centre[159] (The Royal Society 2012a: 68). These are examples of the earth sciences. It seems unlikely that other national domain-centred data archives/centres (e.g. an archive for archaeological research data) would require over 10% of the domain funding of research, considerably less (maybe around 5%) may be sufficient.

Shifting funds for new research to curation of data may raise some concerns, which should be addressed making clear the benefits of open data in several respects (cf. *Section 4.3.1*). This includes exploitation of existing data which may generate more published research results than a couple of projects which are not funded. For the bio-sciences data repository DRYAD[160], launched in January 2008, it has been argued that the investment in its annual costs of about $400,000 (2010) within four years could yield more than 1000 papers based on re-used data; the same amount of money invested in research projects would only produce 16 papers (Piwowar *et al*. 2011)[161].

A survey on the value of data centres in the UK (including the Archaeology Data Service) found that their most widely-agreed benefits relate to research efficiency, i.e. that they *"make research quicker, easier and cheaper, and ensure that work is not repeated unnecessarily".* The acknowledged benefits *"for the most part are a result of the data centre's status as a central and sizeable hub within its field"*

[159] An analysis of the value and impact of the British Atmospheric Data Centre (BADC) has been conducted by Beagrie & Houghton (2013b).

[160] DRYAD, http://datadryad.org; the archive has been supported by National Science Foundation (NSF, USA) grants since September 2008.

[161] The figures are based on a comparison of DRYAD with the GEO - Gene Expression Omnibus, which archives data of functional genomics studies. GEO received 2711 data deposits in 2007 which by the end of 2010 were estimated to have been re-used for 1150 papers published by third-party authors. The comparison was rather speculative, because at the end of 2010 GEO held over 20,000 data sets (Barrett *et al*. 2011) while DRYAD only 1000 data files (http://blog.datadryad.org/2011/01/). But the DRYAD content and use grew rather fast: On the 20th of October 2015 the DRYAD figures (online) were 10,434 data packages (33,577 data files) and the total number of downloads was 2,164,811 (109,894 of which in the last 30 days).

(RIN 2011b, 6-7). In a study of the value and impact of the Archaeology Data Service (ADS) the increase in research efficiency of the users of ADS curated resources was calculated to be worth at least £13 million per annum – 5 times the costs of operation, data deposit and use; including other advantages, the study concluded that £ 1 investment in ADS yields up to £ 8.30 return (Beagrie & Houghton 2013a; Richards 2014).

Eric Kansa of the archaeological data publishing service Open Context sees open data initiatives as an opportunity to address inbred dysfunctions of the academic system which motivate withholding rather than sharing of data. Removing the dysfunctions would allow expanding the scale, quality, diversity and effectiveness of archaeological studies and communication. In particular he stresses, "*If we can find ways to finance costly excavations, we should be able to finance better stewardship and dissemination of the results of those excavations"* (Kansa 2012: 14).

Kansa gives an example of a long-term excavation which suggests that the costs of making archaeological data openly available online are only a fraction of the overall costs of such a project, in this case below 2%: *"In our experience with Open Context, data publishing and archiving with the California Digital Library need not bust the budget of research projects. For example, Open Context recently published data from nearly ten years of excavations at Kenan Tepe, a late Neolithic – Iron Age site in Anatolia (Parker and Cobb 2012)*[162]*. Bradley Parker, the project director, estimated that excavation and documentation required roughly $800,000 in direct costs. Publication of this very large and complex dataset in Open Context cost between $10,000 and $15,000 (mostly labor). Though $15,000 is a large number for archaeologists, it bought comprehensive dissemination of a large, openly licensed and wholly reusable body of data that would otherwise be lost to the larger community and represents only a few percent of the overall research budget"* (Kansa 2012: 14)[163].

Importantly, the costs of this online publication are mainly labor costs for preparing and structuring the data and related editorial work. The figure of 2% of the total excavation and documentation work does not mean just putting data files in a digital repository (which would not cost as much). As noted in the previous section, according to Kintigh & Altschul (2010) the Archaeology Data Service (UK) considers that proper digital curation costs between 1-3% of total project costs, while in the United States the total curation costs for both physical and digital materials on many projects are below 3%.

In summary, we assume that the costs of (post-)project data curation and online publication/access of archaeological projects (field surveys, excavations) are only a fraction of the total project costs, somewhere between 1-3%. Furthermore, that the overall domain funding for archaeological research data curation by appropriate archives might be around 5%. The archives would be national-level community archives, which receive funding directly (basic funding, to maintain data curation and access) and indirectly (from research grants, for new deposits).

Referring to the demonstrated costs/benefits ratio of the Archaeology Data Service (see above), we assume that research funders can enable substantial cost efficiencies by supporting one state-of-the-art national data centre and requesting that all archaeological data is deposited with that centre. This certainly is a better solution than spreading funds across many hard to sustain repository projects of individual institutes. A central digital archive will also be capable to support more easily research evaluation measures based on data sharing and re-use indicators.

[162] Parker B. & Cobb P. (eds., 2012): Kenan Tepe: Digital Data and Media (Released 2012-06-30). Open Context, http://opencontext.org/projects/3DE4CD9C-259E-4C14-9B03-8B10454BA66E

[163] Open Context charges a one-time fee for data publishing services, including archiving with the California Digital Library. The archival curation and publication costs of the Kenan Tepe data are in the high range of Open Context fees. Fees of course vary depending on the complexity and size of the database and related content. Typical seem $5000 for a large project (e.g. a 5-year excavation), $1000 for a medium-size project (e.g. a ceramics database), or $350 for the documentation of a short term archaeozoological contract. Cf. the cost examples at: http://opencontext.dainst.org; Open Context offers a cost estimation form: http://opencontext.org/about/estimate

5.4 Trust in digital archive services

Well-managed and openly accessible data archives are a key success factor of e-infrastructure initiatives, because ultimately researchers want to discover and use data which they need for their research. Therefore e-infrastructure initiatives should help to establish and federate digital archives that store, curate and give access to data in trusted ways. Data archiving required in research grants and other policy-based archiving will raise the demand for trusted and certified data archives.

5.4.1 Requirements for trusted and certified digital archives

Core requirements for a trustworthy digital archive/repository are summarised in the "Ten Principles" defined in January 2007 by representations of the Center for Research Libraries (North America), Digital Curation Center (UK), DigitalPreservationEurope (EU-funded project, 2006-2009) and NESTOR, the German Competence Network for Digital Preservation (Center for Research Libraries 2007). The principles for example include that the repository commits to continuing maintenance of digital objects for the user community, has an effective policy framework, avails of technical infrastructure adequate to long-term preservation and security of the digital objects, and demonstrates organisational fitness (e.g., financing, staffing and routine processes).

A major framework for digital archives is the International Organization for Standardization (ISO) standard 14721:2012 - Open Archival Information System (OAIS) Reference Model[164] (see also CCSDS 2012). The OAIS framework has been adopted by many digital archives, however, for purposes of certification and trust building OAIS compliance remains on a rather generic level (cf. Lavoie 2014). More practical oriented is Trustworthy Repositories Audit and Certification: Criteria and Checklist (TRAC, final version 2007), which has been superseded by the TRAC-based 16363:2012 - Audit and Certification of Trustworthy Digital Repositories standard of the International Organization for Standardization (ISO)[165]. The ISO standard defines attributes repositories must have to meet the certification requirements.

A special mention also deserves the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)[166], jointly developed by the Digital Curation Center (UK) and DigitalPreservationEurope (EU-funded project, 2006-2009). DRAMBORA allows repositories to assess the risks to their operation, from direct physical risk (such as fire and flood) to less tangible risks such as reputational damage. Also a number of other projects have produced valuable guidelines, for example, the UK JISC/NERC-funded PREPARDE project with special emphasis on the data publication function of digital archives (Callaghan *et al*. 2014).

Outstanding in terms of community uptake and support is the Data Seal of Approval (DSA).[167] The DSA guidelines were initially developed by Data Archiving and Networked Services (DANS, Netherlands) and in 2009 handed over to an international DSA board that assigns the seal to archives with proven compliance. The 16 DSA guidelines cover five criteria, which require that the archived data can be identified on the Internet, are accessible (including clear rights and licenses), are in a usable format, are reliable, and can be referred to (persistent identifier).

The DSA community has over 40 members, mostly institutions in Europe and with a focus on the social sciences and humanities, but also archives of other disciplines are present. To illustrate the

[164] ISO 14721:2012: Space data and information transfer systems – Open archival information system (OAIS) Reference model, http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

[165] ISO 16363:2012: Space data and information transfer systems – Audit and certification of trustworthy digital repositories, http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

[166] DRAMBORA, http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora

[167] Data Seal of Approval, http://datasealofapproval.org/en/

scope, the DSA community includes: Archaeology Data Service (UK), Berlin-Brandenburg Academy of Sciences and Humanities (Germany), Inter-university Consortium for Political and Social Research (international), Pacific and Regional Archive for Digital Sources in Endangered Cultures (Australia), Strasbourg Astronomical Data Center (France), U.S. Geological Survey. Notably, several of the community members take part in the CLARIN - Common Language Resources and Technology Infrastructure initiative which has promoted the Data Seal of Approval among its participants.

The Data Seal of Approval is intended as a basic certification scheme with a focus on reliably accessible data, while more ambitious digital archives may aspire for certification based on the ISO 16363:2012 standard. Archive sustainability (e.g. sustained funding) of course cannot be certified, but may result from being a digital archive that is trusted and actively used by the research community it aims to serve.

5.4.2 Factors that promote trust in digital archives

Evidence for the impact of certification on researchers' perception of digital archives is limited, although one would like to know, for example, if certified archives are considered as more stable than others or if they are more likely to attract depositors.

Swauger & Vision (2015) conducted a survey to better understand the factors that influence where researchers deposit their data. The survey participants were 140 researchers who had deposited phylogenetic data in DRYAD or the TreeBASE repository. The respondents were asked to rank the relative importance of eight repository-related factors: prestige, specialisation, trust (in the repository's persistence), ease of data submission, extent of required metadata, direction (by journal, institution or research funder), policy that allows to control re-use (embargoes, licenses), and promotion of scholarly credit (e.g. DOIs for data citation).

The respondents ranked the factors affecting their choice of a repository similarly. Most important were specialisation, ease of use, trust, and direction, although respondents were strongly divided as to whether the latter was among the most or least important factors. Mandates by research funders to deposit specific data in a certain repository could of course trump the other factors. The study authors were especially surprised about the relatively low importance of measures that allow researchers to control re-use. Such measures, especially embargoes, are often considered as vital for promoting open data (e.g. Roche *et al*. 2014).

Yakel *et al.* (2013) report results of in-depth interviews with 44 quantitative social scientists (experts and novices) and 22 archaeologists about factors that influence their trust in digital repositories. The study did not focus on depositing of data but different experiences with repositories. It was understood that the social scientists (at least the experts) were more experienced in this regard. The interviews were recorded, transcribed and coded with the NVivo software for qualitative data analysis. Four categories of factors were distinguished: trust in the organisation managing the repository (e.g. integrity, common values, transparency), data-centred factors (e.g. selection, meta-data, processing), structural guarantees (e.g. preservation/sustainability, third-party endorsement, reputation), and social factors (e.g. opinion/recommendation by colleagues).

The results show different patterns: Both communities associated trust with a repository's trans-parency, but the archaeologists were twice as likely to do so. The study authors thought that this may be *"because they have a culture of not sharing data and little standardized data collection, so clear indications of how the data were collected and managed are vitally important for reuse"* (Yakel *et al.* 2013: 154). Among the closely related data-centred factors archaeologists mainly addressed metadata, whereas this was not an issue for social scientists who discussed more frequently issues of data selection, cleaning and processing. The most important trust factors for archaeologists were guarantees of repository stability and data preservation, while the social scientists mostly pointed to institutional reputation, built over many years, as ensuring trustworthiness. Finally, the influence of colleagues on their trust in repositories was not important for archaeologists and among the social scientists mainly for novices.[168]

5.4.3 Mandated and recommended digital archives

Some research funders mandate one digital archive or suggest using one from a list of recommended archives. Mandated archives are usually domain-based. For example in the UK, the Archaeology Data Service (ADS) is the mandated archive for data of projects funded by the Arts and Humanities Research Council and the Natural Environment Research Council as well as the archive recommended by the British Academy, Council for British Archaeology, English Heritage and the Society of Antiquaries. In the Netherlands archaeologists since 2007 are formally obligated to deposit their data with the Data Archiving and Networked Services (DANS, established 2005), according to the Quality Standard for Dutch Archaeology (Kwaliteitsnorm Archeologie); the DANS-EASY system includes the E-Depot Dutch Archaeology (EDNA).[169]

Mandated archives will of course devote special attention to measures that promote trust, including certification. Both the Archaeology Data Service and Data Archiving and Networked Services (and many other digital archives) are certified according to the criteria of the Data Seal of Approval, indeed the criteria have initially been developed by DANS.

In the absence of a decisive deposit mandate, i.e. deposit is only recommended and several archives are considered as relevant, the certification of a data archive will be one factor in researchers' selection of an archive. The results of the Yakel *et al*. (2013) study suggest that archaeologists in particular appreciate an archive's transparency with regard to the description and management of data collections, and any indications of archive stability (e.g. sustained funding) and long-term data preservation. Recommendation by other researchers appeared as much less important.

The study by Swauger & Vision (2015) in addition highlights that researchers in specialised fields will look for archives that are able to manage their specific data appropriately. Criteria that apply when researchers go about depositing data, e.g. ease of use of deposit mechanisms and guidance, can influence the selection of a data archive only indirectly, through "word of mouth" or testimonials of other researchers on the archive website.

5.5 Suggested actions

Digital archives for long-term curation of and access to data provide core services underlying research e-infrastructures. Archives richly filled by the archaeological research community are one key to the success of the ARIADNE e-infrastructure initiative, which will provide the community with a solution for cross-archive search, access and (re-)use of available data resources.

Researchers are looking for relevant data, but are not necessarily willing to share their own data in an open manner. State-of-the-art, certified community archives allow researchers to publish their data in a secure and trusted way. The publication comes with a price tag, but compared to the total costs of archaeological investigations these costs are relatively small.

Substantial benefits are expected from the availability of ever more open research data. The chances that these benefits materialise appear to be higher if mandated community-level data archives are established. In any case, such archives will reduce the fragmentation of archaeological data and provide a more cost-effective solution than spreading funds across many data repositories of individual institutes.

[168] Kriesberg *et al*. (2013) investigated how re-use of data, including from digital archives, can contribute to the apprenticeship of students/young researchers in archaeology, quantitative sociology and zoology.

[169] DANS: E-Depot Dutch Archaeology (EDNA), http://www.edna.nl

Recognise that the costs of opening up archaeological research datasets are marginal and well spent

The costs of post-project data curation and online publication of archaeological projects are only a small fraction of the total project costs at around 1-3%, depending on the type of investigation and data generated. Compared to the benefits expected from open and re-usable research data this investment seems well spent.

Include the costs of open data sharing and digital archiving in project grants

All surveys on open data sharing show that researchers consider the related effort as a significant barrier. Specifically, this concerns the effort required to prepare shareable data and detailed data description. Therefore research funders should allow inclusion of the costs of this work in project grants. A project data management plan, as increasingly requested by research funders, is the ideal place to present these costs as well as the expected archive charge for long-term data preservation and access.

Recognise the advantages of domain-based community archives

Building and mandating community data archives is the most effective strategy to overcome, or at least reduce, fragmentation and inaccessibility of archaeological data resources. Advantages of central, domain-based archives include clear orientation for all stakeholders, focused mobilisation of data deposits, economies of scale, among others. Substantial return on investment through research efficiency and use also by non-academic groups can be expected from domain archives that provide access to a wider range of relevant data resources. Funding one data centre is very likely a better solution than spreading funds across many hard to sustain data repository projects of individual institutes.

Ensure long-term sustainability of trustworthy data archives

A clear priority for open data initiatives is that archives are available where researchers can share their data in a secure and trusted way. Such archives should be certified and stable in the long-term in order to promote trust that the effort put into re-usable data is well spent. Commitment for sustained support (10+ years) by the main funding bodies would be helpful in this regard. In the current economic climate it seems unlikely that additional funds for open data curation and access can be mobilised. This means that budgets may need to be shifted from already restrained research project funding to appropriate data curation and access solutions. Related concerns should be addressed by making clear the benefits of open research data.

Encourage and support initiatives for data archives in countries where these are currently lacking for archaeologists

The most effective solution for open archaeological data mobilisation, curation and access is a domain-based central archive (e.g. one place to go, community building, cost-effectiveness, and others). But such a data archive is missing in many European countries. Progress towards common solutions may be promoted through knowledge transfer between established data centres and initiatives for new archives in other countries. New entries may “leapfrog” to a state-of-the-art solution by learning from acknowledged benchmarks.

Devise interim archiving solutions for datasets that are inaccessible and at risk of loss

Archaeological research institutions should take stock of valuable old datasets as well as datasets not maintained by researchers actively using the data. Where a data archive for long-term curation and access is missing, devise an interim local or outsourced solution to prevent inaccessibility and potential loss of valuable datasets. A priority of course is excavation data that cannot be re-collected.

6 Focus area 4 – Capacity building for open data sharing

6.1 Introduction and overview

This chapter addresses the need of capacity building for open data sharing at the institutional level and some issues concerning the required quality of data description and review. To support research project data management and sharing, universities and research institutes will have to put in place adequate institutional policies, guidance, training and other support. Strong institutional support of the open data agenda will allow more archaeological data (and metadata) flow into data archives for long-term curation, discovery and access, and (re-)use in further research.

Current situation of institutional data management support

At present the level of data management support by libraries/repositories of universities and other centres of research and education is rather low. The focus is mainly on institutional policy support and advice (e.g. issues of open access, data management plans, IPR/licensing, etc.), and organising training for research students. The main challenges for the research libraries, and institutional document repositories they manage, are limited resources (funds, personnel) and skills gaps with regard to research data curation. Moreover their potential role in managing or providing support in the management of research data is an on-going topic of discussion.

There are considerable doubts about attempts of individual universities to implement research data curation. Proper curation of research data requires specialisation, and leading examples of data archiving services indeed are specialising in research fields. Examples are the data centres funded and mandated by the Research Councils UK (e.g. Archaeology Data Service), or Data Archiving and Networked Services (Netherlands), which centre on social sciences and humanities (including archaeology). Specialisation is difficult to achieve for many different disciplines, e.g. all disciplines present at a university. Also a small or medium-size archaeology institute will hardly be in a position to maintain a data archive, and generally prefer to focus on research rather than data curation.

Support for managing data during project work

Rather than building research data archives, it is preferable that university departments and research institutes focus their efforts on training and support for researchers, so that shareable data emerge from the research process and are provided to appropriate subject/domain-based archives. More efforts could also be devoted to making legacy data, reference collections and other institutional assets available.

It is understood that proper research data management should start and be supported as early as possible, while researchers are working on their projects. Data management plans as requested by ever more research funders may provide a basis for this support. For the management of "active data" various methods are suggested, which range from plug-ins for software already used by researchers to automatic capture of information (metadata), which would reduce the effort required for regular documentation of research activities and results. At a minimum institutions or research networks could offer researchers a safe and controlled environment for storing and sharing data during collaborative project work.

High-quality metadata as required for data re-use

The latest when researchers go about sharing data through an archive/repository the question of metadata comes up. All studies on data sharing through digital repositories (including the ARIADNE survey) found that researchers consider the effort to provide the required metadata as a barrier to open data sharing. While data repositories and users would benefit from high-quality metadata, data sharers face the burden and usually prefer not to invest much effort on providing metadata.

Asking for high-quality metadata is likely to result in fewer contributions, which is one reason why many repositories have rather shallow discovery metadata. Such metadata is insufficient to assist data re-use, which requires data description so that potential re-users can understand the data provenance/context, evaluate if the data is relevant for intended purposes, and use it properly to prevent incorrect conclusions. Requesting high-quality metadata arguably is possible only for domain-based archives that are mandated or recognised as the best place to share valuable data according to community standards. A further question addressed in this context is the effectiveness of metadata for cross-domain data re-use.

Data papers and data review

A novel approach that offers an incentive for researchers to provide rich data description is the peer-reviewed "data paper". Ever more journals are established that invite researchers to describe datasets they have deposited or databases they maintain. Moreover data review has become a topic of much recent discussion. It is anticipated that the open data policies of research funders and journals will bring about a wave of data in need of quality review. Therefore new models of peer-review need to be considered.

6.2 Data management planning and support

6.2.1 Data management and access plans

Data management planning has become an important topic in publicly funded research. A growing number of research funders require a data management plan and the objective is making the data generated in the funded projects openly accessible.

A data management plan should describe the kinds of data that will be created or collected, the methods applied, the data documentation (metadata), where and how long the data will be stored, and ways how the data could be shared and accessed. The grant applicant also has to justify why the proposed measures are considered optimal, and explain which limitations may apply, e.g. due to ethical or intellectual property issues.

The requirement of data management and access planning may bring the individual researcher (PhD candidate, principle investigator of a project) in a difficult situation if there is no national or other community digital archive for his/her data and the home institution (university, research institute) has no capability for long-term data archiving and access.

As more research funders mandate proper management, archiving and sharing of data researchers are turning to their institutions for support in all related matters. Some institutions have already taken steps to provide advice and support (e.g. on how to draw up a data management plan, where to deposit data, legal matters, etc.) while many others are not yet prepared to meet this unexpected demand and the newly assigned responsibilities.

Results of the ARIADNE online survey and interviews indicate that data management support for archaeological researchers by their institute is rather low. Furthermore the burden of maintaining project data rests with the researchers, who are not the most appropriate stewards for long-term data curation and access (see section below). The ARIADNE survey found that the responsibility for maintaining the data after project completion remains with the researchers (54% with the project manager, 27% an appointed member of the research team, and 19% "other")[170]. Major reasons for this arguably are the understanding that the data is owned by the researchers, including the difficult case of collaborative projects, and lack of data curation expertise and resources at the institutional level.

[170] ARIADNE First Report on Users' Needs (D2.1, April 2014), p.100.

Data management planning is not only vital for research projects but also at the institutional level in order to ensure that the necessary policies, resources and skills are in place. Data policies at the institutional level can allow roles and responsibilities with regard to data curation and access to be set forth and monitored internally.

We may expect that archaeological research institutes will increasingly become accountable for data that is produced by the researchers they employ. This will be due to conditions such as grant contracts for research programmes, legal regulations or professional codes of conduct. There is a need for more guidance and support as researchers and research institutes are confronted with requirements imposed by various actors, including public administrations (e.g. heritage management), research funders, professional organisations and others, as well as internal guidelines and audits.

6.2.2 Individual data management does not ensure long-term access

With regard to curation of data for long-term access the scientific enterprise cannot count on the individual researcher or research group. Such curation simply goes beyond the core interest and expertise of researchers, which are focused on carrying out research work.

Many factors such as changes in research groups (e.g. retirement or move of staff elsewhere), shifts in research focus, and others make the long-term curation of data of completed projects for future re-use rather unlikely. Rather funding for new projects needs to be secured, new avenues of research explored, fresh results published, etc. Therefore researchers *“will tend to regard data curation as a set of optional activities to complete once the pressure is off... and it never is!”* (Rusbridge 2007).

The typical process is that data loses its value for the researchers when the project results have been published, the data becomes obsolete and remains on PCs, carrier media, restricted access servers, and is eventually discarded or lost otherwise. Archaeologists, like researchers in other disciplines, will often keep data in ways that involve a high risk of loss, e.g. if a server crashes, carrier media become unreadable, or data are perceived as obsolete and deleted. As noted by the INCREMENTAL project group[171], *“We found that many researchers: (i) organise their data in an ad hoc fashion, posing difficulties with retrieval and re-use; (ii) store their data on all kinds of media without always considering security and back-up; (iii) are positive about data sharing in principle though reluctant in practice; (iv) believe back-up is equivalent to preservation”* (Ward *et al*. 2011).

To prevent data loss and ensure accessibility appropriate and stable data archiving solutions need to be put in place. Below we give some examples of inaccessible or lost data managed by researchers of different disciplines which illustrate the case:

The Consortium for Political and Social Research (ICPSR) in 2007-8 conducted a survey of thousands of US NSF- and NIH-funded quantitative social science studies since the 1970s (Pienta *et al*. 2008). Of the 5848 grantees the investigators could reach via e-mail 2548 (43.6%). 1669 confirmed having produced quantitative data or the ICPSR investigators had previously identified that this was the case. 334 (20%) of the grantees archived the data in a repository, 823 (49.3%) did not but still could find the data, while 512 (30.7%) could not locate and retrieve the data.

Vines *et al*. (2014) conducted a survey on 516 papers published between 1991 and 2011 that reported a discriminant function analysis on morphological data from plants, animals or other organisms. They could collect 385 e-mail addresses of corresponding authors from the articles or by searching online and requested the data used in the analysis. 101 datasets were received (26% of the contacts), another 20 were reported as still in use and not be shared. Of the papers based on these

[171] INCREMENTAL (UK, JISC Managing Research Data programme, 10/2009-03/2011), the project helped building data management capacity within the Universities of Cambridge and Glasgow, http://www.lib.cam.ac.uk/preservation/incremental/

131 datasets, 24 (18%) were published before and 97 (74%) after 2000. Because of lack of working e-mails, non-responsiveness of authors or reported loss or difficulty of retrieving the datasets only 11% of the datasets of all papers published before 2000 (167) could be collected.

A study of astronomy papers published in the years 1997 to 2008 in the four main journals of the American Astronomical Society investigated the stability of links to related data (Pepe *et al*. 2014). A total of 13,447 potential links in 7641 publications were identified. The number of links increased greatly until 2005 when it levelled off at around 1500 per year. In 2011, 44% of the links published in 2001 were broken, while of the links published in 2008 three years later about 10% did not work. Links to personal websites proved to be much more vulnerable than those to collections managed by institutions, surveys or telescope sites, which still showed a roughly constant percentage of broken links between 15% and 20%.

These figures make clear that research data should be deposited in digital archives specifically designed for long-term data curation and access, and likely to stand the test of time. As research funders and publishers increasingly require that data is made available through appropriate archives we may expect a considerable improvement in the long-term accessibility of research data.

Still it will be necessary to rescue valuable datasets and collections from previous research that are not curated and in danger of loss. In such cases the rescuers will often have a difficult job as the data may be stored on obsolete media and/or in unreadable file formats, metadata will very likely not exist, and the original producers may not be available to help with understanding and documenting the data. As expertise in data rescue for long-term archiving may not be available at many institutions, guidance and support by experienced curators is necessary. For instance this includes guidance with regard to media recovery, format migration, data recovery, appraisal and selection, metadata creation, protection of confidential information, among others.

6.2.3 Current data management support by research libraries

Several surveys show that at present the engagement of research libraries in data management is rather low.[172] The focus is mainly on policy support, advisory and training services (e.g. open access, data management plans, IPR/licensing, and organising training for research students). Support to the technical aspects of data curation such as metadata creation is rather limited, according to the mentioned surveys between 10-20%. While this is the general situation, there are of course some more advanced initiatives of research libraries, typically based on a close cooperation with the university's central IT service and/or an informatics and computing centre. The best results arguable have been achieved were dedicated units for research data management support have been set up, pooling available and bringing in new expertise in research data management and access systems.

The main challenges for the research libraries (including institutional repositories) are limited funds, existing skills gaps with regard to data curation, and managing relationships with other parts of the institution, including the library's perceived role and credibility. It can be assumed that a broad engagement of research libraries will only be possible based on a sufficiently large workforce of skilled "data librarians". This professional role however is perceived as "a gap in the market" (Hyams 2008), a "field undefined" (Alvaro *et al*. 2011), and an "accidental" rather than clear career choice (Pryor & Donnelly 2009).

In the United States the DataRes project identified lack of training, certification and other types of professional preparation as one group of deficiencies in academia's readiness for research data management. The others were lack of priority among researchers, lack of institutional mandates and

[172] Cf. Corrall *et al*. 2013 (88 institutions located in Australia, New Zealand, the UK and Ireland); Cox & Pinfield 2013 (81 higher education institutions in the UK); Tenopir *et al*. 2012 (221 members of the US Association of College and Research Libraries).

lack of required organisational structures. Still worse, the project found *"virtually no one in academia perceives that they have a professional responsibility or mandate for research data management functions"* (CLIR 2013: 6). The report notes *"that librarians may be the closest to understanding their role in research data management"*, but needed re-training and appropriate resources. These however were likely to be withhold by university administrations due to an absence of clear guidance from the federal agencies and other funders concerning data management and sharing mandates, and administrations' view *"that there is no point in investing time, money, and staff without a clear return on investment"* (CLIR 2013: 36).

In Europe, the E-Science Working Group of the Association of European Research Libraries (LIBER) in their recommendations for libraries "to get started with research data management" (LIBER 2012) write that data librarianship *"probably is a profession in itself but a lot of work regarding data services can also be done by e.g. (reskilled) information specialists"*. The group considers reskilling as a critical requirement since very few libraries could hire new, specialised staff; the need of re- or up-skilling is also emphasised in many other publications (Auckland 2012; Cox *et al*. 2012; Gow & Molloy 2014 [based on a DigCurV survey[173]]; Henty 2008; Pryor & Donnelly 2009).

Norbert Lossau (2012) notes that the involvement of LIBER members in European infrastructure projects has been low and tended to focus on publications and digitisation of cultural heritage. He thinks that "*Open Access, open science (data), research data infrastructures and management are the catalysts to get research libraries back into the awareness of researchers beyond the humanities and social sciences. (…) The majority of public funding will go into data infrastructures. Research libraries will benefit only marginally from this funding if they stick to research publications and the digitization of cultural heritage"* (Lossau 2012).

6.3 Guides to good practice, capacity building and training

6.3.1 Guides to good practice

There is no lack of guidance material for good practice data management and sharing. A lot of material is available online, providing general guidance, simple rules (Goodman *et al*. 2014; Sandve *et al*. 2013; White *et al*. 2013) as well as advice on good practice concerning specific aspects of research data.[174] There are online courses for data management[175] and online tools which help in writing a data management plan, even including templates according to the structure and content expected by research funders[176].

Among the outstanding resources for archaeologists are the Guides to Good Practice offered by the Archaeology Data Service (ADS) & Digital Antiquity[177]; the ADS DataTrain materials for post-graduate

[173] DigCurV - Digital Curator Vocational Education project (EU, Leonardo da Vinci, 01/2011-06/2013), http://www.digcur-education.org

[174] The Digital Curation Centre (UK) offers a wide range of data management resources: http://www.dcc.ac.uk; see also DataONE (USA): Best practices, https://www.dataone.org/all-best-practices; ANDS - Australian National Data Service: Data Management Planning, http://ands.org.au/resource/data-management-planning.html

[175] For example, MANTRA - Research Data Management Training (University of Edinburgh, UK), http://datalib.edina.ac.uk/mantra/; Earth Science Information Partners (ESIP): Data Management Short Course for Scientists, http://commons.esipfed.org/datamanagementshortcourse;

[176] For example, Digital Curation Centre (UK): DMPonline, https://dmponline.dcc.ac.uk; University of California Curation Center (USA): DMPTool, https://dmptool.org

[177] Archaeology Data Service (ADS) & Digital Antiquity: Guides to Good Practice, online: http://guides.archaeologydataservice.ac.uk

teaching on research data management in archaeology[178]; for archiving of archaeological material and data there is the ARCHES guide to good practice (available in several European languages)[179]; and OpenContext offers brief guidelines for web publication in archaeology[180].

One ARIADNE partner rightly suggested that guides to good practice should include case studies that "*demonstrate how open data can actually emerge from and help various archaeological research activities"*. Because guidance material without such evidence may have little impact with regard to open sharing of re-usable primary research data.

This has also been emphasised by researchers of the archaeological data publication platform OpenContext: *"Over time, publication without full disclosure of underlying evidence will no longer be considered sufficient. Until we reach that point, however, perhaps the greatest barrier to digital dissemination of primary data is the current lack of reference implementations; that is, projects with research outcomes that demonstrate the usefulness of sharing primary content"* (Kansa & Whitcher-Kansa 2009).

6.3.2 Capacity development and training

The responsibility for scaling up data management planning and execution is clearly with the university departments and institutes active in archaeological research. Guides to good practice are available, but need to be implemented in actual practice, at the level of projects and reliable institutional support. This necessitates institutional capacity building, training offers and other measures aimed at ensuring data stewardship and accountability of both research institutions and individual researchers.

Ward *et al*. (2011) offer advice on how to address university-based researchers' needs for information, guidance and training on matters of data management.[181] In the first place this is about helping many researchers find and digest relevant basic information. Brief, practical information must be offered in online formats that make it easy to access the information quickly; extensive policy documents clearly do not meet the demand of most researchers.

Research support staff should be available to contact and get more specific information, for example, on how to meet requirements (e.g. an open access mandate) in their project. It is of course more reassuring to present a case and receive authoritative answers for the particular circumstances, possibly backed up with experiences in similar cases.

Training should be offered in flexible modes, e.g. a mixture of online modules (accessible when researchers need them) and face-to-face sessions with real-life, discipline-specific examples. Hands-on training in data management may be best offered to PhD students and early-career researchers. Furthermore thematic seminars with experts in how to solve issues in intellectual property rights, sensitive data and ethical questions should be offered.

These are valuable suggestions for meeting researchers' needs in digital curation guidance and training. A focus on scaling up data management know-how of researchers has also been emphasised by ARIADNE partners, for example:

[178] ADS: DataTrain - Open Access Post-Graduate Teaching Materials in Managing Research Data in Archaeology, http://archaeologydataservice.ac.uk/learning/DataTrain

[179] ARCHES - Archaeological Resources in Cultural Heritage (2014): The Standard and Guide to Best Practice in Archaeological Archiving in Europe, online, http://archaeologydataservice.ac.uk/arches/Wiki.jsp?page=Main

[180] Open Context (2011): Guidelines for Web-based Publication in Archaeology. E. Kansa & S. Whitcher-Kansa, January 2011, http://ux.opencontext.org/blog/wp-content/uploads/2011/05/Guidelines_Jan2011.rtf

[181] The advice is based on experiences of the INCREMENTAL project that helped building data management capacity within the Universities of Cambridge and Glasgow (the project was funded under the JISC Managing Research Data programme, 10/2009-03/2011), http://www.lib.cam.ac.uk/preservation/incremental/

"*Begin training on an earlier/lower level – creation of data(-bases), structured folders, use of metadata on a file level as part of university curriculum. Or, for those who got out of university some time ago – courses. Regular updates on what is going on, good practices"*;

"Promote and consider offering training in data management/data documentation also for researchers and PhD students (not only for institutional/project data managers), as part of basic PhD training. Fostering a culture of data sharing means we need to make researchers in general to gain knowledge about data documentation".

We think that in addition it is important to foster also capacity building at the institutional level, i.e. focused on data managers who take care for institutional data assets and support collaborative projects. While we argue for the implementation of community-level data archives, the data management practices and tools of archaeological research institutes will still require much attention. Because there is need to support on-going projects (which can extend over many years) as well as to turn legacy data into shareable resources. Moreover the creation of standard metadata and use of common vocabularies will not flourish if research institutes stick to "home-grown" proprietary solutions. Such solutions may work internally but pose difficulties when institutions are asked to open up and share their data in e-infrastructure initiatives.

In this regard we suggest internships of institutional data managers at leading data centres and train the trainer workshops at research institutes, so that the managers can serve as disseminators of good practices. To reach and involve more data managers such workshops could also be offered as a series of webinars with contributions by experts from data centres of different countries.

6.4 Data management from the start

Proper data management along the research project lifecycle can make the on-going research work more effective as well as subsequent data re-use easier. Therefore documentation of research processes and generated data is ideally conducted from the start and throughout the project lifecycle. "Active data", data still in the project loop, is usually not shared beyond the project collaborators, only after deposit and release of the data. Then the shared data can remain active through re-use in further research.

6.4.1 Management of active data and "sheer curation"

Start as early as possible

Proper data management can make on-going research work more effective and facilitate subsequent data re-use (own and by other researchers) by making it easier to organise, understand, compare and combine data. Advocates of systematic research data management therefore emphasise that this management should start and be supported professionally as early as possible, for example:

- *"the earlier in the research process data professionals and proper data management systems appear, the more likely it is that data emerge from research in appropriate form for sharing, reuse, and long-term preservation"* (Salo 2010);
- *"effective data management at the point of creation and use lays a firm foundation for subsequent data publication, sharing, reuse, curation and preservation activities"* (Hedges *et al*. 2012);
- or, as a statement from an ARIADNE partner, "r*esearch data management needs to be considered as a continuous effort throughout the full life cycle of data up to sustainable long-term archiving that goes far beyond mere project funding"*.

This means that data management expertise will need to permeate the research process at large, beginning at the initial planning stages of a research project, running through the execution of the project, and followed by long-term data management and access for public use. In this regard figures of the research data life cycle which show data curation starting after deposit of the data in a digital archive are not favourable for promoting proper data management from the start[182]. Data curation should already be done by the researchers in the course of a project and therefore not presented as a "one-off" activity of data deposit.

"Sheer curation"

Typically researchers are not much concerned about data curation after project completion in order to prevent potential inaccessibility and loss of generated data. Therefore opportunities to improve data management practices should be employed while they produce and work with their data.

Chris Rusbridge notes: *"Ironically, while your project is active, all the information anyone would ever need to re-use your data is all around you. Unfortunately, because 'everyone knows', some is never written down until near or after the end of the project, when the post-doc has left for that great job in the US, and the PhD student has left for a merchant bank (she's a rocket scientist, after all), when the PI [Principal Investigator] realises he hasn't a clue what some of those data files actually contain"* (Rusdbridge 2007). One solution to lack of rich data provenance information, suggested by Rusbridge and others, is workflow-based capture of such information.

The approach of integrating data curation methods and tools into researchers' normal workflows is often called "sheer curation", a term coined by Alistair Miles (2007) in the DCC SCARP / ImageStore project[183]. The approach aims at enabling digital curation activities that are quietly integrated into research practices. The word "sheer" is used to emphasise the intended lightweight and transparent character of such curation activities, conducted at the point of data creation and primary use. This includes building curation into workflows that are a regular part of research work, offering easy-to-use tools and templates for metadata creation, and automatic capture of metadata at least from routine research activities.

The sheer curation approach requires analysing researchers' data practices, "data at work" (Birnholtz & Bietz 2003), in specific domains so that solutions are developed which are fit-for-purpose and likely to be adopted. Such solutions should enhance and complement existing curation practices rather than introduce completely different ways of data management.

Suggested solutions ideally add immediate value to the creators and primary users of the data within local work environments. The enhanced local curation practices should prepare the ground for and ease long-term data archiving and access. The required further work is carried out by curation specialists of the research institution or a community archive.

Work on prototypic "sheer curation" solutions has been conducted with a focus on laboratories. Hedges *et al*. (2012) describe the development of a system for researchers in biophysics and structural biology. Castro *et al.* (2013) report work on an environment for fracture mechanics researchers. This environment provides a data management platform (UPBox), with an interface similar to Dropbox, but full control over the data by the research group, and a semantic wiki (DataNotes) to support data annotations. A general problem in such approaches is motivating

[182] On the other hand, the arguably most prominent curation lifecycle model of the UK Digital Curation Centre (http://www.dcc.ac.uk/resources/curation-lifecycle-model) may scare researchers as it presents almost all data-related activities of researchers as data curation and preservation work.

[183] SCARP - Disciplinary Approaches to Sharing, Curation, Reuse and Preservation (UK, JISC-funded project, 01/2007-07/2009, coordinated by the Digital Curation Centre), http://www.dcc.ac.uk/projects/scarp see also: http://en.wikipedia.org/wiki/Digital_curation

researchers to produce metadata for "active data" already in the research process, although semi-automatic capture may help with the task.

A priority in sheer curation is to ensure that data is stored safely and systematically (incl. versioning) for own and collaborative use among co-workers. Academic data services providers address this priority of research data management by offering scholars a safe online environment to store their data during the research process. One example is the Dutch Dataverse Network[184] that has been initiated in 2008 by the Utrecht University and since 2014 is managed by Data Archiving and Networked Services (DANS).

As reported by many surveys, researchers usually store generated data on personal computers, carrier media and/or storage services, with access restricted to project collaborators. This includes institutional servers but many researchers have adopted services such as Dropbox, arguably because of ease-of-use in personal and collaborative work. Storing data in safe and controlled environments clearly should become one "scholarly primitive" of data management during research work.

6.4.2 Workflow management systems

To make "sheer curation" viable and effective, plug-ins for software already used by researchers could be installed which facilitate documentation of the data generated, i.e. provenance metadata. Automatic or semi-automatic capture of information about activities and results would reduce the effort required for regular documentation of such information. But data-centred research work usually involves chains of activities (workflows), e.g. data capture, cleaning, processing and analysis, or data sharing, discussion, annotation, etc. which may not be covered by one software.

In the last decade many scientific workflow management systems have been developed, ranging from generic to specialised systems for certain fields of study. Most systems cater to the natural sciences, especially the bio-sciences (Curcin & Ghanem 2008; Deelman *et al*. 2009; Talia 2013).

We surveyed some of the less specialised systems, but could not find a system that is applied in archaeological research. Examples of open source systems with users in different domains for example are Kepler[185] and Pegasus[186]. There are of course also several strong commercial systems, for example BIOVIA[187] which supports workflows of different natural and life sciences, both in distributed and laboratory environments. In high-end laboratories information about experiments and analyses is typically recorded automatically at the point of capture, i.e. by the laboratory instruments (this is sometimes called "curation at source").

A clear candidate for applying workflow management systems and automatic capture of information about analyses are archaeometry laboratories. Workflow management systems for laboratories support workflow automation, control of instruments, and capture, storage and management of information obtained in the course of the laboratory work (e.g. analyses of biological material). Thereby reporting of results and documentation / description of data according to domain standards are eased. However, a search of the literature and of laboratory websites suggests that laboratory workflow management systems are not in wide use in archaeometry laboratories. According to one ARIADNE partner that operates a laboratory use of such systems is not common.

The partner also mentioned that metadata for archaeometrical analyses are not standardised or such standards not in wide use, arguably because there are so many types and variants of analyses, which are applied according to particular research questions. However, recommended description and

[184] Dutch Dataverse Network, https://dataverse.nl

[185] Kepler, https://kepler-project.org/users/projects-using-kepler

[186] Pegasus, https://pegasus.isi.edu/applications

[187] BIOVIA, http://accelrys.com

terminology for research objects are generally followed in the research papers and reports, for examples, the *Guidelines of the Prehistoric Ceramic Research Group* (Hamilton *et al*. 1992; Prehistoric Ceramics Research Group 2011) and *International Code for Phytolith Nomenclature 1.0* (Madella *et al*. 2005)

All archaeological research activities in the field and laboratory are of course characterised by some successive steps (workflows). Such activities are usually conceived according to a specific methodology, guided by a protocol, and supported by some templates and/or tools, etc. One line of research work conducted in ARIADNE specifically studies processes that archaeologists carry out when they "do archaeology", the things that they create or use while carrying the work, the temporal organisation of the work, different functions/roles involved, etc.

Methodological practices of archaeological research partners have been analysed based on their documentation of work. References to methods followed, distinct units and stages of work, producers and things involved, etc. in the documents are extracted and modelled with CHARM[188]. CHARM is based on a situational method engineering approach and the ISO/IEC 24744 standard[189], applied to model research practices in the humanities and archaeology (Hug *et al*. 2011; Gonzalez-Perez & Hug 2013). The formalized and shareable descriptions of methodological practices can be used for different purposes, potential enhancement of processes with novel tools, for instance. The developers of CHARM have also proposed it as model for dataset integration (Gonzalez-Perez & Martín-Rodilla 2014).

6.4.3 Data curation profiles and description standards

Data curation primitives and profiles

As noted in a section above, storing data safely during research work (incl. telling file names, versioning, etc.) may be called one "scholarly primitive" of data management. The term "scholarly primitives" was coined by Unsworth (2000) and meant to describe basic activities common to scholarly work across disciplines (e.g. searching, collecting, reading, comparing, referring, etc.). Introduced in the context of humanities scholars' use of ICT, the term found wide adoption especially in this field.

Taxonomies of digital "scholarly primitives" are available, but "curation" is not a core concept in them. Palmer *et al*. (2009: 32-33) subsume "data practices" under "cross-cutting" primitives, together with monitoring, notetaking and translating; "data practices" are only addressed with regard to the sciences. Maybe rightly so, because, humanities scholars do not curate primary data/content they use, which are held by libraries, archives and museums; they may however curate and provide access to an annotated compilation of data/content from these sources (e.g. a scholarly edition of a corpus of texts).

In the elaborated TaDiRAH[190] and NEMO[191] schemes of digital, ICT-based research activities of humanities scholars curation is not a high-level concept either[192]. For instance, the TaDiRAH

[188] CHARM - Cultural Heritage Abstract Reference Model (Cesar Gonzalez-Perez *et al*., Institute of Heritage Sciences [Incipit], Spain), http://www.charminfo.org

[189] ISO/IEC 24744:2007,2014: Software Engineering - Metamodel for Development Methodologies, http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=62644

[190] TaDiRAH - Taxonomy of Digital Research Activities in the Humanities (v. 0.5.1, 05/2014), available on Github, https://github.com/dhtaxonomy/TaDiRAH; developed by researchers of DARIAH-DE and others (Dombrowski & Perkins 2014).

[191] NEMO - NeDiMAH ICT methods ontology, http://nemo.dcu.gr; developed in the NeDIMAH - Network for Digital Methods in the Arts and Humanities project (European Science Foundation, Research Networking Programme, 05/2011-05/2015), http://www.nedimah.eu

taxonomy groups the activities “archiving”, “identifying”, “organizing” and “preservation” under concept 6_Storage, referring *“to the activity of making digital copies of objects of inquiry, results of research, or software and services and of keeping them accessible, without necessarily making them available to the public”*.

Researchers, both scientists and humanities scholars (archaeologists) of course care for their data during research work as well as may better organise and describe it when preparing publications. Some projects have looked into such practices with the aim to conceive approaches and tools which may enhance researchers’ curation work and, subsequently, data and metadata that is being shared through digital repositories. Notably, researchers of ARIADNE partner Digital Curation Unit (a unit of the IMIS-Athena Research Centre, Greece) have proposed a digital curation lifecycle model (Constantopoulos *et al*. 2009) and worked with humanities researchers to better understand their digital curation requirements (e.g. Bernardou *et al*. 2010). Dallas (2015) considers the curation of archaeological knowledge “in the digital continuum” as a grand challenge of digital archaeology.

Extensive work on curation practices of researchers in different domains and specialties has been carried out by the Data Curation Profiles project[193]. The project invited researchers to document how they work with data, different forms and stages of data, tools used, and what they consider as sharable forms of data in their field of study. Also the project investigated if the researchers are willing to share data, when, and with whom, and their requirements for sharing their data through an institutional repository (Cragin *et al*. 2010; Witt *et al*. 2009).

The data curation profiles were created based on the analysis of interviews and worksheets (filled in by the researchers); each profile also presents a specific dataset. Profiles have been created for cases in different fields such as geophysics, plant genetics, demographics and linguistics. A relevant profile in the context of ARIADNE is the epigraphy study (Eaker 2012). The Data Curation Profiles toolkit has also been made available. It has been used for example by librarians of the Cornell University Library for collecting user requirements for their Datastar research data registry (Wright *et al*. 2013).

Developing data description standards

We think that a data curation primitives and profiles based approach could help improving researchers’ curation activities, particularly in fields of study which lack curation support tools in the research process and/or common description standards for the data generated. Instead of trying to define and impose such tools and standards “top-down”, “bottom-up”, but aimed at increasing codification and adoption, may prove to be a better approach.

As noted above, data curation profiles are essentially about what researchers of a domain or specialty commonly see as shareable, fit for re-use data in terms of data formats, description and other requirements such as licensing. The requirements could be specified by researchers in collaboration with curators of data repositories, and exemplified with datasets and dataset records that correspond to the defined requirements. The requirements (guidelines) could be published as a “data paper” (see *Section 6.6*) referring to the datasets which are deposited in an open access repository. Some fields of research may lack common standards for data description, for example, concerning methods and procedures used in the data generation. One case may be the field of archaeometrical analyses (cf. *Section 3.4.7*). Such fields would need what in the bio-sciences is called minimum information standards about investigations, which have been developed in recent years in

[192] For instance, the TaDiRAH taxonomy (see above) groups the activities “archiving”, “identifying”, “organizing” and “preservation” under concept 6_Storage, referring “to the activity of making digital copies of objects of inquiry, results of research, or software and services and of keeping them accessible, without necessarily making them available to the public”.

[193] Data Curation Profiles project (2008-2012, Purdue University / Distributed Data Curation Center and University of Illinois at Urbana-Champaign / Graduate School of Library and Information Science), website, http://docs.lib.purdue.edu/datacurationprofiles/

the MIBBI initiative (Taylor *et al*. 2008)[194]. A specification of the minimum set of information about the methods used in the generation of particular research data(sets) could be one component of a codified data curation profile / guideline.

Some descriptive information about the data, methods, procedures, etc. is typically provided in a more or less standardised form in research papers. One suggestion here is to extract and use this information as metadata in the dataset record. Chao (2015) reports work on this approach. Ideally a tool would allow capturing this information when researchers prepare a publication for which they deposit the underlying data and make it available as dataset metadata. A simple solution might be that repositories ask researchers to insert this information in a field of the dataset record they submit.

6.5 Metadata for re-usable data

In this report we address metadata mainly from the perspective of open data sharing and re-use. Metadata as most commonly used for archaeological and other cultural heritage data/content have been described in other ARIADNE deliverables, including which metadata standards and terminology (e.g. thesauri) project partners employ for their collections and datasets[195]. Below we address general issues of metadata, repository solutions, and the question if and how metadata can support cross-domain re-use of data.

6.5.1 Criteria for fit for re-use data

Proper data management already during project work is required so that shareable, fit for re-use data emerges from the research process. The criteria for fit for re-use data are generally covered by the concept of open data (cf. *Section 4.2*). The basis of open data is the understanding that anyone can freely access, use, modify, and share the data for any purpose. In this report we emphasise the ability to re-use/purpose research data because this enables further research and progress in knowledge.

Moreover re-usable metadata and vocabulary (terminology, ontologies) is essential for interoperable information systems and services, i.e. services which can exchange and interpret information for uses such as data integration and analysis.

Openness of research data requires data licensing with no or only minimal restrictions on re-use, e.g. attribution of the producers to ensure that proper credit is given to data sharers. It also requires preventing technical restrictions of re-use (e.g. non-open, proprietary formats) and, in general, that the data is machine-readable. Furthermore, that software for using the data is commonly available (ideally open source) or that the specific research software is shared together with the data so that others can validate research results and repurpose the data.

Moreover, and in particular, metadata is needed so that potential re-users can discover, understand the provenance, and evaluate the trustworthiness and quality of the data. Research data has special requirements in this regard, because for re-use other researchers need detailed data description and contextual information in order to evaluate if the data is relevant for their purpose and use it properly to prevent incorrect conclusions.

[194] MIBBI - Minimum Information for Biological and Biomedical Investigations, some 40 Minimum Information reporting guidelines, which are included in the BioSharing portal, https://www.biosharing.org/standards/

[195] See the project deliverables ARIADNE D3.1 Initial Report on Standards and on the Project Registry, November 2013, and D3.2 Report on Project Standards, both completed in November 2013, available from the project website, http://www.ariadne-infrastructure.eu/Resources

Fit for re-use thus means that the data can be re-used in principle and fulfils certain criteria of quality as required for the intended particular purpose. If it is "fit for purpose" depends on many aspects, and data relevant for one purpose may not be suitable for another, due to lack of spatio-temporal coverage, measurement characteristics or quality control procedures. Therefore good data description should also make users aware of aspects that limit re-use for certain purposes (i.e. data quality statements or declaration).

6.5.2 Metadata issues and solutions

A conflict of interest

Research data(sets) that are shared through repositories must be provided together with metadata, which are of vital importance for data discovery and access. But discovery metadata may not be sufficient for data re-use. Kunze *et al*. (2011) consider that without proper description the data as such may be unusable, "*'unusable by itself' may be a fundamental characteristic of what it means to be data. In any case it certainly suggests a strong documentation requirement if re-use is a goal"* (Kunze *et al*. 2011)[196].

This raises the question of what kind of data description is required and how detailed it should be to enable proper re-use. In general, description is required so that potential re-users can understand the data provenance/context, evaluate if the data is relevant for intended purposes (e.g. in terms of quality), and use it properly to prevent incorrect conclusions. Re-users may also need specific information about the research design, methods used to generate the data, etc. In some cases re-users may also need access to a specific software, which should be made available together with the data.

The effort to produce rich data description is high and in all surveys on data sharing researchers rank it among the top barriers for open data sharing, including the ARIADNE online survey (cf. *Section 4.3.3*). Metadata in general, and metadata for research data in particular, *"represents a classic mismatch of incentives: while of clear value to the larger community, metadata offers little to nothing to those tasked with producing it and may prove costly and time intensive to boot"* (Edwards *et al*. 2007: 32). The mismatch creates a conflict of interest between repositories, who would like to have "good" metadata, and data depositors, who may not be willing to invest much effort on the task.

As Kathleen Shearer, Executive Director of the Confederation of Open Access Repositories (COAR) notes: *"There is an inherent tension that exists in the repository community. On the one hand, we aim to make the deposit process as easy as possible so that creators will contribute (or repository staff costs are manageable); on the other hand, we want to assign good quality metadata (which takes time and effort) because we know it will enable greater interoperability and improve discoverability of content. So far, the former has been a greater priority"* (Shearer in: Poynder 2014).

There is a trade-off between ease of data deposit and the quality of a repository's metadata. High-quality metadata will allow a repository better data management and services than shallow metadata, and users to find and evaluate relevant data more easily. But requiring high-quality metadata from depositors is likely to reduce the number of contributions, if there are alternatives. Good support from repository staff may help countering this effect but increases costs. Therefore state-of-the-art repositories, especially domain-based archives, work hard to make the provision of sufficient metadata as easy as possible, lowering the barrier to deposit data with good enough metadata, while keeping their costs manageable (e.g. reduce submission guided by curators).

[196] The authors suggest "data papers" to receive detailed data description and link it to data records; the "data paper" approach is described in *Section 6.6*.

Metadata solutions

All studies on data sharing through digital repositories found that researchers consider the related effort, especially to provide the required metadata, as a barrier to depositing data. While data repositories and users would benefit from high-quality metadata, data sharers face the burden and prefer not to invest much effort on providing metadata. How does this work out in practice?

Repository scope, context, and forms of data

While several factors will determine the metadata solutions of data repositories, three dimensions allow distinguishing current solutions rather easily. The dimensions are repository scope (all subjects – domain-based), contexts of data deposit (informal – formal), and forms of data (single item – project data). Each dimension presents a spectrum of possible variants but clear endpoints.

Concerning repository scope there are two main variants, domain-based repositories (one, some or all domains of a discipline) and repositories which invite content from all disciplines/subjects. Into the latter category fall the repositories of universities and other general-purpose repositories.

The spectrum of deposit context has at the informal end the researcher who simply wants to use a freely available repository for storing some content data, instead of using a solution like Dropbox. A repository such as Figshare, which has a "research flavour" to it, might be just the right solution in this case. At the formal end of the spectrum there is the domain archive, specialised to curate data(sets) of a research community, and mandated or highly-recommended by research funders and journals. In between, there are repositories which are commonly used by researchers for depositing certain data, but without any formal obligation to do so.

Concerning the forms of data again there is a spectrum. At the low end there is the single piece of data, e.g. spreadsheet, figure, image, video, etc. This category also includes documents and presentations that are shared on platforms such as Academia.edu or SlideShare. At the high end, there is the package of data of a research project, which contains all data considered as necessary to document and support the project results, and for others to (re-)use, if relevant for their purposes. In between, various forms can be distinguished, e.g. some figures put together as supplementary material for a paper, or a deposited set of images documenting an experiment.

Impact on repository metadata

The factors addressed above determine to a high degree the quality of repository metadata. The "catch all" repository (all subjects, all forms of content/data, no formal obligation) has poor metadata, while the domain-based, specialised and mandated repository can excel in metadata quality. Below we illustrate these variants with some examples.

"Catch all" repositories

One example of a "catch all" repository is Figshare, launched in 2011 by the Digital Science brand of Macmillan Publishers. The repository is presented as a place *"where users can make all of their research outputs available in a citable, shareable and discoverable manner"* (Figshare website[197]). Figshare does not charge depositors (up to certain limits), accepts all kinds of data, assigns a DataCite DOI to each object, and invites depositors to apply a Creative Commons license. Also a tool is offered that allows upload of content and metadata directly from the researcher's computer. Well over 1.5 million items have been deposited. In terms of metadata Figshare asks for little: title, author/s, domain/subject categories (from a list of 14), adding some own tags, a description of the material, and users can add links to external sources.

[197] Fishare: About, http://figshare.com/about

Another example is the Zenodo repository[198], manged by CERN in the context of OpenAIRE. Zenodo invites *"all research outputs from across all fields of science"*, with a focus on so called "long-tail" (small-scale) data. Actually the majority of items are publications (over 17,000) while datasets are much less present (about 1600). Uploaded items get a DOI and the metadata is stored in the MARC bibliographic format. Zenodo is a low barrier repository and, consequently, the metadata is relatively simple[199]. Recently Zenodo announced that it will employ automated metadata extraction to ease depositors' provision of basic information such as title, description, author/s and publication date[200]. For a more detailed description and comparison of Figshare, Zenodo see Amorim *et al*. (2015).

Domain-based repositories

Domain-based repositories are specialized to curate data(sets) in a field of research, valued by the respective research community, and often mandated or highly-recommended by research funders and journals. For example, the *Journal of Open Archaeology Data*[201], an e-journal for peer-reviewed data papers started in 2012, recommends Archaeology Data Service (UK), Data Archiving and Networked Services (NL), Mappa (Italy), Swedish National Data Service, Open Context (USA) and tDAR (USA). Figshare is mentioned as *"suitable for small to medium sized archaeological datasets that do not require specialised curation"*; Zenodo is suggested for such data of projects funded by the European Commission[202].

Domain-based repositories can set high requirements for metadata, which the depositors will accept and follow, guided by archive curators, if necessary. One example is the E-Depot for Dutch Archaeology (EDNA)[203], which is provided by Data Archiving and Networked Services (DANS, Netherlands), as part of the DANS-EASY archiving system. Since 2007 Dutch archaeologists are formally obligated by the national Quality Norm Archaeology to deposit their data in DANS-EASY.

Looking into the metadata we find qualified Dublin Core and other elements, providing information about the research project, description of materials and artefacts, information about temporal and spatial coverage, etc. Codes and terminology of the Archaeological Basic Register (ABR), the national codebook/thesaurus for archaeology are employed (Gilissen 2014).

Use of qualified Dublin Core with domain-specific vocabularies seems to allow a good balance of (relative) simplicity, usefulness, and interoperability. Also repositories of other research communities follow this approach. Examples are the DRYAD repository for bio-sciences data, with extensive vocabulary support (Farnel & Shiri 2014)[204]; the GESIS Datorium repository for social and economic sciences data, which uses Dublin Core with some extensions and controlled vocabularies for subject area, data collection method, etc. (Wira-Alam 2014)[205]. For the description of social sciences datasets both on the study level and the variable level Data Documentation Initiative (DDI) can be use[206].

A survey by Qin & Li (2013) confirms the wide use of Dublin Core metadata elements with additional elements (i.e. a DC-based application profile) by dedicated research data repositories. Andrade &

[198] Zenodo, http://zenodo.org

[199] Simply searching Zenodo for "archaeology" in the datasets segment brings up 211 items, most of which 3D scans of vessels deposited by one American researcher; as an example of the limited metadata available in this case see https://zenodo.org/record/28310?ln=en#.ViuPmysjE2A

[200] Zenodo: Features, https://zenodo.org/features

[201] Journal of Open Archaeology Data (Ubiquity Press), http://openarchaeologydata.metajnl.com

[202] But not used for such datasets. We found only publications that may have resulted from EU-funded projects related to archaeology.

[203] DANS: E-Depot Dutch Archaeology (EDNA), http://www.edna.nl

[204] DRYAD: Metadata Application Profile, http://wiki.datadryad.org/Metadata_Profile

[205] GESIS (Leibniz Institute for the Social Sciences): Datorium, https://datorium.gesis.org

[206] Data Documentation Initiative (DDI), http://www.ddialliance.org

Baptista (2015) surveyed managers of repositories registered in OpenDOAR, which are mainly document repositories. Also in their sample of over 400 repositories Dublin Core is dominant, followed by MARC21 (bibliographic metadata) and METS (Metadata Encoding and Transmission Standard), but few repositories use domain-specific extension. A major field of development of both document and data repositories is connecting/interlinking publications and data which often are stored in different repositories (Hoogerwerf *et al*. 2013; Manghi *et al*. 2014; Ritze & Boland 2013).

6.5.3 Communities of practice, metadata and re-use

Archaeology is a multi-domain field of research and therefore care must be taken to enable the re-use of data in a wide array of research activities. Metadata is one important requirement for data discovery and re-use beyond institutional and disciplinary divides, promoting use for different and often not anticipated uses. It is hardly possible to anticipate all possible uses, but thinking about the potential reach of data resources can support their active "marketing".

Studies into requirements for data re-use point out that sufficient provenance and research context information is needed so that potential re-users can understand, evaluate, trust and re-use the data properly. This has also been confirmed for archaeologists by the DIPIR project (cf. *Section 4.4.3*). The main requirement for considering re-use of data generated by others is sufficient contextual information, while other criteria (e.g. background of the data creators, data archive) are much less important. To re-use data the archaeologists need detailed research context information, especially about the research design and procedures of data collection and analysis (e.g. recovery, sampling, identification methods, etc.).

Detailed metadata is not necessary when researchers in a field of study share data directly with colleagues, which is the most common data sharing practice (Fecher *et al*. 2015b). In specialties where such sharing dominates, there is little incentive to produce metadata or, in the first place, codify what metadata should be produced. However, if research data is meant to be openly shared through a data repository, detailed metadata is required. The metadata is meant to enable others, who have not been in involved in the data generation, re-use the data for some other purpose.

Huggett (2014) notes a critical issue in open research data: *"Increasing access to increasing amounts of data has to be set against greater distance from that data and a growing disconnect between the data and knowledge about that data".* Indeed, distance, or: not being familiar with the data of a particular domain of research can be a strong barrier to data re-use.

It is worth to consider issues of "distance", actually distances in several respects, with regard to requirements of data sharing and re-use, such as the level of metadata detail, for instance. Below we address this theme proceeding from the individual research group to the community of practice in a (sub-)domain of research, and further to the cross-domain level. We do not consider distances between research data producers and the interested public, which will be most unfamiliar with scientific data. However, dedicated "citizen scientists" can of course acquire the necessary literacy regarding scientific topics, theories and data, including meaningful re-use of data.

Re-use within the domain community of practice

Closest to a data resource are of course the data producers, the research group that has created the data and use it regularly during a project. They have first-hand knowledge about the research data, know how it has been produced, organised, processed, analysed and interpreted. Therefore re-use of the data is most easy for this research group. But even in this group the ease of re-use may decline over time, e.g. when researchers move elsewhere or retire, others cannot make sense of the data documentation, if available, etc.

In any research domain or specialty there is community of practice with a good understanding of how the research data is typically produced, analysed and interpreted. This allows members a good chance to re-use of data created by other researchers in their domain. They are familiar with the data types, instruments and software, measurement protocols, modelling procedures, analysis and presentation of results. Members of the community of practice may not need metadata and see no need to produce it in their everyday work. When data is shared with colleagues, which is the most common case of data sharing, information and explanation is given ad hoc, in a communicative exchange rather than through detailed metadata.

Ethnographic studies of data sharing emphasise that formal metadata can help data discovery but are often not sufficient for re-using data produced by others. Empirical research is always situated, purpose-laden and highly context-dependent. Therefore not codified, informal and tacit knowledge plays a critical role in making re-use possible. Such knowledge is acquired through training, apprenticeship, own data collection experiences and discussion with colleagues, e.g. about problems that arise in the data acquisition and interpretation. It allows members of the community of practice to understand how the data was produced, consider what is required to re-use the data properly, and develop sound conclusions. In case of doubt, when they need to ask the data producers (if still available) the communication will be eased by the common background and expertise in their domain of research.

Most literature which raises the issue of required informal/tacit knowledge when re-using research data refers to the ethnographic work of Zimmerman (2003, 2007, 2008) on practices in ecological research. Data of ecological research is typically characterised as a "long tail" data, being small in volume (as individual dataset), dispersed, heterogeneous, highly context-sensitive (both in terms of conditions in the field/ecology and techniques used), and therefore difficult to re-use – *"not by metadata alone"* (Zimmerman 2007). A lot of archaeological data shares characteristics with ecological data, and some archaeological and ecological research have common aspects such as the focus on "sites"[207], environmental conditions, etc. (for example, archaeobotany or environmental archaeology may compare well with ecological research). It seems this is a case well worth exploring with regard to similarities in standards and data infrastructure development.

The notion of *"not by metadata alone"* has been taken up by Edwards *et al*. (2011) who emphasise "metadata-as-process", alongside metadata records as structured products. In the process view metadata – information about data – is given through other channels (face to face, e-mail, skype, etc.) to overcome frictions around or in metadata. The authors illustrate such frictions around the Ecological Metadata Language (EML) standard which has been adopted by the US Long Term Ecological Research (LTER) Network in 2002, with a rather slow uptake over many years. Millerand *et al*. (2012) explain that the implementation of EML as the common LTER sites standard was only successful by making the trouble with EML experienced by the site data mangers an issue of the whole community/network. The issue was reframed as "success to come" and addressed with a toolset, good practice examples and other measures (on the development of the LTER data infrastructure and practices see also Karasti & Baker 2008; Baker & Millerand 2010).

It seems appropriate to add that the ecological research and monitoring community has made big progress through the Long Term Ecological Research (LTER) Networks, started in the 1980s in the United States. The international community, ILTER[208] involves 780 sites worldwide, of which LTER Europe[209] brings together over 400 sites. LTER Europe also manages the international DEIMS - Ecological Research Site and Dataset Registry[210], and recently the 4-year eLTER H2020 research infra-

[207] A Long Term Ecological Research (LTER) site typically covers an area of 1 to 10 km²; but ecological conditions and processes are investigated not only locally but also in larger spatial as well as over long temporal scales.

[208] ILTER - International Long Term Ecological Research Network (founded in 1993), http://www.ilternet.edu

[209] LTER Europe (initiated in 2003), http://www.lter-europe.net

structure project has started[211]. Notably, archaeological sites with good conditions of organic preservation are also seen a "distributed long-term observing network of the past" (IHOPE 2015).

Cross-domain re-use of research data

Given the requirements for proper re-use of research data within a domain community of data producers, considerable difficulties in cross-domain re-use can be assumed. In any case, easy re-use cannot be expected. Adding to the difficulty will that re-use of data from another domain typically means re-purposing of the data, i.e. data collected for one purpose is applied to address another research question. Such re-use is of course common already within a domain of research. But in cross-domain re-use of data the re-purposing will be done to tackle other research questions than are common in the originating domain of research.

This raises the question of what is necessary to allow such re-purposing (or "secondary use") to be both scientifically sound and technically feasible (e.g. extraction and integration of data into other datasets). Cross-domain re-use/purposing of data will certainly require detailed data description and other provenance metadata, ideally also data quality declarations which spell out limitations of data quality, so that users can make informed decisions about fitness for re-use/purposing. Enabling cross-domain re-use by way of metadata therefore implies a high effort on the side of the data sharers. At the same time, cross-domain re-use of the data may not be expected or in the horizon of the data producers. Rather they will be focused on matters of their research group and domain community of practice. However, quite some may be interested in cross-domain collaboration and inter-disciplinary projects, based on contacts with researchers in other research domains and/or awareness of complementary data.

Interesting thoughts and suggestions for enabling cross-domain use of data have been developed by one working group of the EarthCube geosciences e-infrastructure initiative[212]. The cross-domain interoperability working group in their roadmap suggests considering the *"conceptual knowledge distance"* between domains of research (EarthCube 2012b). This distance is determined by many factors such as different research intents, assumptions and methods, different data-related practices, and different ontologies and terminology. The authors present the lifecycle of a typical scientific dataset which progresses from original collection to eventual retirement, and posit that intra-domain based forms of data sharing may not allow re-use by researchers of conceptually distant domains. Rather, cross-domain readiness requires significant extra effort by the data producer, curator and/or value-added service provider do document and package data *"for unforeseen and opportunistic discovery and reuse"*.

Not surprisingly, given the complexity of the matter, the authors suggest clarifying the requirements for such discovery and re-use, support by enhanced documentation, archiving and discovery mechanisms, and tools that allow for integration of datasets with regard to semantic, spatio-temporal and other aspects. As a first step in this direction the authors consider the creation of templates and examples for specifically anticipated cross-domain use cases to characterize interoperability challenges, for instance, challenges concerning data formats and encodings, interpretation of concepts, and terms or (re-)using software of other domains.

[210] DEIMS - Ecological Research Site and Dataset Registry, http://data.lter-europe.net/deims/

[211] eLTER - European Long-Term Ecosystem and Socio-Ecological Research Infrastructure (EU, H2020, 06/2015-05/2019), http://www.lter-europe.net/projects/eLTER

[212] EarthCube, http://earthcube.org, is an e-infrastructures initiative of the US National Science Foundation since 2011 with an at least 10-year horizon. The initiative involves the geosciences research and user communities. EarthCube (2014) presents the initiative's development until end of 2014 and outlook for the future.

If such cross-domain data use cases are identified, they will most certainly face the challenges described in a survey report by Volk *et al*. (2014 [118 respondents]). The online survey focused on data sharing in collaborative natural environment & resources research and management but the identified issues will just as well surface in other domains (e.g. cultural heritage research and management).

The survey asked respondents about their "the top three" of over ten listed issues when giving and receiving data. The major issues when providing data were lack of clarity in data requests, unclear or issues concerning formats, unreasonable expectations and/or timeline; as less critical appeared lack of credit after data was provided. When receiving data, insufficiencies of the data documentation and data appeared as most difficult for data use and integration. This included no data collection description / no protocol; data aggregated or summarised without explanation; failure to define zeros, empty cells, NA, or placeholders; column headers are absent or undefined; missing data without explanation (incomplete dataset).

The authors of the survey report suggest strategies on how to tackle the issues which are relevant for any data sharing initiative, in particular initiatives involving distributed research teams. The suggestions are well worth reading as they range from the overall team process down to use of data dictionaries, defining null values, etc.

6.6 Data papers

An important recent development is "data papers" published by domain journals either as the main or as a special category of the journal content (Chavan & Penev 2011; Kunze *et al*. 2011; Candela *et al*. 2015). A data paper is a peer-reviewed publication that describes a data resource (dataset, database or other), the methods and standards used to create it, its structure and size, and where and how it can be accessed. In particular, the paper should also describe the re-use potential of the data for further research or other purposes. Generally not intended is that a data paper presents research methods, results and conclusions.

Data papers are sometimes characterised as a comprehensive metadata record, but they are published in journals (not as repository metadata records) and therefore part of the scientific literature. This is particularly important for the researchers, because a data paper is peer-reviewed, can be included in the list of publications, and receive citations just like traditional journal articles. Hence data papers offer an incentive for researchers to make their data available through repositories and provide rich description of the data intended to promote and enable re-use. This can increase the likelihood of data re-use and that the data and data paper are cited by other researchers. Instead of recognition only for a research paper, researchers can receive additional credit for shared and well-described data.

The data paper approach also provides a solution for issue that research papers are not meant to describe in detail datasets and repositories often lack such description. The latter is due to the effort required for sufficient data description, which researchers are not always willing to invest, and the intention of the repositories to not overburden researchers by asking for too much work. Domain-specific repositories usually ask for an appropriate effort while others are often limited to rather simple discovery metadata.

In the context of a research paper the deposited data is typically seen as an appendix to the research paper, linked from the record of the paper, to call it up directly from the repository (or the landing page of the dataset record). If researchers search a repository to find relevant data, the search options are limited to the typically simple discovery metadata. If they find possibly relevant datasets, there is often not sufficient information about the dataset. There may be a link to a related research paper, provided as a proxy for dataset metadata, but research papers do not serve this purpose particularly well. In short, the data paper provides a solution for appropriate dataset description, at least more than is available in many repositories or a typical research paper (Callaghan 2013).

Special attention should of course be devoted to stable cross-linking between the data paper, the described dataset, and any research papers associated with the dataset. Included in the data repository, the data paper might serve as the landing page of the dataset, and be automatically updated any time a new paper references the data.

The rapid uptake of the data paper concept by many publishers (Candela *et al*. 2015 [survey]) strengthens the recognition of available data and their producers.[213] It also strengthens the cooperation of journals with existing and new data repositories. In Elsevier's open access *Data in Brief* journal[214] authors can describe data supporting a paper in currently 120 of their journals or any stand-alone data accessible in *"an appropriate public repository which is respected in your field"*. The intention with regard to the research paper related data appears to be doing away with the classical supplementary material. Instead of such material, "all data" associated with *Data in Brief* papers should be deposited in repositories (including Elsevier's DataVerse Network based repository), and the authors of course pay for the data papers, which they can put on their list of publication.

One point that is sometimes discussed is if instead of data journals it would not be better to have data papers in established disciplinary journals. For example, Lin & Strasser (2014) in their recommendations for publishers concerning data access state, *"Allow the publication of papers that describe high-value datasets as a regular stream within existing disciplinary journals rather than segregating such papers into specialized data journals"*. It seems that quite some data journals have been pushed by publishers as a distinct new product for some disciplines like the thriving bio-medical and life sciences. We think there is nothing wrong about this. Rather editors of existing journals of other disciplines may consider introducing the category data paper and actively promote such papers.

In archaeology, one e-journal for peer-reviewed data papers is the *Journal of Open Archaeology Data* (started 2012)[215]. JOAD aims to publish "papers describing archaeology datasets with high reuse potential". Such datasets should be deposited in professionally managed and accessible domain, general-purpose or institutional repositories[216]. JOAD highlights that "the data and the papers are citable, and reuse is tracked". Also the e-journal *Internet Archaeology* has initiated a series of peer-reviewed data papers[217], in collaboration "with several trusted repositories, including ADS, tDAR and Open Context". Internet Archaeology highlights, *"You have put a lot of effort into creating your data and a data paper allows you to get credit for it, to publicise and share it with the community"*.

[213] Similarly, "software papers" may raise the academic profile of software developers. One journal that aims to improve the recognition, re-usability and sustainability of scientific software is the Journal of Open Research Software, http://openresearchsoftware.metajnl.com; Chue Hong *et al*. (2013) describe the review criteria of the journal.

[214] Elsevier, http://www.journals.elsevier.com/data-in-brief/policies-and-guidelines/data-in-brief-faq/

[215] Journal of Open Archaeology Data (Ubiquity Press), http://openarchaeologydata.metajnl.com

[216] JOAD runs an own Dataverse based repository; among their other recommended repositories are Archaeology Data Service (UK), DANS (NL), Mappa (Italy), SND (Sweden), Open Context (USA), tDAR (USA). Furthermore Zenodo and Figshare are mentioned which accept data/content from all subject areas and may be appropriate for archaeological datasets that do not require specialised curation. Only one institutional repository is included, the Discovery repository of the University College London.

[217] Internet Archaeology, http://intarch.ac.uk/authors/data-papers.html

6.7 Data review

Peer reviewed journal papers are still the most trusted research output, and even more so in a digital environment of scholarship that is flooded with ever more information, including from a plethora of social media outlets (Nicholas *et al*. 2015 [international study, about 4000 academic researchers]). In recent years peer review of publications has also been innovated by models of open peer review, which include editor-mediated review, crowd-sourced review and post-publication review, among others (Ford 2013; Grossmann 2014; Tattersall 2014).

The requirement set by ever more scientific journals that the data underlying the reported research should be available has raised the question of how the quality of the data should be evaluated. One answer is of course data peer review. In the literature this is addressed by distinguishing different objects and forms of data review (Lawrence *et al*. 2011; Mayernik *et al*. 2014). This includes evaluation of the information given about the analysed data in a traditional scientific paper; consultation of a data description document (e.g. a detailed database description, dataset record or data paper), including simple checks of data accessibility and correspondence to information given in the research paper; and more extensive checks of the actual data against claims made in the paper. With the recent boom in data repositories, journals have specified their data evaluation criteria in more detail and defined criteria for new formats such as data papers[218].

Kratz & Strasser (2015a) report results of a survey conducted in 2014 on researchers' perspectives on data publication and peer review (about 250 qualified responses). Asked *"What would you expect data peer review to consider"*, 90% of respondents expected that methods are appropriate, 80% enough metadata for replication, 70% a technical details check out, 61% that the data is plausible, 39% that metadata is properly standardised, and only 22% that data are novel/impactful. The report authors highlight that most respondents *"sidestepped examination of the data itself"*.

Data review related to a research paper is fundamentally different from a review of a data paper or data source. As Kratz (2014) explains: *"The fundamental question in peer review of an article is 'did the authors actually demonstrate what they claim?' This involves evaluation of the data, but in the context of a particular question and conclusion. Without a question, there is no context, and no way to meaningfully evaluate the data".* The review of a research paper should in principle also evaluate if it contributes to the scientific record (i.e. not replicates what is established knowledge), as well as prevent redundant or "salami publication" (Šupak-Smolčić 2013).

While reviewers of a research paper will focus on the evidence provided for knowledge claims, they should also have an eye on the documentation (metadata) of the data, because shallow description will impede re-use, which is one major reason for open access to research data. The request of journals that data papers should point out the re-use potential of described datasets supports this agenda. Reviews of a data papers allow a general audit whether researchers are following established professional standards in describing their data, and data archives which require solid metadata have an important role in setting and maintaining such standard. Conversely, repositories that accept "data dumps" with little or no metadata may degrade professional conduct; usually journals recommend acknowledged domain-based repositories but some also have low-standard providers on their list.

Some authors hold that data repositories can only conduct a technical review of submitted data, which is not correct given the extensive data quality assessment and editing some repositories carry out. The value of the curatorial work of leading data repositories with researchers on the data, metadata, controlled vocabulary, etc. is often undervalued (Weber *et al*. 2013). Peer *et al*. (2014) insist that research repositories play a key role with regard to *"the ability to engage in informed*

[218] Mayernik *et al*. (2014) includes the data paper review guidelines of three journals, Earth System Science Data (Copernicus Publications), Geoscience Data Journal (Wiley), and Scientific Data (Nature.com).

reuse, which requires that data are independently understandable". They discuss the quality review of different types of repositories which ranges from no review (e.g. Figshare) to extensive data assessment and improvement (e.g. the ICPSR social sciences repository).

Outstanding with regard to curatorial review is the archaeological data publishing service Open Context. They indicate the editorial status of projects which ranges from "Peer-reviewed" to "Minimal editorial acceptance".[219] "Peer-reviewed" means at least two external reviews in addition to internal reviews and edits from the managing editor and the Editorial Board. "Minimal editorial acceptance" means the managing editors accepted a project's dataset despite some lack in important documentation and supplemental information. The latter is also applied with the addition of "Demonstration" to Open Context proof-of-concept applications or demonstrations which use such datasets. The editorial status of a project data publication is displayed on every record belonging to the project.

Research data archives could benefit greatly from promoting reviews by dataset (re-)users. Obviously researchers and other archive users who employed or tried to employ data for different purposes can give valuable feedback on fit-for-purpose of the data description and data. In the literature user reviews are often considered as the most reliable, quasi an acid test of fit-for-purpose.

A questionnaire-based survey of Dutch professors and senior lecturers in different disciplines included comments by re-users as one measure for improving the quality of collections of research data (361 responses). Over 70% of respondents said as a re-user they would give comments on the quality of the dataset, over 80% that as data producers they would welcome such comments, and over 80% considered these comments to be valuable for future re-users of the datasets (Waaijers & van der Graaf 2011).

Data Archiving and Networked Services (DANS) conducted pilot studies for an online data review system. DANS e-mailed service users who had downloaded data from their archive in order to get a rating of and comments on the datasets through an online questionnaire. Also information about the users' background, their purposes, the archive website, etc. has been collected (Grootveld & van Egmond 2012). From three e-mailings (2010, 2011 and 2012) to in total 3600 downloaders DANS received about 400 sufficiently filled questionnaires.

The main reason for downloading data (352 responses, open question) was research (62%), followed by a study or educational purpose (9%), general interest (9%), and miscellaneous (20%). Overall the respondents evaluated the downloaded data as "good", followed by "very good"; on average all criteria such as data quality, quality of documentation, completeness of the data, etc. scored at around 4 of maximal 5 points[220]. Evaluation of the structure and of the consistency of datasets was often "not applicable", 40% and 50%, respectively. With regard to data quality respondents in free text responses recognised mostly that the data *is* accessible online and, even more so, if it is complete and comprehensive. 70% of respondents said that the downloaded data was helpful in answering their research questions, 18% opted for "not applicable", and 12% answered in the negative; in a follow-up question e.g. "not what I expected" and "not relevant enough" was each selected by roughly a third of 50 respondents. About 15% of respondents said that they had "used the data set for a publication", and about 60% intended to use it for this purpose.

[219] Open Context: Data Publication Guidelines for Contributors, http://opencontext.org/about/publishing

[220] The DANS website shows the results for these dataset review criteria for over 80 datasets with at least two reviews: Reviews van datasets in EASY, http://datareviews.dans.knaw.nl

6.8 Suggested actions

To promote the open data agenda, universities, research institutes and other stakeholders should put in place policies, guidance and training. Institutional capacity building and support for researchers in the management of data is necessary so that open and re-usable data flows into data archives for long-term curation, access and (re-)use. The research community should also consider novel approaches to data description and review.

Ensure that adequate institutional policies, guidance and other support are in place

A growing number of funding bodies require research projects to implement a data management plan, with the objective that the generated data becomes openly available. Therefore researchers are looking for advice, guidance and support from their institution. Many institutions may not be prepared to meet this demand. Because the responsibility to manage and maintain data has traditionally been assigned to the researchers, based on the understanding that they own the data. We recommend that research institutions address issues of proper data management and sharing pro-actively and put adequate policies, guidance and other support in place.

Step up capacity building and training for data management and sharing

There is no lack of guidance material for good practice data management. For archaeological projects particularly the guides offered online by the Archaeology Data Service / Digital Antiquity and the ARCHES guide for data archiving (available in several European languages) merit highlighting. The challenge for publicly funded research projects and institutions now is to implement and support data management in view of open data sharing. This necessitates institutional capacity building, training offers and other measures aimed at ensuring data stewardship and accountability.

In practical terms, this means a range of assistance including, but not limited to, hands-on training in data management for PhD students and early-career researchers, help in drawing up a solid data management plan for discipline-specific data, expert advice concerning sensitive data and ethical issues, IPR/licensing (incl. rights clearance). Addressing “the data issue” therefore means more than pointing researchers to an appropriate data archive, institutional capacity building is necessary to provide the kind of assistance mentioned.

Below we give some specific suggestions on how research institutions and archives can foster proper data management and sharing:

- *Require that data management and subsequent open sharing of data are considered already in the project planning phase, e.g. data management and access plans.*
- *Promote the preparation of shareable data through dissemination of good practice guides and expert advice on specific matters (e.g. sensitive data, licensing of databases).*
- *Bring project data managers in contact with data archives/centres, not only in view of data deposit, but skills development (e.g. work on legacy data, use of specific standards).*
- *Emphasize the need for appropriate description of the methods used to collect, analyse and present the data, including technical and other requirements for data re-use (e.g. software).*
- *Suggest use of established open data formats, metadata standards as well as common terminology/vocabularies.*

Provide support for managing data during project work

Advocates of proper research data management suggest that it should start and be supported professionally as early as possible. This would make on-going research work more effective and data re-use easier. Suggested approaches include offering researchers lightweight curation tools and/or automated mechanisms integrated into researchers' normal workflows (so called "sheer curation"). Solutions ideally add immediate value to the creators and primary users of the data as well as prepare the ground for long-term data archiving and access.

This is still a field of research & development with some prototypic solutions. Also use of workflow management systems, for example, in archaeometry laboratories seems uncommon. Rather than looking for automation, we suggest that institutions or research networks offer researchers a safe and controlled environment for storing and sharing "active data" during project work. Moreover it would be beneficial if institutional data experts are available who support on-going projects, with regard to metadata generation and use of common vocabularies, for instance.

Recognise high-quality metadata is required for data re-use

Research datasets that are shared through repositories must be provided together with metadata, which are necessary for data discovery and access. But discovery metadata may not be sufficient for data re-use. Data description is required so that potential re-users can understand the data provenance/context, evaluate if the data is relevant for intended purposes, and use it properly to prevent incorrect conclusions. The effort to produce such data description is among the top barriers for open data sharing.

Metadata presents a conflict of interest between repositories, who would like to have good metadata, and data depositors, who may not be willing to invest much effort on the task. In practice this means that there are many repositories which have only shallow metadata, in particular general-purpose and university-based repositories. Both invite deposits of material from many disciplines hence cannot support specific domains like archaeology specifically. Domain-based, specialised and mandated archives can set high-quality metadata standards, which depositors will accept and follow, guided by archive curators, if necessary.

Promote data papers for archaeological datasets

A data paper is a peer-reviewed publication that describes a data resource (dataset, database or other), the methods and standards used to create it, its structure and size, and where and how it can be accessed. In particular, the paper should also describe the re-use potential of the data for further research or other purposes. Peer-reviewed and citable data papers offer an incentive for researchers to make their data available through repositories and provide rich description of the data intended to promote and enable re-use. One example of a dedicated, online and open access data journal is the *Journal of Open Archaeology Data*; the e-journal *Internet Archaeology* has initiated a series of data papers, and others may follow suit.

Explore novel approaches to data peer review

The open data policies of research funders and journals have raised the question of how the quality of the data should be evaluated. Peer reviewed data papers provide one solution, furthermore novel models recently developed for publications may be trialled. Such models are editor-mediated review, "crowd-sourced" review and post-publication review. For example, Data Archiving and Networked Services (DANS) collect ratings and other feedback from people who have downloaded data from their archive system, which includes the E-Depot for Dutch Archaeology.

7 Focus area 5 – Providing e-infrastructure services

7.1 Introduction and overview

Archaeological research in Europe (and elsewhere) lacks common and integrated e-infrastructure and services for data curation, discovery, access and (re-)use. The consequence of this lack is a high fragmentation of archaeological data and limited capability for collaborative research across institutional and national as well as disciplinary boundaries.

Common and integrated e-infrastructure basically means accessible digital archives with structured, interoperable and re-usable data resources (e.g. collections, datasets of projects) at the bottom and discovery, access and other services on top. ARIADNE develops e-infrastructure that allows for interoperability of existing and newly built archaeological data archives and, based on this interoperability, cross-archive search, access, and (re-)use of available data.

ARIADNE will not replace any of the underlying infrastructures (e.g. institutional repositories and community-level data archives), but provide integrating functionality and services on top of them. Thereby ARIADNE will help to make currently isolated archaeological data more accessible and useable for the research community as well as other groups such as heritage management agencies and citizens. The main intended user group of the ARIADNE e-infrastructure and services is the archaeological research community.

Identifying relevant services

Much care has been devoted to identifying the services end-users such as researchers and data managers expect from the ARIADNE data infrastructure and portal. This identification has been carried out through an online survey on user needs and a survey of existing data portals.

In brief, the online survey (with over 500 qualified responses) made clear that most researchers expect from ARIADNE to provide a data portal that allows an overview of existing archaeological data resources, and to search across the resources, using novel mechanisms for data discovery and access. Much less interest exists for typical features of Web 2.0 platforms such as content filtered based on tags or ratings of other users. However researchers appreciate effective mechanisms that save time in identifying relevant data (e.g. data preview mechanisms, clear licensing information).

Much further and more detailed insights for the development of the ARIADNE data portal have been acquired through a survey of existing portals and suggestions for the project’s portal given by a panel of 23 archaeologists and data managers (reported in the project deliverable D2.2, February 2015). The 34 suggestions of the survey report have been evaluated by 28 experts of 21 partners in order to focus on the most relevant portal services. Hence, the project has a solid basis for developing a data portal that will serve the immediate and evolving needs of the archaeological research community.

In general, the service portfolio of the ARIADNE infrastructure and portal should meet core requirements of data overview, search and access. This includes data search based on geo-location (maps) and date-ranges/chronologies which the evaluators appreciated most. Not considered as a priority are personalized services and support of expert networking and discussion on the data portal. There is little scope to invest limited funds on specific services that are not appreciated, are provided by other portals, or may run ahead of the needs of broad user segments.

High relevance has been attached to deploying Linked Open Data to integrate information within the portal and to link to external resources. Also providing interfaces to allow external applications exploit available (meta)data was considered as very important. Clearly the ADRIADNE data infrastructure and portal should not be an “island” but enable added value in the wider information ecosystem of archaeology and beyond.

A “front-runner” category are services that support online research work (e-research), which is not an immediate concern, but may emerge in the 10 year horizon of the innovation agenda. We assume that the needs and requirements of the archaeological research community will evolve towards e-research capability when more open data becomes accessible through state-of-the-art community archives.

ARIADNE e-infrastructure architecture and services

This chapter includes an overview of the ARIADNE e-infrastructure architecture and service implementation. The architecture presents a layered approach with various components. A core role plays the data registry component. The component has been implemented based on the Data Catalog Vocabulary (DCAT) standard/recommendation of the World Wide Web Consortium (W3C), adapted for describing archaeological data resources. This adaptation is the ARIADNE Catalogue Data Model - ACDM.

The ARIADNE registry allows archaeological data providers, large and small, to describe their resources (collections, datasets), following a common model (the ACDM), which will allow for interoperability of the resources. This means that ARIADNE services can build on consistent information about how the resources are structured, what they contain, how they can be accessed, etc. The information will be available as Linked Open Data also to other developers to enable data interlinking and creation of additional, external services.

Data search, visualization and access on the ARIADNE data portal will be possible based on thematic, spatial and temporal information contained in the metadata of the data resources. Further integration of data can be achieved if providers map their databases to the extended CIDOC Conceptual Reference Model, as intended by several ARIADNE partners. Access to the data resources on the portal will be enabled in support of different study purposes, e.g. access to only one object (e.g. a 3D model of a building), an aggregated virtual reference collection of objects to allow for comparison, one or more datasets allowing for extraction of numeric data, etc.

For example, visual media services are already available that allow easy web-based publication, visualization and exploration of high-resolution images, reflection transformation images (RTI), and 3D models, including 3D landscapes. Other, more experimental stage services will exploit Linked Data (based on the CIDOC Conceptual Reference Model) or use data mining techniques to suggest patterns of relevance between data resources.

A major step towards integrated archaeological data and e-research

The overall objective of the ARIADNE e-infrastructure and data portal is building an environment for services that act as brokers between archaeological data providers and users. The creation of such an environment is a substantial step forward in the archaeological domain, in particular it provides a common platform where dispersed data resources can be uniformly described, discovered and accessed.

The ARIADNE e-infrastructure will allow overcoming some of the idiosyncrasies of the underlying different infrastructures of the data providers which currently prevent the collaborative exploitation of available data. At the same time, it represents an essential step towards the more ambitious goal of providing integrated services and tools capable to support web-based research aimed at creating new knowledge (e-archaeology). This may be achieved by future Virtual Research Environments (VREs) built on top of the ARIADNE e-infrastructure and services.

As noted above, support of online research work is not an immediate concern of ARIADNE, but may become an important topic in the medium to long-term as more open data and useful tools for data integration and exploitation become available.

Fostering data integration and added value services beyond ARIADNE

ARIADNE should foster data integration and added value services beyond the ARIADNE data portal. This can allow the initiative to play a significant role in the data service environment in and beyond the different domains of archaeological, cultural heritage and other humanities research.

A core element of this scenario is Linked Open Data (LOD) which ARIADNE generates and employs for internal service provision, but can also be used for interlinking with external data resources as well as provided to developers for producing added value services. Such applications may combine data from different sources (ARIADNE-mediated data and others) and promote cross-domain use of data resources.

The ARIADNE e-infrastructure could also benefit from using services provided by other major e-infrastructures, including from domains such as earth, environmental, biological and other sciences, which are present on the multi-disciplinary map of archaeological research (cf. *Section 3.3.4*).

Deep interlinking of the data from such services would, however, require integration of conceptual knowledge of the different domains (e.g. terminologies, ontologies). Such integration could be sought based on use cases with a clear added value for archaeological research communities. While this is not a priority for ARIADNE in the 5-year horizon, we assume that in the 10-year horizon such integration might be attempted.

7.2 Different levels of e-infrastructures

The ARIADNE Users Framework[221] explains where the ARIADNE e-infrastructure and services sit in the complex environment of archaeological research data creation, curation and access. The framework distinguishes between direct stakeholders (e.g. researchers and research institutes) and indirect stakeholders (e.g. research funders). The direct stakeholders are the intended users of the research e-infrastructure and services, and the main intended user group is the archaeological research and data management community in Europe and beyond.

The ARIADNE Users Framework distinguishes 4 levels of data-related workflows and respective data managers/curators and service providers:

- Level 1: Research projects: archaeologists, subject experts, project data managers,
- Level 2: Research institutions: research directors, managers of institutional repositories and databases,
- Level 3: Subject/domain based data archives: data managers of community archives that provide deposit, long-term preservation and access services,
- Level 4: E-infrastructure and integrated services: managers of ARIADNE and other high-level e-infrastructures and services.

The levels of course represent also different perspectives, responsibilities and requirements of the experts who carry out tasks at the different levels. The general data workflow scheme across the levels is:

- The data is produced and managed first by research projects or in the context of other work carried out on archaeological sites and objects, e.g. heritage management (level 1),
- the data/datasets together with metadata is then deposited in institutional repositories (level 2) or a community data archive (level 3), and

[221] ARIADNE D2.1 First Report on Users' Needs, April 2014, pp. 51-58.

- the metadata from several repositories and archives is collected (e.g. harvested) into a common metadata pool, and search and other services are provided based on the metadata (level 4).

The ARIADNE e-infrastructure and integrated services span the levels 2 to 4, hence, do not aim to support the acquisition of primary archaeological data in field surveys or excavations. Level 4 represents the integrating functions of the ARIADNE e-infrastructure and services. These functions include dataset registration, metadata and controlled vocabulary services, search across the distributed repositories/archives, and other services for the archaeological research community.

ARIADNE will not replace any of the underlying infrastructures (e.g. institutional repositories or community-level data archives) but provide integrating functionality and services on top of them. Thereby ARIADNE will help to make currently isolated archaeological data more accessible and useable for the research community and other groups, e.g. heritage management agencies and citizens.

7.3 Required core e-infrastructure services

7.3.1 Current situation and expectations from ARIADNE

Archaeological research in Europe lacks common and integrated e-infrastructure and services for data curation, discovery, access and re-use. The consequence of this lack is a high fragmentation of archaeological data and limited capability for collaborative research across institutional and national as well as disciplinary boundaries. Common and integrated e-infrastructure basically means accessible digital archives with structured, interoperable and re-usable data (e.g. collections, datasets of projects) at the bottom and cross-archive discovery, access and other services on top.

While large archaeological research institutes may implement their own data repository, small institutes, units and individual research projects certainly need community-level data archives, which provide the required long-term data curation and availability. The ARIADNE e-infrastructure will enable interoperability of data resources of existing and newly built digital archives/repositories and, based on this interoperability, to cross-search, access and re-use the available data resources. Sharing and integration of data from many projects will also pave the way towards novel web-based research methods and practices (e-archaeology).

Lack of common e-infrastructure and expectations from ARIADNE

Below we present results of the ARIADNE online survey which make clear the consequences of the current lack of common data infrastructure and services, and summarise expressed needs and related expectations from ARIADNE. The full results of the survey are presented in the First Report on Users’ Needs (ARIADNE D2.1, April 2014).

Consequences of the current lack of e-infrastructure and services

The consequences are evidenced in the results of the ARIADNE online survey. Overall the results show that archaeological researchers lack services for finding and accessing relevant data. The selected results presented below are based on between 470 and 590 survey responses per result:

- *Need to find possibly available but scattered data:* 87% of respondents agreed fully or partially that they often do not know what is available, because research data are scattered across many places and different databases. Consequently 95% considered as very or rather important having a good online overview of available data(sets).
- *Need of online access to research data:* 94% of respondents considered as very or rather important that data(sets) are available online and in an uncomplicated way. Among the

perceived barriers are that access is often “limited to specific persons/communities” or “kept in private collections of other researchers”.

- *Need of access to international data:* 74% thought that it is very or rather important having easy access to international data(sets). This is a particularly encouraging result for ARIADNE as most archaeological researchers arguably work in a national or regional context. It signals high interest in data that allows for integrative research, comparative analysis and broad synthesis.
- *Need of appropriate data archives/repositories:* A lack of international archives where archaeological data sets could fit into was perceived by 66% as a very or rather important barrier to sharing data with colleagues. 60% of the researchers also said that their organisation (university, research institute or other) does not have an institutional repository that is managed by dedicated staff. As described in our survey of digital resources, most institutional repositories, general as well as subject-specific, manage only documents (cf. *Section 9.1.1*). Consequently the survey found that data was made available through an institutional repository only in a few projects or not at all by 67% of the researchers. The figures for national and international repositories are 76% and 83%, respectively.

Expectation from ARIADNE

In the ARIADNE online survey most of the researchers expected from the project to create a data portal that allows an overview of existing archaeological data resources and to search across the resources, using novel mechanisms for data discovery and access. Asked from which services they would benefit most (“very helpful”), the responding researchers considered:

- 79% a portal that makes it more convenient to search for existing archaeological data that is stored in different archives/repositories,
- 63% a portal enabling innovative and more powerful data discovery mechanisms,
- 58% a directory of European archaeological databases and repositories,
- 52% services for geo-integrated data,
- 29% data recommendations based on collaborative filtering, rating and similar mechanisms.

Thus capability to search and “mine” distributed digital archives for relevant data was appreciated most. Much less interest found typical features of Web 2.0 platforms such as content filtered based on tags or ratings of other users. Researchers appreciate effective mechanisms that save time in identifying relevant data, what they typically do not like is resources pre-culled by others.

Also an additional online survey for managers of data repositories was run and is reported in the First Report on Users’ Needs (ARIADNE D2.1, April 2014). Due to the small sample of 52 respondents the results are only indicative. The results show that the main concern of data managers is the quality of metadata. Also they would appreciate a higher awareness among researchers of good practice in data management (e.g. available guides and recommendations). Moreover the data managers expected much more than the researchers better data access through improvements in data/metadata extraction and indexing as well as Linking Data. But Web 2.0 features also ranked last.

Results of the lead user survey for the ARIADNE data portal

Going beyond the user expectations collected through the online survey, many further insights for the development of the ARIADNE data portal have been acquired through a survey of various portals by a panel of archaeological researchers and data managers. This included looking for good practices and giving suggestions for the ARIADNE data portal[222]. The 34 suggestions of the survey report have been evaluated by 28 experts of 21 partners in order to focus on the most relevant services. Hence,

[222] Reported in the ARIADNE D2.2 Second Report on Users’ Needs, February 2015.

the project has a solid basis for creating a data portal that will serve the immediate and evolving needs of the archaeological research community.

Details of the portals survey, evaluation of the results and final outcomes are presented in *Section 9.3*. On a general level the outcomes can be summarised as follows:

Clear expectations for the ARIADNE portal are:

- A highly functional as well as attractive portal
- Good overview of existing/available data resources
- Powerful search mechanisms, especially for cross-searching repositories
- The highest score received search based on geo-location (maps) and date-ranges/chronologies

In the middle to high level range:

- Terminology support
- Linking of data and publications to enable integrated access
- Data preview mechanisms and licensing information
- Various data export and download options
- Interfaces to allow external applications data exploitation
- Support of online research work (e-research), but not within ARIADNE

Not appreciated are:

- Personalized portal services (e.g. alerts on possibly relevant new data)
- Linking of online professional information (e.g. researchers' profiles)
- Support of expert networking and discussion on the portal

Surprises were:

- High relevance attached to deploying Linked Open Data to integrate information within the portal and to link to external resources
- Relatively high scores for providing interfaces to allow external applications exploit available data, metadata and conceptual knowledge

Concerning the latter, the ARIADNE data portal should indeed not be an "island", the wider *information ecosystem* and how to enable added valued beyond the portal will be considered (cf. *Section 7.4* below)

Some suggested portal features were not appreciated. These features concern personalized services, professional information (e.g. researchers' profiles), networking and discussion on the portal. Portals for the latter exist (e.g. Academia.edu, ResearchGate and others) and are also used by many archaeological researchers.

Clearly the service portfolio of the ARIADNE portal should meet core requirements of data overview, search and access. There is little scope to invest limited funds on specific services that are not appreciated, are provided by other portals, or may run ahead of the needs of broad user segments. In the "front-runner" category we see support for *online research work (e-research)*, which is not an immediate need, but may emerge in the 10-year horizon of the innovation agenda (cf. *Section 8.3*). We assume that the needs and requirements of the archaeological research community will evolve towards e-research capability when more open data becomes accessible through state-of-the-art community archives.

7.3.2 Overview of ARIADNE e-infrastructure architecture and services

In the development of a data infrastructure and portal it is of utmost importance to understand user needs and requirements. At the same time, ARIADNE must consider what actually is feasible due to various constraints such as existing shortcomings in the organisation of many data resources (e.g. lack of metadata or of common terminology). Such constraints may impede meeting user expectations until the data infrastructure can build on state-of-the-art resources of many data providers. Building a data portal therefore requires much support of data holders with regard to describing and providing their (meta)data. At the same time, this enables active involvement of the community in the development of the data infrastructure and services, aimed at usefulness and usability.

The ARIADNE e-infrastructure and services

ARIADNE aims at making currently isolated archaeological datasets and collections more accessible and useable for the research community and other user groups. This will be made possible by a 4-layer e-infrastructure architecture comprising various components (see figure below).

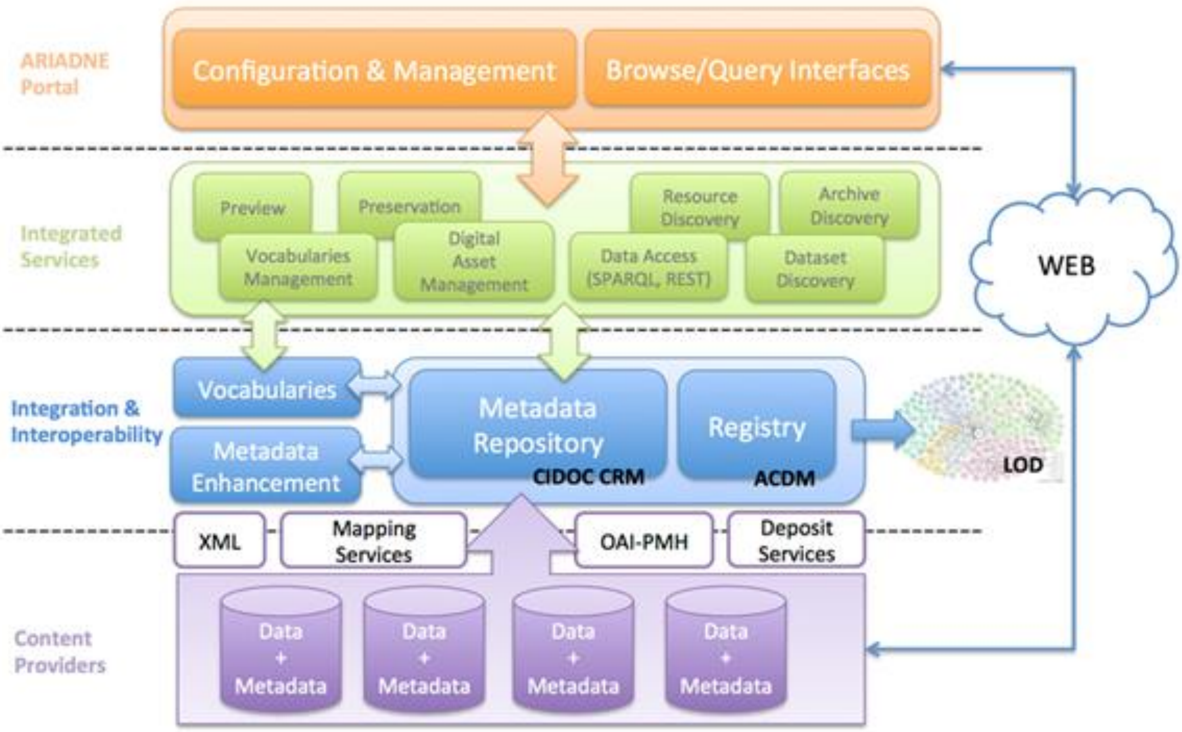


Figure 5: Architecture of the ARIADNE e-infrastructure and services.
ARIADNE, PIN, 11/2014

The components and services follow established standards of which we highlight the standards deployed in the Integration & Interoperability layer:

The Registry component uses the ARIADNE Catalogue Data Model (ACDM)[223]. The ACDM is based on the W3C Data Catalog Vocabulary (DCAT)[224] standard, adapted for the description of archaeological

[223] ARIADNE Registry, http://registry.ariadne-infrastructure.eu; ARIADNE Dataset Catalogue Model (ACDM) support website, http://support.ariadne-infrastructure.eu

[224] W3C (World Wide Web Consortium) - DCAT - Data Catalog Vocabulary, http://www.w3.org/TR/vocab-dcat/

data resources. The ACDM also allows description of the associated services, e.g. an available OAI-PMH target or SPARQL endpoint[225].

Content/data archives describe their holdings (collections, datasets, etc.) and associated services in the ARIADNE Registry, and the Metadata Repository aggregates the metadata of the archived material[226]. As several ARIADNE services will work based on Linked Data, the metadata is enhanced semantically, i.e. converted to Resource Description Framework (RDF) format[227] (if necessary) and linked with widely used vocabularies. Therefore data providers using a proprietary thesaurus (or term list) should map terms employed in their dataset descriptions to matching terms of the Art & Architecture Thesaurus[228], which the Getty Research Institute has made available as Linked Open Data.

The Integrated Services layer contains several modules that provide backend services for data management and end-user applications. These applications will be accessible through the ARIADNE portal, e.g. end-user applications for data discovery and access. Data integration, facetted search, visualization and access are foreseen to be possible based on thematic, spatial and temporal information contained in the metadata.

ARIADNE strives to enable a high level of semantic interoperability of datasets/collections, based on (meta)data in RDF and knowledge organization systems (e.g. thesauri) in Simple Knowledge Organization System (SKOS) format[229]. Furthermore the project uses the CIDOC-CRM[230], the reference ontology for the interchange of cultural heritage information, which is an official standard of the International Organization for Standardization (ISO 21127:2006). Also several recent extensions of the ontology are available, CRMsci (scientific observation model) and CRMarcheo (excavation model), for instance[231]. The extended CIDOC-CRM allows advanced data integration and exploration which, however, depends on the interest of data providers to map their databases conceptual reference model, which several ARIADNE partners consider and some already do.[232]

Summary of the core and additional services

Core services

ARIADNE aims to implement a state-of-the-art data portal that provides three core services which are data registration, search and access. The development of these services is based on a pragmatic approach which recognises the heterogeneity of the existing data sources and access methods offered by the providers:

1) *Registration:* The registry allows data providers to describe their resources (collections, datasets, etc.) following a common standard (the ACDM), which enables Linked Data based interoperability of the registered resources. Interoperability here means that the registry holds

[225] Open Archives Initiative - Protocol for Metadata Harvesting, http://www.openarchives.org/pmh/; W3C - SPARQL Protocol and RDF Query Language, http://www.w3.org/standards/techs/sparql#w3c_all

[226] The MoRe (Metadata & Object Repository) aggregator is employed for this task, http://more.dcu.gr; in the case of a small collection also provision of the metadata in an Excel file may do.

[227] W3C - RDF - Resource Description Framework, http://www.w3.org/standards/techs/rdf#w3c_all

[228] A matching tool is provided by ARIADNE partner University of South Wales (Binding & Tudhope 2015); Heritage Data: Vocabulary Matching Tool, http://heritagedata.org/vocabularyMatchingTool

[229] W3C – SKOS - Simple Knowledge Organization System, http://www.w3.org/2004/02/skos/

[230] CIDOC Conceptual Reference Model (CIDOC-CRM), http://www.cidoc-crm.org

[231] CIDOC-CRM extensions, http://www.ics.forth.gr/isl/CRMext/

[232] Examples of already conducted mappings are datasets of the Institute for Oriental and European Archaeology (OREA) at the Austrian Academy of Sciences (e.g. dFMRÖ coin database, Franzhausen-Kokoron graves dataset) and of the Italian national Central Institute for Cataloguing and Documentation (e.g. the Archaeological Monuments/Archaeological Complex [MA/CA] catalogue form).

information about how the resources are structured, what they contain, how they can be accesses, etc. The registry provides this information to the data search and other services.

2) *Search:* Data search and exploration services will be provided for different modes, e.g. querying (simple and structured queries) and browsing, also considering different levels and targets, i.e. whole data collections or individual items of certain types of data. Semantics based search & browse (e.g. facetted, based on thesauri) will play a core role in allowing users to discover relevant data. The search options of course also include map and timeline based search.

3) *Resource access:* Access to the data resources discovered on the portal will be enabled in support of different study purposes, e.g. access to only one object (e.g. a 3D model of a vase or building), an aggregated virtual reference collection of objects to allow for comparison, one or more datasets to extract numeric data, etc. Thus the access and possible use of data resources will correspond to the specific formats and interaction modalities offered by the resource providers (e.g. a 3D model may be rotated, zoomed etc.).

Advanced visual media services

Major steps towards advanced services have already been taken in ARIADNE in the area of visual media[233]. Based on their long-standing expertise and tools the Visual Computing Lab of CNR-ISTI (Pisa)[234] has implemented novel online services for publication, visualization and exploration of high-resolution images, Reflectance Transformation Imaging (RTI), and 3D models[235]. Also web services specifically for large terrain datasets generation, 3D landscape composing and 3D model processing are provided based on open source frameworks and tool[236]. Employing such visual media services in research projects or for content deposited in digital archives will greatly enhance researchers' capability to access, visualize and study archaeological objects online.[237]

Experimental use cases / services

The project also explores other, experimental stage services, which include use cases aimed at exploiting Linked Data based on the extended CIDOC Conceptual Reference Model. This requires mapping of databases to the conceptual reference model, which is under way. Another area of experimental development and testing of feasible services is employing data mining techniques to suggest patterns of relevance between data resources.

A major step towards integrated archaeological data and e-research

In summary, the overall objective of the ARIADNE e-infrastructure and data portal is building an environment for services that act as brokers between archaeological data providers and users. The creation of such an environment is a substantial step forward in the archaeological domain. In particular, it provides a common platform where dispersed data resources can be uniformly described, discovered and accessed.

This will allow overcoming some of the idiosyncrasies of the underlying different infrastructures of the data providers which currently prevent the collaborative exploitation of available data. At the

[233] A high demand for such services has been expressed in ARIADNE workshops and with regard to training offers; cf. ARIADNE 3D and Visualisation SIG, http://www.ariadne-infrastructure.eu/Community/Special-Interest-Groups/3D-and-Visualization, in particular the documentation of the Special Interest Group meeting 7-8 October 2013.

[234] Visual Computing Lab (CNR-ISTI, Pisa), http://vcg.isti.cnr.it

[235] ARIADNE: Visual Media Server, http://ariadne1.isti.cnr.it

[236] ARIADNE: Landscape Services: http://landscape.ariadne-infrastructure.eu

[237] For example, the Archaeology Data Service (ADS) has already implemented an 3D viewer for accessing and exploring 3D models deposited in their digital archive; the greater ambition is to provide an interactive 3D web-based working environment for data (cf. ADS 2015a).

same time, it represents an essential step towards the more ambitious goal of providing integrated services and tools for e-research, i.e. web-based research aimed at creating new knowledge (e-archaeology). Such services and tools may be implemented as Virtual Research Environments (VREs) offered on top of the ARIADNE data infrastructure and portal.

7.3.3 Archaeological research data and CIDOC-CRM based integration

This section addresses differences between archaeological research data and typical cultural heritage collection, and the use of the extended CIDOC Conceptual Reference Model (CIDOC-CRM) for integrating the rich information of research project archives and databases.

Archaeological content/data and context

It is important to recognise significant differences between archaeological research data and typical cultural heritage digital content collections. Such a collection of a museum for example is a database of photographs and description of pottery or, in the case of an archive, a database of historical photographs. The metadata of such collections is often made available to content federation networks and portals, like Europeana, for instance.

Such digital collections are less common in the domain of archaeological research. Indeed, the concept of homogeneous content collections with metadata that can be easily harvested, aggregated and presented on a portal does not fit well for archaeology. Because bringing together for example photographs from different excavations in a collection would mean to rip them out from their contexts. The context of an archaeological photograph or other content is the particular excavation or other archaeological investigation, and without much information about this context it is difficult to interpret what it means in scientific terms.

In short, the content/data of an archaeological investigation should in principle stay together as a project archive. Such project archives need to be described with sufficient metadata, while metadata for the various content items in an archive (e.g. site maps, photographs, find lists, subject experts' analysis of finds, etc.) may or may not be produced in practice.

A description of how excavation data is managed at one ARIADNE partner institution can illustrate the point about keeping together what belongs together: *"Documentation is kept together according to excavations and all documentation related to that excavation goes into the main folder of the excavation. If there is an analysis that includes materials from two or more sites, the same report (+raw data, pictures, maps etc.) is also copied in the main folders of those sites as well. This means duplicating and triplicating data but since we have no database yet it is easier to keep data wherever it belongs to."*

The project archive in this case includes, each in a subfolder, site assessment documentation (description of features, maps, excavation photos etc.), finds inventory (e.g. an excel or database file), finds drawings and photographs, documentation of different archaeometry analyses (e.g. reports, raw data, tabular data, photographs, etc.), documentation of conservation work, publications (articles, presentations). The example illustrates what "context" means with regard to archaeological content: the photographs (as addressed above) are firmly embedded in folders (finds folder, archaeometry folder) within the project archive.

Many projects of individual institutions and research communities of course attempt to create a database of information from different archaeological investigations. For example, a database of all field surveys conducted in a region or analyses of specific physical or biological materials conducted in different laboratories. Some outstanding finds of excavations (e.g. pottery, coins, jewelry, etc.) will also find their way into collections of museums (as described above).

Metadata schemas for cultural heritage objects

Common metadata schemas enable consistent description and sharing of information about collections of different providers. Basically a metadata schema describes how data are organised in digital libraries, collections or repositories and supports the search & retrieval of the stored digital content.

Due to the variety of the metadata standards and many proprietary solutions that are in use at cultural heritage organisations, typically schema-level mapping to a common general data model is employed to enable a basic level of interoperability in content federation and portal initiatives. A widely used model for such mappings is Dublin Core which is the main standard in Open Archives Initiative based metadata harvesting (OAI-PMH).

But Dublin Core is not adequate for sharing information about cultural artefacts as held by museums. For such artefacts richer schemas are required such as LIDO - Lightweight Information Describing Objects. In the creation of LIDO elements of existing museum standards (CDWA Lite, Spectrum, MuseumDat) as well as the CIDOC-CRM were considered.[238] The new standard has been promoted by Athena and other major museum projects in Europe and beyond (AthenaPlus 2014).

As a museum standard LIDO focuses on moveable objects, hence lacks options to document information about architecture (e.g. monuments) and archaeological sites. Therefore the CARARE project created a schema based on elements drawn from LIDO, MIDAS Heritage (the UK Historic Environment Data Standard) and CIDOC-CRM to mediate between the native metadata collected from project partners.[239]

While LIDO and CARARE metadata basically are intended to exchange and aggregate information about cultural heritage objects, metadata for research data presents an even greater challenge. Because researchers need rich additional metadata (contextual information) to evaluate and re-use the data for their research purposes; concerning archaeologists see *Section 4.4.3*.[240] A challenge therefore is balancing the requirement of common metadata for data discovery services and the need of detailed metadata to support data evaluation and decisions on re-use, maybe requiring a compromise between these objectives.

The CIDOC-CRM as common Conceptual Reference Model

ARIADNE uses the CIDOC-CRM as standard Conceptual Reference Model and promotes the adoption of compliant metadata schemas (such as CARARE, LIDO and others), thereby leveraging the metadata richness and standardisation in the cultural heritage sector. At the same time, the CIDOC-CRM is intended to enable exchange and integration of scientific documentation of finds, sites and monuments at the level of detail and precision required by researchers of the heritage sciences[241]. Recent extensions of the CIDOC-CRM cover scientific observations and argumentation (CRMsci, CRMinf) and there is also a special extension for archaeological excavations (CRMarchaeo).[242] Thus CIDOC-CRM based modelling of scientific processes and documentation of observations can enable integration of scientific information and argumentation (knowledge claims).

---

[238] LIDO was released by the ICOM-CIDOC Working Group Data Harvesting and Interchange in November 2010, http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/

[239] CARARE - Connecting Archaeology and Architecture in Europeana: The CARARE metadata schema [v1.1, 2011-11; v2.0, 2013-02], http://www.carare.eu/bul/Support/CARARE-metadata-schema

[240] A case in point is also that humanities researchers perceive the metadata of archives and museums as aggregated by Europeana as insufficient for research purposes, due to lack of required context information (cf. *Section 9.1.2*).

[241] Cf. Definition of the CIDOC Conceptual Reference Model. Version 6.1, February 2015, pages i-ii, http://www.cidoc-crm.org/docs/cidoc_crm_version_6.1.pdf

[242] See the overview and description of the CIDOC-CRM extensions at: http://www.ics.forth.gr/isl/CRMext/

In the perspective of the traditional information aggregation and integration approach the need of richer information would be addressed by conceiving a large set of fixed “core” metadata fields each federation member has to provide for aggregation. Recently members of the CIDOC-CRM community have criticized this approach as fundamentally inadequate for archaeological and other cultural heritage as it entails artificial generalizations and does not tie in with the actual curatorial practices and contextual knowledge of the data providers (Doerr & Oldman 2013; Oldman *et al*. 2014).

Instead, a model of data provision and aggregation is proposed in which the data providers employ the CIDOC-CRM to adequately describe the meaning and context of their information objects and data aggregators integrate and provide homogeneous access to the information, in a way that retains its original meaning and proper context. The proponents argue that this approach increases the utility of the information services for cultural heritage research, education and engagement activities and offers research e-infrastructure a better chance to be sustained into the future. But the approach requires properly managed processes and working relationships between data providers and aggregators as well as novel methods and tools, which are being defined and developed by members of the CIDOC-CRM Special Interest Group (Doerr *et al*. 2014a).

In the ARIADNE initiative, set in a horizon that goes well beyond the current project, providing access to archaeological data is not about presenting collection objects (like a museum collection website) or pointing to a database where some specialist information may be found. ARIADNE, among other goals, aims to lay the foundation for the integration of rich, structured information from heterogeneous archaeological resources that are relevant for answering research questions. The resources such as databases of archaeological institutes contain a multitude of facts that have been established with various methods and in different contexts of research. Therefore a common way to describe archaeological information is required that allows semantic integration and addressing questions beyond the local context of data creation and use.

This objective is addressed by the development of the “ARIADNE Global Schema” which is based on the CIDOC-CRM and its latest extensions. The aim of semantic integration of research data requires that the participants conduct a conceptual mapping of their database structures to the extended CIDOC-CRM. The mapping enables the conversion and export of the databases in a CIDOC-CRM compatible Resource Description Framework (RDF) format which can be shared as Linked Data on the Web or directly transferred to the ARIADNE platform for data integration.

The challenges of enabling effective mapping are addressed by an innovative solution, the SYNERGY Reference Model.[243] SYNERGY is intended as a modular environment composed of different instruments which will perform individual tasks of the mapping process, including also a knowledge base of re-usable mapping cases. The mapping definitions and conversion operations are still in progress within ARIADNE, based on the experiences of several project partners who are defining complex correspondences between the entities contained in their databases and the conceptual classes provided by the extended CIDOC-CRM[244].

7.4 Support for services beyond ARIADNE

The ARIADNE data infrastructure and portal will not be an “island” but is intended to serve as a node in the ecosystem of e-infrastructures and applications for the archaeological research community in Europe and beyond. To serve as such a node support for, and interoperability with, external services is required.

[243] The SYNERGY Reference Model of Data Provision and Aggregation. Current contributors: M. Doerr, G. de Jong, K. Konsolaki, B. Norton, D. Oldman, M. Theodoridou, T. Wikman. Draft, June 2014, http://www.cidoc-crm.org/docs/SRM_v0.1.pdf

[244] They use the available Mapping Memory Manager module of SYNERGY, http://www.ics.forth.gr/isl/3M.

Therefore the service portfolio of the ARIADNE data infrastructure and portal includes interfaces and services which allow provision of data for external uses. These are interfaces for re-use of (meta)data which developers may receive through harvesting (e.g. OAI-PMH), querying the ARIADNE Linked Data store, or simpler mechanisms of information syndication, RSS feeds to community websites, for instance. The purpose of such services is to allow the ARIADNE e-infrastructure play a significant role in the data service environment in and beyond the different domains of archaeological, cultural heritage and other humanities research.

The ARIADNE e-infrastructure could also benefit from using services provided by other major e-infrastructures, including from domains such as earth, environmental, biological and other sciences, which are present on the multi-disciplinary map of archaeological research (cf. *Section 3.3.4*). Deep interlinking of the data from such services would, however, require integration of conceptual knowledge of the different domains (e.g. terminologies, ontologies). Such integration could be sought based on use cases with a clear added value for archaeological research communities. While this is not a priority for ARIADNE in the 5-year horizon, we assume that in the 10-year horizon such integration might be attempted.

7.4.1 Support for integrating services

In the evaluation of the portals survey the participants (28 experts of 21 project partners) assigned relatively high importance to the following recommendations:

- Deploy Linked Open Data (LOD) to integrate information within the portal and to link to external resources which follow LOD principles (e.g. HTTP URIs and RDF);
- Provide interfaces to allow external applications exploit available data, metadata and conceptual knowledge (e.g. well-documented API, OAI-PMH target, SPARQL endpoint)[245];
- Enable integrated access to data and publications (i.e. include metadata of document archives and publishers).

Thus the evaluators acknowledged that the portal should not be an "island" but serve as a node in an ecosystem of data and services, which may well extend beyond core areas of archaeological research. Below we briefly address the three recommendations to make clear some requirements and opportunities.

Deploy Linked Open Data (LOD) to integrate information within the portal and to link to external resources which follow LOD principles

The last 10 years have seen much progress in Linked Open Data (LOD) know-how, i.e. what is required to produce, publish and interlink LOD of archaeological and cultural heritage collections/ databases. In practice, however, not many LOD datasets have been produced and interlinked as yet. A wider uptake and effective use of the LOD approach in the domain is necessary (cf. *Section 8.5.6*). Overall two core requirements must be met: effective interlinking of Linked Open Data requires use of common or mapped vocabularies (thesauri, ontologies), and LOD resources need to be curated to ensure reliable interlinking.

We envisage that the initially relatively small web of Linked Open Data emerging from ARIADNE resources can be expanded further through interlinking of data and conceptual knowledge both within cultural heritage and humanities domains as well as with other disciplines. This would support interdisciplinary research involving researchers in archaeology and other domains, natural history and environmental change, for instance. As this will go beyond ARIADNE developer capacity, related

[245] API: Application Programming Interface; OAI-PMH: Open Archives Initiative - Protocol for Metadata Harvesting, http://www.openarchives.org/pmh/; SPARQL Protocol and RDF Query Language (W3C recommendation), http://www.w3.org/standards/techs/sparql#w3c_all

developer communities should be invited to expand the LOD with link-sets (i.e. sets of links between LOD resources) and build applications to explore and use the interlinked resources.

Provide interfaces that allow external applications exploit available data, metadata and conceptual knowledge

ARIADNE services that allow external applications consume available data, metadata and conceptual knowledge could be an OAI-PMH target for data harvesting, a SPARQL endpoint of an RDF triple-store (to query and consume RDF data) and, for small-scale information purposes, RSS feeds (e.g. notification about newly registered data).

Such services for external application developers should of course be documented very well. Good documentation is required to allow for ease of use and thereby increase the likelihood that new added value services are created. This could be data “mashups”, i.e. combination in novel ways of data from different sources. The service interfaces may also allow developers of domain-specific Virtual Research Environments (VREs) to ingest relevant data into computational or other applications.

It is foreseen that some ARIADNE services will allow accessing data for research purposes like comparing objects of an aggregated virtual reference collection or extracting numeric data from datasets for computation. In this area cooperation with external application developers or “data scientists” would be beneficial in order to promote the creation of applications that support such specific research tasks.

Local and online workspaces of researchers may both consume as well as provide data and conceptual knowledge from/to ARIADNE. As one partner suggests, the project should “*provide rich and easy-to-use APIs for researchers, so that they can effectively query, import and integrate data and conceptual knowledge from ARIADNE and work with them in their own research workspace”.* Such workspaces can include relatively simple applications for extraction and combination of “small data” or conceptual knowledge (e.g. a statistical tool, terminology list generator), whereas in the 10-year horizon advanced Virtual Research Environments (VREs) could emerge that allow online collaborative work with larger if not “big” data or merged Linked Open Data (LOD) resources.

Enable integrated access to data and publications (i.e. include metadata of document archives and publishers)

Because of ever more open data mandates of funding agencies we expect that in the coming years the publication of research papers with underlying data will become more widespread also in the area of archaeological research. ARIADNE aggregates metadata of a larger number of archaeological datasets. Aggregating also metadata of publications from document repositories and publishers is not considered. But the connection of research publications and underpinning data is of vital importance in the scientific enterprise.

Therefore the project should investigate how interlinking of and integrated access to publications and data on the ARIADNE portal might be achieved, e.g. based on established data citation standards such as DataCite. If a comprehensive corpus of linked metadata of both publications and datasets emerges, an advance towards tracing the re-use of data might be possible in the future (10-year horizon). The immediate goal in this scenario is to establish open data sharing through state-of-the-art archives as a standard practice.

7.4.2 Services of lower relevance

In the evaluation of the recommendations for the ARIADNE data portal two services were not appreciated by most project partners. The first recommendation, relating to social networking platforms, is clearly not in the scope of ARIADNE, while the second may be considered if there is a demand for the suggested service.

Implement mechanisms for linking up with academic/professional networking and communication platforms

Such mechanisms were not appreciated by the majority of project partners. But some had previously suggested implementing a dedicated professional exchange platform, providing social networking features on the data portal, or finding appropriate ways to link up with external platforms. For example, one ARIADNE partner thought of a web portal that "*would allow the direct contact between persons with the same professional competences, the exchange of ideas and information, and also the awareness of the necessity in standardisation of the administration work (in managing the archaeological heritage) at the level of each institution across Europe"*. Another partner suggested *"taking advantage of social networking in the portal to be created in order to raise the site's readability among users and explore the potentials of various social media"*.

ARIADNE core and advanced services will primarily provide capability to discover, browse, access, visualize and (re-)use data with tools available online or in the local workspace of the researcher. Offering a state-of-the-art social networking platform within the ARIADNE portal will not be attempted. This would duplicate existing academic / professional communication platforms like Academia.edu, Mendeley and ResearchGate where archaeological researchers are already present in large numbers (cf. ARIADNE D2.2 Second Report on Users' Needs, February 2015).

The researchers use the platforms for functions such as presentation of a personal profile, sharing of literature references and documents, and networking with colleagues. The platforms offer a welcome alternative to institutional services (e.g. a department/research group website, document repository), which many researchers see as ineffective for their own needs. But with regard to institutional services ARIADNE is only interested in networking of available data repositories, not to help remove perceived shortcomings with regard to the presentation of individual researchers.

Help enrich specialised community sites (thematic or focused on specific types of data) with relevant information, e.g. RSS feeds on newly available data

Such services were not appreciated by most project partners. But some had previously suggested that ARIADNE should support the creation or enrichment of community sites, for example, a site for *"the web publication of field survey data"* or *"aggregating satellite data sources for archaeologists".*

There is a concern that this would add to the proliferation of specialised websites that are often not maintained and enriched with user contributions. Therefore, in general, discovery and access to rich content/data should be offered by the ARIADNE portal, where the data is also interlinked semantically with various related resources.

However, established websites of research communities could benefit from ARIADNE information feeds based on RSS or other information syndication mechanisms. This could allow them enrich their coverage of particular subjects, e.g. notification of newly available datasets for a certain geographic region. If there is such a demand of community websites, ARIADNE might well feed (meta-)data to such sites which could provide research context and relevant tools for the data. This would also allow enlarging ARIADNE's footprint in and beyond the sector and may stimulate (re-)use of available data.

7.5 The ARIADNE portal as a data mobilisation tool

In the evaluation of the recommendations of the portals survey report a relatively high importance has been assigned to promoting open data sharing via the portal (average score 3.6 of max. 5). It seems that several evaluators could not envisage how a portal might serve as a tool for data mobilisation. The main recommendations in this regard were to make users aware of a "give and take" obligation, point to guides to good practice for open data, and suggest community repositories (including archives in the ARIADNE federation) where data can be deposited and made accessible.

We believe that the ARIADNE portal can serve as a useful tool for data mobilisation. As the portal will allow for cross-searching data archives, it can (and should) also contribute to the awareness of and capacity building in open sharing of re-usable data through archives. In this regard the portal can align with the emergence of many more open data policies/mandates of research funders and their request for data management and access plans.

The ARIADNE portal can support open data "evangelism" by highlighting and pointing to

- Authoritative and supporting sources (in the European context, for example, the Horizon 2020 Open Research Data Pilot and OpenAIRE2020),
- Guides to Good Practice in archaeological data management, open licensing and access,
- Open data licenses such as Creative Commons (CC0, CC-BY) and Open Data Commons (ODC-PDDL, ODC-By),
- Existing and newly developed state-of-the art data archives, and criteria for such archives (e.g. Data Seal of Approval),
- Standard metadata and vocabularies for archaeology and other cultural heritage, which will be included in the ARIADNE Registry (e.g. CIDOC-CRM extensions for archaeology, SENESCHAL open data thesauri and others),
- Initiatives that apply Linked Open Data principles to allow interlinking of data (e.g. Pelagios in the field of ancient geography and history),
- Inspiring examples of research projects that re-use data from accessible archives such as the ARIADNE partner archives,
- The option to register and share own accessible data resources through the ARIADNE catalogue.

Notably, some of these measures relate to and can be combined with other project activities, for example Guides to Good Practice for specific archaeological data, trans-national access/visits to research centres or summer schools (e.g. on making legacy data accessible).

7.6 Suggested actions

The suggested actions in this focus area concern two main goals of the ARIADNE data infrastructure and portal: to provide the archaeological research community with core and high-value additional services, and to enable the e-infrastructure act as a node in the wider information ecosystem of archaeology and beyond. Furthermore it is recommended to use the ARIADNE data portal as a tool supporting mobilisation of open research data.

Suggested actions: Providing core and additional services

ARIADNE has taken great care to identify the services which the archaeological research and data management community expects from the project. This includes results of a large online survey on user needs, evaluated service suggestions of a panel of archaeologists and data mangers, and regular consultation with project partners and other institutions. Therefore the project has a solid basis to provide data infrastructure and services according to current expectations, and take into account that user needs will evolve, for example, in view of the opportunities offered by ever more available open data resources.

In brief, the archaeological research and data management community expects from ARIADNE:

- A data portal that allows an overview of available but dispersed archaeological data resources,

- Capability to search across different digital archives/repositories which hold such resources (i.e. data collections, databases, datasets of projects, etc.),
- Effective data discovery, browsing and filtering mechanisms, in particular based on geo-location (maps) and date-ranges/chronologies, but also other advanced options such as facetted search,
- Data access methods according to the different data access levels, types/products and interaction modes offered by the providers.

These requirements will be met by providing the user community with an online facility to register and describe accessible data resources based on a common model, semantic integration of the information, and offering portal services that provide the required set of data discovery and access functionalities. We note that the data access levels of providers vary, so that the data portal can in general enable collection-level but in many cases not direct item-level access.

On the other hand, advanced online access and use of data items may be enabled with regard to products for which a strong demand exists, for example, visual media such as 3D models of objects, built structures and landscapes. The already available ARIADNE services in this area demonstrate that the project actively responds to perceived demand.

There is also clear evidence for services which the research community does not expect from the ARIADNE data portal. This concerns typical features of Web 2.0 platforms such as content filtered based on tags or ratings of other users, academic/professional profiles of users, and expert networking and discussion on the portal. Such services are provided by widely used portals such as Academia.edu and others.

Clearly the service portfolio of the ARIADNE portal should meet core requirements of data overview, search and access. There is little scope to invest limited funds on specific services that are not appreciated, are provided by other portals, or may run ahead of the needs of broad user segments.

Provision of the described services (registry, portal) will be a major step forward in the archaeological domain, allowing overcome current limitations of discovery, access and re-use of available data. At the same time, the ARIADNE e-infrastructure may open up future opportunities for online collaborative research in Virtual Research Environments (VREs), offered on top of the e-infrastructure and accessible data resources.

Suggested actions: Support for services beyond ARIADNE

The ARIADNE data infrastructure and portal should not be an “island” but serve as a node in the ecosystem of e-infrastructures and applications for the archaeological research community in Europe and beyond. To serve as such a node, support for, and interoperability with, external services is required. This concerns the question of how the ARIADNE data infrastructure can enable added value beyond the portal services.

In general this will be through the ARIADNE Linked Open Data (LOD). LOD are employed for internal service provision, but can also be used for interlinking with external data resources as well as provided to developers for producing added value services. The suggested actions below are evaluated recommendations of the ARIADNE portals survey.

Deploy Linked Open Data (LOD) to integrate information within the portal and to link to external resources which follow LOD principles

Using Linked Open Data (LOD) for information integration has been among the highest valued recommendations of the portals survey and is already followed in the development of the ARIADNE services. The development focuses on the information available in the data registry and employs widely used LOD vocabularies such as the Geonames gazetteer for geo-information, the Art & Architecture Thesaurus for subjects, and others. Use of such vocabularies will also allow interlinking with external resources which follow LOD principles. Rich interlinking with relevant resources within the domains of cultural heritage and humanities as well as of other disciplines will require dedicated efforts. To expand the initial web of ARIADNE LOD, related developer communities could be invited to create link-sets (i.e. sets of links between LOD resources).

Provide interfaces that allow external applications exploit available data, metadata and conceptual knowledge

This recommendation can be followed easily by allowing external application developers access and use the ARIADNE LOD (i.e. the SPARQL endpoint of the Linked Data [RDF] triple-store or a "dump" of the whole dataset). Available and well-documented interfaces can promote experimentation and generation of novel applications. Such applications may combine data from different sources, e.g. data "mash-ups" or assembled virtual collections of research objects. In the medium to long term we may also expect Virtual Research Environments (VREs) which use the ARIADNE LOD and others to support cross-domain collaborative research.

Enable integrated access to data and publications (i.e. include metadata of document archives and publishers)

ARIADNE focuses on data resources while aggregation of metadata of publications from document repositories (i.e. self-archived papers) and publishers is not considered yet. But the connection of research publications and underpinning data is of vital importance in the scientific enterprise. Therefore the project should investigate how interlinking of and integrated access to publications and data on the ARIADNE portal might be achieved. Established data citation standards (e.g. DataCite) will be instrumental in this regard.

Help enrich specialised community sites (thematic or focused on specific types of data) with relevant information, e.g. RSS feeds on newly available data

Such services were not appreciated by several project partners. There is a concern that this would add to the proliferation of specialised websites that are often not maintained and enriched with user contributions. However, established websites of research communities might benefit from ARIADNE information, e.g. notification about relevant new data. This would also allow enlarging ARIADNE's footprint in and beyond the sector and may stimulate (re-)use of available data.

Suggested actions: Data mobilisation support

The ARIADNE data portal should also support data mobilisation through raising awareness of the importance of open research data and assisting capacity development. Thereby the ARIADNE initiative aligns with open data policies, required data management and access planning, and requirements for high-value data in general.

The ARIADNE data portal, for example, could highlight and point to authoritative and supporting sources (e.g. the Horizon 2020 Open Research Data Pilot, OpenAIRE2020 and others), Guides to Good Practice in data management, open data licensing, available state-of-the art data archives, and inspiring examples of research projects that re-use archived data. Also promoted should of course be the option to register and share curated data resources through the ARIADNE Registry.

8 Innovation agenda – 10-year horizon

"I think there is a world market for as many as 5 computers."
Thomas Watson, Head of IBM, 1943

"I can assure you that data processing is a fad that won't last out the year."
Editor of business books for Prentice Hall, 1957[246]

8.1 Introduction and overview

For the 10-year horizon of the Innovation Agenda and Action Plan this report outlines the overall goals, intended methods, and selected topics which will be addressed in greater detail in the final agenda report. The overall goal is to create a roadmap towards potentially transformative innovations in archaeological research and communication, in particular digital archaeology (e-archaeology). With a roadmap at hand, new approaches to digital, ICT-enabled archaeological research and communication can be explored, challenges addressed, and feasible routes taken.

The ARIADNE Innovation Agenda and Action Plan are not about incremental innovations through small improvements of research practices and existing tools. The goal, in the long term, is transformative innovations, substantial changes in practices, which may be triggered, among other factors, by new technologies. But technology is only one factor in innovation and not necessarily the most important. Therefore the endeavour is not primarily about technology, although some topics in the 10-year horizon will require addressing questions of technology and data in greater detail.

Overall, however, transformative innovation will concern the whole ecosystem of archaeological research and communication, including institutional, technical and socio-cultural dimensions. As an example: if we think of novel forms of digital publication, the core challenge is very likely not technical but institutional. Because publication is tied to the system of scientific review, recognition and reward, and what does not fit with the established system will find it difficult to advance from prototypic solutions to wide adoption by the research community.

The current debate about the system of scholarly research and communication is all about "open", including, among others, open access (publications), open data, linked open data, open science and, of course, open research infrastructures. As many proponents suggest, "openness" provides much potential for novel forms of research collaboration, including participation of citizens, and innovative generation and publication of new knowledge. Thus the request for "openness" (or "open science", "science 2.0" and other labels) is closely tied to the expectation of transformative processes which could lead to favourable outcomes as envisioned by different proponents, advances in knowledge and relevance of science in/for society, for instance.

Approach and brief overview

This chapter first explains the envisaged shift from the 5-year perspective of the innovation agenda, which centres on immediate innovation needs such as open research data, to a long-term view of 10 years. One major question in the long-term perspective is what new opportunities open research

[246] Source: "Hopelessly wrong predictions", http://www.audiencedialogue.net/predict.html

data, data infrastructure and e-research methods will offer for innovative digital, ICT-enabled research in archaeology.

The ARIADNE data infrastructure and portal services will be one element that connects the different horizons in terms of time and perspective. For example, some results of the ARIADNE data portals survey are available that provide insights into current expectations for e-research enabled by an archaeological data portal. This does not mean, however, that all or even a large part of the innovation agenda for the 10-year horizon will relate to the ARIADNE data infrastructure and services. But considering their potential and further development will be part of the exploration of the 10-year horizon.

Next we broaden the perspective and ask which innovation topics could become particularly relevant in the 10-year horizon. As noted above, the interest is not in incremental changes but innovations that could considerably enhance and transform archaeological research and communication practices. Expectations of a transformative impact of "digital science" or "science 2.0" are particular relevant in this regard and therefore addressed in greater detail.

The main part of this chapter then addresses selected topics which seem relevant for the 10-year horizon of the research agenda. These topics have been identified through a scanning of recent sources (literature, reports, project websites and other online sources), mostly of authors of the field of digital archaeology in its various forms. In the further work on the 10-year horizon some topics may appear as less relevant and others added.

Scanning and summarisation of current topics is a first step in developing a roadmap towards innovative digital, ICT-enabled archaeology. Use of other methods will be required to shape and evaluate the roadmap, and suggest actions stakeholders could take to bring about the envisioned innovations. Methods considered for this purpose such as scenario building and expert panels are addressed in the final section of this chapter.

Some of the topics for the 10-year horizon of the research agenda have been discussed in the ARIADNE Expert Forum on Digital Futures of Archaeological Practice 2020-2025, 2-3 July 2015 in Athens. The forum has been organised by the Digital Curation Unit of the IMIS-Athena Research Centre on behalf of the ARIADNE Special Interest Group (SIG) on Archaeological Research Practices and Methods[247]. But the content of this chapter does not result from this forum, hence, ideas and opinions expressed in this chapter should not be understood as the view of participants, except the main author of this report. The published results of the forum will of course feed into the work on the final ARIADNE Innovation Agenda and Action Plan, which will become available in November 2016.

The 10-year horizon goes beyond the formal lifecycle of the current ARIADNE project. The elaboration of a roadmap and action plan towards 2025 is an opportunity to stimulate a broader reflection and discussion among stakeholders of long-term needs in innovative digital, ICT-enabled research and communication practices.

[247] ARIADNE Expert Forum: Digital Futures of Archaeological Practice 2020-2025 (Athens, 2-3 July 2015), http://summerschool.dcu.gr/?page_id=19#expert-forum; the expert forum has been recorded and a report or paper manuscript, based on the summarised recording and further discussion, is expected before the end of 2015.

8.2 Shift from data- to research-focused innovations

The development of the 10-year horizon (2025) of the innovation agenda builds on and extends the framework of the work on the 5-year horizon. The objectives and suggested actions for the first horizon (2020) primarily concern immediate innovation needs with regard to open archaeological research data, data archives, e-infrastructure and services. The 10-year horizon of the innovation agenda can focus more on innovations in digital, ICT-enabled practices of archaeological research.

We assume that without substantial achievements with regard to the objectives of the 5-year further progress in digital archaeology and collaboration across current boundaries of research domains, organisational and national settings is unlikely. Alternatively, if the immediate goals are met sufficiently, future-oriented perspectives will be strengthened or emerge in the 10-year horizon. These perspectives will be less dominated by the need of growing and integrating the stock of open and interoperable research data. This will still require much attention and improvement. But instead of being blocked by seemingly unsurmountable barriers, further progress in digital, ICT-enhanced archaeological research (e-archaeology) can be foreseen, explored and targeted.

Scenarios in the long-term horizon of 2025 can look more into potential research-focused innovations, explore emerging new perspectives and capabilities, and suggest pathways towards innovative and potentially transformative e-archaeology. The figure below illustrates the shift from the current focus of enabling open access to re-usable and interoperable data to enabling novel forms ICT-enabled archaeological research:

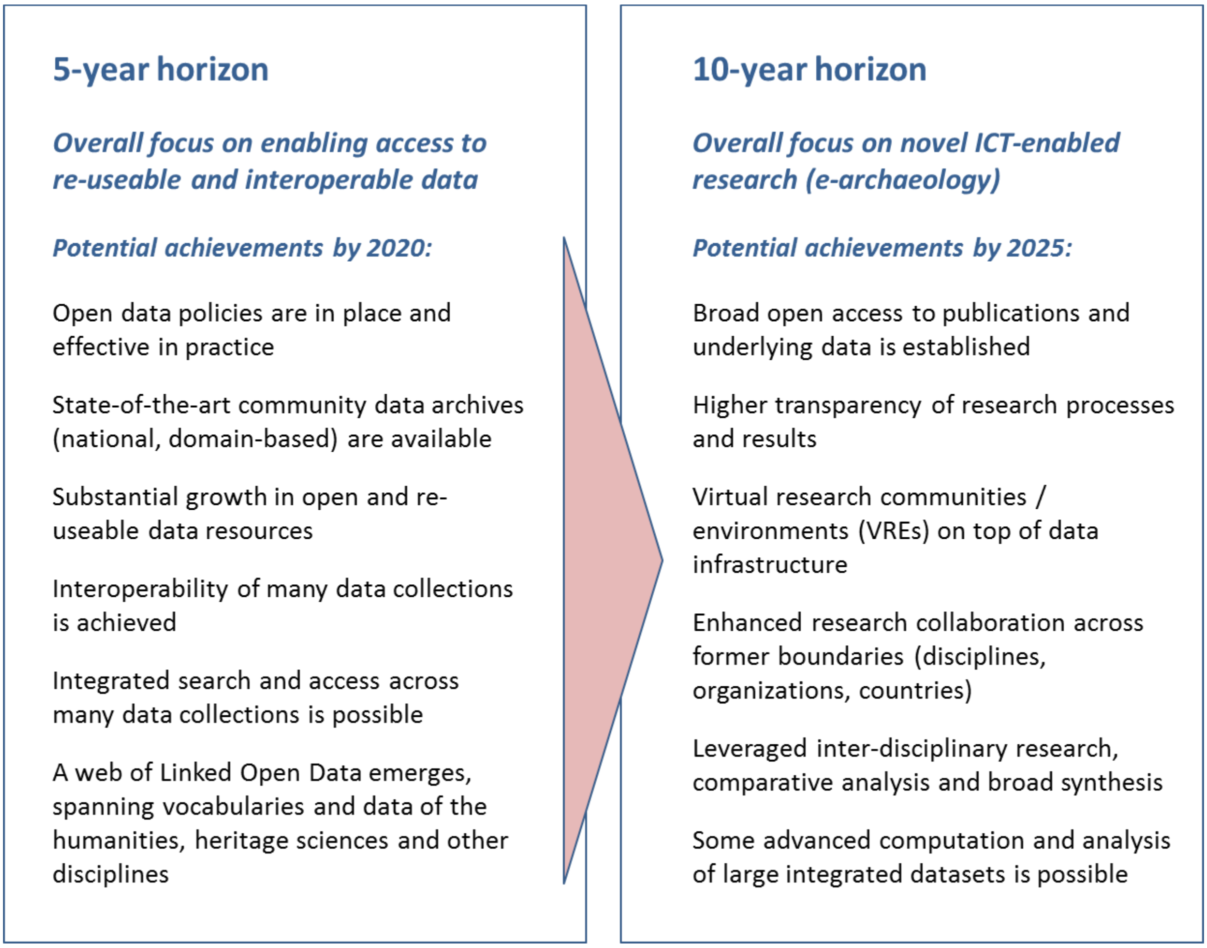


Figure: Towards open and transformative e-research. ARIADNE, SRFG 2015

The overall focus of the 10-year horizon thus is digital, ICT-enabled research practices, with special emphasis on e-research that is conducted online. We assume that the needs and requirements of the archaeological research community will evolve towards e-research capability when more open data becomes accessible and re-usable through community archives, data infrastructure and services.

We do not expect, however, that such novel research practices will be driven mainly by advances concerning data and technology, changes in research collaboration and methods will be just as or even more important.

The figure above contains potential achievements which may require some explanation:

The 5-year horizon includes community-level archaeological data archives. At present a common data archive is missing in many European countries, hence achievement of this goal will be rather demanding. Some large research institutes may be tempted to implement their own data repository, however a community-level solution will allow bringing together the research data in one place and will be more cost-effective than maintaining several repositories (cf. *Section 5.2.4*). Countries where the research community lacks a common data archive may "leapfrog" to a state-of-the-art solution by learning from acknowledged benchmarks such as the Archaeology Data Service in the UK.

Another difficult to achieve element in the 5-year horizon is the emergence of a web of Linked Open Data (LOD) that spans vocabularies and data of the humanities, heritage sciences and other disciplines. As a multi-disciplinary field of research archaeology could benefit greatly from such a LOD web as it allows discovery, browsing and retrieval of semantically related data and knowledge. But not many archaeological LOD resources have been published and interlinked as yet. The ARIADNE data catalogue will become available as LOD, which may promote some further LOD publication and interlinking. However, in the coming years the domain of archaeological and other heritage research would have to adopt the LOD approach much more so that a rich web of LOD emerges (cf. *Sections 7.3/7.4* and *Section 8.5.6* below).

In the 10-year horizon we expect that virtual environments for archaeological research will become available, built on top of the ARIADNE and related data infrastructures and services. The archaeological research communities currently lack such environments that are tailored to their specific requirements (cf. *Section 9.1.4*). In the figure some caution is expressed with regard to expectations of advanced computation and analysis of large integrated datasets. Archaeological data is difficult to integrate for such computation and "big data" mining technologies may not be applicable to many data resources (cf. *Section 8.5.4* and *Section 8.5.7* below)

In the 10-year horizon there will still be a need to broaden the access to various data resources required for archaeological research. As a multi-disciplinary field of research, archaeology depends on a high-level of openness of all involved disciplines and research specialties. Furthermore required will be good access to data which are usually not produced by archaeologists or in an archaeological context, airborne or satellite remote sensing and imaging data, for instance.

8.3 The ARIADNE data portal and e-research services

The shift from data-focused to research-focused innovations allows exploring opportunities for ICT-enabled archaeological research which will be opened by outcomes of the ARIADNE project, in particular the data infrastructure and services. This does not mean, however, that all or even a large part of the innovation agenda for the 10-year horizon will relate to the ARIADNE data infrastructure and services. But we consider them as a core element for connecting the different horizons and perspectives.

Results of the ARIADNE portal survey

The ARIADNE data services will be integrated in the data portal that is being built. Results of the ARIADNE portal survey suggest that some portal support for e-research, i.e. research that is performed mainly online, would be appreciated by archaeologists. Survey participants gave ideas for relevant e-research services, but in the final evaluation such services were not seen as a priority for the ARIADNE data portal at this point. The average relevance score for such services was rather good (3.6 of max. 5), yet most evaluators saw them as beyond the immediate goals for the portal[248].

The suggestions of survey participants can illustrate expectations for e-research practices which may be supported by a data portal beyond discovery and access to data of different digital archives. Participants envisioned e-research tools for several research tasks which included: mining, extraction, integration, measurement, comparisons, statistical and other analysis of numeric data or digital surrogates of research objects. Some examples are *"extraction of raw data from several (dispersed) project archives"*, *"comparison tools, statistical tools"* or *"measurement tools (e.g. digital calliper).*

The examples suggest two ways in which a research portal could support e-research tasks: tools which enable the extraction/creation, combination and integration of data and, building on this, tools to process and analyse the data. Both numeric data (e.g. isotope values) as well as content (e.g. images of finds) were considered. Most suggestions related to capability for working, in an integrated way, with various content/data as required for archaeological research questions.

Three recommendations for the ARIADNE data portal concerning support for online research work (e-research) were formulated and then evaluated by 28 archaeological and technical experts of 21 project partners. Below we note the average relevance score (ARC, max. 5 possible) and what the evaluators envisaged with regard to the realization of the recommended services. This could be within the duration of the ARIADNE project (W.A., until January 2017) or beyond (B.A.) by 1-2 years or 3+ years (average percentages are given for these estimates):

- *Support integrated access as required for studying various research resources online (e.g. linking and comparing content)* – ARC 3.8; W.A. 36%, B.A. 1-2 years 54%, B.A. 3+ years 11%.
- *Provide or link to tools which enable researchers to extract and combine data (e.g. images from different databases, numeric data to produce a derived dataset)* – ARC 3.7; W.A. 29%, B.A. 1-2 years 64%, B.A. 3+ years 7%.
- *Provide or link to tools for data processing and analysis (e.g. statistical analysis, image data processing and analysis)* – ARC 3.4; W.A. 21%, B.A. 1-2 years 64%, B.A. 3+ years 14%.

Evaluators of data providers assigned considerably higher relevance scores (ARC) than technological partners while the time-horizons with regard to the likely realization of the services were roughly the same.

Work on e-research capability within ARIADNE

While the ARIADNE data portal will mainly focus on enabling data discovery and access, research & development activities of the project also address capability for advanced archaeological research practices. Some of this work has an experimental character and includes exploring use cases for the data that will be integrated by the e-infrastructure, pilot deployment experiments using CIDOC-CRM extensions for archaeology and data that is accessible and interlinked through the e-infrastructure.

Furthermore the ARIADNE services for enhanced online exploration of different types of visual media may promote novel forms of e-research. This will be based on the ability to publish, access, visualize and study digital archaeological research objects online using web services. Services for high-

[248] *Section 9.3* describes the two-step approach of 1) idea generation as part of screening existing data portals and services, and 2) evaluation of the derived recommendations.

resolution images, Reflectance Transformation Imaging (RTI), 3D models of objects and landscapes have already been made available for trial users[249]. The services can be employed for exploring digital research objects stored in digital archives or as part of virtual research environments.

Going beyond this work of project partners according to the project workplan, activities that promote and support the development of e-research capability may be:

- Collaboration with developers of e-research tools and VREs for archaeology,
- Integration in the ARIADNE data infrastructure of open data resources which are particularly relevant for archaeological "data scientists", e.g. large machine-processible datasets ("big data"),
- Promotion of the use of such ARIADNE-mediated data in hackathons for e-research tools.

Concerning broad training of archaeologists in the use of novel e-research tools and environments, we understand that this is not the task of e-infrastructure and service providers but should be offered by dedicated research centres. Such centres could allow hands-on user experience in novel e-research tools and environments in the context of research projects.

8.4 Towards transformative open digital science

Initiatives for innovative digital, ICT-enabled research in a 10-year horizon should be embedded and driven by broader frameworks of "open science", which are relevant also for archaeological research and communication. The concept of "open science" (or "science 2.0") emphasizes openness of research practices and results and that digital, Internet-based research and communication can greatly extend the societal reach and impact of science. Moreover it expects a transformation of the scientific enterprise in many favourable ways. This includes for example open sharing of knowledge and data, novel forms of research collaboration and publication, and greater relevance of science in/for society.

The transformative character of the request of openness, especially open research data, by governments, research councils, funding agencies and leading research institutions (e.g. UNESCO, the G8 States, European Commission, European Science Foundation, national research councils and others) is already felt widely in the ecosystem of research. Because compliance requires efforts such as negotiation of open data mandates, implementation of appropriate digital archives, intricate questions of IPR & licensing, training of researchers (e.g. data management planning) and more.

Open data is understood as a pre-requisite for envisioned transformative innovations which are subsumed under the concept of "Science 2.0". The overall vision is making the whole research lifecycle as transparent and accessible as possible, including a deeper involvement of citizens in science and research activities (see Fecher & Friesike 2014 on different perspectives). The vision especially builds on the new opportunities offered by digital environments and processes of research.

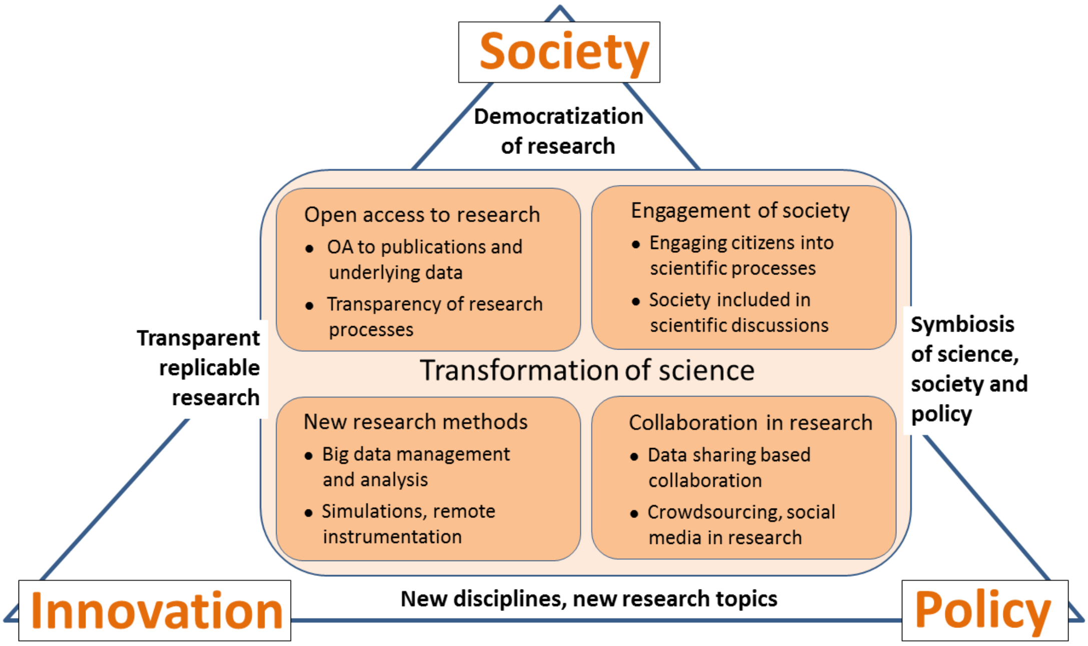
Below we address this line of thinking as exemplified in the European Commission's "Digital Science" strand of the Digital Agenda for Europe. Within the Horizon 2020 programme this strand relates to the overall goal of "Excellence in Science", which includes e-Infrastructure.[250] In the Digital Science strand the European Commission promotes *"transforming science through ICT tools, networks and media, to make research more open, global, collaborative, creative and closer to society"*[251].

[249] ARIADNE: Visual Media Server, http://ariadne1.isti.cnr.it; Landscape Services: http://landscape.ariadne-infrastructure.eu

[250] European Commission, Directorate General: Communications Networks, Content and Technology, Units: Digital Science and eInfrastructure, http://ec.europa.eu/dgs/connect/en/who-we-are

[251] European Commission: http://ec.europa.eu/digital-agenda/en/digital-science

While this seems to nourish a "technology-push" perspective (ICT tools, networks, media), the responsible unit's webpage notes: *"Digital science relies on the combined effects of technological development and cultural change towards collaboration and openness in research"*. The wider vision is depicted in the "Digital Science Triangle" figure below and described in the concept paper "Digital science in Horizon 2020" (European Commission, DG Connect 2013a).[252]



Digital Science Triangle. Source: http://ec.europa.eu/digital-agenda/en/science-and-technology/digital-science

The Triangle summarises the idea that the use of digital technologies for open access to scientific information (both publications and data), transparency of research processes, engagement and participation of citizens, enhanced research collaboration and application of new methods will change and transform how research is done. The uptake of "digital science" thus should benefit both the research community and society at large.

Some of the points in this figure may appear as somewhat outside the research environment and data resources of most archaeologists. But the overall directions will also fit for archaeological institutes and researchers, if read in more familiar terms like building larger database for computational archaeology ("big data"), public and community archaeology ("engaging citizens"), while "crowdsourcing" may not strike a chord. Such sourcing means to employ a web application for collecting many inputs (data, ideas) from researchers and/or non-experts to a project; examples of this approach in archaeology are rare as yet.[253]

It is also worth to note that the request for "engaging citizens into scientific processes" asks for trans-disciplinary research while archaeology today, at best, strives for more inter-disciplinarity. Trans-disciplinary has several meanings, overcoming academic and cognitive boundaries between

---

[252] See also the background document for the public consultation on "'Science 2.0': Science in Transition" (European Commission, DG Connect 2014).

[253] One example of such engagement of non-experts in archaeology is the identification of potential archaeological sites in Mongolia on satellite images to help direct the survey team on the ground (Lin *et al*. 2014; http://exploration.nationalgeographic.com/mongolia). The articles in Ridge (2014) present examples of crowdsourcing in cultural heritage, and Franzoni & Sauermann (2014) provide a systematic analysis and many examples of "crowd science" projects.

disciplines, for instance. But truly involving citizens in research would mean to go beyond scientific / academic concerns and take full account of citizens' own understanding of issues (Andrén 2010; Hirsch-Hadorn *et al*. 2008; Wiesmann *et al*. 2008).

8.5 Topics in the 10-year innovation horizon

The shift in focus from archaeological data to digital, ICT-enabled research and communication (e-archaeology) extends the framework of discussion, going beyond data-centred topics. Through a scanning of recent publications we identified and explored a number of other topics which seem relevant in the 10-year horizon of the innovation agenda. This section addresses some of the identified topics. In the further work on the 10-year horizon topics may appear as less relevant and others added.

Brief introduction

The ARIADNE innovation agenda does not nourish a "technology-push" view of innovation. Nevertheless, it is important to consider if discussion about digital, ICT-enabled research and communication in archaeology relates to technologies that are available or in development, or if future, maybe emerging technologies are envisioned.

Therefore, the first section below addresses types and timeframes of technological research & development (R&D). These types and timeframes include: available solutions (which may require some further engineering), market-near, prototypic systems/tools (1-3 years), medium-term, applied R&D (3-5 years), long-term, basic R&D (5-10 years), and "grand challenge" focused research (>10 years).

The next section presents two detailed suggestions of ARIADNE partners on digital archaeology perspectives in the 5-10 year horizon. One illustrates an innovation perspective focused on systems/tools that are well within research (available or market-near) and allow increased cost-effectiveness in archaeological data capture and processing. The other contribution requests considering the full panorama of schools (theories/methodologies), contexts and objects present in the field of archaeological research.

We also include results of an exercise of scoping "e-science" needs of archaeologists carried out 10 years ago. E-science here was understood as employing advanced digital technologies, in particular computational methods and tools. The suggestions from the exercise are thoughtful as well as pragmatic, and still highly relevant.

The presentation then proceeds from topics of a more technical character to broader concerns such as public archaeology and how to ensure the societal relevance of archaeology in general.

8.5.1 Types and timeframes of technological R&D

Attempts at envisioning or foresighting the future of fields of research or industrial sectors have often been driven by a technology-push approach, in which some future technical capabilities remove current needs and concerns. Such an attempt may also assume that the emergence and uptake of technological innovations would quasi automatically drive the research community towards new questions and breakthroughs in analytical capacity and understanding.

The ARIADNE innovation agenda does not nourish a primarily technological perspective of innovation. But it will be helpful to bear in mind the typical timeframes of different types of technological research & development as summarised in the table below:

Type of research	***Time horizon***	***Characteristics / focus***
"Grand challenge" focused research	>10 years	Visionary technological concepts which require breakthroughs in capability in several respects
Basic R&D	Long-term: 5-10 years	Aimed to generate new technological capability Basic research prototypes as prove of theoretical concept
Applied R&D	Medium-term: 3-5 years	Undertaken to acquire new knowledge with a specific application in view, e.g. a substantial enhancement of an existing technology Prototypes as demonstration of enhanced capacity
Market-near R&D	Short-term: 1-3 years	Aimed to turn prototypic solutions into market ready applications Pilots and tests applications under "real world" conditions Provides interoperability with existing systems and most required features for users (technical or non-technical)
Applications in use	Present	Further technical improvement, e.g., to add features, ancillary tools or services Make applications "off the shelf", well maintained and serviced

Table: Types of research & development. ARIADNE, SRFG(gg), 2015.

With regard to applications of information and communication technology (ICT) for research purposes some implications are important to note:

- Grand challenge focused and basic R&D in ICT does not aim to serve particular domains of research, but may have required capability of certain domains in mind (e.g. bio-sciences).
- Applied R&D takes domain-specific needs into account but the results typically are prototypes, in need of a better match with the requirements of intended users;
- Market near R&D aims at effective solutions and therefore involves more engineering and user-centred work.

In this widely used scheme the distinction between basic and applied research has been debated since long (e.g. with regard to its role in research funding schemes). Technologies can have an impact on disciplinary frameworks, including the distinction basic/applied. For instance, in genetics sequencing and other technologies have turned the initially basic into an applied field of research, and the established methods and techniques became the basis of commercial bio-technology.

With regard to archaeology the question is what can be expected from novel ICT-based methods and practices beyond enhanced capability or novel forms of research collaboration, i.e. do they make a real difference in terms of generating new insights and knowledge. For instance, if we assume that "big data" computing can be realized in archaeology, would the results challenge established theoretical or disciplinary frameworks or at least suggest new research questions?

8.5.2 Suggestions of ARIADNE project partners

In the consultation on emerging digital archaeology perspectives and needs in a 5-10 years horizon project partners gave valuable suggestions on what should be considered in scenarios. Below we reproduce and comment on the two most detailed contributions.

A baseline scenario of enhanced digital practices

One contributor suggested: "*Digital innovation has to allow a faster, more cost-effective archaeology but cannot – and must not – replace field investigations and direct links between researchers. The field investigations must become strongly computer-assisted, with a strong effect on archaeological research methodology, such as: less time spent for data entry, a possible closed-loop between field investigation and interpretation*".

Moreover the colleague suggested that this could be achieved with the following:

1) Promotion of techniques such as GIS at archaeological site level, drone-based photography, 3D scanning, photogrammetry, … to allow direct on-site digital representation of the field, enhance its understanding and have an immediate retroaction with excavation process.

2) Create an extension of an open source imagery software, dedicated to archaeological purposes.

3) Create (or give easy access to) computing infrastructures for heavy digital data processing such as 3D reconstruction. That means the use of a common software and data modelling.

4) Create an open source software for field data entry, implementing a standard data model and to be used on a handheld device such as a rugged smartphone or tablet.

4) The benefits of having RFID tags on each artefact should be evaluated."

The suggested digital innovation approach and tools seems achievable, well within reach, but nevertheless would require substantial work to allow for the sought for advances in archaeological methods and tools. The focus is clearly on cost-effectiveness, open source tools (which might reduce licensing costs and allow for better interoperability of tools and data), and standardisation of data modelling.

In the 10-year horizon we would consider this as a baseline scenario of enhanced digital practices in fieldwork and data processing, which could be expanded by e-research activities with its results (e.g. collaborative interpretation, comparison with results of other investigation, etc.). The baseline scenario illustrates that nowadays archaeological research can hardly be done without digital methods and tools. It also raises the question if the described enhancements (or others) may open up new avenues of scientific/scholarly research.

A panorama of archaeological schools, contexts and objects

A group of contributors suggested that studies on potential future e-archaeology have to take account of the complexity of archaeological research in terms of different theories, approaches and settings. This would require approaches and considerations as follows:

- *"Address digital infrastructure, methods and practice developments affecting the full lifecycle of the archaeological research process, from excavation and survey to publication;*
- *Cover the theoretical and epistemological diversity of different archaeological research approaches (from positivism to postmodernism), especially as in the context of post-disciplinarity they converse with traditions of diverse disciplines;*
- *Manage the methodological diversity of archaeological research (different countries, periods, types of archaeological sites, epistemic concerns and established traditions and practices), including both hypothesis-driven nomothetic research such as common with regard to prehistoric*

demography, economy and social structure, and hermeneutic, idiographic research such as common with regard to historical and Classical archaeology;

- *Address how diverse kinds of information objects, from 'raw' data to argumentation and discursive structures in scholarly objects, and everything in-between, will be curated in a future digital environment, including both the 'active' process of current field research and the re-interpretation of existing, poorly curated and studied archaeological data from past excavation and survey, as well as museum archaeological collections and their documentation;*
- *Consider the embedding of archaeological research within archaeological resource management activities (including contract, urban and rescue archaeology), of increasing importance for the construction and management of the archaeological record in years to come; and,*
- *Consider the interactions of future archaeological research within academic education and training in archaeology, and within increasingly relevant practices of public archaeology, such as those involving public access, interpretation and co-curation of archaeological resources, interaction with source communities, and ethnoarchaeological and phenomenological approaches."*

The contribution outlines a panorama of diverse "schools" in archaeology, distinct with regard to theoretical/epistemological and methodological frameworks; different contexts such as academic archaeology, archaeological resource management and related contract archaeology, and public and community archaeology; and highlights the need to support the full lifecycle of the archaeological research process, involving the curation of diverse information objects.

The diverse "schools" arguably will require different digital, ICT-enabled environments and tools. Archaeological resource management and related contract archaeology very likely will follow the baseline scenario above. All groups will need support in digital lifecycle curation although some of the products of the different "schools" will be different. The topic of public and community archaeology is addressed in *Section 8.5.11* below.

Suggested actions

- *Recognise that different archaeological schools of thought and research practices require different digital, ICT-based research environments and tools.*
- *Focus on phases in the lifecycle of archaeological research in which significant progress in knowledge may be achieved. In the last decades data generation has seen enormous progress; in the future other phases may require more attention.*
- *Recognise that issues of standardisation (e.g. data models) and cost-effectiveness are relevant for future research practices.*

8.5.3 Scoping archaeological e-science needs

Still relevant are the results of an exercise of scoping e-science needs of archaeologists carried out 10 years ago. The background to this exercise is that an E-Science Programme was available in the UK since 2001 but generally not well understood and used by scholars in the arts and humanities. Therefore the Arts and Humanities Research Council (AHRC) sponsored as series of workshops aimed at understanding better the needs of researchers in the arts and humanities concerning advanced digital technologies. The scoping work was managed by Sheila Anderson, a researcher at the Centre for e-Research at King's College London, with support from acknowledged researchers in different fields of the arts and humanities, including archaeology (Anderson 2007; project documentation: AHDS 2006).

Anderson defined e-science as *"the development and deployment of a networked infrastructure and culture through which resources – be they processing power, data, expertise, or person power – can be shared in a secure environment, and in which new research questions will arise, new forms of collaboration can emerge, and new and advanced methodologies explored"* (Anderson 2007: 6).

For each discipline covered, an expert seminar involving 10-12 researchers was run, prepared, moderated and reported by a lead researcher. Below we summarise some results of the expert seminar on archaeological e-science needs (Kilbride 2006):

- *"Bigger" is not the answer:* The expert group agreed that doing what some archaeologists were already doing at a bigger scale was not the answer to the e-science challenge. Examples were: remote sensing, imaging, surface modelling, volumetric modelling, numeric simulation modelling, geo-spatial and geo-temporal processing and text mining.
- *Archaeological e-science "way-finding" initiatives:* The expert group felt that "way-finding" was needed to help archaeologists along the road of e-science development, and suggested the following initiatives:
  - Studies of large scale data gathering and delivery,
  - Data mining to extend data sharing beyond metadata,
  - Ontology development and testing as a support to data mining and integration,
  - Development of simple data exposure and integration tools,
  - A review of web services and their relation to e-science,
  - Reconnaissance of e-science tools for archaeological sciences,
  - Transition from research frameworks to grand challenges.
- *Going beyond resource discovery:* The group also agreed that one challenge was unlocking the great potential of the many data sets and tools that had already been produced. Resource discovery was seen as important but to exploit existing potential more sophisticated search and retrieval would be required.
- *Preparing and linking for computation:* There was consensus that a lot of work was needed to make archaeological datasets fit for advanced computational processing. Given appropriate datasets, persistent identifiers and deep and dependable linking between datasets and tools was required beyond file level to items within files.
- *Generic principles:* The expert group also considered and recommended 12 principles for e-science activities in archaeology, e.g. that they should focus on the needs of archaeology; be recognised as fundamental research; assessed on the basis of archaeological values; pushing the boundaries in the use of information technology in research; involve the whole community or sector (not only the academic segment).

We agree to all of the above and with regard to developing the 10-year horizon of the innovation agenda consider the suggestions below.

Suggested actions

- *Promote collaborative way-finding for e-science approaches, methods and tools relevant to archaeological researchers.*
- *Focus on e-science needs specific to archaeological research, which may differ from those of other humanities as well as basic natural sciences.*

8.5.4 Data infrastructures and computing environments

In the 10-year horizon we may envisage an increase of *advanced* computing in some areas of archaeological research, based on larger databases (if not "big data") and novel computing paradigms. If such computing is conducted online in a *distributed* setting, interfaces between data and computing infrastructures are required which allow a seamless flow from data to computing infrastructure and *vice versa*. While this may seldom be "big data", there is the expectation that aggregated archaeological datasets with thoroughly detailed contextual information and sophisticated modelling could enable innovative large-scale comparative and synthetic research (cf. *Section 8.5.9* below).

The ARIADNE e-infrastructure is not intended to directly support advanced and distributed computing, but it will be worthwhile to keep an eye on and link up with developments in this field. Indeed, the interplay of data infrastructures with advanced research environments is one major area in the current development of digital or e-science solutions, including computing facilities and methods (e.g. Grid/Cloud based high-performance computing), virtual laboratories and remotely accessible data capture instruments. The EUTEMA "Next Generation Computing Roadmap" (2014) presents an overview of technological challenges in this area (see figure below).

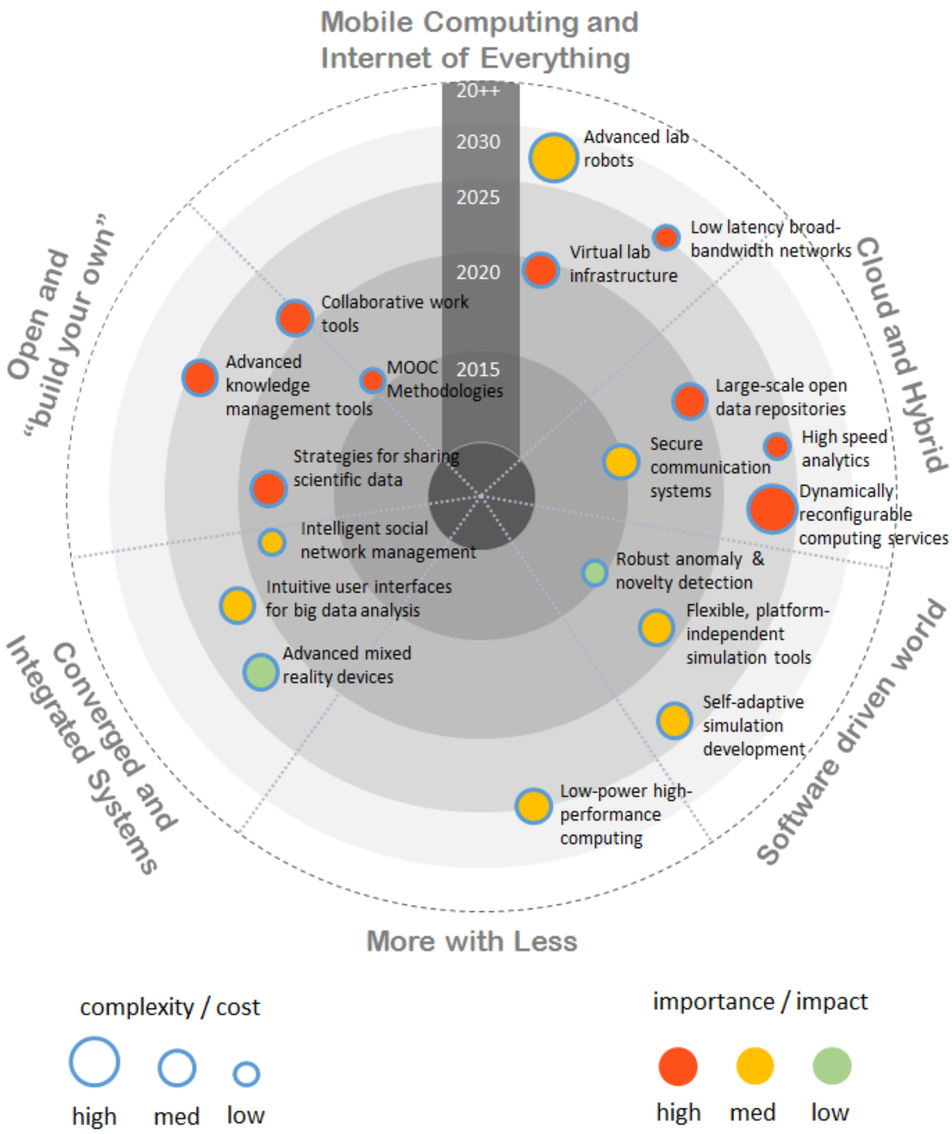


Figure: EUTEMA (2014: 79): Scenario "Education and Research: Connected brains"

The EUTEMA "Next Generation Computing Roadmap" study has been conducted for the European Commission - DG Communications Networks, Content & Technology. The roadmap positions perceived challenges in this field on the time scale 2015-2030(+) and evaluates them with regard to "importance/impact" and "complexity/costs" (see legend in the figure above).

The scenario underlying this overview (EUTEMA 2014: 33-34 and 50) emphasises the need for e-infrastructure that enables research collaboration based on massive archives of shared data that can be queried and processed with novel tools, involving libraries of modelling and simulation software. The e-infrastructure supports also virtual laboratories, telepresence and remote use of instruments.

Overall the EUTEMA roadmap confirms our emphasis on strategies aimed at driving open data sharing and the implementation of data archives in the time-horizon of 5 years (2015-2020). We also see as milestones in the 10 year horizon (2020-25) advanced knowledge management tools and breakthroughs in the availability of mature virtual collaboration environments. Advanced knowledge management we understand as semantics-based systems, tools and services not only for managing information but as productive applications that suggest and support new research. Virtual collaboration environments (VREs) are addressed in the next section (some background is provided in *Section 9.1.4*).

Archaeology and humanities computing

We question if "next generation computing" is essential for a large segment of the archaeological research community. Distributed Computing Infrastructures (DCIs), although available and promoted to all research communities since long, have not seen much adoption by archaeological researchers. DCIs provide a "Science Gateway" to Grid/Cloud infrastructure and software applications (virtual machines) for data processing, storage and transfer. Major impediments for using Grid or Cloud based data processing in archaeology arguably are the diverse and complex types of datasets, lack of consistent data structures, etc. (Hedge 2009). The limited adoption may also be due to a low affinity for and, consequently, little expertise in advanced and distributed computing. However archaeological researchers may increasingly use Cloud-based research support services for other purposes than computing, transfer, temporary storage and access during research projects, for instance.

"Next generation computing" projects therefore may remain rare cases in mainstream archaeological research. Given the characteristics and limited aggregation of archaeological datasets we would also not expect much application of "big data" mining and other methods (Sharma *et al*. 2014; for background on the "big data" topic see Bollier 2011; Halevi & Moed 2012; Lagoze 2014).

Examples of on-going work on "Big Humanities Data" centre on mining historical text corpora, modern media analyses (including social media), and audio-/visual recordings (music, film material, interviews)[254]. Historical text mining and processing of visual media (e.g. digitised historic maps) may also be of interest to archaeologists in some cases.

The projects funded in the Digging Into Data Challenge[255] for humanities and social sciences may represent the current state-of-the-art in these fields. One current example of "digging"

[254] cf. the papers of the "Big Humanities Data" workshops at the recent IEEE Big Data conferences (Hedges *et al*. 2013 and 2014)

[255] Digging Into Data Challenge, http://diggingintodata.org; Williford & Henry (2012) provide an analysis of the first round of funded projects (2009); Bajcsy (2010) presents some illustrative examples concerning requirements of storage, processing and visualization.

archaeological information (grey literature, image metadata) is DADAISM[256], building on the know-how developed in projects such as Archaeotools[257], STAR[258] and STELLAR[259].

Archaeological "big data" has only recently been addressed by some researchers, with regard to the volume of not yet digitised information (Wesson & Cottier 2014), its complexity and intricacy (Cooper & Green 2015), or how "big data" computing techniques might be used in archaeology (Gattiglia 2015).

Suggested actions

- *Look for uses of low-level Grid/Cloud based services and emerging examples of archaeological applications of "big data" mining and other methods.*
- *Encourage use of available distributed computing infrastructures (DCIs) by bringing together archaeological research groups with providers of DCI, especially with their developers of data processing software and services.*

8.5.5 Future virtual research environments (VREs) for archaeology

Research e-infrastructures typically are conceived as a multi-layered architecture with many data repositories at the bottom, a layer of common e-infrastructure and services in the middle, and flexible combinations of more specific technologies and services on top. The top level are VREs tailored to the needs of research communities of different disciplines and specialities or cross-domain, interdisciplinary research projects. VREs may support functions such as professional networking and information exchange, collaborative digital collection formation and use, and tools for data processing and analysis. Ideally research communities could flexibly select and combine solutions for these functions according to their specific needs.

Visions of VREs see researchers carrying out their work fully online, collaborative and supported by powerful integrated tools, and services that provide access to any kind of data and in whatever volume (e.g. "big data") the researchers may need for the research. But current VREs typically do not span the whole workflow of researchers, only a few tasks, and the researchers may be happy enough if they can carry them out more effectively with tools and data not available locally.

The VRE topic has been around since many years (e.g. David & Spence 2003; Bos *et al*. 2007; Carusi & Reimer 2010; Candela 2010). The creation and use of VREs in a discipline can indicate if and how far digital research tasks are already conducted online and in a collaborative manner, i.e. provide a good indicator of the overall situation of e-research in the discipline.

Therefore we conducted a background study on VREs in the field of archaeology (*Section 9.1.4*). The results suggest that at present there are no advanced, multi-functional VREs in use. Several digital environments have been developed for Classical Studies with a focus on digital scholarly editions (Babeu 2011). Arguably the main research support system of archaeologists is a Geographic Information System.

[256] DADAISM - Digging into Archaeological Data and Image Search Metadata (Digging Into Data Challenge, round 3, UK, AHRC/ESRC, 06/2014-12/2015), http://dadaism-did.org

[257] Archaeotools (UK, e-Science Research Grant, 2007-2009); Archaeotools - Data mining, facetted classification and E-archaeology, http://archaeologydataservice.ac.uk/research/archaeotools

[258] STAR - Semantic Technologies for Archaeological Resources (UK, AHRC, funded project, 2007-2010), http://hypermedia.research.southwales.ac.uk/kos/star/

[259] STELLAR - Semantic Technologies Enhancing Links and Linked data for Archaeological Resources (UK, AHCR funded project, 2010-2011), http://hypermedia.research.southwales.ac.uk/kos/stellar/

In short, the e-research scenario of VREs has not yet reached the archaeological research community. This could change in the 10-year horizon so that some fields of archaeological research may benefit from e-research environments. VREs may range from loosely coupled tools and services to tightly integrated workbenches for specific research communities, implemented on top of data infrastructures and underlying repositories. Hence the ARIADNE data infrastructure may contribute to the development of VREs for archaeological researchers. The current ARIADNE project does not include the development of VREs. To prepare the ground for future archaeological VREs some suggestions are given below.

Suggested actions

- *Look into VREs developed for other domains to conceive environments relevant for e-research in specific archaeological domains as well as in cross-domain collaboration.*
- *Consider cases where researchers use data mediated by ARIADNE as well as data infrastructures and services of other disciplines (e.g. geo, environmental, biological data).*

8.5.6 Web of archaeological Linked Open Data for research

The topic of Linked Open Data (LOD) is addressed in the chapter on ARIADNE's data infrastructure and services (*Sections 7.3/7.4*). ARIADNE will make available the information of the dataset registry as LOD so that other developers can interlink it with other datasets and develop additional applications. The multi-disciplinary field of heritage sciences research could benefit greatly from a rich web of LOD to allow discovery, browsing and retrieval of semantically related data. So far, however, not many archaeological and other heritage LOD datasets have been published and interlinked. Therefore we include the topic of a Web of such LOD for research in the 10-year horizon. But activities to promote and support further publication and interlinking of datasets would be required already in the coming years so that the envisioned Web of LOD can emerge.

A Linked Open Data vision

In 2010, Christian Bizer, a leading researcher in Linked Data methods and applications, outlined a 10 year vision for *"extending the Web with a global scientific data space"* (Bizer 2010). Bizer observed an already wide and increasing adoption of Linked Data technologies for sharing library, government and scientific data, as well as a first generation of applications that exploit interlinked datasets for novel information services. His vision for the next 10 years, quoted in full, was:

- *"Linked data will develop into the standard technology of sharing scientific data on global scale and for interconnecting data between different scientific data sources.*
- *The emerging Web of linked data will contain scientific data as well as data from other domains and might become as omnipresent in our daily lives as the classic document Web is today.*
- *Most open-license scientific data sets will be directly available as linked data on the Web. For extremely large data sets from astronomy or physics for which it is inefficient to generate an RDF representation, the Web of linked data will contain detailed metadata that will enable the discovery of these data sets.*
- *All scientific work environments will have linked data import and export features and will provide for publishing scientific data directly to the Web of linked data. Disciplinary repositories of scientific data as well as data archives will provide linked-data views on the archived data and will thus make their content available on the Web.*

- *Scientists will navigate along RDF links between different scientific data sets as well as between publications and supporting experimental data. They will use linked-data search engines to discover all data on global scale that is relevant to their question at hand".*

As one critical requirement for such Linked Data empowered research Bizer highlights discipline-specific vocabularies (e.g. thesauri, ontologies), which need to be integrated so that a searchable web of scientific data can emerge. Furthermore he notes that integration of Linked Data tools in scientific work environments is missing. The Bizer vision has five further years to materialise. A Web of LOD that spans vocabularies and data of archaeological, cultural heritage and other humanities research will very likely need 10 years (2025).

Progress in archaeological LOD

Efforts for heritage sciences LOD have so far been invested mainly on publishing datasets of artefact collections, e.g. museum collections[260] and special collections such as the numismatics databases that participate in the Nomisma initiative[261]. Also some archaeological data collections have been published as Linked Data, for example, in the STELLAR project Linked Data of Archaeology Data Service collections has been produced[262]. Special mention deserves that the Getty Research Institute has published their major cultural heritage thesauri as LOD[263], and also other widely employed international vocabularies and national vocabularies have become available as LOD (e.g. Iconclass[264]; UK thesauri made available by the SENESCHAL project[265]).

The last 10 years have seen substantial advances in LOD know-how, i.e. what is required to produce, publish and interlink LOD of archaeological and cultural heritage collections/databases (cf. Hyvönen *et al*. 2005; Aroyo *et al*. [eds.] 2007; Kollias & Cousins [eds.] 2008; Isaksen 2011; Tudhope *et al*. 2011; Elliott *et al*. 2014). In total, however, not many have been produced and effectively interlinked as yet. If there is a substantial increase in published and interlinked LOD datasets, semantic search and browse applications will allow discovery and retrieval of related content/data[266].

This will require a community of LOD curators who ensure reliable availability and interlinking of LOD datasets and vocabularies (Blumauer 2013). One example of good LOD curation practices is the Bio2RDF initiative in the field of bio- and life sciences, which created and/or interlinked 35 datasets[267]. On the Linked Open Data Cloud diagram Bio2RDF datasets are one of the densest clusters[268].

While there have been great advances in the field archaeological and cultural heritage LOD, progress as yet has been mainly towards search and access, use of LOD for research purposes is not implied. By use for research purposes we mean capability to address research questions and validate or scrutinize knowledge claims. This has not gone unnoticed by researchers and data managers who

[260] For example, the British Museum - Semantic Web Collection Online, http://collection.britishmuseum.org

[261] Nomisma, http://nomisma.org/datasets; several coin datasets of the American Numismatic Society and institutions in Europe have been made available in RDF format; the Nomisma project also provides an ontology for describing coins.

[262] Archaeology Data Service: The STELLAR project, http://archaeologydataservice.ac.uk/research/stellar/; ADS Linked Open Data, http://data.archaeologydataservice.ac.uk

[263] Getty Vocabularies as Linked Open Data, http://www.getty.edu/research/tools/vocabularies/lod/; ARIADNE uses their Art & Architecture Thesaurus for integrating subjects related information.

[264] ICONCLASS as Linked Open Data, http://www.iconclass.org/help/lod

[265] Heritage Data - Linked Data Vocabularies for Cultural Heritage, http://www.heritagedata.org

[266] Harpring (2014), Managing Editor of the Getty Vocabulary Program, gives a scenario in which their LOD vocabularies aid discovery of related heritage information.

[267] Bio2RDF: Linked Data for the Life Sciences, http://bio2rdf.org

[268] Cf. the Linking Open Data cloud diagram, http://lod-cloud.net

expect relevance of the LOD approach also in this dimension. For example, ARIADNE colleagues with regard to employing the LOD approach in archaeology note: *"Important that these concepts and technologies continue to be developed, but the next five years really need to start showing its usefulness for answering research questions".* Also researchers of the data publication platform Open Context emphasize, *"Archaeologists need to see more direct research applications in order to better justify the added cost and effort required to publish Linked Open Data"* (Kansa & Whitcher-Kansa 2013: 9; see also Kansa 2015).

The need to advance from LOD of data collections to research-focused applications is also emphasised by the e-science community that wants to see LOD support the process of research, including scientific workflows, computing and analysis (Bechhofer *et al*. 2010 and 2011). Indeed, novel LOD based models and applications that demonstrate considerable advances in research processes and outcomes may be decisive in fostering uptake of the LOD approach by research communities.

CIDOC-CRM based research applications

Expectations of research-focused applications of LOD in the field of archaeology and other cultural heritage research mainly relate to the CIDOC-CRM as integrating framework. Oldman (2012) explains that the Linked Data publication of the British Museum online collection data in CIDOC-CRM format *"comes from a concern that many Semantic Web / Linked Data implementations will not provide adequate support for a next generation of collaborative data centric humanities projects. They may not support the types of tools necessary for examining, modelling and discovering relationships between knowledge owned by different organisations at a level currently limited to more controlled and localized data-sets".* The ResearchSpace project[269] (led by the British Museum) is developing an online collaborative environment for humanities and cultural heritage information sharing and research that builds on CIDOC-CRM based methods.

Oldman (2012) also notes that since some years the CIDOC-CRM has been adopted by many projects "*but it has also reached a 'chicken and egg' stage needing the implementation of public applications to clearly demonstrate its unique properties and value to humanities research"*. This is about more than semantic search of related content/data based on the CIDOC-CRM or other ontologies.

The CIDOC-CRM is intended to enable exchange and integration of documentation of finds, sites and monuments, at the level of detail and precision required by researchers of the heritage sciences[270]. Data collections/databases contain a multitude of facts that have been established with various methods and in different contexts of research. Therefore a common way to describe the information is required that allows semantic integration and addressing questions beyond the local context of data creation and use.

The CIDOC-CRM developer community invites providers of data collections/databases to use the ontology to describe the meaning and context of their information so that research e-infrastructure and services in the field can provide integrated access to the information, in a way that retains its original meaning and proper context. The proponents argue that this is the way forward to relevant heritage research and other applications (Oldman *et al*. 2014).

It requires that many data collections/databases are conceptually mapped to the CIDOC-CRM and integrated through LOD methods. Recent extensions of the CIDOC-CRM cover scientific observations and argumentation (CRMsci, CRMinf), and there is also a special extension for archaeological

[269] ResearchSpace - Creating the Cultural Heritage Knowledge Graph project, http://www.researchspace.org

[270] cf. Definition of the CIDOC Conceptual Reference Model. Version 6.1, February 2015, pages i-ii, http://www.cidoc-crm.org/docs/cidoc_crm_version_6.1.pdf

excavations (CRMarchaeo)[271]. Thus CIDOC-CRM based modelling may also allow integration of research data (observations) and argumentation.

At large scale this approach will allow reaping the expected benefits only in the long term, when many data resources are mapped to the extended CIDOC-CRM. However, mapping of a few related collections/databases may demonstrate significant advantages of CIDOC-CRM based integration in the short-term, possibly promoting LOD publication and interlinking.

Suggested actions

- *Promote publication of LOD datasets (collections, databases) by more archaeological and other cultural heritage institutions, especially based on mappings to the extended CIDOC-CRM.*
- *Foster a community of LOD curators who ensure reliable availability and interlinking of LOD resources (datasets and vocabularies).*
- *Develop LOD-based applications that demonstrate advances in heritage science research to foster uptake of the LOD approach by research communities.*

8.5.7 Data integration for comparative and synthetic research

The need for data integration to enable broad comparative and synthetic research is a long-standing topic in archaeology since long. Eric Kansa notes, *"Historically, archaeologists became interested in computing and databases to control huge quantities of excavation data. They looked to the computer as a tool to retrieve and analyze information across multiple data sets and excavations to create broad syntheses. Unfortunately, the promise of digitally based meta-analysis has not panned out. The mass of digital material generated by archaeological activity is geographically distributed, fuzzy, incomplete, inconsistent, and often hard to access. The resulting complexity deluge presents a whole new set of problems for archaeology"* (Kansa 2011: 10-11).

This resonates a paper by Harrison Eiteljorg II (2004) who wrote that there was (and still is) the hope of archaeologists for data storehouses that allow to retrieve and analyse information from many excavations to produce broader synthesis. Due to difficulties with common terms and data structures this remained "an illusive goal", but a goal still worth to keep and strive for. *"Combining disparate datasets may not be a realistic near-term goal, but preserving datasets for future access is a necessity now"* (Eiteljorg 2004). This means counting on concerted efforts to gather more open datasets in repositories and the development of novel methods and tool that allow effective mining, comparison and synthesis.

Richards *et al*. (2013) note that a move towards data synthesis has been impeded because archaeological research projects made only summarised results of their research available. Moreover technical solutions remained at the level of data handling and data processing capability: *"Researchers have to handle an enormous amount of information, but it is not storage resource or processing performance that are required. The purely technological approach of providing more petabytes or guaranteeing more teraflops is insufficient; the diversity of archaeology requires fundamental research encompassing many disciplines, and developing innovative approaches. Integration also represents a challenge when considering the diversity of contexts, collecting protocols, relevance and goals under which data are collected"* (Richards *et al*. 2013: 315).

In recent years archaeological research has taken aboard ever more data capture methods and produced growing volumes of field survey and excavation data for documenting individual sites and areas. On the other hand, data integration for broad comparative and synthetic research is lagging

[271] See the overview and description of the CIDOC-CRM extensions at: http://www.ics.forth.gr/isl/CRMext/

behind ever more. As Kintigh *et al*. (2010) emphasise, *"We desperately need to foster synthetic research that transcends the spatial and temporal scales of individual research projects, both within and across traditional disciplinary boundaries"*. Obviously a stronger emphasis on such research and novel approaches and tools to gather, integrate and synthesize data collected in different investigations are required.

The lack of integration and wide-ranging comparative and synthetic research is evident when reading the proceedings of major conferences in the field, the CAA - Computer Applications and Quantitative Methods in Archaeology conference, for instance. The bulk of the presented applications are small-scale data integration tools such a 3D models, geographic information systems used at site-level or regional scale, and various systems and databases waiting to be filled with data (cf. *Section 9.1.4*). The title of the latest CAA conference (Siena 2015) was *"keep the revolution going"*, but there is no revolution, at least not with regard to masses of data being mobilised for broad synthesis or path-breaking frontier research. There are of course outstanding data collections/databases in different fields of archaeology, the challenge however is data integration (and, in many cases, data mobilisation).

Enabling broad comparison and synthesis obviously is difficult, because it requires bringing together and integrating data from different sources, including data from other disciplines that may be unfamiliar to archaeologists. One solution that could drive progress in this area is competence centres specifically set up to carry out or support such tasks. In the United States the National Science Foundation funds centres which look into existing data created by many projects in a larger field of research and help to integrate them for addressing "big picture" questions, which typically require a cross-disciplinary approach. Examples are the National Center for Ecological Analysis and Synthesis (NCEAS)[272], National Evolutionary Synthesis Center (NESCent)[273] and Socio-Environmental Synthesis Center (SESYNC)[274]. Development of an archaeological synthesis centre has recently been suggested by Kintigh *et al*. (2015: 11).

Rodrigo *et al*. (2013) describe requirements, approaches and experiences of such centres. The core point is that they incubate and support cross-disciplinary synthetic research through facilitating the exchange of expertise and collaboration between researchers with domain-specific knowledge (e.g. theories, methods, terminology, data resources) and data scientists with expertise in data modelling, analytical methods, advanced computation etc.

Suggested actions

- *Foster the development of novel methods and tools that allow researchers to bring together and work with the variety of data required for cross-domain, interdisciplinary research.*
- *Promote competence centres and programmes aimed at data integration for comparative and synthetic archaeological research.*

8.5.8 New forms of scientific publication

In the 10-year horizon we may also expect some advances with regard to new forms and ways archaeological researchers publish outcomes of their projects. This could be even more challenging than breakthroughs on the next generation computing and VRE fronts. Because publication is tied to the system of scientific review, recognition and reward, in which publications in "high-impact journals" and "the scholarly monograph" are traded for tenure and promotion.

[272] National Center for Ecological Analysis and Synthesis (NCEAS), United States, https://www.nceas.ucsb.edu

[273] National Evolutionary Synthesis Center (NESCent), United States, http://www.nescent.org

[274] Socio-Environmental Synthesis Center (SESYNC), United States, http://www.sesync.org

Challenging current publication practices

The established publication system and practices compare badly to the visions and prototypes of new forms of online publications, with embedded tools and deeply integrated open datasets, for instance (e.g. Aalbersberg *et al*. 2012; Attwood *et al*. 2010; Breure *et al*. 2011; Bardi & Manghi 2014; Jankowski *et al*. 2013)[275]. Without some adjustment in the scholarly credit system, new forms of scholarly/scientific publication may find it difficult to progress from prototypic implementations to wide uptake by publishers and authors.

In the Taylor & Francis international online survey on open access publications in 2014 one question was about the researchers' expectations for the future of such publications (6,749 responses): 63% thought that the academic journals will remain as the principal publication outlets, 21% that a significant proportion of research papers will be published only in subject or institutional repositories, 5% envisioned that subject or institutional repositories will become the primary home for research papers, and 11% that a new kind of publication outlet accommodating new types of research output will become dominant (Taylor & Francis 2014: 19).

The results concerning repositories may signal that researchers expect repositories to take on added functionalities in the research life cycle, instead of being the last in the queue. The distinction between "journal" and "repository" may become increasingly blurred if repositories become publishing platforms and value-added services for scholarly communication.

Concerning the adoption of new forms of scientific publication we assume that the uptake will be higher for moderately enriched or enhanced familiar ways of publication rather than approaches which depart significantly from established practice. A likely candidate for wider adoption could be "embedding" explorable digital objects in online papers, e.g. 3D models or RTI-based images of archaeological objects. For example, a recent paper in *Internet Archaeology* (Riris & Corteletti 2015) includes Reflectance Transformation Imaging (RTI) based images of engravings at a Brazilian rock art site which can be dynamically explored with the WebRTIViewer developed by the Visual Computing Lab of CNR-ISTI in Pisa[276]. Citable objects with stable unique identifiers embedded in online publications will allow also crediting the creators of the published objects, if such objects are shared for further research by other scholars.

A future for the monograph, and book-"published excavation"?

One of our background studies addresses the composition of the social sciences and humanities literature in comparison to the natural sciences (*Section 9.2.1*). The print-on-paper monograph stands out as particularly characteristic for the humanities, including some fields of academic archaeology. Despite the existing financial and publishing constraints the product still keeps its role as "the gold standard". There is some interesting experimentation on digital open access monographs funding and publishing. In general, it supports the established system of academic publication and credit, however publisher such as Open Book have found ways to greatly increase free online access to monograph content[277].

While a large number of open access books from different publishers are already available relatively few appear to be archaeological publications. Archaeologists are aware that the finally "published

[275] See also: Elsevier's projects Article of the Future project, http://www.articleofthefuture.com and Executable Papers, http://www.executablepapers.com; winners of the latter were Van Gorp & Mazanek (2011) and Nowakowski *et al*. (2011).

[276] Visual Computing Lab (CNR-ISTI, Pisa): WebRTIViewer, http://vcg.isti.cnr.it/rti/webviewer.php

[277] Open Book Publishers, http://www.openbookpublishers.com; started in 2008, currently offer 65 books, and many new titles are forthcoming. Their website reports *"Over 700,000 book visits; readers from 207 countries; 400 readers per title every month"*. The high figure of readership concerns the freely accessible HTML version Open Book produces of each title.

excavation" in book form is not the optimal solution. One example of a novel approach is the "Last House on the Hill" (LHotH) online publication of the Berkeley Archaeologists at Çatalhöyük (BACH) project[278]. With LHotH they aimed *"to holistically reconstitute the rich multimedia and primary research data with the impressive texts of the monograph"*.

Enhanced publications

Enhancement of a publication can mean many things including, for example, enhancing standard sub-components (e.g. images, tables, references), adding supplementary material files, creating relationships with research datasets or external objects, including executable parts, i.e. software and data to run a simulation (cf. the overview in Bardi & Manghi 2014). Among the high-end solutions are research publications with tables and figures that auto-update as new data becomes available. The first example of a "living" figure in a peer reviewed online research paper (Brembs & Colomb 2015) has recently been published in the open access F1000 journal (Ingraham 2015).

The project "Enriched Publications in Dutch Archaeology" explored expectations of archaeologists with regard to enhanced publications (Adema 2011a/b). The project included interviews with archaeologists, the enhancement of some papers published in the *Journal of Archaeology in the Low Countries* of the Amsterdam University Press, and collecting online feedback on the enhancements. More advanced enhancements, for example, were tables with filters on an underlying dataset or dynamic presentation of a burial with different layers and objects, shown or hidden according to users' choice. Breure *et al*. (2011) rate these enhancements as "low end".

Interviews with 14 archaeologists before the implementation explored expectations and concerns with regard to enriched publications. Adema (2010b) reports that the expectations were mostly basic and concerned addition of material that would not fit in a printed publication, e.g. adding a set of images or database; including GIS maps in publications represented the high end of the expectations. Interviewees expressed concerns about extra work, ownership of data, interoperability, peer review and sustainability of enhancements.

The project's conclusion was that for enhanced publications *"a lot of missionary work still needs to be done"*, especially *"to relieve fears and uncertainties regarding the new format"* (Adema 2010b). Experiments should take traditional forms of scholarly publication as starting point, and clear policies and trustworthy infrastructure for enhanced forms of publications were considered as crucial to ensure buy-in of the research community.

Future research objects

De Roure (2010) lists characteristics of research objects which in the future might replace the classical paper, at least in experimental and data-driven fields of research with intensive use of information technology and computing. The characteristics include that research objects should be retrievable, reproducible, repurposable, reviewable, referenceable. As computational research objects they would be *"designed for reuse and to drop easily into the tooling of e-Research, and better suited to the emerging practices of data-centric researchers"* (De Roure 2010; see also De Roure 2012 and 2013).

This will require new forms of formalized, semantic and dynamic provenance metadata which support automated re-use of resources (e.g. models, workflows, software) in network-based and data-intensive science. Van de Sompel & Lagoze envisioned some years ago "*that eventually provenance information will be exposed so it can be leveraged by a variety of tools for discovery, analysis, and impact assessment of some core products of new scholarship: workflows, datasets, and processes"* (Van de Sompel & Lagoze 2009: 197).

[278] Last House on the Hill (Centre for Digital Archaeology, University of California, Berkeley, CA), http://lasthouseonthehill.org

New forms of publishing research are likely to be driven by a process-view of research rather than static documents like open access publications in PDF format. In the pursuit of "open science" such publications would be tied in much stronger with the research context from which they stem. We might for example think of the "digital record" of an on-going excavation as a stream of data from the field and laboratories, continuously made available, analysed and discussed by subject experts. In such a setting "publications" would be snapshots of the state of knowledge at a certain time, instead of the annual excavation report and some papers by researchers in need of taking care for their academic record.

Nano-publications

Nanopub.org defines nano-publications as *"the smallest unit of publishable information: an assertion about anything that can be uniquely identified and attributed to its author"*.[279] This new approach typically builds on Linked Data standards. Nano-publications may be used, for example, to collect statements about a research objects, annotate online databases, link pieces of information, etc. Such nano-publications are considered as a way to allow credit to be given to contributors of small but important pieces of information (Mons *et al*. 2011). Assertions for nano-publications may also be extracted from published sources employing NLP technology.

The concept of nano-publications has seen some success in the bio-informatics community. Frequently referenced examples of development towards productive solutions are Open PHACTS[280] and DisGeNET[281]. Page (2015) explains how nano-publications could allow providing annotations for biodiversity databases like GBIF (e.g. flag up issues in their taxonomic indexing of species). In the Humanities one project with an implemented experimental system aims to integrate contributions into the bibliographical database EMTO (Early Modern Thought Online).[282] The technical developers of the PeriodO[283] gazetteer project in the field of ancient geography and history also aim to implement nano-publications (Golden & Shaw 2015; Rabinowitz 2014).

Suggested actions

- *Promote novel forms of digital publication that could "work" for archaeological projects in terms of enhanced access to research outcomes as well as academic credit.*
- *Start with moderately enriched or enhanced familiar ways of publication (e.g. embedding explorable digital objects in online papers), and make approaches which depart significantly from established practice adoptable as easy as possible.*
- *Investigate fields of "data-driven" archaeological research and publication in which accessible datasets and executables (software, dynamic figures, etc.) could play an essential role.*
- *Explore repositories as platforms for media/data-rich archaeological publications.*

8.5.9 Grand challenges as drivers of digital archaeology

Experts of the archaeological informatics and computing community suggest seeking "grand challenges" which are research-directed and contribute to the development of new theories and methods. Thereby the community could play a stimulating and transformative role rather than simply support well-established research practices. Grand challenges for digital archaeology should

[279] Nanopub.org: What is a Nanopublication, http://nanopub.org/wordpress/?page_id=65

[280] Open PHACTS (Open Pharmacological Space), https://www.openphacts.org

[281] DisGeNET (focuses on human gene-disease associations), http://www.disgenet.org

[282] EMTO Nanopub, http://emto-nanopub.referata.com/wiki/EMTO_Nanopub

[283] PeriodO - Periods, Organized, http://perio.do

go beyond what seems feasible in the short to medium term through applied research and engineering. The challenges would inspire the research community to push boundaries and explore new avenues of research with potential revolutionary impact, i.e. a shift in established paradigms, theories and methods of archaeological research (cf. Beale & Reilly 2014; Huggett 2012a, 2015a).

An online survey of asked archaeologists *"What are archaeology's most important scientific challenges?"* (Kintigh *et al*. 2014a). The survey was conducted in a project funded by the US National Science Foundation, and the authors mention: *"The question arose as we sought to develop recommendations for investments in computational infrastructure that would enable the discipline to address its most compelling questions."* The respondents were invited *"to identify problems of broad scientific and social interest that could drive cutting edge research in archaeology for the next decade and beyond".*

The survey received 181 responses, primarily from the United States (79%) and Europe (12%); the respondents were split across academic (45%), consulting (32%), and government (14%) employment sectors. The responses have been organised, enhanced, prioritised and refined by an editorial team. The published outcome is 25 grand challenges, grouped in 5 large categories. Below we list these categories each with one selected question:

(A) *Emergence, communities, and complexity*, e.g. Why and how do social inequalities emerge, grow, persist, and diminish, and with what consequences?

(B) *Resilience, persistence, transformation, and collapse*, e.g. What factors have allowed for differential persistence of societies?

(C) *Movement, mobility, and migration*, e.g. Why does migration occur and why do migrant groups maintain identities in some circumstances and adopt new ones in others?

(D) *Cognition, behaviour, and identity*, e.g. How do people form identities, and what are the aggregate long-term and large-scale effects of these processes?

(E) *Human–environment interactions*, e.g. How do humans perceive and react to changes in climate and the natural environment over short- and long-terms?

The survey report emphasises that many of the challenges concern cultural processes and involve complex, nonlinear relationships. Therefore they will require both sophisticated modelling and large-scale synthetic research. Furthermore, the challenges necessitate far more comprehensive online access to thoroughly documented research data with the required contextual information essential for comparative analyses. Moreover the authors note that many of the research questions *"will require intensive, cross-disciplinary collaborations. Although those collaborations will be demanding and time consuming, they have the potential to yield transformative results with cascading impacts far beyond archaeology".*

The report does not address specific ICTs or forms of collaborative e-research which may be necessary to tackle the described challenges[284]. The concern is primarily that large-scale and integrated datasets will be required for most of the challenges; related to one challenge under the Human-Environment Interactions (E) group for example: *"The challenge is to join disparate efforts into a broad-based initiative that can integrate existing and new sets of archaeobiological, geomorphological, paleoenvironmental, demographic, and other relevant data to model human/environmental interactions through time".*[285]

[284] In this regard see Kintigh *et al*. 2015.

[285] In an effort of *"looking forward through the past"*, Seddon *et al*. (2014) identified 50 priority research questions in palaeoecology. One of the six themes used to structure the questions is *"combining and synthesizing information from multiple records"*.

Similarly, contributors to the ARIADNE innovation agenda of one data archive and service provider wrote: *"As a data-archive that works interdisciplinary the usage of data across academic fields is something we promote and are working towards. For example: take archaeological data from a geographically limited area where excavations have been done, combine it with data about agriculture, climate changes, population changes, historical data from tax records, church books etc. and you can follow the historical/social/geographical development for that area up to modern times".* They also envisioned that in the future there might be a way to cross-search on the ARIADNE portal (meta)data of providers of different domains, combine the data, and work on it with special data analysis tools.

However, the challenge of bringing together and working effectively with the variety of data required for interdisciplinary research (e.g. in fields such as historical ecology) should not be under-estimated. One example of a project that develops research tools for working with disparate data from multiple sources is Terra Populus[286]. But the project works with available recent survey data from fields such as demographics, land use, environmental change, epidemiology, and others. For archaeological and historical research much effort will be necessary first to produce and bring together datasets for the envisioned interdisciplinary research.

Suggested actions

- *Seek grand challenges that inspire the research community to push the boundaries of digital archaeology.*
- *Suggest challenges that promote mobilisation and integration of datasets for domain and cross-domain, interdisciplinary research.*
- *Bring together domain experts and developers to create methods and tools for such research.*

8.5.10 Societal and environmental issues

Archaeology could benefit greatly from connecting to current societal and environmental issues and contribute its long-term, "deep history" perspective of social systems and changes in human behaviour, adaptation to environmental and climate changes, among other topics.

One important research field in this regard is human-environment interactions or dynamics of coupled human and natural systems (Dearing 2006). This field allows archaeologists to connect their knowledge of long-term developments as well as sudden changes in the past to current concerns of societal and environmental sustainability (e.g. Van der Leeuw & Redman 2002; Fisher *et al*. 2009; Smith *et al*. 2012; Wells 2011).

The Joint Programming Initiative (JPI) on Cultural Heritage and Global Change in their strategic research agenda asks for improvements in the modelling of the impact of climate change on cultural heritage (JPI-CH 2014: 21; see also Sabbioni *et al*. 2010; Climate Change and Heritage Management, EAA-2015 conference session). Here archaeologists can contribute their models and insights into the long-term transformation of landscapes, sites and objects.

Guttmann-Bond (2010) argues that archaeological research on ancient sustainable agricultural practices (e.g. prevention of soil loss, water harvesting methods) could help addressing issues of agricultural sustainability and food security.

Van de Noort (2011, 2013) conceptualized "climate change archaeology", and Rockman (2012) emphasises "the necessary role" of archaeology in climate change mitigation and adaptation. Lane

[286] Terra Populus, http://www.terrapop.org

(2015) provides a critical assessment of claims being made of archaeology's potential to help overcome the challenges posed by climate change and other environmental developments.[287]

Novotny (2010) points out that many societal challenges, like climate change, require bringing to bear research perspectives and insights of different disciplines: *"Most of today's major issues cannot be clearly categorized as belonging to either the natural or the social order. They are the result of complex, mutual interdependencies. Typically they emerge through a process of co-production which privileges neither social nor natural science. Climate change is the latest and perhaps most potent example: a natural phenomenon caused at least partly by anthropogenic intervention in the natural environment. Humanity has reached the planetary limits for numbers and resources, and must confront hard choices: how to discount the future, the cost for future generations, and the price a society is willing to pay in order to decrease carbon emissions. The scales of space and time found in nature need to be re-conceptualized in order to accommodate human spans and the human spatial environment"* (Novotny 2010: 319).

O'Brien (2010) urges all social sciences to participate in research efforts for solutions that respond to the global environmental change, while Balstad (2010) notes difficulties of social sciences researchers to engage in interdisciplinary climate change research because of a low affinity with and no background in the natural sciences. The strong relation of many fields of archaeological research with the natural sciences will be helpful in collaborative efforts in the field of climate and other environmental change (Smith 2009).

There are also examples where archaeological and earth & environmental research are neatly combined to address questions of climate change, for instance at the Centre for Past Climate Change of the School of Archaeology, Geography and Environmental Sciences (SAGES) at University of Reading.[288] Papers of archaeologists in the journal *Climate of the Past* of the European Geosciences Union[289] exemplify the spectrum of valuable contributions; some examples which address the impact of environmental change on past cultures are Hole (2007), Chepstow-Lusty *et al*. (2009), Tsonis *et al*. (2010), Sun & Li (2012) and Schittek *et al*. (2015).

Core areas in the field of environmental change are land cover and vegetation/biome reconstructions, site formation processes, implications of climatic and environmental trends for human occupation, organisation and subsistence (e.g. sea level change, floods, historical rainfall variation, desertification, etc.), in turn, human impacts on flora and fauna, landscapes, etc. Essential research resources are ice cores, geomorphology/chemistry data, lake sediment records (e.g. pollen), stable isotope and tree-ring analysis (dendroclimatology), among others. Concerning submerged landscapes and prehistoric archaeology the SPLASHCOS[290] project merits special mention.

In addition to the addressed climate change and environmental issues there are other fields where archaeologists could contribute a distinct long-term view of human cultures. Guidi & Armitage (2014) in their "History Manifesto" point out the need of *"long durée"* perspectives which they see largely absent in the political discourse.

[287] We also note the contribution of archaeology to the debate on the "Anthropocene" (Edgeworth 2014a; Edgeworth *et al*. 2014). The Anthropocene debate is about the long-term human transformation of the global environment, which has reached a scale and intensity that would deserve recognition by introducing, retrospectively, a new epoch in planet Earth's history. On a "geological" distinction of the Anthropocene from the Holocene see Waters *et al*. (2014). Nowviskie (2014) addresses the debate from a humanities perspective.

[288] Centre for Past Climate Change, https://www.reading.ac.uk/sages/research/CPCC/

[289] Climate of the Past - An interactive open-access journal of the European Geosciences Union, http://www.climate-of-the-past.net

[290] SPLASHCOS - Submerged Prehistoric Archaeology and Landscapes of the Continental Shelf (EU, COST Action, 2009-2013), http://www.splashcos.org; see also: European Marine Board (2014).

Suggested actions

- *Seek significant contributions of archaeological research to tackling societal, environmental and other issues (e.g. mitigation and adaption to climate/environmental change).*
- *Engage in collaborative, interdisciplinary research on issues; contribute archaeology's distinct long-term view of human culture, human-environment interactions, sustainability/resilience as well as sudden change.*

8.5.11 Public / community archaeology

Ironically, at the same time as archaeology finds it difficult to make it into the public's perception of relevant sciences, "archaeology is a brand" (Holtorf 2007) in popular media. There is a multitude of websites that present the "greatest", "latest", "incredible" or "most intriguing" (or "mysterious" or "terrifying") archaeological discoveries. While this is archaeology for people fascinated by what archaeologists unearth, participatory "public archaeology" or "community archaeology" still seems to have to prove its case.

An important point to bear in mind in this context is that a large part of archaeology is tied to the heritage management framework. This framework is all about protection or, in the first place, rescue of remains from destruction by development work. As Henson (2011) notes: *"Archaeology as academic study, and as professional practice, has become increasingly intertwined with heritage management. The chief concern so far for archaeologists has been to protect and conserve the heritage of the past, usually on the grounds that it is being preserved for future generations of archaeologists to study with better scientific techniques than ourselves. Built into the nature of archaeology therefore is an ethos of protectionism and self-serving. It can be argued that this has provided poor grounding for archaeology as public service, and makes it hard for the discipline to find an accepted role within society."*

Protection of sites for the public good is of course important, but usually does not involve the public, and also does not yield research results that might leverage citizens' appreciation, understanding and support of archaeology. The same arguably can be said of most academic publications. Moreover, documentation and data of the archaeological record, if publicly available, is not something many non-experts may easily understand and use for own work ("citizen science"). No wonder therefore that archaeology, not as "incredible" finds, but painstaking research work to establish facts and knowledge, requires a lot of mediation.

The Working Group on Public Archaeology of the European Association of Archaeologists emphasises: *"An activist approach is crucial for the successful impact of archaeology and public archaeology, however difficult it can be to position oneself, especially in the private sector. Public archaeology goes far beyond the remit of what we understand to be archaeological practice, into the daily lives of people, and if our work is not going to make a real difference, we need to rethink our strategies"* (Almansa-Sánchez & Richardson 2014: 16)

Simpson (2009) suggests that the motivations for the boom in community archaeology projects include the desire to meet a range of educational and social values of increasing public awareness and understanding of the relevance of archaeology as well as the need to secure funding for research through demonstrating "impact" beyond academia. The author argues that appropriate criteria and methods for evaluating the effectiveness of community archaeology projects have yet to be designed, applied and validated. In general, this field of archaeology would suffer from short-term, project-centred funding limiting its ability to sustain the outcomes of individual projects.

There appears to be a large gap between what is expected from involving citizens in the research process and what is actually possible in such involvement. As Fairclough notes: *"Attempts to 'involve'* *or 'engage' the public, to encourage 'participation', to 'share the excitement', seem often to end with 'us' telling them what 'we' have discovered - or persuading them to act as archaeologists in an approved manner. All well and good, but is it enough?"* (Fairclough 2014: 5).

Maybe for most interested citizens it would be enough to understand what it means to be an archaeologist, engage in a local investigation, and learn about how archaeologists produce knowledge claims. Some might even sponsor an excavation on a crowd-funding platform like DigVentures[291], with a higher contribution for being part of the excavation team for some time.

But there are further, far-reaching expectations by some of the archaeological community. For instance, Marshall posited *"that the kind of collaborative research fostered by community archaeology will be crucial if archaeology is to have a future"* (Marshall 2002: 218). Ten years later Beale (2012) sees potentials in open data to enable community archaeology *"to achieve further its objectives"*. She notes that many community archaeology projects have found it difficult to engage the communities they claimed as stakeholders in the ways described in theory, e.g. non-hierarchical and "rooted" in the community. There have also been profound changes in the make-up of modern societies that necessitate reconsidering the concept of "community", which becomes all the more illusive in the online environment.

Still Beale expects that the Web could allow novel, more open forms of public engagement in archaeology. Cases of web-based community archaeology using open data are – understandably – absent in her paper. The examples given of engagement in "crowd sourcing" are not archaeological projects but Open Plaque[292] (online documentation of commemorative plaques) and the Old Weather[293] project that invites people to transcribe ship logbooks producing data for historical research and climate model projections.

There are not many examples of direct online engagement in archaeology, but some recent ones are:

- *Field expedition: Mongolia*[294]: identification of archaeological features (potential sites) on satellite images of Mongolia to help direct the survey team on the ground (Lin *et al*. 2014);
- *HeritageToGather*[295]: invites people to take and upload photographs of heritage artefacts and environments from which 3D models are produced;
- *ACCORD - Archaeology Community Co-Production of Research Data*[296]: similar to HeritageTo-Gather but with more on the ground work with volunteers;
- *Ancient Lives*[297]: transcription of Egyptian papyri (Williams *et al*. 2014); Ancient Lives is part of the Zooniverse[298] network of projects;
- *MicroPasts*[299]: transcription of museum object cards that document Bronze Age metal artefacts found mostly in Britain from the nineteenth century onwards, an archive of 30,000 cards housed in the British Museum; another MicroPasts project invites volunteers to "photo-mask" artefact images which allows to produce 3D models (Bonacchi *et al*. 2014);

[291] DigVentures - Archaeology in your hands, http://www.digventures.com

[292] Open Plaque (community-based project), http://openplaques.org

[293] Old Weather (UK National Maritime Museum), http://www.oldweather.org

[294] Field expedition: Mongolia, http://exploration.nationalgeographic.com/mongolia

[295] HeritageToGather, http://heritagetogether.org

[296] ACCORD (led by the Glasgow School of Art's Digital Design Studio), https://accordproject.wordpress.com

[297] Ancient Lives (University of Oxford), http://ancientlives.org

[298] Zooniverse ("platform for people-powered research"), https://www.zooniverse.org/projects/

[299] MicroPasts (University College London, Institute of Archaeology), http://micropasts.org

- *UrCrowdsource*[300]: transcription of documents (e.g. field notes, letters, reports) related to excavations of the ancient town of Ur in Mesopotamia (1922-1934); part of the project "Ur of the Chaldees: A Virtual Vision of Woolley's Excavations"[301]

Most of these examples present a one-way model of participation, although the MicroPasts platform has a broader scope of volunteer engagement and participation, including that people can propose and/or fund new collaborative research projects. Concerning the co-creation of 3D models of historic objects, places and monuments the ACCORD project is an example of good practice.

Many papers now address "crowd sourcing" and "open data" as ways to engage non-experts in archaeology (e.g. Bevan 2012; Bonacchi 2012; Morgan & Eve 2012; Richardson 2013a), but with few examples at hand the thoughtful discussions remains somewhat theoretical. The themes are likely here to stay for long and deserve to be explored further, hopefully based on many archaeological examples. This should, however, be conducted in a broader framework of online activities. "Crowd sourcing" is a rather directed form of engagement, and "open" in "open data" seems illusive concerning archaeological research data; it may be open access but without expert assistance it may not be useable for citizens.

The digital business consultancy Gartner included "citizen data science" in their 2015 "Hype Cycle for Emerging Technologies" report, with the expectation that it would take two to five years to become a productive field of business (Gartner 2015). As one business observer notes, *"This could turn out to be the most optimistic prediction in this year's report"* (Press 2015).

A overly data-focused approach to citizen engagement may also bind attention of archaeologists which may be needed more to critically engage elsewhere with the online social web. Perry & Beale (2015) think that many in the profession appreciate "Web 2.0" or social media mainly as dissemination channels rather than reflecting it as "*a paradigm-shifting system".* They consider this as a "*failure to critically engage with its dimensions as one of the most profound challenges confronting archaeology today"*.

The question may be about powers that direct, strengthen or disempower ideas, opinions and activities on the social web, including activities that concern archaeological matters. Beer (2013) cautions, *"The politics of circulation that underpins social media may serve to give the impression of democratisation and decentralisation, whilst actually working to obscure and silence some important visions of the social world. Some ideas will gain visibility, others will be lost".* Public archaeology needs to be aware of and address the politics of circulation (e.g. exclusion/silencing of alternative views), and critically reflect its own social media practices.

In any case the shift of engagement activities into the digital realm aggravates conceptual issues of public/community archaeology, in particular, the concept of a "heritage community" becomes illusive. Archaeology 2.0 may not be about the meaning of the remains of a community of the distant past for a present-day virtual community (or several thereof).

It may be argued that most European public/community archaeology on the Web 1.0 has been centred on connecting citizens with their (assumed) heritage and ensuring support for the heritage protection agenda. Many citizens seeking roots, identity and community have taken up the agenda and actively contributed to the preservation of the historic environment. Web 2.0 social media allow novel participatory approaches which, however, will not make a real difference if disconnected from critical social, environmental and other issues of today's society.

[300] UrCrowdsource, http://urcrowdsource.org

[301] UrOnline (University of Pennsylvania Museum and British Museum), http://www.ur-online.org/about/

Suggested actions

- *Conceive and engage in participatory approaches based on online platforms (e.g. social media) in a highly reflective way.*
- *Explore the concept and practicalities of "open research communities" that involve archaeologists and citizens in the production, dissemination and re-use of open data.*
- *Instead of seeking "roots", focus on relevance of archaeology for societal, environmental and other issues, at regional as well as global scale.*

8.5.12 Increase of the societal relevance of archaeology

It is felt widely that the research community has no strong voice in current debates. Gaining such a role is about demonstrating relevance for societal, environmental and other concerns. However, there seems to be a large gap between what matters to citizens and politics and what is going on in archaeological heritage management as well as mainstream academic research. The archaeological heritage management, and its adjunct contract archaeology, are mainly focused on the protective function of archaeology. And within the academic framework, archaeological research seldom attempts making knowledge count in the political sphere, as this is not part of the academic agenda or simply not reasonably possible. Still, if there are relevant insights for current issues, there is the question of how to make them heard in the political arena.

An excellent discussion of the matter is a recent, 5-page "opinion editorial" by Carsten Paludan-Müller (General Director of the Norwegian Institute for Cultural Heritage Research) in the newsletter of the European Association of Archaeologists (Paludan-Müller 2015). Paludan-Müller stresses that the community must invest more effort in demonstrating societal relevance. He notes a "weakened standing" of archaeology which should not be explained solely by the current economic crisis. As other, arguably deeper causes, he lists:

"Too much of the archaeology that is going on has been reduced to a bureaucratically legitimized exercise, where sites are excavated only because they are under threat from construction projects, and because excavations can then be prescribed by law. Too often – though not everywhere and not always – the end product is a stock of artefacts, samples, documentation and grey literature, left to oblivion in storage vaults and archives, because funding for the rest of the 'food chain', the production of new knowledge and giving it back to society, has been cut out (…). And too often, cultural heritage management has failed to work in more inclusive ways with local communities, and with sufficient awareness of issues beyond those strictly pertaining to the sphere of cultural heritage.

As ways to set the record straight and increase archaeology's significance Paludan-Müller suggests, *"to explicitly reconnect archaeology and cultural heritage to big issues occupying people in contemporary society: Climate, conflict, globalizations and origins – for instance"*, as well as *"reconsider the ways in which we articulate and practise our profession in contemporary society"*, and *"think hard about how heritage can become a relevant part of a new equation, instead of a reminder of better days and the set pieces in 'Heritage Theme Park Europe'."*

As signs and potential ways in which archaeology might regain relevance Paludan-Müller notes the current debate in the domain of history about how to the focus back to the "longue durée" (Guldi & Armitage 2014); that geo-politics has re-emerged as an important domain of knowledge (Paludan-Müller 2013b), and that "big data" based research could allow archaeology focus on long-term and global patterns of societal change, with regard to growth, sustainability of collapse of cities, for instance (Ortman *et al*. 2014).

Therefore Paludan-Müller thinks, *"There is no reason why archaeology should not weigh in from its privileged position for analysing big data to understand long-term developments over vast spaces, be it in the development of global power and economic structures and of empires, the shifting patterns of migration or the response of human society to changing climate"*. Finally Paludan-Müller asks for a *"determined push to reinsert our field as highly relevant and respected in society"*, activities that could best be undertaken under the umbrella of the European Association of Archaeologists (EAA), the association's Executive Board, supported by thematic Committees that engage the community *"in a creative dialogue about how we can support the EAA's real purpose"*.

There is little to add to this situation analysis and the plea for connecting archaeology stronger with current concerns and affairs, e.g. climate and environmental change, globalization and geo-politics, urban agglomeration, sustainability, conflicts and migration, and others.

One point to consider is the current economic crisis which since 2008 has also affected archaeology (Aitchison 2009a/b; Schlanger & Aitchison 2010). Organisations employing archaeologists have typically become smaller, the number of jobs shrinked, and jobs are more likely to be part-time and/or for shorter contractual periods (DISCO 2014). Also an impact on archaeologists at universities has been observed (Aitchison 2011), but not as severe as in the commercial sector. This hardly is a good situation for expanding archaeology's societal relevance through relevant existing and new knowledge.

Shallow "marketing" of archaeology or a call to "integrate" it better in society will not work. For example, Sakellariadi (2014) criticises a volume of papers from the International Conference on the Social Role, Possibilities and Perspectives of Classical Studies (Frankfurt, 2012), organised by the Archaeology in Contemporary Europe project (EU, Culture Programme, 2008-2012). Some out-standing papers aside, Sakellariadi missed reasons why archaeology should be more "integrated" in society, which was a common theme of most contributions, and how to address funding cuts. Instead, she notes, "*there is a lot of talk about promotion, marketing, and audience research, but nothing about the past and its remains as a common good that belongs to all and as such is investigated by archaeologists and other experts with the primary aim to open access to knowledge to all. Additionally, there is no mention of demonstrating the role of archaeology in society as a means to oppose the cuts"* (Sakellariadi 2014: 757).

The wider domain of cultural heritage research benefits from the many fields in which results can be claimed to be relevant, be it regional/national or European cultural identity, cultural tourism, or social inclusion. But translation of archaeological insights for the political arena and decisions on current issues is a difficult undertaking. Results of the social sciences have much more currency and promotion in this regard. As Smith (2015), an archaeologist who conducts comparative research on ancient cities, notes, *"Most commonly in the social sciences, this is through intermediaries: entrepreneurs, advocacy groups, think tanks, and media consultants. But these things just don't exist for archaeology"*. He suggests that archaeologists should relate stronger to social sciences topics and make relevant insights known to scientists who have a certain standing in political fora on societal topics.

There is also the question on what empirical basis archaeological knowledge would be accepted as relevant contributions to addressing societal issues. Barton (2013) posits that inductively constructed archaeological narratives dominate the archaeological discourse, and that such narratives would not count much in current debates: *"if insights about the human past are to play a meaningful role in understanding human society today and inform policy decisions for the future, narrative fictions are insufficient for such contributions, no matter how plausible"* (Barton 2013: 156). He suggests a stronger computational modelling approach as a way for archaeology to move forward and re-inforce its standing. It seems unlikely that a large segment of the archaeological research community shares this view. But the question of how to raise archaeology's relevance and standing in essential debates may be one of the crucial questions of the domain in the coming years.

Suggested actions

- *Leverage archaeology's societal role and relevance, for example through connecting archaeological research and knowledge with current concerns and affairs such as climate change, environmental sustainability, urban agglomeration, globalization and geo-politics, regional conflicts, migration, and others.*
- *Consider contributions which could allow archaeology a stronger voice in current debates.*

8.6 10-year innovation agenda road mapping

The 5-year horizon of the ARIADNE Innovation Agenda and Action Plan centres on immediate innovation needs concerning open research data, digital archives, data infrastructure and services. These needs are rather clear, however, there are of course many intricate issues (e.g. in data sharing and re-use) which need to be addressed thoroughly.

The extension of the innovation horizon to 10 years and change of focus to future digital, ICT-enabled archaeological research and communication necessitates using appropriate methods that enable the road mapping for potential innovations. The work on the 10-year horizon so far has been conducted mainly by scanning current topics and discussion. This is a first step in developing a roadmap towards innovative digital, ICT-enabled archaeology, but use of additional methods will be required to shape and evaluate the roadmap, and suggest actions stakeholders could take to bring about the envisioned innovations.

The focus on innovative digital, ICT-enabled archaeology is motivated by ARIADNE's remit which is to provide an e-infrastructure and services that allows the research community (and other users) share, discover, access and (re-)use available data. With such an e-infrastructure in place, archaeological research and communication can be enhanced and new, innovative approaches to digital archaeology explored and developed. These are vital goals for the research community, however embedded in perspectives of archaeology which may change or, as some would contend, should change to keep archaeology relevant in the long term.

8.6.1 General considerations

The roadmap towards innovative archaeological practices, in digital environments and elsewhere, is not primarily about new technologies for archaeologists and other researchers in the heritage sciences. Existing and new technologies are tools for doing and communicating research, but purposes and desirable outcomes need to be considered first. Questions of relevance are of vital importance in this regard, for example, what role archaeology can and should play in the society, in Europe and beyond.

Therefore it will be crucial to question and expand or change established frames of reference, instead of thinking about incremental improvements of existing practices and outcomes. However it could mean reinforcing what have been vital purposes of archaeology in the pursuit of knowledge about humankind, but lost in academic specialisms and the next "least publishable unit" or feature of a software application.

Instead of a loss of agency, the endeavour should be driven by vital questions, for example: What would increase the relevance of archaeology in our societies and cultures, offering knowledge as well as inspiration and critical perspectives? What would be desirable future contributions of archaeological research to tackling societal, environmental and other challenges? What kind of research – digital, ICT-enabled or not – might enable such contributions? What is required to allow for such research, e.g. in terms of researchers' perspectives, knowledge, methods, technology, etc.

If desirable future contributions of archaeological research, and the required means and capacities are identified, possible targets for realising them can be defined, at least generically. For the targets then potential pathways with chains of required and interdependent achievements can be elaborated and targeted.

Such a roadmap is not about some purely analytical studies of possible futures without connection to possible actions. It should be action-oriented, aimed at helping stakeholders – the archaeological research community, policy makers, research funders and others – to actively bring about and shape the desirable future. Thereby it can provide orientation and impetus to move in directions which are preferable than the current state-of-affairs in the sector.

8.6.2 Road mapping methods

An innovation agenda for innovative and potentially transformative digital, ICT-enabled archaeological research and communication practices in a 10-year horizon is of course explorative. This does not mean "in the wild", rather certain methods will be used for this road mapping for innovation. Below we present relevant methods, but also available others may be used or new ones developed, if required.

Scanning the innovation landscape

Some scanning of various sources will be necessary to acquire a good understanding of the current landscape of innovation, specifically with regard to ICT. Archaeological research already uses ICTs intensively, in many research fields specific technologies are essential, i.e. the research could not be done without them. Examples where technologies have opened up new lines and ways of carrying out research will be noted as well as perceived current drivers and inhibitors of innovative practices in archaeology in general as well as in selected fields.

The scanning will aim to identify what is considered as innovative and may already have shown some impact (e.g. examples of changes in/through digital practices). Furthermore, where are significant changes in archaeological digital practices expected, and what kind of changes – e.g. more effective work, at a larger scale, in a very different way? Also weak signals or pockets of expected future practices in the present may be identified. Such expectations and signals may be found in areas where innovators see advantages of collaboration across disciplinary boarders which is often accompanied by adopting or jointly developing new theoretical frameworks, methods and tools.

Some relevant sources could be special issues of journals, review articles, domain reports and surveys, conference sessions, expert workshops, online forums, weblogs, etc. with a focus on recent innovative developments. However, also seemingly old material can allow for valuable orientation, at least with regard to identifying changes in perspectives or if there has been some progress in what experts' expected or suggested. One example is the archaeological e-science scoping work conducted in 2006 which suggested "way-finding" initiatives to enable the community adopt and apply e-science methods and tools according to their objectives and requirements (cf. *Section 8.5.3*). It appears that such ways have not been found (or sufficiently sought) as yet for archaeological research.

Moreover, *"The future is already here – it's just not very evenly distributed"* (William Gibson[302]). Hence it is also important to look what is going on in other disciplines. What have been drivers of advances/changes in digital practices there? Might they also work in archaeology?

[302] Wikiquote: William Gibson: https://en.wikiquote.org/wiki/William_Gibson

Examples of current advances towards innovative e-archaeology

Following a "birds-eye view" of the innovation landscape, some outstanding examples of current e-archaeology could become the subject of case studies. Such cases represent "innovators" areas according to the classical Diffusion of Innovations curve (Rogers 1962). Here it is particularly relevant to understand what advantages researchers expect from still prototypic tools, how immature and incomplete prototypes are evolved and maybe taken up by others.

The work would be done in an ethnographic spirit, looking deeper into actual e-archaeology, at least based on the richest available description of what archaeologists actually do digitally. Some material like documented workshops, online expert forums or weblogs might also allow identifying what researchers perceive as necessary to advance, required capacity, barriers to innovation, and how they might be overcome.

The particular areas of ICT-supported archaeological research considered need to be selected carefully. The set of selection criteria could include areas that are perceived of vital importance for archaeological research, in terms both of advancing knowledge and archaeology's relevance in society. Relevant areas may be societal, environmental and other challenges where archaeologists could contribute their long-term, "deep history" perspective of social systems and changes in human behaviour, adaptation to environmental and climate changes, for instance. Examples of efforts to build, combine and analyse large datasets for addressing the research questions in a cross-disciplinary approach may be of particular interest.

Another example could be the role of (future) "Web 2.0" approaches and tools in public and participatory archaeology. The focus would not be primarily on the technologies but questions such as: who are the participants?, what are their interests?, how do they relate to archaeology?, and what actually is at stake in such participation? A study of these questions could be conducted online in a participant-observational form of "netnography" (Kozinets 2009). Insights might inform novel concepts, practices and tools.

Scenario building

The scenarios of future e-archaeology could be "chain scenarios". Such scenarios depict chains of events along a timeline (5 years, 10 years, beyond) that would lead to a desired future situation. Necessary events on the innovation path could be novel ways of organizing and doing research, tools with certain capabilities, new types of data resources, etc.

The chain scenarios thus require envisioning of future e-archaeology, an understanding of what might be needed, ideas about how it could be realised. The innovation path would start from the present situation, as identified in the landscape scanning and case studies. With regard to the chaining it is important to consider portfolios of interdependent tools/services that are necessary and need to proceed together in order to allow for the envisioned future situation.

Such portfolios might be framed as Virtual Research Environments (VREs) for different fields of archaeological research. The VREs would build on common underlying e-infrastructures and include some specific tools and services as required for specific domains. VREs should not be seen as monolithic systems but range from loosely coupled tools and services to tightly integrated environments for specific research communities. VREs may centre on one archaeological site, but overall we would expect more from VREs for integrative research, comparative analysis and broad synthesis. A higher integration of vocabularies (terminology, ontologies) and data resources across disciplinary boundaries will be instrumental in this regard.

Scenarios of VREs for future e-archaeology should have a solid grounding in what represents the current state-of-the-art of VREs. This state is given by VREs of other domains, which can be covered by case studies[303]. Case studies could inform concepts of future environments for archaeological e-research, which may share common features, but very likely be different in other respects.

Scenarios of future e-archaeology would generally be informed by the double lens of emerging "grand challenges" of archaeological research, and required capabilities of digital infrastructures, tools and services for tackling such challenges. Special attention needs to be devoted to the necessary data resources as no scenario would hold if they are not considered thoroughly.

The scenarios should be action-oriented, provide specific suggestions for the different stakeholders including, but not limited to, the archaeological research community, policy makers, research funders, e-infrastructure and software developers, and data providers.

Technology road mapping

Innovative ICT-enabled research practices for addressing "grand challenges" may need the application of new technologies which do not exist already or are available but not readily available for archaeological purposes due to difficulties which need to be overcome. There could be lack of necessary functionality, bottlenecks concerning the capability of some existing technologies, or the required type of technology does not even exist as research prototypes. Furthermore there could be a lack of necessary knowledge and tools to create envisioned new research environments.

Required new types of technologies are typically addressed by elaborating a technology roadmap. Such a roadmap is difficult enough to produce for most sectors, but a roadmap towards transformative research in the multi-disciplinary field of archaeology may necessitate using (or first developing) additional and complementary methods.

We might think of roadmaps or scenarios concerning multi-domain semantic integration of research data and services, multi-dimensional modelling and computing of societal, environmental and other changes in the past, or environments for integrative interdisciplinary research. If required advances are identified, the road mapping can serve to define and set targets. For the targets then potential pathways with chains of required and interdependent achievements may be elaborated and aimed at.

The ARIADNE community as expert panel

The ARIADNE project involves leading centres of archaeological research, data archives, technology developers and others who participate in the framing and elaboration of the Research Agenda and Action Plan. The community also is active in Special Interest Groups (SIGs, ARIADNE Task 2.2) that survey the state of the art of research in their field of interest and will provide suggestions for the innovation agenda and action plan. The SIGs have been established by project partners and involve experts in their professional network of research and other work. Furthermore the ARIADNE summer schools provide the opportunity to involve a younger generation of scholars in the road mapping. For example, the ARIADNE Expert Forum on Digital Futures of Archaeological Practice 2020-2025, 2-3 July 2015 in Athens[304], which was part of a summer school, brought together 22 young and senior researchers interested in the future of ICT-enabled archaeological research.

[303] Examples are the VREs built by the projects myExperiment (sharing of biological research workflows), http://www.myexperiment.org; ViBRANT (e-taxonomy/biodiversity), http://vbrant.eu; BioVel (ecological modelling), https://www.biovel.eu; there are also some examples in the humanities (Babeu 2011).

[304] ARIADNE Expert Forum: Digital Futures of Archaeological Practice 2020-2025 (Athens, 2-3 July 2015), http://summerschool.dcu.gr/?page_id=19#expert-forum

8.7 Summary of topics and suggested actions

This section summarises the discussion of the selected roadmap topics and suggested actions. In the further work on the 10-year horizon topics may appear as less relevant and others added.

8.7.1 An explorative undertaking

In line with ARIADNE's focus on data infrastructure and services for the archaeological research community, the work on the 10-year horizon of the innovation and action plan focuses on innovative digital, ICT-enabled archaeological practices. Emerging new perspectives, innovation potentials and development paths towards "e-archaeology" in 2025 are being explored, and possible actions for enabling envisioned innovative practices, methods and tools considered.

In comparison to the rather clear objectives and requirements of the 5-year horizon, the 10-year horizon is more of an open field. It must count on considerable achievements in the 5-year horizon, with regard to open and re-usable data, for instance. What archaeological researchers, tool developers and others will do based on the achievements is much less clear.

The drive towards innovative e-research in archaeology will come less from research policies (as in the case of open data), rather visionary researchers, data scientists, tool developers and others will pave the way towards further achievements. However the ARIADNE innovation roadmap for the 10-year horizon may provide some orientation, point towards interesting avenues and, thereby, help in bringing about and shaping desirable futures.

As a first step for developing a roadmap towards innovative "e-archaeology" current future-oriented topics and discussion have been scanned, described, and some tentative suggestions added. The next step will be scenarios of innovative progress, involving consideration of required approaches and means (e.g. tools, data and others). Scenarios may also be provided by members of the ARIADNE expert community (e.g. the ARIADNE Archaeological Research Practices and Methods SIG). Based on an evaluation of the scenarios, actions will then be suggested on how the envisioned future digital, ICT-enabled practices and required means could be brought about.

As the work on the 10-year horizon will be conducted until November 2016, it is foreseen interim results (e.g. scenarios) will be issued in the form of working documents, inviting suggestions and contributions by the wider ARIADNE stakeholder community.

8.7.2 Suggested actions

The suggested actions below concern the development of scenarios of innovative progress in digital, ICT-enabled archaeology (e-archaeology) in the 10-year horizon of the ARIADNE innovation roadmap. They are, however, phrased as actions stakeholders could already take today.

Address the question of archaeology's societal relevance

The 10-year road mapping focuses on innovations in archaeological practices of research and communication. In this context questions of relevance are crucial. Such questions concern the role archaeology can and should play in society, with regard to societal and environmental challenges, for instance. What would be desirable future contributions of archaeological research to tackling such challenges? What kind of research – digital, ICT-enabled or not – might enable such contributions? What is required to allow for such research, for example in terms of researchers' perspectives, knowledge, methods, technologies and data?

Concepts of digital science will be futile in archaeology (and elsewhere) if not connected to vital current debates. It is felt widely that the archaeological research community has no strong role and voice in such debates. Gaining such a role is about demonstrating relevance for societal, environmental and other concerns. This may leverage citizens' appreciation, understanding and support of archaeology.

Suggested actions

- *Leverage archaeology's societal role and relevance, for example through connecting archaeological research and knowledge with current concerns and affairs such as climate change, environmental sustainability, urban agglomeration, globalization and geo-politics, regional conflicts, migration, and others.*
- *Consider contributions which could allow archaeology a stronger voice in current debates.*

Take account of the diversity of archaeological research practices and methods

Archaeology is a multi-disciplinary field of research in which researchers in various domains address their research questions with different theories, methods, data and tools, e.g. classical studies versus environmental archaeology, for instance. The diversity of "schools" and research practices in archaeology should be considered when trying to identify and conceive pathways towards innovative digital, ICT-enabled research that makes a difference. One common feature is the research lifecycle from project idea to publication of results (incl. open sharing of the generated data). Also issues of standardisation (e.g. data models, terminology) and cost-effectiveness will remain common concerns.

- *Recognise that different archaeological schools of thought and research practices require different digital, ICT-based research environments and tools.*
- *Focus on phases in the lifecycle of archaeological research in which significant progress in knowledge may be achieved. In the last decades data generation has seen enormous progress; in the future other phases may require more attention.*
- *Recognise that issues of standardisation (e.g. data models) and cost-effectiveness are relevant for future research practices.*

Target data integration for comparative and synthetic research

Open and re-usable data will allow easier combination of data for comparative and synthetic, cross-disciplinary research. Enlarged and integrated datasets may allow new insights through applying models of social systems and behaviours, adaptation to climate and environmental change, for instance. In recent years archaeological research has taken aboard an arsenal of data capture methods and produced growing volumes of field survey and excavation data for documenting individual sites and areas. On the other hand, data integration for comparative and synthetic research is lagging behind ever more.

There is a need of novel approaches and tools that allow researchers bringing together and working effectively with the variety of data required for cross-domain, interdisciplinary research, in fields such as historical ecology or urban archaeology, for instance. One solution that could drive progress in large-scale data integration could be setting up programmes and competence centres that support this task, for example, through facilitating collaboration between researchers with domain-specific knowledge and data science expertise.

Suggested actions

- *Foster the development of novel methods and tools that allow researchers to bring together and work with the variety of data required for cross-domain, interdisciplinary research.*

- *Promote competence centres and programmes aimed at data integration for comparative and synthetic archaeological research.*

Explore relevant future VREs for archaeological research

The e-research scenario of virtual research environments (VREs) build on top of e-infrastructure and underlying data repositories has not yet reached the archaeological research community. The ARIADNE data infrastructure and services may inspire VRE developers to create environments for archaeological researchers. Such environments can range from loosely coupled tools and services to tightly integrated virtual workbenches for specific research communities. Relevant VREs for archaeological research should be explored, taking account of the state-of-the-art in other disciplines and particular requirements of archaeological researchers in different domains as well as in cross-domain collaboration.

Suggested actions

- *Look into VREs developed for other domains to conceive environments relevant for e-research in specific archaeological domains as well as in cross-domain collaboration.*
- *Consider cases where researchers use data mediated by ARIADNE as well as data infrastructures and services of other disciplines (e.g. geo, environmental, biological data).*

Identify e-science practices based on data infrastructure and computing facilities

E-science has become an ever broader topic that now spans all forms of research activities conducted online, even involving “citizen scientists”. The classical, still vital variant of e-science is use of advanced and distributed computing by researchers in the natural sciences and other disciplines. Research in the heritage sciences, i.e. archaeology, cultural heritage and other humanities research seldom employs such computing. Available Grid/Cloud-based Distributed Computing Infrastructures did not find much use by researchers in the heritage sciences so far. The ARIADNE data infrastructure and services may promote e-science activities. In future scenarios this could involve distributed computing based on a seamless flow of data to computing infrastructures and *vice versa*.

There are expectations that mining of “big data” can allow also archaeologists novel insights, e.g. relevant patterns in data that suggest new research questions. However, many aggregated and integrated large archaeological datasets may not be available for quite some time. But archaeological researchers may increasingly use Cloud-based research support services for other purposes than computing, data transfer, temporary storage and access during research projects, for instance.

Suggested actions

- *Promote collaborative way-finding for e-science approaches, methods and tools relevant to archaeological researchers.*
- *Focus on e-science needs specific to archaeological research, which may differ from those of other humanities as well as natural sciences.*
- *Look for uses of low-level Grid/Cloud based services and emerging examples of archaeological applications of “big data” mining and other methods.*

Propose grand challenges for the digital archaeology community

Experts of the archaeological informatics and computing community suggest seeking “grand challenges” which are research-directed and contribute to the development of new theories and methods. Thereby the community could play a stimulating and transformative role rather than simply support well-established research practices. Grand challenges for digital archaeology should go beyond what seems feasible in the short to medium term through applied research and engineering. The challenges would inspire the research community to push boundaries and explore new avenues of research with potential revolutionary impact, i.e. a shift in established paradigms, theories and methods of archaeological research. Meanwhile also grand challenges for archaeology have been identified, most of which require large-scale and integrated datasets. It appears that the sought for advances require bringing together domain experts (e.g. theories, methods, data) and developers (software, computing) to create novel research tools, and a strong focus on cross-domain, interdisciplinary research.

Suggested actions

- *Seek grand challenges that inspire the research community to push the boundaries of digital archaeology.*
- *Suggest challenges that promote mobilisation and integration of datasets for domain and cross-domain, interdisciplinary research.*
- *Bring together domain experts and developers to create methods and tools for such research.*

Evolve a Web of archaeological Linked Open Data for research

The last 10 years have seen much progress in Linked Open Data (LOD) know-how required to produce, publish and interlink LOD of archaeological and cultural heritage collections/databases. In practice, however, not many LOD datasets have been produced and interlinked so far. As a multi-disciplinary field of research archaeology could benefit greatly from a web of LOD that spans vocabularies and data of the humanities, heritage sciences and other disciplines. This would allow discovery and retrieval of semantically related data and knowledge of different domains of research within and beyond archaeology.

The ARIADNE data catalogue will become available as LOD, which may promote some further LOD publication and interlinking. However, a much wider uptake of the LOD approach for semantic interoperability in the archaeological and other domains is necessary so that a rich web of LOD can evolve. Two core requirements must be met: effective interlinking of LOD requires use of common or mapped vocabularies (thesauri, ontologies), and the LOD resources need to be curated to ensure reliable interlinking and access. A central role in the LOD scenario plays the CIDOC Conceptual Reference Model, which recently has been extended for scientific observations and argumentation, and domain-specific modelling (e.g. archaeological excavations). It is expected that mapping of archaeological data collections/databases to the extended CIDOC-CRM will enable enhanced search as well as research-focused applications.

Suggested actions

- *Promote publication of LOD datasets (collections, databases) by more archaeological and other cultural heritage institutions, especially based on mappings to the extended CIDOC-CRM.*
- *Foster a community of LOD curators who ensure reliable availability and interlinking of LOD resources (datasets and vocabularies).*
- *Develop LOD-based applications that demonstrate advances in research capability, which may motivate a wider adoption of the LOD approach by research institutions and projects.*

Promote new forms of scientific/scholarly publication

The next 10 years will very likely see some advances with regard to new forms and ways archaeo-logical researchers could publish project outcomes. However, publications are tied to the scholarly review and credit system. Without some adjustments in this system new forms of online publication may find it difficult to progress from prototypes to wide adoption by publishers and authors. Uptake will be higher for moderately enhanced familiar forms (e.g. embedding explorable objects such as 3D models in online papers). Advanced forms will be driven by a process-view of research rather than static documents, for example, figures that auto-update as new data becomes available. This will require deep interlinking of publications and datasets. Also the distinction between journal and repository may become increasingly blurred if digital repositories become publishing platforms and value-added services for scholarly communication.

Archaeologists are aware that the finally "published excavation" in book form is not the optimal solution. Novel repository-based forms of publication may allow significant steps towards media/data-rich reports of investigations. We may also envision the "digital record" of an on-going excavation as a stream of data from the field and laboratories, continuously made available, analysed and discussed by subject experts. In such a setting "publications" would be snapshots of the state of knowledge at a certain time, instead of the annual excavation report and some papers by researchers in need of taking care for their academic record.

Suggested actions

- *Promote novel forms of digital publication that could "work" for archaeological projects in terms of enhanced access to research outcomes as well as academic credit.*
- *Start with moderately enriched familiar ways of publication (e.g. embedding explorable digital objects in online papers), and make new approaches as easy as possible.*
- *Investigate fields of "data-driven" archaeological research and publication in which accessible datasets and executables (software, dynamic figures, etc.) could play an essential role.*
- *Explore repositories as platforms for media/data-rich archaeological publications and value-added services for scholarly communication.*

Foster participatory and reflective online public/community archaeology

Many community/public archaeology projects have found it difficult to engage the communities they claimed as stakeholders in the ways described in theory, i.e. non-hierarchical, participatory or, even, "rooted" in the community. There appears to be a large gap between what is expected from involving citizens in the archaeological research process and what is actually possible in such involvement. The shift of engagement activities into the digital realm aggravates conceptual issues of public/community archaeology, for example, the (sociological) concept of community becomes illusive. Online environments specifically built for "crowd sourcing" contributions by non-experts often present a one-way participation approach. Concerning expectations from open research data, archaeological documentation and data is not something many non-experts may easily understand and use for own research work ("citizen science"). Mediation by archaeologists might again reproduce hierarchical and expert-directed involvement. Archaeological institutions and projects already use social media to increase visibility and disseminate information. Conceiving novel, participatory approaches based on such media will require a highly reflective usage.

Suggested actions

- *Conceive and engage in participatory approaches based on online platforms (e.g. social media) in a highly reflective way.*
- *Explore the concept and practicalities of "open research communities" that involve archaeologists and citizens in the production, dissemination and re-use of open data.*
- *Instead of seeking "roots", focus on relevance of archaeology for societal, environmental and other issues, at regional as well as global scale.*

9 Annex: Background studies

The Annex includes three background studies:

Section 1 – Landscape of digital resources for humanities and archaeological research

- This study looked into the current landscape of digital resources for archaeological research within the wider area of humanities research. The results confirm the perception of a high fragmentation and inaccessibility of relevant data resources, as expressed by many researchers who participated in the ARIADNE online survey on expectations from project's e-infrastructure and services.
- The study has been conducted to gain a better understanding of the situation of different resources, which may help to conceive and put into place a more targeted approach of resource development and access.
- A summary of the results and suggested actions are given in *Section 3.4*.

Section 2 – Scholarly publication culture, citations and altmetrics

- (Re-)use and citation of research results by other scholars are core elements in the system of scientific/scholarly recognition and reward. In the quest for open data sharing through digital archives the system needs considerable adjustment to accommodate data.
- The background study presents information and empirical results on citation-related topics and issues, for example, the citation advantage for open access publications (papers and data), altmetrics (alternative metrics), and the importance of usage data for data archives. Special attention has also been devoted to specificities of the humanities publication culture, which should be taken account of in the development of comparative metrics of data sharing, re-use and citation.
- A summary of the results and suggested actions are given in *Section 4.4.5*.

Section 3 - ARIADNE data portal – Results of the lead user survey

- To support the development of the ARIADNE data portal a survey of various archaeological websites (giving access to content/data of more than one institution or project) and some existing data portals of other domains has been conducted. The survey participants looked for good practices and gave recommendations for services of the ARIADNE data portal. The results are reported in the ARIADNE Second Report on Users' Needs (D2.2, February 2015).
- The report in this chapter presents the method and results of the evaluation in April 2015 of the 34 survey recommendations by 28 experts of 21 partners. The evaluation has been carried out to focus the development of the services of the ARIADNE data portal on the most relevant services in the short to medium term.
- The evaluation results are taken account of in the focus area "Providing e-infrastructure services", *Chapter 5*.

9.1 Landscape of digital resources for humanities and archaeological research

To support the elaboration of the ARIADNE Innovation Agenda and Action Plan, we looked in greater detail into the current landscape of digital resources and infrastructures for humanities and archaeological research. We distinguish between[305]:

- *Digital archives (repositories):* Offer deposit, long-term curation and access services for the research and other user communities;
- *Digitised or born-digital collections:* Various cultural and scientific reference and other collections that provide access to study objects (e.g. digital surrogates of museum specimen), including description and other information;
- *Common research e-infrastructure and services:* Allow search, access and re-use of information across several archives and collections;
- *Virtual research environments (VREs):* Are built on top of the entities above and enable online research work according to the needs of specific research communities.

The ARIADNE online survey presented in the First Report on Users' Needs (ARIADNE D2.1, April 2014) revealed a high dissatisfaction of researchers with the difficulty of finding and accessing relevant digital resources. The difficulty is due to the fact that the resources are highly dispersed and that many are not yet openly accessible, or not available in digital form. For the innovation agenda we surveyed archaeological and related humanities collections, repositories, e-infrastructures and services and found that they are indeed not easy to find (e.g. not many are documented in the existing online registries and directories) and that many existing content/data resources are difficult to access or not accessible at all. A summary of the results and suggested actions are given *Section 3.4*.

9.1.1 Archaeological open access repositories

Distinctions and general situation

There is an extensive literature on digital repositories, including various studies and surveys (e.g. Adamick & Reznik-Zellen 2010a/b; Armbruster & Romary 2010; Björk 2013; Burns *et al*. 2013; COAR 2013; e-SciDR 2008; Haak-Marcial & Hemminger 2010; SURF 2009). The literature addresses different types, setups and various challenges repositories are facing, mobilisation of content/data deposits and sustainability, for instance.

The basic distinctions are between document repositories and data repositories and if they are "institutional", "subject-based" or "governmental". Governmental repositories provide access to public sector information, which can also include information and data for research purposes. Subject-based repositories contain material of one scientific discipline or sub-domain. Furthermore there are general, multi-subject repositories which allow deposit of material from all or several disciplines. The largest category is institutional repositories which hold material of only one

[305] In this survey we do not cover Current Research Information Systems (CRIS) of universities and research centres, national and other funding agencies. CRIS support the management of information about human resources (researchers and others), facilities (e.g. laboratories, equipment, etc.), research, education and knowledge transfer activities, project grants, etc. CRIS typically are closed systems, accessible only by the information managers and other authorised persons. CRIS of funding agencies however may allow searching more or less detailed information about research grants and outcomes. The core European organisation in this field is EuroCRIS (http://www.eurocris.org). EuroCRIS maintains the Common European Research Information Format (CERIF) which is the official European standard in this field and employed by major research funders and research institutions.

institution deposited by affiliated researchers. These repositories are mostly university-based and therefore multi-subject repositories *en miniature*.

Studies on document repositories of universities found a largely disengaged faculty, unwilling to self-archive papers and other scholarly output (e.g. Troll-Covey 2011; Wacha & Wisner 2011; Xia 2007; Xia *et al*. 2012). In the Taylor & Francis international online survey on open access publications in 2014 one question was about where the researchers had deposited their last article (6,888 responses): 52% did not deposit the article (or give permission for someone to do this); others made it available through one or more of the following channels: 23% institutional repository, 23% personal / departmental website, 12% subject-based repository, and 8% Data repository (Taylor & Francis 2014: 13). Personal motivations to do so were much stronger drivers than an institutional or funder requirement to deposit the article.

The Supporting Research project (2010-2011) identified a high dissatisfaction of researchers with institutional repositories. The scholars perceived them as more in the service of the institution rather than research, offering no functionality for e-research and little incentives for sharing material (MacColl & Jubb 2011: 3-4). To be fair, however, the difficult situation of university-based repositories should be taken account of. They have to cater to all subjects and deal with very different kinds of content that derive from the academic activities, including course material, lectures, thesis & dissertations, publications, etc. in different areas of research and teaching. No wonder therefore that providing special services for research purposes or consider needs of individual researchers is not on top of their priorities.

Consequently many researchers turn to the major professional platforms Academia.edu, Mendeley and ResearchGate[306] that allow them to set up a personal research profile, share and promote own research results as well as network with other researchers (Van Noorden 2014; Thelwall & Kousha 2014a/b). In the Webometrics world rank of top online repository portals, ResearchGate holds position one and Academia.edu position three, while Mendeley ranks at position 61.[307] The largest attractor of people who are active or interested in fields of archaeological research is Academia.edu; on 19 September 2015 the platform had 321,107 such users of whom many were interested in specific fields like Funerary Archaeology (13,521) or Bronze Age Europe (7878). In total 53,217 documents were referenced and a lot could also be downloaded (the exact figure could not be identified). An analysis of the three professional platforms is included in the ARIADNE D2.2 Second Report of Users' Needs.

OpenDOAR - Directory of Open Access Repositories

The best available source for a "birds-eye view" on existing repositories is the Directory of Open Access Repositories (OpenDOAR, www.opendoar.org). OpenDOAR does not require that a repository makes all content available without access control, but some should be.

On 16 September 2015, OpenDOAR documented 2973 repositories around the world (up from 2603 in March 2014). Most of the repositories are institutional repositories (82.4%), followed by disciplinary or subject-based repositories (9.8%), while governmental and other repositories have only a small share (6.4%). The bulk of the content are documents (journal articles, theses, book chapters, conference papers, learning material, grey literature, etc.), but 23% of the repositories include also multimedia and audio-visual material. Only 146 repositories (4.9%) also contain some datasets.

The results for archaeological content/data in the repositories documented by OpenDOAR are meagre. 237 of the 2973 repositories (8%) have "History and Archaeology" among their subjects; 120

[306] Academia.edu, https://www.academia.edu; Mendeley, http://www.mendeley.com; ResearchGate, http://www.researchgate.net

[307] Webometrics: Top Portals, http://repositories.webometrics.info/en/top_portals

(50.6%) of these repositories are located in Europe. The pattern of available content across the 237 repositories is somewhat different than in the total, e.g. multimedia and audio-visual materials are present in 54% of these repositories. But again only few (9) report also holding datasets; a clear case concerning archaeological datasets is the Archaeology Data Service (UK).

In most repositories “History and Archaeology” is but one among several subjects. Repositories with only this subject mainly hold historical content, e.g. Archive Server DEPOSIT.D-NB.DE (German National Library), Biblioteca Virtual de Aragón (Spain) or Bolivarium (Venezuela). The few repositories with archaeology, classics or other relevant material are: Archaeology Data Service (UK), Digital Karnak (University of California), Gothic Past (Trinity College, Ireland), English Heritage ViewFinder (UK), Propylaeum-DOK (Heidelberg University Library), and the document repository of the Norwegian Directorate for Cultural Heritage (Riksantikvarens vitenarkiv). Furthermore there are the Acropolis Educational Resources Repository (Acropolis restoration service), the Parthenon Frieze Repository (Greece), and DIGIMOM (Maison de l’Orient et de la Méditerranée, France), but their collections are rather small (below 300 items).

The Long-tail of Data Interest Group of the Research Data Alliance[308] conducted a survey on current practices for discovery of research data in “long tail repositories”, i.e. repositories holding mainly small-scale data deposits. The survey was run in February/March 2014 and received 60 responses from repositories located in different countries, 30 of the responses were complete (Shearer 2014). The survey focused on the use of metadata and other methods for enhancing discovery of the “long tail” content/data. 16 repositories employed Dublin Core, seven together with other standards. Among the other standards used by those and other repositories were DataCite, Darwin Core, DDI (Data Documentation Initiative), geo/wgs84 (World Geodetic System 1984), ISO19115 (Geographic Information Metadata), MARC21 (bibliographic), MODS (bibliographic), RIF-CS (Registry Interchange Format - Collections and Services).

Most respondents indicated that the metadata was sufficient for users to find datasets within the repository, e.g. if it is a small-scale repository or users are looking for the data supporting a specific study. However several thought that they would need to implement additional mechanisms to improve data discovery, through external search services or data federation, for instance. Need to provide richer, domain-specific metadata was mentioned by some respondents. Some suggestions made in the context of the survey are: Use DOIs, link publications to enrich dataset metadata, provide landing pages that describe datasets, incorporate different disciplinary metadata schemas in “long tail” repository platforms.

ROAR - Registry of Open Access Repositories

ROAR (http://roar.eprints.org) allows for searching repositories similar to OpenDOAR, but offers more specific search categories like “Archaeology”, “History of Civilization”, “Ancient History”, “The Greco-Roman World" and “Medieval History”, but on 16 September 2015 they returned only 15 results. These included two collections already mentioned above, the Acropolis Educational Resources and the Parthenon Frieze Repository. Further possibly relevant repositories include Human Origins (University of Southampton, UK), the Encyclopedia of Iranian Architectural History (IranShahrPedia), the Faculty Scholarship at the Claremont Colleges (California), and the Historical Philological Journal of the National Academy of Sciences of Armenia. The other search results were repositories holding academic and learning material on various topics and were located in El Salvador, Indonesia, Ireland, Peru, Saudi Arabia, Turkey and Ukraine.

[308] RDA Long-tail of Data Interest Group (accepted as an RDA Interest Group in Summer 2013; over 90 members from around the world), https://rd-alliance.org/groups/long-tail-research-data-ig.html

Re3data.org registry of research data repositories

The re3data.org registry (www.re3data.org) indexes over 1300 research data repositories from all over the world making it the largest online catalogue of such repositories on the web (it includes the former DataBib list of repositories). Re3data.org is funded by the German Research Foundation (DFG) and the Institute of Museum and Library Services (IMLS) in the United States.

The main category for archaeological data repositories is “Ancient Cultures” which includes the sub-categories Prehistory, Ancient history, Classical archaeology, Egyptology and Ancient Near Eastern Studies. Searching “Ancient Cultures” however delivered only 14 results (17 September 2015). The repositories covered by re3data.org are:

- The Archaeology Data Service (UK), ARACHNE IDAI.objects (Germany) and DANS-EASY / eDNA (NL) in Europe (all three partners of ARIADNE) and the US-based repositories Open Context and tDAR - The Digital Archaeological Record.
- The Data Catalogue Service of the Natural Environment Research Council (NERC, UK), which holds data records of geosciences, geology, palaeontology, geophysics, geodesy and geography, some of which are relevant for ancient cultures and historical studies.
- OAGR - Online Ancient Genome Repository (Australia): captures and catalogues open ancient human genome and microbiome data, including raw sequence and processed data, along with metadata about its provenance and production.
- The Wolfenbütteler Digital Library (Germany), containing sources (manuscripts, incunabula, imprints and digital editions) relevant for studies of the medieval and early modern period. Three other entries from Germany are databases and services for various humanities content: BABS - Bibliothekarisches Archivierungs- und Bereitstellungssystem; Prometheus - Das verteilte digitale Bildarchiv für Forschung & Lehre, and HeidICON - Heidelberger Bilddatenbank (an institutional repository).
- Two further entries from the United States are the Wittliff Digital Collections Repository (Texas State University) and the Berman Jewish Data Bank that is presented as the central online address for quantitative studies of North American Jews and Jewish communities.
- Furthermore there is the Codex Sinaiticus project that aims to reunite the entire manuscript of the codex in digital form and make it accessible online.

In summary: Re3data.org confirms that there are only few archaeological data repositories online; a repository the registry could add is the Swedish National Data Service which holds some archaeological data.

BASE search results

The strongest search service for open access academic repositories worldwide is the BASE - Bielefeld Academic Search Engine (www.base-search.net). BASE has an index of over 78 million records of 3725 sources aggregated through OAI-PMH harvesting. About 70% of the indexed items are accessible as full-text documents.

A search on “archaeology” in the subject headings of the worldwide resources yields 387,402 results (17 September 2015): 147,464 are Primary Data, 89,165 Texts, 59,878 Articles/Journals, 35,026 Images, 26,695 Unknown, 19,079 Books, 11,037 Reports, Papers, Lectures, 4432 Reviews, 3074 Theses (only few are videos, audio recordings or maps).

The main source of the 147,464 Primary Data records is the DataCite Metadata Store with 141,171 records, of which about 33,600 from the Archaeology Data Service (UK) and 1200 from the Swedish National Data Service (others could not be identified). Other data providers listed by BASE are the Data Archiving and Networked Services (DANS, NL) with 3439 results and PANGAEA (Earth &

Environmental Science) with 2804 results. BASE limits the displayed search results for each of the sources to 1000 records, but items of interest can be identified by using several filters.

OpenAIRE results

OpenAIRE (https://www.openaire.eu) was started in December 2009 to provide the e-infrastructure to support the Open Access Pilot for publications of projects funded under the European Union's 7th Framework Programme (in seven thematic areas of the programme) and to engage a network of multiple stakeholders, in particular research project coordinators and researchers, repository managers, research administrators at universities and others. The follow-up OpenAIREplus (started in December 2011) expanded the objectives by linking publications to research data and other associated information (e.g. funding programmes). OpenAIRE2020 (01/2015-07/2018) now assists the European Union's open access policy including the research data that will be created by projects funded under the Horizon2020 programme. At present OpenAIRE holds information about 12,225,631 publications 9,757 datasets from 5,933 repositories and OA journals (17 September 2015). A search for "archaeology" yields 5569 publications and 166 data resources. The data resources are 138 packages of 3D scans of vessels of the Caddo culture (United States) deposited by Robert Z. Selden Jr with the Zenodo repository; 24 resources deposited with the Archaeology Data Service by different providers; and 4 published by the Faculty of Archaeology at Leiden University.

Repositories and databases of other domains

Our survey focused on repositories, collections and other resources that might be characterised as "archaeological" or contain related content/data, because they are registered and/or can be found with common categories such as "archaeology", "prehistory", "history and archaeology", "ancient cultures", "ancient history", "classical archaeology", "medieval history", etc.

In addition to resources addressed above we also found others, contained in repositories and databases which are not "archaeological" but belong to other domains of research. To name but a few examples:

- PANGAEA - Data Publisher for Earth & Environmental Science[309] (hosted by the Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Germany), includes about 3100 datasets related to "archaeology";
- NASA Global Change Master Directory (GCMD)[310], covers more than 34,000 earth & environmental data set and service descriptions including, for example, World Data Center for Paleoclimatology and other relevant datasets;
- GenBank[311], includes about 10,000 records of ancient DNA of humans, animals, plants, bacteria, etc.; GenBank participates International Nucleotide Sequence Database Collaboration;
- DRYAD[312], is a major repository for data of papers in the bio-sciences, among the most downloaded data packages in 2014, in the 5th place, was a paper about livestock farming since the Neolithic using lake sediment DNA[313].

Thus there is more data of/for archaeological research around when the scope of relevant resources is extended to repositories and databases of all empirical disciplines. These resources can hardly be

[309] PANGAEA, http://www.pangaea.de

[310] Global Change Master Directory (GCMD), http://gcmd.nasa.gov

[311] GenBank, http://www.ncbi.nlm.nih.gov/genbank/

[312] DRYAD, http://datadryad.org;

[313] Giguet-Covex *et al*. (2014), data of their data package was downloaded 453 times (Morovati 2015). The paper in Nature Communications was not published open access while the data is freely available from DRYAD under the CC0 ("no rights reserved") license (http://dx.doi.org/10.5061/dryad.h11h7).

integrated by one, usually domain-based data infrastructure. Rather cooperation of e-infrastructures is required to enable data discovery and access across the different domain resources.

9.1.2 Humanities, cultural heritage and scientific collections

This section summarises surveys on digital cultural heritage and scientific collections which are a major component of the humanities research *e*-infrastructure landscape. Two are European-level surveys conducted 2006/7 which presented some first insights into the general situation, while the results of the 2014 ENUMERATE survey provide many details of the current status of digital collections. Special attention is given below to scientific collections as they are one of the blind spots in the e-infrastructure environment (cf. Geser & Niccolucci 2012).

Virtual cultural heritage collections as research infrastructure

In 2006/7 two European-level surveys on research infrastructures (RIs) have been conducted. Following the broad definition of RIs by the European Strategy Forum on Research Infrastructures (ESFRI) in their Roadmap 2006 (ESFRI 2006: 16), the surveys looked for entities of European relevance irrespectively if they are single-sited facilities or "virtual", i.e. providing online services. With regard to the humanities and cultural heritage the surveys found mainly "virtual" RIs which provide access to digital resources, but most of the services were managed by renowned bricks-and-mortar cultural heritage institutions (archives, libraries and museums). In other words, virtual RIs in most of the cases meant access to digitised collections of these institutions.

The survey Trends in European Research Infrastructures of the European Commission's DG Research and the European Science Foundation (EC & ESF 2007) aimed to cover all research infrastructures of European relevance. It identified 598 such infrastructures, both physical and virtual, which included 64 of the humanities sector. The survey report includes a list of these Humanities RIs; some examples of relevance to archaeology, classics, medieval studies, art history and history in general are: Beazley Archive of Classical Archaeology and Art (Oxford), Danish Runic Inscriptions of the Nordisk Forskningsinstitut (Copenhagen), Musée du quai Branly (Paris), Manuscripta Medievalia database (Marburg), Archivo General de Indias (Sevilla) and International Institute for Social History (Amsterdam).

The survey found that the social sciences and the humanities had much higher percentages of virtual RIs with a European dimension than other sectors: 41% and 25% compared to 12% in the total. With regard to visitors, most RIs were used on-site: across all RIs 60% reported less than 10% and some no remote users at all. But of the 64 Humanities RIs 57% had over 50% remote users.

The survey Infrastructural Research Facilities and Practices for the Humanities in Europe (2006) of the HERA-NET - Humanities in the European Research Area project discovered 405 humanities "Infrastructural Research Facilities" (IRF), both physical and virtual. The survey report does not contain a list of the IRFs, but some illustrative examples from 43 IRFs reported by Austrian respondents are: Conservation department and image database of the Austrian Archaeological Institute; Database of Greek ceramic fabrics in the Western Mediterranean of the Institute of Classical Archaeology, Vienna University; Database of historical images (HIBIDAT) of the Institute of History and Ethnology, University of Innsbruck; Austrian Exile Literature since 1933, Department of German Language and Literature, University Salzburg.

Most of the "Infrastructural Research Facilities" served the domain of cultural heritage (51%). Sixty per cent of all IFRs were digital resources and "experimental facilities"; 92.2% of those were digital resources, including digitised texts, maps, artefacts etc., language and linguistic corpora, and federated databases. Experimental facilities (7.8%) included Technical instruments (5.1%), Audio-visual laboratories (2%), and Virtual Laboratories (0.7%). One example of the latter is the large online database of the European Social Survey that allows users to select variables and conduct statistical analyses on the available data. Concerning access, 64% of the digital resources were freely accessible via the Internet, 6% required membership or subscription, 20% a special appointment with the host institution (e.g. due to sensitive information, specific technical requirements, etc.); 10% were "other" access modalities.

With regard to the further development of humanities research facilities, the survey respondents perceived as most critical issues copyrights and free access, better coordination among funding agencies and stable, long-term funding. The respondents also perceived a particular need for new Cultural Heritage research infrastructures. Concerning content and services, most respondents wanted to see digitisation of more resources, standardisation of metadata, and ability for cross-domain federated search.

Situation of scientific reference collections

Object-based scientific reference collections of museums, university institutes and other centres of research since long have been an important part of the scientific enterprise in many disciplines. The collections gather, document and present man-made artefacts such as pottery or coins as well as natural specimen, geological or biological samples, for instance. Reference collections are organised sets of such objects that are used to identify similar objects, comparative study as well as teaching. Reference collections are kept in museums and other specialised centres because they need curation and often specific conservation methods to ensure their longevity. The best known examples probably are the reference collections of species as curated by major natural history museums.

Since several decades scientific collections face severe challenges with regard to proving their continued relevancy and sustain their operation despite shrinking budgets. In the on-going inventorying and evaluation of research infrastructures the collections have been recognised, however the difficulty to sustain and, in many cases, re-vitalize them will remain. One driver for renewed interest in the collections is the progress in scientific methods, for example DNA sequencing methods (cf. Wissenschaftsrat 2011b: 13), but there are many ways in which individual specimens or whole collections can help tackling research questions (cf. the examples in Suarez & Tsutui 2004).

The Joint Programming Initiative (JPI) on Cultural Heritage and Global Change in their strategic research agenda (JPI-CH 2014: 21) note the importance of bringing together and making better accessible collections of reference material, measurement data and models (e.g. data of surveys, remote-sensing, modelling of material decay). Such collections could help improve the understanding of the impact of climate change on cultural heritage landscapes, sites and objects.

Difficult situation of many scientific collections

The OECD Global Science Forum (2008) addressed policy issues related to scientific collections and stressed that they should be understood as research infrastructure. The report however notes that often they have not been considered, managed and funded as such due to "*extreme heterogeneity of scientific collections, their age, their structure, their funding and their association with non-scientific ministries or institutions*". As specific difficulties the report notes that the research collections "*have very often grown from older collections and may still reflect outdated structures and practices. They operate in diverse ways, they are often hosted in old buildings, they are financed with diverging mandates from different ministries, agencies, foundations, universities etc.*"

Moreover the report emphasises that there has been a loss of skilled curators, lack of training programs as well as distinct career paths. Not surprisingly the report also perceived lack of internationally sharing of data and physical objects in support of research activities as "*the most compelling testimony to the independent evolution of many scientific collections and the need for a culture of collections as a shared research infrastructure*".

The Interagency Working Group on Scientific Collections of the US National Science and Technology Council's Committee on Science in their survey of 366 collections held by 14 federal agencies found that *“collection databases are not widely developed and Web access to collection information is still in its infancy”* (IWGSC 2009). Documentation and online access was highly variable: 60% of the 366 collections reported that none of the specimen information was accessible online, among the other collections 14% had made half of the specimen information and 5% all accessible online. The sample included eight archaeology collections of the Smithonian – National Museum of Natural History. Today the Museum allows searching some 450,000 records of their archaeology and ethnology collections which document all of their reportedly 2.5 million individual objects (including a small fraction of about 250,000 ethnological objects).[314]

The Smithonian is an example of a very large and prestigious museum whereas other, medium-size or small museums and collections are struggling to survive. Especially difficult is the situation of scientific collections at universities which often tend to be small and linked to individual institutes and departments. Due to limited resources many collections are not curated anymore and sometimes perceived as an obsolete legacy; outsiders often know little about the objects they hold.

A study of the German Council of Science and Humanities (Wissenschaftsrat 2011b) reported 1051 collections at German universities, including collections held by university museums. Of this large stock of collections 292 (28%) were broken up, lost or the whereabouts unknown. History & archaeology collections, in total 109 collections, presented the best situation with only 5 such cases. These figures were from the German collections database of the Helmholtz Centre for Cultural Techniques, as of 27 September 2010.[315] The Centre also maintains the UMAC Worldwide Database of University Museums & Collections that provides information compiled by the ICOM Committee for University Museums & Collections (UMAC).[316]. At present the online database gives access to information about 315 “History & Archaeology” museums or collections worldwide, of which 173 in Europe (96 in Germany, documented by the Helmholtz Centre).

Making scientific collections available for e-research

Making scientific reference and other collections accessible in a systematic way would be beneficial, especially with regard to the identification and comparison of artefacts and natural specimen. Attempts at promoting an online system of archaeological reference collections (e.g. Lange 2004) have not been successful yet. The lack of such a system arguably is due to the wide scope of objects studied by archaeologists as well as the unfavourable situation of many small and medium size museums and their collections. They may have their holdings catalogued, but with limited visual documentation, covering only exhibited special or exemplary specimens.

Digitisation and online accessibility of many archaeological collections could allow use of advanced computer supported methods (e.g. content-based analysis) for identifying different features of artefacts or natural specimens, comparing and classifying them. Instead of visiting reference collections one by one (online, if available), aggregation of the digital objects would allow researchers to select and create various (virtual) samples for their specific research questions. In turn, research results could enrich museum collections and exhibitions with new knowledge acquired based on objects of different collections, thereby strengthening the research function of museums and enhancing the scientific basis of exhibitions.

At present, however, the situation of most scientific reference and other collections does not allow for the realization of this scenario. The aggregation of specimen information into a shared online

[314] Smithonian - Anthropology Collections search, http://collections.nmnh.si.edu/search/anth/

[315] Helmholtz-Zentrum für Kulturtechnik (Humboldt-Universität zu Berlin): Universitätssammlungen in Deutschland, http://www.universitaetssammlungen.de/dokumentation/statistik; at present 1145 collections are documented of which 330 (29%) have been broken up, lost, or where the whereabouts unknown (details at the subjects level were not accessible).

[316] UMAC Worldwide Database of University Museums & Collections, http://publicus.culture.hu-berlin.de/collections/

collection database largely depends on efforts of some research communities. Among the advanced communities for example are the paleoontologists who collect fossil specimen information in the Paleobiology Database.[317]

In archaeology, collaborative building of online reference collection is not common, although there are many websites of research institutes, museums, laboratories and others that make collection information accessible online. As these resources are not integrated, numerous lists have been put together to help scholars find relevant resources for particular subjects like archaeozoology[318] or archaeobotany[319], for instance. Through bottom-up collaborative initiatives some scholars also aim to produce an open community resource, for example Paleobot.org (Warinner *et al*. 2011)[320], however often with little success concerning the number of entries.

In the UK, English Heritage has commissioned a project to collect information about the archaeological reference resources used by researchers in the study of artefacts and ecofacts.[321] The resources considered are reference collections, corpora, typologies and synthetic studies. Published, online and physical reference resources will be covered. The project aims to create a database of available resources, and identify gaps in and issues with current provision. The final report was expected for May 2015, but in September documentation of resources and asking specialists about experiences of using such resources was still on-going.

ENUMERATE survey on digitisation in European cultural heritage institutions

The ENUMERATE survey 2014 received input from 1372 cultural heritage institutions of 33 European countries. The results allow a good overview of the current situation of digitised collections of libraries, museums, archives, records offices and other institutions. The survey includes responses from 548 museums, 180 of which in the category "archaeology or history". Because of this mix, the exact status of the digitisation of archaeological museums' content cannot be evaluated. But some relevant results for "archaeology or history" museums in comparison to the average (av.) of all institutions are:

- Analogue collections (items) catalogued in a database: 54% (av. 54%, n=1.179);
- Have a digital collection: 85% in 2012 (av. 83% in 2012; 87% in 2014, n=1.179); note: a digital collection can be large (i.e. comprise of several digitised collections) or rather small (e.g. one special collection);
- Digitised collections of all 548 museums (details for "archaeology or history" are not available): 24% already digitised, 57% not yet, 19% no need to digitise (average of all institutions: 17%, 52%, 30%, respectively; n=1.179); the further details below are again for archaeology or history museums;
- Collect also born-digital heritage: 46% (av. 53%, n=1.262),
- Have a written digital preservation strategy that is endorsed by the management: 20.7% (av. of all museums 23.6%; av. of total 26.1%, n=905),

[317] Paleobiology Database, http://paleobiodb.org, assembled by hundreds of paleontologists from around the world; Fossilworks, http://fossilworks.org, provides search and data analysis tools for the large database.

[318] University of Sheffield, Zooarchaeology Laboratory - Useful Links and Resources https://www.shef.ac.uk/archaeology/research/zooarchaeology-lab/web-links

[319] WikiArc - Wiki Archaeological Information Resource, mainly Palaeoecological Reference Collections, http://www.wikiarc.org/Archaeological-and-Palaeoecological-Reference-Collections-Online

[320] Paleobot.org (collaborative web resource for archaeobotanical research), http://www.paleobot.org

[321] Archaeological Reference Resources Project, http://www.archaeologicalreferenceresources.uk

- Digital collections stored in archives that follow international long-term preservation standards: 51% no solution yet (av. 48%, n=905), 22% a professional public archive, 8% a privately managed archive; a few (5 or 6) have an own archive, and other respondents did not know.

The figures above may also be representative for institutions (museums and others) that hold mainly or a significant amount of material of/about Archaeological Sites, Monuments and Buildings, and Landscapes (characterised by the survey as “geography-based resources”). The survey gives the following figures for 239 such institutions, of which the percentage of digital material corresponds to, or is somewhat below, the 24% already digitised collections of all 548 museums:

	Not in the collection	***Analogue***	***Digital***
Archaeological sites	61%	33%	18%
Monuments and buildings	21%	73%	24%
Landscapes	67%	20%	18%

The ENMERATE survey also provides figures for the accessibility of the digital material of the cultural heritage institutions, however, not broken down according to different categories. Given that the situation of “archaeology or history museums” in all respects roughly corresponds to the average across all institutions, we may assume the same with regard to accessibility. The table below presents average percentages for all digital objects of respondents (n=905) which are or will become accessible and in which form:

Form of access	***Accessible now [11/2013]***	***Accessible 2 years from now***
Offline	42%	42%
Institutional website	34%	42%
National aggregator	14%	12%
Europeana	10%	18%
Other aggregator	5%	8%
Institutional API	5%	7%
3rd party API	3%	4%
Wikipedia	2%	3%
Other social media platforms	3%	5%
Other access channels	2%	3%

The figures show that the institutions primarily focus on providing access to the digital information on their own website and in-house (“offline”), e.g. at computer stations or other interactive installations. Also quite important is participation in content federations at the national or European level (e.g. Europeana). There is considerably less interest in allowing service developers direct access to (meta-)data through an Application Programming Interface (API). Also there is little, but somewhat growing interest in making information available through Wikipedia or social media channels.

Europeana as a research tool?

Europeana is the main gateway to digital cultural heritage content of European libraries, archives and museums. The gateway provides search functionality on over 30 million items, which are a large part of the estimated 10% of European cultural heritage digitised as yet (ENUMERATE 2014). Europeana and many of the underlying resources have been developed through numerous projects funded under European Union Framework Programmes (FP6-IST/FP7-ICT) and eContent(+) Programmes.

Europeana is not branded as a research e-infrastructure; the gateway falls outside the scope of the established RI schemas, although it has been included in the MERIL inventory of European RIs, coordinated by the European Science Foundation (cf. *Section 9.1.3* below). Researchers are not the focus of Europeana's strategic plan 2011–2015, but they are mentioned as one target group (Europeana Foundation 2011).

A policy document "Europeana for Research" (June 2014) issued by Member States' policy makers and sector experts notes that *"Europeana is not yet perceived as a platform that systematically serves research"*, and invites the platform *"to accelerate, subject to funding, the implementation of tools and services that will serve the needs of researchers, provide good practices of the use of the platform for research, and increase its collaboration with relevant initiatives and e-infrastructures at the European level"* (Europeana 2014).

The document does not explain why humanities and cultural heritage researchers perceive Europeana as insufficient for research purposes. Europeana was not intended as a platform for researchers but developed as a content search portal for the European public at large. It depends heavily on the information (metadata) which the libraries, archives and museum can provide. This information is seen by researchers as often too limited with regard to the required provenance/context and other description of the objects.

It is also difficult to envision how text or image based analysis of the content might be conducted effectively. The content items would have to be retrieved one by one and extracted to allow for processing; collecting images in sufficient quality would require much extra effort, for example, calling up the original providers' web pages of items, extract the content there, and so forth. The circumstances make effective Europeana tools and services for research as requested in the "Europeana for Research" paper rather unlikely, though not impossible.

The described issues of Europeana as a source for researchers also impede its use by archaeologists. Nevertheless some major providers of archaeological, classicist and related content have provided records to Europeana, arguably with the goal to ensure the presence of their resources on the main European cultural heritage search platform. A search of Europeana for "archaeology" yields 1,295,712 results (19 September 2015). The largest providers are the Swedish National Heritage Board (812,971 items), the German Archaeological Institute (183,683), the Ministère de la culture et de la communication, services régionaux de l'archéologie (72,969), the Statens historiska museum / Sweden (51,925) and the Archaeology Data Service (34,197).

Among the strengths of Europeana are clear frameworks for licensing and metadata as well as the ability to use the Application Programming Interface (API) to produce datasets for exploring the metadata. However few scholars are proficient in using APIs. Scholars want to work with the actual content and the challenge for Europeana therefore is to offer them useful tools.

The Europeana Cloud (eCloud) project[322], among other objectives, includes the goal to provide researchers with such tools. The project therefore has explored for which research subjects and how scholars might use the Europeana content. The exploration included expert fora, web surveys and a study of collections within Europeana (Benardou & Garnett 2015; Europeana Cloud 2015).

[322] Europeana Cloud project (EU, ICT-PSP, 02/2013-01/2016), http://pro.europeana.eu/web/europeana-cloud

Not surprisingly a study of the 100 largest collections found a large variety of content in terms of subjects and artefact types and some common themes such as art history, social history and the World Wars I and II. Social historians, for example, could "mine" the extensive collections of newspapers, photographs, political pamphlets and census records that are available. Problems that limit the use of Europeana content by scholars were identified through surveys of potential users and discussing with experts the requirements for the research platform the project aims to build. Identified core problems were:

- Navigating the Europeana portal and identifying relevant content,
- Lack of tools for building study corpora and conducting research such as fine-grained textual analysis, study of contextual information, etc.
- Need of an environment that allows members of research groups to work collaboratively, exchange information, link data and study results, etc.

The expert fora organised by Europeana Research produced a number of recommendations concerning the metadata and application programming interface (API), content corpora, and scholars' study approaches. Special emphasis was placed on (pre-)building specific, re-usable and sufficiently large corpora for scholars, and supporting different use cases in these thematic areas with the required tools.

Candidates to help Europeana in building or adapting such tools could be developers of the two humanities projects DARIAH and CLARIN. DARIAH (Digital Research Infrastructure for the Arts and Humanities) did not produce an e-infrastructure or research platform for humanities research, hence might be interested in opportunities offered by Europeana. Through CLARIN (Common Language Resources and Technology Infrastructure) Europeana can offer scholars access to advanced text analysis tools. It will also be worthwhile considering the experiences of a number of recent projects in which scholars have used novel tools to work with specific content such as manuscripts and correspondence, for instance[323].

9.1.3 European research (e-)infrastructures and services

This section addresses the two humanities research infrastructures on the roadmap of the European Strategy Forum on Research Infrastructures (ESFRI), and analyses the coverage of relevant facilities and resources in the MERIL inventory of European research infrastructures. A special study is presented of those natural sciences facilities identified in MERIL that provide analyses for archaeologists, among other clients. The study evaluates the online accessibility of archaeometrical and other data.

CLARIN and DARIAH – ESFRI humanities research infrastructures

The first Roadmap of the European Strategy Forum on Research Infrastructures (ESFRI) was released in 2006 and included 35 new research infrastructures or major upgrades of existing ones. Two of the new infrastructure initiatives were CLARIN - Common Language Resources and Technology Infrastructure and DARIAH - Digital Research Infrastructure for the Arts and Humanities.

The difference to other humanities-focused projects funded under the European Union's Framework Programmes is that the ESFRI initiatives CLARIN and DARIAH are expected to become sustainable based on funding commitments by Member States. Both are set up as European Research Infrastructure Consortiums (ERIC), a legal form that has been designed specifically to facilitate the

[323] European Correspondence to Jacob Burckhardt (EUROCORR, 06/2010-05/2015, funded by the European Research Council), http://burckhardtsource.org; DM2E - Digital Manuscripts to Europeana (EU, ICT-PSP, 02/2012-01/2015), http://dm2e.eu

joint establishment and operation of research infrastructures of European interest (the legal form exists since 28 August 2009).

CLARIN centres on language resources (in written, spoken, video or multimodal form) and tools to explore, annotate, combine and analyse them (Hinrichs & Krauwer 2014). CLARIN resources and tools are understood as useful for scholars of all disciplines, but in particular for language branches of the humanities, hence not for core purposes of archaeological research. Languages of course are important for archaeologists, e.g. when consulting written sources (e.g. papyri, medieval manuscripts) or deciphering engravings or specific sources such as cuneiform tablets.

DARIAH in principle addresses any kind of cultural content, historic as well as contemporary (e.g. works of art and architecture, music, film, etc.). Both initiatives received European Union funding for the preparation and implementation of the research infrastructure through the 7th European Union Framework Programme and further implementation support is expected under the Horizon 2020 programme.

On the current ESFRI Roadmap CLARIN and DARIAH are still the only two humanities RIs alongside 44 RIs of other disciplines, including as one "e-Infrastructure" the Partnership for Advanced Computing in Europe (PRACE). Thus most ESFRI projects are not dedicated e-infrastructures or "virtual" research infrastructures, but they need information technologies for their operation and several allow researchers virtual access to research databases and tools.

The first ESFRI report emphasised the difficulty of providing e-infrastructure and resources for the humanities: *"The complexity of the record of human cultures – a record that is multilingual, historically specific, geographically dispersed, and often highly ambiguous in meaning – makes digitisation difficult and expensive. (…) Data, information and knowledge are scattered in space and divided by language, cultural, economic, legal, and institutional barriers."* (ESFRI 2006: 20) Consequently the report perceived as a priority to provide RIs that allow for discovering and accessing available and newly created digital resources for humanities research.

This begs the question if CLARIN and DARIAH, the only ESFRI humanities RIs developed such e-infrastructures. CLARIN built a central registry of partners' repositories of language resources which have to comply with certain requirements (e.g. support for metadata harvesting, metadata in a defined format, persistent identifiers, federated login). A number of CLARIN centres have been certified (Data Seal of Approval) and invite deposits of data and tool. The CLARIN portal offers several search, access, and other services. DARIAH did not implement such a research e-infrastructure but organised a network of national centres where most activities, resource mobilisation and service provision takes place. Building a research e-infrastructure for any kind of humanities content and tools would indeed be a tremendous task.

Cultural heritage and archaeology resources & facilities in the MERIL inventory

MERIL - Mapping of the European Research Infrastructure Landscape (http://portal.meril.eu) is an initiative to provide a comprehensive European inventory of Research Infrastructures (RIs) of all types and across all scientific domains. The RIs have to be of more than national relevance and open to external users physically or online. The MERIL initiative is coordinated by the European Science Foundation (ESF) and receives financial support of ESF's member organisations; from October 2010 to September 2012 it was funded by the European Commission through a grant under the Seventh Framework Programme.

Inclusion of a Research Infrastructure (RI) in MERIL is understood as a label of quality, because each entry is evaluated based on criteria of quality of resources, access policies and principles of RI management. The MERIL records provide rich information which is based on parts of the Common European Research Information Format (CERIF) and RI specific attributes.

We conducted free-text searches of the MERIL database for "archaeology" and "heritage" across the records of all registered RIs; "archaeology" yielded 9 and "heritage" 24 entries (18 September 2015). An additional search of the category "Humanities and Arts" produced 70 results, including most, though not all of the already identified RIs. After removing doubles a total of 84 entities remained which (seemingly) appeared to be related to humanities and arts, heritage and archaeology. From those we kept only RIs which might provide at least some content/data relevant to archaeological research, and physical research facilities, e.g. heritage conservation or archaeometry laboratories. The selection required going through all 84 results to identify if this was the case.

The largest group found in the "Humanities and Arts" category with 18 results were Super & High-performance computing centres, Grid/Cloud infrastructures, Research and Education Networks and various information technology research centres. We did not include those results in our sample. The next large presence under the "Humanities and Arts" category and when searching for "archaeology" and "heritage" were natural sciences facilities (17) that apply various methods of material analysis; we included all in our sample but below name only some illustrative examples.

Not included in our sample are 9 linguistic resources (e.g. CLARIN and individual centres of which most participate in the CLARIN initiative); 8 various digital libraries of academic, educational and other material (e.g. the Wittgenstein Archives at University of Bergen); 6 national records and contemporary history centres/resources (e.g. the Historical Sample of the Netherlands); 4 Social Sciences archives (e.g. the European Values Study).

Also excluded are some humanities and other resources which we searched for "archaeology" (or the term in the main language of the resource) but produced little or no results, e.g. ECHO - Cultural Heritage Online (2 results), Huma-Num - la TGIR des humanités numerique (4 results), Repositório Científico de Acesso Aberto de Portugal (4 results); Central Library for African Studies, Portugal (about 200 results), e-Varamu - Estonian e-Repository and Conservation of Collections (0, under construction).

We included the ilissAfrica - Internet Library Sub-Saharan Africa (Germany) which aggregates records of several databases. A search for "Archäologie" (excluding results from the BASE engine, described in *Section 9.1.1* above) produced about 1300 literature references, 38 websites, and 22 photographs of the Colonial Picture Archive (Frankfurt).

But we excluded DARIAH – Digital Research Infrastructure for the Arts and Humanities as it does not provide a searchable database; and of course we excluded the Oden Icebreaker. The final result is the 31 resources/facilities presented in the table below:

Type of resource / facility	***Number of RIs***	***Examples***
Natural sciences facilities (relevant for archaeometry)	17	Examples: European Spallation Source (Sweden); Hertelendi Laboratory of Environmental Studies (Hungary); IPANEMA - Ancient Materials Research Platform / Synchrotron SOLEIL (France), National Research Centre on Human Evolution (Spain)
Heritage conservation and protection	2	Opificio delle Pietre Dure (Italy); Unique Infrastructures for Civil Engineering (CEDEX), Spain
Archaeology Digital Archives	3	Archaeology Data Service (UK); DANS-EASY (incl. the E-Depot for Dutch Archaeology); Swedish National Data Service (incl. archaeological deposits)
Museums	3	Germanisches Nationalmuseum (Germany); Museum - Royal Belgian Institute of Natural Sciences (Belgium);

		Royal Museum for Central Africa, Belgium (incl. ethnographic collections)
Various cultural heritage libraries and archives	4	EUROPEANA; Herzog August Library Wolfenbüttel, Germany (incl. mediaeval manuscripts); ilissAfrica - Internet Library Sub-Saharan Africa (Germany); Phonogrammarchiv, Austria (incl. ethnographic recordings)
Other	2	Réseau National des Maisons des Sciences de l'Homme, France (e.g. documentation centres and technical facilities); PANGAEA - Data Publisher for Earth & Environmental Science (3099 datasets found searching for "archaeology")
Total	**31**	

The table presents the research infrastructures which the MERIL inventory considers to be of more than national relevance and support research on archaeological and related other cultural heritage. In our selection we considered also important heritage categories like museum objects, mediaeval manuscipts, ethnographic recordings, etc. However in MERIL few institutions are present that hold such material.

Natural sciences facilities relevant for archaeometry

The majority of the RIs (17) are laboratories that may support archaeologists and other heritage researchers through physical, chemical and biological analyses of objects and materials. This is only a small sample of relevant laboratories that arguably are included in the MERIL inventory because they are large ones and avail of special facilities such as synchrotron radiation and accelerator mass spectrometry instruments. Such instruments are not available at typical archaeometry laboratories of university institutes and major museums of which there are many more.

We examined the websites of the MERIL sample of laboratories to see if they provide access to documentation (metadata) and data of their analyses of archaeological and other cultural heritage material (i.e. may have a data repository). We found that online documentation is hard to come by, access to data is offered nowhere, and even some illustrative webpages of conducted analysis are missing in most cases[324]. Lists of papers are typically presented, but few entries link to an open access journal paper or self-archived, pre-print copy.

The number of relevant publications we could identify on the websites ranged from 0, 1, a few within some years, up to 40 per year. The latter figure is for the National Research Centre on Human Evolution (CENIEH, Spain) where the researchers conduct studies on prehistoric human and other remains. While in the years 2012-2015 they published over 200 papers and other publications[325], on the centre's website we did not find documentation or data of these studies.

One example of good practice is the Center of Accelerators and Nuclear Analytical Methods – CANAM (Czech Republic) that provides annual reports of experiments and a webpage informs that

[324] As one example of an illustrative webpage, European Spallation Source - ESS (Sweden): Inside cereal grains planted under ESS 6000 years ago, http://europeanspallationsource.se/inside-cereal-grains; at the ESS we could identify five relevant publications since 2012 which concern wooden material (Lehmann & Mannes 2012), charred cereal grains (Fedrigo *et al*. 2015), and 17th century Japanese armour, helmets and swords (three studies, e.g. Salvemini et al. 2013).

[325] National Research Centre on Human Evolution (CENIEH), Spain, http://www.cenieh.es/en/research/recent-publications

results can be requested from their different laboratories (including contacts)[326]. However of 149 publications of CANAM researchers in the years 2012-14 only one is relevant in our context (Kameník *et al*. 2013; analyses of remains of the famous 16th-century astronomer Tycho Brahe).

At the SOLEIL synchrotron facility (France), researchers have supported studies of ancient and historical materials since 2007. The Ancient Materials Research Platform (IPANEMA) has been developed since 2011 and opened as a separate unit of SOLEIL in September 2013. Their website lists over 60 publications since 2007[327] but also does not provide access to documentation and data. The same situation at the Budapest Neutron Centre (Hungary) with a record in 2013-2014 of in total 62 publications of which 20 described or discussed experimental studies of mostly archaeological material[328].

The main reason for the lack of study documentation and data access at the laboratory facilities arguably is the understanding that a study is available if it is published in a paper[329]. Documentation on the laboratory websites is avoided or only used to illustrate research at the facility with some description and images. Data of the studies is generally not made accessible and may also not be catalogued. We think that documentation (metadata) of analyses conducted at laboratory facilities could greatly improve the discovery of data of such analyses and allow researchers to request data from repositories where they are deposited, which need not necessarily be at the laboratories.

It may be expected that within some years e-infrastructure initiatives in this field promote and facilitate a better accessibility of study data. One initiative has been the PaNdata Open Data Infrastructure[330] of thirteen major European facilities. PaNdata ODI developed a federation of user and data management systems of the photon and neutron facilities.

As one example of a national effort in this field we note the ArchLab Consortium of Swedish Archaeological Research Laboratory[331], a collaboration of six archaeological research laboratories which aims to create a common resource of methods and data. Some of their data is included in the SEAD - Strategic Environmental Archaeology Database[332].

Data of cultural heritage restoration and conservation centres

In addition to the survey above we looked into the situation of access to data produced by cultural heritage restoration and conservation centres. In this field EU-funded projects have enabled the networking of a large number of centres and, as one objective, to make their data better available. This initiative started with EU-ARTECH[333], 2004-2009, involving 12 partners, followed by CHARISMA[334], 2009-2013, involving 21 partners. Among the CHARISMA partners were natural sciences facilities, for example, the Budapest Neutron Centre and Synchrotron SOLEIL/IPANEMA (addressed above). A newly started venture is IPERION-CH[335], 2015-2019, that involves 23 partners

[326] Center of Accelerators and Nuclear Analytical Methods (CANAM, Czech Republic), http://canam.ujf.cas.cz/en/open-access-to-data/scientific-results/results-of-experiments

[327] Ancient Materials Research Platform (IPANEMA) / SOLEIL, France: Publications, http://ipanema.cnrs.fr/spip/publications/article/publications-d-ipanema?lang=fr

[328] Budapest Neutron Centre (BNC, Hungary), http://www.bnc.hu (see publications, articles 2008-2014)

[329] As examples of such publications, the Journal of Analytical Atomic Spectrometry, Issue 30, presents a themed collection of 28 papers in “Synchrotron radiation and neutrons in art and archaeology” (Bertrand, Dillmann & Reiche, eds., 2015).

[330] PaNdata ODI (EU, FP7, 2011-2014), http://pan-data.eu

[331] ArchLab Consortium of Swedish Archaeological Research Laboratories, http://archlab.se

[332] SEAD - Strategic Environmental Archaeology Database, http://www.sead.se

[333] EU-ARTECH (EU, FP6, 2004-2009), http://www.eu-artech.org

[334] CHARISMA (EU, FP7, 2009-2013), http://www.charismaproject.eu

[335] IPERION-CH (EU, Horizon2020, 2015-2019), http://www.iperionch.eu

and aims to establish a distributed European research infrastructure for restoration and conservation of cultural heritage.

EU-ARTECH and CHARISMA have been coordinated by the Centre SMAArt (Scientific Methodologies applied to Archaeology and Art) [336] at the University of Perugia (Italy). The Centre combines research groups from six departments of the university and several laboratories, including one mobile laboratory for non-invasive analyses of art historical and other objects. In the area of archaeology a focus on paleo-environmental research, geophysical prospection, and analyses of stone, metal, ceramics and glasses is mentioned. However, only two documents describing work conducted in the years 2004-2008 with the mobile laboratory are available from the website.

The CHARISMA project included the creation of a portal[337] that allows searching information about over 81,322 artworks held at six institutions, none of them natural sciences facilities. Almost 90% of the information records are from the Centre de Recherche et de Restauration des Musees de France (Paris), the other five partners have percentages between 1-3.5% (820-2800 records), for example the Opificio delle Pietre Dure (Florence) 2,16% (1754 records).

The intention of the portal is *"to guide users to institutions having documentation archives recording the scientific analysis of cultural heritage objects"*. This does not mean that the documentation of the scientific analyses is accessible through the portal, it is only indicated what data is available at the institute. The portal statistics indicate that for the 81,322 artworks at the institutes images are available for 46.87% (38,114), spectra data for 4.39% (3569) and colour data for 2.14% (1738); "detailed documentation" figures at a surprisingly low 0.18% (139).

9.1.4 Virtual research environments (VREs)

Virtual Research Environments (VREs) is an umbrella category of e-research environments which combine at least two of the following functions: professional networking and information exchange, collaborative digital collection formation and use, and programming and data processing tools and services (cf. David & Spence 2003; Bos *et al*. 2007; Carusi & Reimer 2010; Candela 2010). Ideally research communities could flexibly select and combine solutions for these functions according to their specific needs. The use of VREs arguably is a good indicator of the overall situation of e-research in a discipline. Therefore we include the results of a small study on VREs in the field of archaeology.

AVROSS survey on virtual infrastructure projects

The online survey of the AVROSS - Accelerating Transition to Virtual Research Organisations in Social Science project received information on 218 "virtual infrastructure" projects in Europe and beyond (AVROSS 2007). The survey provides some results on 26 projects in the field of archaeology or involving archaeologists. The study authors note that these projects "*show some very specific features: high bandwidth, frequent use of virtual/3D environments and innovative data collection methods distinguish these projects from the others"* (AVROSS 2007: 69).

It is worth comparing some of the figures for Archaeology to the whole AVROSS sample of disciplines, which includes Economics and business, Geography & regional science, Sociology, Linguistics (and Archaeology). Most projects created or used as "e-Infrastructure item" a research collaboration tool/system: Archaeology 76%, all projects 79.7%; with regard to distributed data and/or a repository the figures were 87.5% and 77%, respectively (AVROSS 2007: 34). High-bandwidth data transfer was required in 84% of the Archaeology projects and 61.3% in all projects. Virtual/3D environments were produced in 41.7% of the Archaeology but only 15.7% in all cases. Finally, innovative data collection methods presented the largest difference with 68.8% and 25.3%, respectively.

[336] SMAArt centre, University of Perugia, http://www.smaart.it

[337] CHARISMA art database portal, http://archives-charisma-portal.eu

Geographic Information Systems (GIS) are not included in the AVROSS list of "e-Infrastructure items" but appear under project outcomes which the respondents considered as new method or tool. Only respondents of nine projects thought that this was a GIS and six of these projects involved Social Sciences disciplines (AVROSS 2007: 37), whereas Archaeology was subsumed under the Humanities. Innovative or advanced methods/tools mentioned by survey respondents included "Geospatial cultural GIS-data", "GIS mapping, data mining, place name extraction", "3d gis, web service", "interactive scenarios (3-D, GIS-based with add on reports, tables, filters, etc.)" and "participatory gis" (AVROSS 2007: 186, 189, 195).

In summary, the figures make clear that research collaboration tools and distributed data were common features of projects in all disciplines surveyed. Innovation in data collection was (and still is) vital for Archaeology but much less so in other domains. The reason for this can be found in the variety of data that need to be produced for documenting an archaeological site and its environment. The transfer and work on some of this data can require high-bandwidth more often than in the other disciplines surveyed. The subject and data of archaeological research projects often allow producing Virtual/3D environments while not in other disciplines. A geographic information system (GIS) would not be considered as an innovative "e-Infrastructure item" of an archaeological VRE, but as a standard tool that is used in many projects.

Mainly single-function VREs

In the literature the most often described example of an archaeological VRE is the VERA - Virtual Research Environment for Archaeology developed for the large-scale excavations of the Silchester Town Life Project (UK)[338]. Based on the Integrated Archaeological Database (IADB) the project achieved an improvement in data organisation and access of the excavators and subject experts involved in the project; also data entry with digital pens was trialled (Rains 2011 and 2015).

The publication "Archaeology 2.0" (Kansa *et al*. 2011) beside VERA presents some other examples, which mainly support one function: Archaeology Data Service – supports data deposit and access; Open Context – data publication; BoneCommons[339] – information exchange; or UCLA Encyclopedia of Egyptology – access to scholarly articles[340]. Alison Babeu (2011) presents many digital environments developed for Classical Studies. The largest part is focused on digital scholarly editions (function: collaborative collection formation), with a strong presence of epigraphical resources. But these appear to be scattered and it is hoped for that the EAGLE - Europeana Network of Ancient Greek and Latin Epigraphy project[341] will achieve some integration of major resources.

Worthwhile, but not possible here, would be an evaluation of different platforms that could be used for archaeological VREs. For example, in the biological research community many Wiki-based environments have been developed to cope with a growing volume of laboratory data in need of proper annotation, analysis and publication (Waldrop 2008); one often mentioned example is TOPSAN (Weekes *et al*. 2010). A typical example of Wiki use in the humanities may be Digital Classicist[342], which is a website that links together several information resources. For research purposes a collaborative environment based on a semantic Wiki (e.g. Semantic MediaWiki[343]) could

[338] Silchester Roman Town, http://www.reading.ac.uk/silchester/

[339] BoneCommons, http://www.alexandriaarchive.org/bonecommons/

[340] The UCLA Encyclopedia of Egyptology presents 137 peer-reviewed articles published between 2008 and 2015, https://escholarship.org/uc/nelc_uee; the "full version" of the Encyclopedia is intended to provide various functionality but is in beta version and requires registration, http://uee.ucla.edu.

[341] EAGLE project (EU, ICT-PSP, 04/2013-03/2016), http://www.eagle-network.eu

[342] Digital Classicist, http://www.digitalclassicist.org

[343] Semantic MediaWiki, https://semantic-mediawiki.org

provide a useful solution in many cases. The WissKI project[344] developed a VRE inspired by the wiki way of dealing with information and collaboration and cultural heritage research as use case.

Most archaeologists arguably see a Web-based geographic information system (GIS) as their main digital research environment or platform. GIS allow them to bring together data of individual excavations as well as regional analyses (e.g. settlement patterns). Consequently archaeologists have acquired great mastery in the use of GIS, much more so than other humanities scholars who's usage of GIS *"can seem very limited, even simplistic, to archaeological eyes"* (Huggett 2012b). However, the use of GIS in archaeology has also not shown an impact with regard to advanced archaeological theory (Hu 2011, specifically addressing landscape archaeology).

Beside the widespread use of GIS, virtual representation of archaeological sites based on 3D models has become a major field of archaeological information technology. The "word cloud" below illustrates these developments. It has been produced based on the most frequently used 80 words in the titles of the 266 online available papers of the proceedings of the Computer Applications & Quantitative Methods in Archaeology (CAA) conferences in the years 2007 (136 papers), 2008 (83 papers) and 2009 (47 papers).[345]

Figure: Word cloud of the most frequently used 80 words in the titles of 266 online available papers of the CAA proceedings 2007-2009

In the word cloud as big as GIS and 3D is Analysis which of course serves as the main reason for deploying the technologies.

Web 2.0 as a threat to VREs

A special bread of "research environments" is Web 2.0 or social media solutions. The authors of the eResearch2020 study note that their *"Many-Headed Beast"* scenario, *"is based on the backdrop of a*

[344] WissKI - Scientific Communication Infrastructure (Germany, DFG-funded project, 2009-12, 2014-), http://wiss-ki.eu

[345] CAA - Online Proceedings, http://proceedings.caaconference.org

growing bottom-up Web 2.0 (or 3.0) tools and datasets" (eResearch 2020: xiii). This growth is motivated in part by a lack of useful e-infrastructure and services, or that specific needs are not met.

Web 2.0 solutions are perceived as bottom-up, flexible and arguably more readily accessible for many researchers and practitioners than VREs developed in the context of e-science infrastructure. However they mainly serve the need for immediate and informal professional networking and information exchange (Colley 2013; Dunn 2011; Kansa 2011), while the building and curation of digital collections and data processing require other systems.

Eric Kansa notes: *"While Web 2.0's impact is far reaching, it does seem to have limits. Web 2.0 platforms and services mainly facilitate informal communications among archaeologists. Web 2.0 systems are simple to use, fast, and geared to content that requires relatively minimal investment to create. Archaeologists tend not to use Web 2.0 platforms as the primary dissemination channel for forms of content that take a great deal of effort and expertise to create. In this light, data sets and sophisticated scholarly manuscripts see less circulation in Web 2.0 channels."* (Kansa 2011: 5; cf. Dunn 2011).

Web 2.0 solutions are not a threat to e-infrastructure development but to VREs, if such environments do not adequately serve also professional networking and information exchange. Some findings of substantial surveys on the use of Web 2.0 tools by academics are that they are mainly perceived as an informal supplement to academic communication channels, are actively used only by a small segment of researchers and, surprisingly maybe, young scholars are not among the avid users (cf. Procter *et al*. 2010a/b; RIN 2010a). One international survey also found that across all participants, social media users were 1.27 times more likely to be found in the arts & humanities and social sciences, and 0.67 times less likely in biosciences and health (UCL & Emerald 2010).

Distance to Distributed Computing Infrastructures

Arguably the largest distance of most archaeological research projects exists to Distributed Computing Infrastructures (DCIs). DCIs provide a so called "Science Gateway" to Grid infrastructure and software applications for data processing, storage and transfer. User groups can share data resources and computing applications ("virtual machines") and, thereby, form a virtual research community. The leading promoter of DCIs is European Grid Infrastructure (EGI.eu).

DCIs in the first place have been built in response to the need of some natural and engineering sciences for massive networking and processing of data. The DCI providers have sought to expand their user base, not least to legitimate the high investments in Grid and high-performance computing. In response to the market development and user demand, recently also Cloud services have been included in the DCI offer (Curtis+Cartwright 2010; EGI-InSPIRE 2011; e-IRG 2012).

But the demand for DCI has been limited in many disciplines as yet, especially in the humanities. Research groups in the humanities seldom deploy Grid-based datasets and processing, with the exception probably of some groups that avail of large corpora of texts. The major impediments for using Grid or Cloud based data processing in archaeology arguably are the diverse and complex types of datasets, lack of consistent data structures, incomplete, isolated and often not openly available data sources (cf. Hedge 2009).

However, many research institutions and projects may be attracted by (academic) Cloud services for other purposes. Reliable and affordable Cloud services might be a solution to many problems of resource-poor institutions and collaborative projects. Such services could allow them minimize the effort needed for local management of technology, repository/collection related tasks, as well as the need for dedicated technical staff, that are often not available. Such expectations however may underestimate factors such as still required technical knowhow, mandatory control over own or third-party content, etc. In any case, adoption of Cloud-based services, in particular, content management & curation systems, would require high trust of users in the service provider.

9.1.5 A selection of e-infrastructure related projects and resources

The table below presents an illustrative selection of completed and on-going projects for e-infrastructures, digital archives and other resources in the fields of archaeology, cultural heritage and humanities research. Furthermore some related general and domain-specific resources are included.

Acronym/name	Brief description	Link
3D-COFORM	3D-COFORM (EU, FP7-ICT, 12/2008-11/2012); created tools and a repository for 3D replicas of cultural heritage and archaeological artefacts	http://www.3d-coform.eu
3D-ICONS	3D-ICONS (EU, ICT-PSP, 02/ 2012-01/2015); produced 3D models and other content of archaeological and other architectural structures and provided the content to Europeana	http://3dicons-project.eu
Archaeolandscapes (ArcLand)	ArcLand (EU, Culture Programme, 09/2010-09/2015); focused on outreach and training programmes in the use of aerial, remote sensing and surveying techniques	http://www.arcland.eu
Archaeology Data Service (UK)	Data archiving and access service (ARIADNE partner and data provider)	http://archaeologydataservice.ac.uk
ARENA project and search portal	Archaeological Records of Europe - Networked Access project (EU Culture Programme, 2001-04; further work on the search portal was conducted 2009-10 in the context of DARIAH); the ARENA "what-when-where" search service approach has been adopted by ARIADNE	http://ads.ahds.ac.uk/arena/search/
ARACHNE (Germany)	Central objects-database of the German Archaeological Institute (DAI) and the Cologne Digital Archaeology Laboratory (DAI is one of the ARIADNE partners and data providers)	http://arachne.uni-koeln.de
ARCHES	Archaeological Resources in Cultural Heritage: a European Standard (EU Culture Programme, 06/2012-05/2014); made available a guide to good practice in archaeological archiving which is available in several languages	http://archaeologydataservice.ac.uk/arches/
Artefacts - Online Encyclopedia of Archaeological Small Finds (France)	Collaborative database and search portal for archaeological finds in Europe; the portal interface is available in French, English, German, Italian and Spanish	http://artefacts.mom.fr
CARARE	Connecting Archaeology and Architecture in Europeana (EU, ICT-PSP, 02/2010-01/2013); created content and CIDOC-CRM compliant metadata for monuments and other historical sites and provided the content to Europeana	http://www.carare.eu

CENDARI	Collaborative European Digital Archive Infrastructure (EU, FP7, 02/2012-01/2016); integrates digital archives for the medieval and World War One eras	http://www.cendari.eu
CHARISMA	Cultural Heritage Advanced Research Infrastructures: Synergy for a Multidisciplinary Approach to Conservation/Restoration (EU, FP7, 10/2009-09/2013); includes ArchLab, a search portal of scientific laboratories that conduct analyses of cultural heritage objects	http://www.charismaproject.eu
Civic Epistemologies	Civic Epistemologies (EU, FP7, 08/2014-03/2016); promotes participation of citizens in humanities and cultural heritage research, including the use of e-infrastructures	http://www.civic-epistemologies.eu
CLARIN (ERIC)	Common Language Resources and Technology Initiative / Infrastructure: Provides networked access for scholars in the humanities and social sciences to digital language data and tools of repositories and research centres; the initiative has been an ESFRI roadmap project, developed with EU financial support, and became an ERIC (legal entity); EU support for activities is continued through CLARIN-PLUS -Strengthening the CLARIN Infrastructure (EU, H2020, 09/2015-08/2017)	http://www.clarin.eu
CLAROS	Search portal for major ancient and classical arts collections developed since 2000 in a collaboration of research centres, archives and museums in the UK, Germany and Greece, led by the University of Oxford; employs CIDOC-CRM and semantic web technologies	http://www.clarosnet.org
DADAISM	Digging into Archaeological Data and Image Search Metadata (06/2014-12/2015), aims at allowing more effective search and analysis of large volumes of images and grey literature	http://dadaism-did.org
DANS / E-depot for Dutch Archaeology (Netherlands)	Data Archiving and Networked Services (DANS): The DANS-EASY system includes the e-depot for archaeological data, http://www.edna.nl (ARIADNE partner and data provider)	http://dans.knaw.nl/en/content/data-archive
DARIAH (ERIC)	Digital Research Infrastructure for the Arts and Humanities: aims to enhance ICT-supported research and teaching across the humanities and arts; the initiative has been an ESFRI roadmap project, developed with EU financial support, and became an ERIC (legal entity). EU support for activities is continued through the HaS-DARIAH project (see below)	https://www.dariah.eu

DASISH	Data Service Infrastructure for the Social Sciences and Humanities (EU, FP7, 01/2012-12/2014); involved research infrastructures of the humanities (CLARIN, DARIAH) and social sciences (CESSDA, ESS and SHARE)	http://dasish.eu
DCCD	Digital Collaboratory for Cultural Dendro-chronology; measurement series of different wood species derived from over 5000 objects and sites dating between 6000 BC and the present, ca. 50% of the DCCD collection is derived from archaeological sites and structures, including maritime archaeological sites (coordinated by ARIADNE partner DANS)	http://dendro.dans.knaw.nl
DCH-RP	Digital Cultural Heritage Roadmap for Preservation - Open Science Infrastructure for Digital Cultural Heritage in 2020 (EU, FP7, 10/2012-09/2014); mobilised stakeholders to establish a common roadmap for the preservation of digital/digitised heritage	http://www.dch-rp.eu
DISCO	Discovering the Archaeologists of Europe project (EU, Leonardo da Vinci II, 2006-8; Lifelong Learning Programme, 2012-14)	http://www.discovering-archaeologists.eu
DRI	Digital Repository of Ireland, national repository for social and cultural data	http://dri.ie
EAGLE	Europeana Network of Ancient Greek and Latin Epigraphy project (EU, ICT-PSP, 04/2013-03/2016)	http://www.eagle-network.eu
EHRI	European Holocaust Research Infrastructure (EU, FP7, 10/2010- 03/2015); developed a portal and tools for the Holocaust research community; extended with a Horizon2020 grant for the period 05/2015-04/2019	http://www.ehri-project.eu
EUROPEANA	The gateway to digital content of archives, libraries and museum across Europe. Research.Europeana promotes use of the content by digital heritage and humanities researchers (http://research.europeana.eu)	http://www.europeana.eu
eCloud	Europeana Cloud project (EU, ICT-PSP, 02/2013-01/2016); focuses on novel data provision services and tools for researchers to enrich and use the data	http://pro.europeana.eu/web/europeana-cloud
FAIMS (Australia)	Federated Archaeological Information Management System; provides a suite of tools for all stages of the digital data collection lifecycle	https://www.fedarch.org
Fasti Online	Fasti Online is a project of the International	http://www.fastionline.org

	Association for Classical Archaeology (AIAC) and the Center for the Study of Ancient Italy of the University of Texas at Austin (CSAI). Fasti Online provides access to about 12,000 reports and site summaries of excavations across the Mediterranean and other countries since 2000 (ARIADNE partner and data provider).	
Guides to Good Practice	Archaeology Data Service (ADS) & Digital Antiquity: Guides to Good Practice	http://guides.archaeologydataservice.ac.uk
Humanities at Scale (HaS-DARIAH)	Led by DARIAH ERIC the project (EU, H2020, 09/2015-08/2017) aims to enlarge the digital humanities community and develop a platform that provides integrated access to the output of the different DARIAH constituencies.	http://dariah.eu/about/humanities-at-scale.html
IANUS (Germany)	Research Data Centre for Archaeology and Ancient Studies (in development), coordinated by ARIADNE partner DAI	http://www.ianus-fdz.de
INNET	Innovative Networking in Infrastructure for Endangered Languages (EU, FP7, 10/2011-09/2014); developed a network of language archives and strengthened their relations with the CLARIN research community	http://www.innet-project.eu
INSPIRE	Infrastructure for Spatial Information in the European Community (based on the EU Directive 2007/2/EC and subsequent regulations); the INSPIRE spatial data themes includes protected sites	http://inspire.ec.europa.eu/theme/ps/
IPERION-CH	Integrated Platform for the European Research Infrastructure on Cultural Heritage (EU, H2020, 2015-2019); develops a distributed European research infrastructure for heritage science, restoration and conservation	http://www.iperionch.eu
JPI Cultural Heritage	Joint Programming Initiative on Cultural Heritage and Global Change: initiative of European ministries and other agencies (since 10/2011 supported by projects under the FP7 Environment and ERA-Net programmes); developed an action plan for research on cultural heritage and climate/environmental change and funds research projects	http://www.jpi-culturalheritage.eu
Linked Heritage	Linked Heritage (EU, CIP-PSP, 04/2011-09/2013) employed Linked Data technology, metadata enrichment methods and multilingual terminologies for providing new digital cultural heritage content to Europeana	http://www.linkedheritage.eu
MAPPA (Italy)	Data repository at the University of Pisa, includes MAPPA Open Data (invites deposits	http://mappaproject.arch.unipi.it/?lang=en

	also by researchers not affiliated with the university)	
NeDIMAH	Network for Digital Methods in the Arts and Humanities (European Science Foundation, Research Networking Programme, 05/2011-05/2015)	http://www.nedimah.eu
Open Context (USA)	Archaeological data publication service offered by the Alexandria Archive Institute (USA); published data are deposited with the California Digital Library	http://opencontext.org
PARTHENOS	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies (EU, H2020, 05/2015-04/2019); builds on existing e-infrastructures of CLARIN, DARIAH, EHRI and others to enable cross-disciplinary research in the fields of humanities and cultural heritage	http://www.parthenos-project.eu
Pelagios	Platform and collaboration of many projects aimed to create and integrate information about ancient places, based on a lightweight Linked Data approach; uses gazetteer URIs of Pleiades (http://pleiades.stoa.org) to join up a variety of online resources that refer to the ancient places	http://pelagios-project.blogspot.co.at
SND (Sweden)	Swedish National Data Service (ARIADNE partner and data provider)	http://snd.gu.se
tDAR - The Digital Archaeological Record (USA)	Data archive and access service of the Digital Antiquity consortium, hosted at Arizona State University	http://www.tdar.org
APA / APARSEN	Alliance for Permanent Access to digital information; APARSEN - Alliance for Permanent Access to the Records of Science in Europe Network (EU, FP7, 01/2011-12/2014)	http://www.alliancepermanentaccess.org
COAR	Confederation of Open Access Repositories (about 100 members and partners)	http://www.coar-repositories.org
Costs4C	Collaboration to Clarify the Cost of Curation (EU, FP7, 02/2013-01/2015); the project has established the Curation Costs Exchange (CCEx, http://www.curationexchange.org)	http://4cproject.eu
DCC	Digital Curation Centre (UK) offers a wide range of data curation/management resources	http://www.dcc.ac.uk
DigCurV	Digital Curator Vocational Education (EU, Leonardo da Vinci, 01/2011-06/2013)	http://www.digcur-education.org

DRYAD	Repository for bio-sciences research data made accessible under journal data policy	http://datadryad.org
e-IRG	e-Infrastructures Reflection Group, the main advisory body for European e-infrastructures	http://www.e-irg.eu
EGI.eu	EGI Foundation: coordinates a pan-European platform of distributed computing infrastructures (incl. the EGI Federated Cloud) and their users; the EGI-Engage project (EU, H2020, 03/2015-08/2017) promotes open science commons and supports research communities with technical and data solutions	http://www.egi.eu
ERF-AISBL	Association of European-level Research Infrastructure Facilities: unites 15 major research infrastructures / facilities in this field, including the Central European Research Infrastructure Consortium (CERIC-ERIC, http://www.ceric-eric.eu) with facilities in nine countries	http://www.erf-aisbl.eu
ESFRI	European Strategy Forum on Research Infrastructures: coordinates the development of major European research infrastructures through the ESFRI roadmap	http://ec.europa.eu/research/esfri/
EUDAT	European Data Infrastructure (EU FP7, Infrastructure, 10/2011-09/2014); developed common data services, e.g. storage, persistent identifiers, and metadata solutions; continued under H2020, 03/2015-02/2018, to maintain and extend the services as well as reach more user communities	http://www.eudat.eu
FOSTER	Facilitate Open Science Training for European Research (EU, FP7 project, 02/2014-01/2016)	http://www.fosteropenscience.eu
GenBank	GenBank includes about 10,000 records of ancient DNA of humans, animals, plants, bacteria, etc.	http://www.ncbi.nlm.nih.gov/genbank/
Global Change Master Directory (NASA)	Provides access to more than 34,000 Earth & Environmental data set and service descriptions; includes the World Data Center for Paleoclimatology and other relevant datasets for archaeology	http://gcmd.nasa.gov
MERIL	Mapping of the European Research Infrastructure Landscape; provides an inventory of research infrastructures of all types and across all scientific domains	http://portal.meril.eu
OpenAIRE	OpenAIRE (12/2009-11/2013), OpenAIREplus (12/2011–05/2014), OpenAIRE2020 (01/2015-07/2018); assists the EU open access policy	https://www.openaire.eu

	which first focused on research publications and now also includes data created in EU funded projects (Research Data Pilot)	
PaNdata Open Data Infrastructure	PaNdata ODI (EU, FP7, 11/2011-10/2014) developed a federation of user and data management systems of European photon and neutron facilities (which may be used for archaeometrical studies)	http://pan-data.eu
PANGAEA	Data Publisher for Earth & Environmental Science (hosted by the Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Germany); the repository holds about 3100 datasets related to archaeology	http://www.pangaea.de
re3data	Registry of Research Data Repositories: indexes over 1300 data repositories from all over the world making it the largest online catalogue of such repositories on the Web	http://www.re3data.org
RECODE	Policy Recommendations for Open Access to Research Data in Europe project (EU, FP7-SiS, 02/2013-01/2015)	http://recodeproject.eu
RDA	Research Data Alliance – "Research data sharing without boundaries"; involves working and interest groups on many research data related topics; see also the RDA Europe project (http://europe.rd-alliance.org)	https://www.rd-alliance.org
RItrain	Research Infrastructures Training Programme (EU, H2020, 09/2015-08/2019)	http://ritrain.eu
THOR	Technical and Human Infrastructure for Open Research (EU, H2020, 06/2015-12/2017); builds services based on the identifier systems of DataCite and ORCID (identifiers for researchers) aimed to support institutions, researchers and research objects of different disciplines	http://project-thor.eu
Zenodo	Data repository at CERN, related to OpenAIRE; invites deposit of "long-tail" (small-scale) data of all disciplines	http://zenodo.org

9.2 Scholarly publication culture, citations and altmetrics

This background study presents information and empirical results on citation-related topics and issues, for example, the citation advantage for open access publications (papers and data), altmetrics (alternative metrics), and the importance of usage data for data archives. Special attention has also been devoted to specificities of the humanities publication culture, which should be taken account of in the development of comparative metrics of data sharing, re-use and citation. A summary of the results and suggested actions are given in *Section 4.4.5*.

Open access of publications is already well established in the research community. Proponents of the open data movement request that in addition also the data underpinning the published research results should be accessible online and at no charge for the users. While there are many good arguments for open access data (cf. *Section 4.3.1*) most researchers appear to be unwilling to share their data openly through a data repository due, among other factors, to the additional effort and a potential disadvantage. Open data advocates argue that instead of a disadvantage open data will often be re-used and cited, bringing recognition and reward to data sharers (e.g. in decisions on career advancement).

The citation advantage for open access papers is very well supported by many studies, although less so for papers in the humanities. For papers that are published with data there is little evidence yet for an advantage. This is due to several factors, for instance, low data sharing in many disciplines as well as that empirical evidence for a citation advantage is difficult to establish.

Below we present and discuss available results of both traditional metrics (citations) and “altmetrics” (alternative metrics) for advantages of openly accessible publications and data that are shared through repositories. But any metrics of such advantages must take full account of the specificities of the publication culture in the humanities. Therefore we first look in greater detail into aspects which distinguish the humanities from other disciplines and may also play a role in the context of data sharing, re-use and citation.

9.2.1 Specificities of the humanities publication culture

Bibliometric and other studies point out specificities that set the humanities apart from other disciplines. It is important to consider the specific aspects of the publication culture in the humanities and take them into account with regard to indicators of usage and citation of publications and data. Below we summarise the specific aspects of publication in the humanities. As the bibliometric literature often treat the social sciences and humanities (SSH) as one block, we present results for the SSH, where available for the humanities, and indicating observed differences also between them and the social sciences. However, what is not possible based on the available literature is taking account of the multi-disciplinarity of archaeology, for example, that archaeological research results are also published in journals of the applied natural sciences.

Under-representation of humanities journals and English-language bias

The humanities are known to be a difficult terrain for bibliometric studies because the main citation indexes Web of Science (Thomson Reuters) and Scopus (Elsevier; available since 2004) do not well cover these disciplines. This is due to a number of specificities of the humanities and social sciences which are not adequately reflected in the indexes.

The Web of Science Core Collection[346] incorporates the three discipline based indexes (Science, Social Sciences, Arts & Humanities) as well as the Conference Proceedings Citation Index and the Book

[346] Web of Science - Core Collection, http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/web-of-science-core-collection.html

Citation Index. In terms of journals, the WoS website mentions coverage of in total 13,200 "high impact" journals worldwide, including Open Access journals. According to the website, the Science Citation Index covers 8500 journals across 150 disciplines, the Social Sciences Citation Index 3000 journals across 50 disciplines, and the Arts & Humanities Citation Index 1700 journals also of some 50 disciplines. While 1700 arts & humanities journals seems a lot, the Ancient World Online maintains a list of currently 1577 open access journals and other serials in the field of archaeology, classics and other domains of ancient world studies (up from about 1100 in July 2012)[347]. As this is but one segment of the variety of scholarship in the arts & humanities, the figures make clear the low coverage of humanities journals in Web of Science (or, that the majority is not considered as "high impact" publication outlets).

Mongeon & Paul (2014) compared journals indexed in Scopus (21,918) and Web of Science (16,957 across the WoS Core Collection) with Ulrich's Periodicals Directory that covers 61,745 active "academic/scholarly" journals worldwide. They could match the majority of journals indexed in Scopus and WoS, 20,346 and 13,605 respectively, with Ulrich's list. In comparison to Ulrich's, the Social Sciences (SS) and Arts & Humanities (A&H) journals were underrepresented in both Scopus and WoS; in Scopus: SS -28% and A&H -36%; in Web of Science: SS -45% and A&H -37%.

Comparing the coverage by publishing country, Social Sciences and A&H journals published in the Netherlands, UK and USA were over-represented in both Scopus and the Web of Science indexes (in the A&H also in Belgium and France), while journals published in all other countries were under-represented. Concerning languages, over-represented were SS journals in English and Dutch and A&H journals in English, French and German, while all other languages were underrepresented, most by 50-100%.

While the "hegemony of English" (Ammon 2010) is unfavourable for authors native in other languages there are also many positive aspects of having one *lingua franca*. However, Gingras & Mosbah-Natanson (2010) show that according to the Web of Science data the social sciences (including anthropology and archaeology) appear to be mainly produced in North America and Europe, a situation which they see as fatal due to non-presence of other perspectives on societal issues globally.

Fewer citations of papers but over a longer period

Despite the limits of bibliometrics for the humanities, based on extensive studies that used Web of Science data for the period 1980-2007 some major differences compared to other disciplines can be summarised (cf. Archambault & Larivière 2010):

a) Humanities papers on average cite few other papers, in 2007 about 8% (with a small increase since 1980), whereas papers in the social sciences cite others around 50%, natural sciences & engineering slightly less than 70%, and medicine over 80%; the social sciences and the natural sciences & engineering disciplines show a considerable increase in this regard.

b) Humanities papers cite considerably older literature, in 2005 the median age of the literature was 14 years, whereas the figures for the other disciplines are between 5 and 8 years; in this regard the social sciences are similar to the natural sciences & engineering disciplines.

Thus humanities scholars who publish papers indexed by Web of Science (Arts & Humanities Index) cite more often other literature than journal papers and most of the referenced literature is much older than in other disciplines. For instance, in Greek and Latin philology on average 25 years old literature, not including the original texts (Crawford 2013, not based on WoS).

[347] AWOL - The Ancient World Online: Alphabetical List of Open Access Journals in Ancient Studies, http://ancientworldonline.blogspot.co.at/2012/07/alphabetical-list-of-open-access.html

The more local, context-specific nature of SSH research

It can be argued that research in most natural sciences is not, or at least much less, context-specific as most of the SSH research: *"Whereas the problems identified in the NSE [Natural Sciences and Engineering] tend to be universal by nature, social science and humanities research topics are sometimes more local in orientation. (…) In many cases, the concepts and subjects covered in social sciences and humanities can be expressed and understood only in the culture that shapes them"* (Archambault & Larivière 2010: 253). Therefore natural scientists can address the international research community, ideally reached through a major international journal. In comparison the share of "local" publications, i.e. in national journals, addressing the readership in the native language, is much higher in the SSH. Furthermore articles in non-scholarly media represent interaction of the research with local contexts of application (Hicks 2005).

The issue has not gone unnoticed in the indexing industry. At Thomson Reuters the Web of Science Editorial Development Team since 2007 examined some 10,000 "regional journals" for potential inclusion. "Regional journals" are defined as *"journals that typically target a regional rather than international audience by approaching subjects from a local perspective or focusing on particular topics of regional interest"*. The integration of these new journals to WoS is understood to build *"a bridge between significant regional studies and the global research community – increasing the visibility of topics of regional concern"*. In 2011, already 1675 regional journals from 76 countries were added, 879 from Europe, 375 from the Asia Pacific region, 209 Latin America, 174 Middle East and Africa, and 38 North America. The additions represent over 45% of new journals included in WoS from 2000 to 2010 (cf. Testa 2011; Web of Science 2011).

Language boundaries in archaeology

Concerning archaeology, citation analyses reported by Kristiansen (2001) indicate a decrease in the previous 10 years of references in selected journals published in various European countries to publications in other languages. The trend toward "a single language research environment" was clear in English, French and German journals, while journals of smaller languages and countries (e.g. Hungary and Sweden) were more international with regard to references. A similar pattern showed book reviews. Kristiansen notes that an increasing preference for regional and local studies may have added to the overall trend.

The survey confirms the distinction between "mainstream" and "minority" European research communities introduced by Evžen Neustupný (1998). Willems (2008) summarises the distinction as follows: *"The larger groups (English, French, German, Russian, Spanish) he terms 'mainstream'. These are the communities that are big enough to have a full internal discourse on all relevant topics in the discipline, so there is no immediate need to refer to outside sources or to participate in other discourses. For his 'minority' communities that is not possible. Indeed, the Scandinavian, Dutch, or Czech 'minority' communities traditionally have a very outspoken international orientation."*

Willems (2008) assumes that the trend described by Kristiansen (2001) is still dominant: a decrease in scope with regard to research community and language and, therefore, in the number and diversity of international references. He considers that this happened also in the smaller countries, due the growth of archaeological heritage resource management following the implementation of the Valetta Convention.[348]

Composition of the SSH literature

The composition of the scholarly literature produced by SSH researchers is substantially different from the natural sciences in that monographs, book chapters, conference papers, reports and non-

[348] Council of Europe: European Convention on the Protection of the Archaeological Heritage (Revised), Valetta, 16.I.1992, http://conventions.coe.int/Treaty/en/Treaties/Html/143.htm

scientific literature play a greater role in the SSH. However, the largest part of the citation indexes is based on journal papers, which may be adequate for medicine and natural sciences, but less so for the social sciences and certainly not for the humanities. As noted above, about 50% other literature is referenced in social sciences papers and over 90% in humanities papers. An example from Austria may illustrate the case: Mutz *et al*. (2012) analysed the output of 1742 projects funded by the Austrian Science Fund in the years 2002 to 2010. 31.4% were in fields of the SSH. The output was grouped in profiles of which one was "Book and Non-Reviewed Journal Article". 35.8% of all projects had this output profile, including almost all of the SSH projects.

In recent years Web of Science and Scopus have improved their coverage of scholarly products, however many imbalances remain. Zuccala (2013) presents an overview of different types of arts & humanities literature that have been indexed in Scopus and Web of Science (A&H Index) in the years 2008-2012. WoS (A&H Index) covered 558,823 items of which 40.64% were book reviews and 33.88 journal articles; Scopus included 318,996 items of which 24.05% were reviews and 56.05% articles.[349] In comparison, the presence of conference papers was low: Scopus indexed 11,577 (3.63%), while in WoS (A&H Index) such papers together with book chapters, biographical texts, scripts, music scores and various reviews (film, theatre, records, etc.) amounted to 53,264 (9.53%).

The different composition of the literature also affects the shares of self-archived publications of the disciplines in institutional repositories. Two US-based studies suggest that the amount of Sciences & Engineering (S&E) content in such repositories is triple that of the SSH: Jantz & Wilson (2008), based on a survey of 49 repositories of members of the Association of Research Libraries (about 5000 items listed under faculty/researcher names) found: S&E 67%, Social Sciences 27%, Humanities 5%; Dubinsky (2014), based on 107 repositories of different types of higher education institutions (about 64,000 items): Sciences 61%, Social Sciences 18%, Humanities 21%. In both surveys anthropology, which in the United States usually includes archaeology, was subsumed under Social Sciences.

Conference proceedings

Conference proceedings are an important communication channel in all disciplines but their scientific impact in terms of citations to articles in the proceedings is low in most areas of research. A study of the citation of conference proceedings in the Web of Science databases (Lisée *et al*. 2008) found that in 2005 in the field of natural sciences & engineering approximately 1.7% of all literature referenced proceedings, and in the social sciences and humanities 2.5% (in some SSH subfields proceedings up to 5%). The average age of cited proceeding was highest in the humanities with about 18 years, while in the social sciences slightly below 8 years, like in most of natural sciences and engineering disciplines.

Since the 1980s the relative importance of proceedings diminished in all areas except in engineering and computer sciences. In engineering the average share of references to proceeding grew from 8% to 10% and in the computer sciences it reached almost 20%. The study concludes that the scientific impact of papers published in proceedings is limited, except in the mentioned disciplines, where they more often represent the final publication, i.e. are not followed by a paper in a scientific journal. In archaeology, the proceedings of the Computer Applications and Quantitative Methods in Archaeology (CAA) conferences are the most prominent outlet for reporting advances in the application of information and communication technologies for archaeological purposes.

[349] The large amounts of book reviews make clear the importance of this type of literature in the arts & humanities. A detailed study of humanities book reviews in WoS 1981-2009, among other results, found that reviews referencing the book and other publications are cited more frequently (Zuccala & Leeuwen 2011).

Books / monographs

In the arts & humanities books are, and will very likely remain, a major form of publication, actually the "gold standard" to present scholarly work. The monograph is still also considered a strong requirement for career progression in the humanities, e.g. in the American "tenure and promotion" process. As expressed by one anthropologist interviewed in an American study on scholarly communication, *"Anthropology is primarily a 'book discipline,' meaning that faculty must write books, at least one 'great' book for tenure, and another book to be promoted to full professor, or as one administrator put it: 'Anthro is just sort of irreconcilably book fetish-ized... This is a book discipline. You can have great articles, you hit a ceiling, potentially very low, if you don't have a book, and the book has to be significant, it has to get reviewed in the right places…'."* (King *et al*. 2006: 54)

In the United States anthropology covers the main anthropological subfields (i.e. physical, linguistic, cultural and social anthropology) as well as archaeology. The focus on the scholarly monograph may be somewhat less in physical anthropology that has undergone a shift towards the bio-sciences, where a consistent record of publications in peer-reviewed journals is a key to advancement. In other subfields important books, peer-reviewed papers, and chapters in edited volumes will pave the way from assistant to associate and on to a full professor position. In this academic value system, focused on "high impact" publications, the curation of data and creation of digital products is perceived as rather irrelevant (cf. the anthropology and archaeology case studies in King *et al*. 2006: 48-63; Harley *et al*. 2010b).

A project at the University College London investigated the role and future of monographs in arts & humanities research (Williams *et al*. 2009; based on a background study and 17 in-depth interviews with A&H academics). The results confirmed that monographs continue to be seen of great academic value and essential for career progression in the A&H. Monographs were seen as the single most valued means of scholarly publication and also the main currency of the A&H in the UK Research Assessment Exercise. The latter was perceived as a cause of concern as it exerts pressure to produce monographs fast and may reduce their quality. Furthermore reservations were expressed about moving towards digital versions of the monograph (e.g. concerning longevity), while print-on-demand was considered to be a viable option for the publication of specialist works.

Thus despite the existing financial and publishing constraints (Ward 2010) the print-on-paper monograph keeps its role as the "gold standard" in the field of A&H. One major element of the "gold standard" approach is that the authors strive to be published by a renowned academic publishing house or university press, which adds prestige and may increase visibility and recognition. However, the publishers find it difficult to sustain monograph publishing due to the high costs while over the last 30 years average sales declined from 2000 to 200 (Milloy 2013).

Experimentation with Open Access monographs

The difficult situation of scholarly monographs publishing has stimulated experimentation with new business models involving Open Access to digital versions, which allow scholars and publishers to align with OA objectives and, arguably, to keep the willingness of sponsors (research funders, institutions and others) to support scholarly monographs. Indeed, after a long period of crisis the scholarly monograph "phoenix" seems to gain some energy in the digital environment (Steele 2003 and 2008).

Researchinformation.info (2015) summarises key findings of a recent survey of the Publishers Communication Group on OA monographs publishing. The survey sought input worldwide from both publishers who are active in or consider publishing OA books as well as libraries that may support such books. Reported findings of the not publicly available survey are:

Publishers: One third of the respondents stated that they publish OA monographs but such works accounted for less than 5% of their book collections. 44% of the publishers felt that their OA program is growing, albeit modestly. Among those not yet publishing OA books about 30% thought that starting an OA program is likely within the next five years. Publishers feared unrealistic funding expectations in the academic community and resemblance of OA books to "vanity publishing".

Libraries: 20% of the library respondents reported participation in OA funding initiatives such as Knowledge Unlatched and others. 53% of the funds for supporting OA book publishing were taken from the existing materials budget. From available OA funds 74% were used for journals (authors' article processing charges), 26% for OA book publishing. 57% of the libraries catalogue OA books of which 81% use established criteria such as relevance to curriculum. Library respondents felt they should advocate for OA publishing within their institutions, but feared diversion of existing funds.

In the survey funding sources for OA book author fees were identified variously as the authors themselves (23%), academic departments (21%), library funds (15%) and outside grants (26 per cent).

Ferwerda (2014) and Milloy (2013) provide overviews of innovative business models and examples in this field, and JISC & OAPEN (2013) report the discussion at a major conference dedicated to OA monograph publishing. The business models are varied but the products typically hybrid (print and digital), including also print-on-demand.

One example of an innovative humanities and social sciences publisher is Open Book[350]. Open Book started in 2008, currently offer 65 books, and many new titles are forthcoming. Their website reports *"700,000 book visits; readers from 207 countries; 400 readers per title every month"*. The high figure of readership concerns the freely accessible HTML version Open Book produces of each title.

The Knowledge Unlatched[351] project invites libraries to sponsor proposed OA books that a publisher is willing to produce for a fixed fee, with the PDF downloadable freely. Thus the more libraries commit the lower the cost per library. Nearly 300 libraries from 24 countries have pledged their support for making the initial collection of 28 new books produced by 13 scholarly publishers available.

The Directory of Open Access Books (DOAB)[352] of the OAPEN Foundation allows searching OA books. The directory has been initiated by the consortium of the European eContentplus project OAPEN (09/2008-02/2011)[353]. As current sponsors the website notes Brill, OpenEdition and Springer. At present the directory records 3080 peer-reviewed academic books from 107 publishers. We found 51 archaeological books, 20 of which from the ANU Press (Australian National University) and 1-4 titles from 16 other publishers located in Austria, Canada, Estonia, France, Germany, Italy, Netherlands, UK and USA[354].

The web portal Open Edition[355] (funded by the French Government) integrates catalogues of humanities and social sciences resources (books, journals, weblogs and events). The book catalogue at present features 1926 titles, 650 in the category history & archaeology of which 48 are archaeological books. 39 have been published Open Access Freemium, which means in HTML only, two are available in PDF, HTML and ePub formats, and seven are restricted access. The books have

[350] Open Book Publishers, http://www.openbookpublishers.com

[351] Knowledge Unlatched, http://www.knowledgeunlatched.org

[352] Directory of Open Access Books (DOAB), http://www.doabooks.org

[353] OAPEN - Open Access Publishing in European Networks project, http://project.oapen.org; see Adema & Ferwerda (2014) on this and other humanities open access publishing initiatives.

[354] Four books: University of Tartu Press (Estonia); three books: Brill (NL), Verlag der Österreichischen Akademie der Wissenschaften (Austria); two books: Athabasca University Press (Canada), Collège de France (Open Edition, France), De Gruyter (NL), Open Book Publishers (UK), Österreichisches Archäologisches Institut (Austria), Presses de l'Ifpo (France), Ubiquity Press (UK), Universitätsverlag Göttingen (Germany); one book: Böhlau (Austria), Ledizioni/LediPublishing (Italy), Leiden University Press (NL), Manchester University Press (UK), Newfound Press (University of Tennessee Libraries, USA).

[355] Open Edition, http://www.openedition.org

been made available by 13 publishers, for example, Presses de l'Ifpo (10 books), l'École française de Rome (8), Collège de France (7), and one non-French publisher, the Centro de estudios mexicanos y centroamericanos in Mexico (7).

In summary, while a large number of open access books are already available relatively few appear to be archaeological publications.

Book coverage by citation indexes

In 2011, Thomson Reuters launched the Book Citation Index with currently over 50,000 indexed books from 2005 to present, 10,000 new books are added each year (the majority in the arts, humanities and social sciences fields). The index introduced more than 15.7 million new cited references to Web of Science.[356] An analysis in 2013 revealed considerable issues in the citation data (e.g. inflation and inconsistency of publication counts) which impede its use for research evaluation purposes (Gorraiz *et al*. 2013).

Elsevier started their Scopus Book Titles Expansion Program in 2013 with the goal to index 75,000 titles by 2015, and 10,000 each year thereafter. The motivation to include books reportedly was that Scopus "more effectively meets both the discoverability and evaluation needs of book-based disciplines in the social sciences and arts and humanities". In June 2014, the program reached over 40,000 books from more than 30 major publishers of which about 25% were in social sciences and arts & humanities subject fields. In March 2015 the goal of 75,000 books was reached with social sciences and arts & humanities representing more than 55% of the titles; indexing another 45,000 books has been intended for this year. (cf. Elsevier 2013, 2014, 2015)

A recent venture of Springer in collaboration with Altmetric is Bookmetrix[357] which aims to track scholarly and social media activities around Springer books, including downloads, tweets, reviews, citations and readers (based on Mendeley data).

Brief summary of the specificities

Bibliometric and other studies point out specificities that set the publication culture of the humanities apart from other disciplines, in particular the natural sciences, but also in some respects the social sciences. They include that the humanities scholarly literature

- is under-representation in the main citation indexes Web of Science (Thomson Reuters) and Scopus (Elsevier), and that the indexes are biased towards English-language publications,
- presents a substantially different composition than other disciplines in that monographs, book chapters, conference papers, reports and non-scientific literature play a greater role than journal publications,
- usually cites few journal papers (in 2007 over 90% of the cited literature were not papers), and more often much older literature than other disciplines,
- has a more local, context-specific character than natural sciences publications, i.e. are more often published in national journals, addressing the readership in the native language.

The specificities of the humanities publication culture may also play a role in the context of data sharing, re-use and citation, and need to be taken account of in metrics which are being developed for these practices. For example, important drivers of open data are journals which request that data underpinning papers is made publicly available. As journal papers are not the main form of publication in the humanities this development will have a much lower impact in this sector. But

[356] Thomson Reuters: Book Citation Index, http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/book-citation-index.html

[357] Springer: Bookmetrix, http://www.bookmetrix.com

archaeology is a multi-disciplinary field of research, involving branches of the applied natural sciences, so that some impact can be expected via journals of these branches. Another specificity of the humanities which should be considered is their stronger national focus, which in archaeology will mean the preference of national archives where researchers can deposit their data, with metadata in the native language.

9.2.2 Citation advantage for open access papers

The Open Access advantage is the hypothesis that papers which are published Open Access (OA) will receive more citations than papers published in subscription-based journals (cf. Swan 2010; Swan & Chan 2010). The main factor that contributes to the assumed advantage is that OA papers are made freely available online and, thereby, available to researchers (and other users) who do not have access to relevant literature through journal subscriptions.

The advantage therefore in the first place is an accessibility advantage in that the full text of OA papers can be accessed, downloaded, used and cited more easily by more researchers. Notably, the same applies to self-archived papers that are available from open access repositories ("green OA"). A further assumption is the early access advantage which means that papers which become available at preprint stage, immediately upon publication or after a short embargo period are cited more often than others.

It is of course not expected that OA produces an automatic citation boost for every paper. Rather that high-quality OA papers, like non-OA papers, will accrue more citations than papers of poorer quality. Furthermore it has been assumed that the OA advantage varies among different domains, e.g. be greater in the natural sciences and fast-moving fields or research than in the humanities, for instance.

The citation advantage for open access papers is now well established. The SPARC Europe "Open Access Citation Advantage Service"[358] documents 70 studies (2001 to present) across many disciplines of which 46 found a citation advantage, 17 did not confirm the advantage, and 7 were inconclusive, found non-significant data or measured other things than citation advantage for articles.

The advantage in terms of more citations for OA papers reported in the various studies is relatively low for papers in philosophy (45%), legal scholarship (49 to 56%) and political science (86%), also relatively low for biology (-5 to 36%) and mathematics (35 to 91%), while reaching high scores in computer science (157%), economics (308%) physics/astronomy (170 to 580%), medicine (300 to 450%) and agricultural sciences (200 to 600%). (cf. Swan 2010; SPARC Europe – OA Citation Advantage Service).

While these figures are impressive it should be noted that several studies did not find an advantage in some fields of research (e.g. distance education, orthopaedic research, ophthalmology, physiology) or even a negative result (e.g. pharmacy & pharmacology).

Unfortunately, none of the studies reported by the SPARC Europe – OA Citation Advantage Service relates to archaeology and only four addressed the humanities among other disciplines. These studies show that open access allows reaching a wider audience (i.e. more downloads) but may not

[358] SPARC Europe: The Open Access Citation Advantage Service, http://sparceurope.org/oaca/; the service continues the work of the Open Citation (OpCit) project: Reference Linking and Citation Analysis for Open Archives. The effect of open access and downloads ('hits') on citation impact: a bibliography of studies [2004-2013], http://opcit.eprints.org/oacitation-biblio.html

generate more citations to OA papers (Davis 2010 and 2011[359]) or books (Snijder 2010[360]). In specific cases non-OA humanities publications may even perform better with regard to citations. The study of Xu *et al*. (2011) analysed 12,354 papers OA and non-OA papers that were published in 2009 across 93 Oxford Open journals of several disciplines. OA humanities papers received fewer citations than papers published non-OA (a negative result of 49%), whereas OA social sciences papers showed a citation advantage of 88% (other disciplines 52% to 83%).

One major question is what will happen if OA publishing and self-archiving becomes ever more widespread. It seems clear that the competitive advantage of OA papers will decrease while still exerting pressure on non-OA journals to convert to OA (e.g. the author pays model). The early access advantage will remain in that papers which are openly disseminated faster than others (e.g. preprints or published without embargo) are likely to accrue citations faster, typically in the first years after publication.

Thus more pressure to make papers available openly and quickly can be expected. Overall we might also expect that high-ranked OA journals in terms of impact will trump low-ranked ones. For instance a study of over 200,000 papers published between 1996 and 2005 in 100 journals in botany, ecology, biology and multidisciplinary sciences (McCabe & Snyder 2014) found an average OA citation advantage of 8%. However, citations to high-ranked journals increased significantly with OA while low-ranked OA journals experienced a significant decrease in citations. Thus there will be increasing competition between OA journals for readers' attention, authors and references.

This will be a point of concern for the many OA journals that have been launched in recent years, especially also in the field of archaeology, classics and other domains of ancient world studies. Ancient World Online offers a list and links to over 1500 OA titles, up from about 1100 in July 2012.[361] Many of these OA journals and other serials will be difficult to maintain if they do not occupy a well-defined niche in terms of sponsors (e.g. a professional association), authors and readers. That OA journals and authors must promote their products in order to raise visibility, drive downloads and, eventually, citations, was one of the major themes of the ARIADNE Open Access Session at the Annual Meeting of the European Association of Archaeologists (EAA) on the 3rd of September 2014 in Istanbul (Geser 2014b).

9.2.3 Citation advantage for papers with underlying data

One expectation for open research data is that publications with such data will receive more citations than others for which the data is not accessible. The main assumption that this will be the case is that such publications are more transparent and reliable, as reported research results and conclusions can be scrutinized and, therefore, are more often consulted and cited. Furthermore, the data of the author/s as such may attract interest, for example as a source for further research, if it can be re-used easily. However there is little evidence as yet that data sharing drives citations of publications and/or the underlying data. Indeed, evidence is mainly available for some fields of biomedical research.

[359] Davis (2011, based on the PhD dissertation 2010) compared OA and non-OA papers across science, social sciences, and humanities journals. No OA citation advantage was found but the OA papers were downloaded more often and thus reached a broader audience within the first year.

[360] Snijder (2010) conducted an experiment on OA academic books that compared the effects of making books available through a repository, Google Book Search, and both channels, with 100 books in each sample. Google Books performed much better than the repository in promoting access. During the rather short nine months experiment no relation could be found between OA publishing and citation rates.

[361] AWOL - The Ancient World Online: Alphabetical List of Open Access Journals in Ancient Studies, http://ancientworldonline.blogspot.co.at/2012/07/alphabetical-list-of-open-access.html

Studies have shown that mandated data archiving by journals greatly improved the availability of data underlying publications in the field of population genetics (Vines *et al*. 2013) and biomedical microarrays research (Piwowar & Chapman 2010; Piwowar 2011). Widely referenced by open data advocates are two studies on the field of biomedical microarrays research that provide evidence that open data sharing is associated with a paper citation advantage. Piwowar *et al*. (2007) examined 85 microarray clinical trial publications and found that 41 trials with publicly available data received 85% of the aggregate citations. Publicly available data was significantly associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin.

In a more extensive survey Piwowar & Vision (2013) compared 10,557 microarray studies published in the years 2001-2009 that had citation counts in Elsevier's Scopus database and either did not make their data available or deposited it in one of the two most commonly used digital archives for such studies (ArrayExpress and Gene Expression Omnibus). Papers with openly available data (2617) received on average 9% more citations than those with unpublished data; those with data deposited in 2004-2005 had a citation advantage of about 30%. A manual review of 138 citations revealed that 6% were attributions for data re-use. Actually the survey found that a substantial fraction of archived datasets are re-used by third-party investigators and that the intensity of re-use has steadily increased since 2003. The re-use was distributed across a broad base of datasets, and in a very conservative estimate 20% of the datasets deposited between 2003 and 2007 had been re-used at least once by third parties. Authors published most papers based on their own data within two years whereas data re-use papers of third-party investigators continued to accumulate for at least six years.

Research Trends, a quarterly magazine that provides insights into scientific trends based on bibliometric analysis, conducted a study of references to data repositories in research papers (Huggett 2014). The study was motivated by the Force 11 community's Joint Declaration of Data Citation Principles which stressed the importance of data being *"considered legitimate, citable products of research"* (Force 11, 2014). Research Trends extracted from the DataBib catalog[362] the web addresses of 971 data repositories in different research fields and of various sizes; 48% of the repositories were located in the United States, while repositories in the UK and Germany each had a share of 9% (Canada 5%, India 3%, Australia 2%, China 2%, other or blank 22%).

To identify papers citing these repositories the Scopus database was used; the advanced search function of the database allows searching the reference fields of papers for websites. All URLs of the DataBib list were searched and the records of relevant publications extracted. It must be noted that the sample does not include publications in which the repository URL is included in the full text instead of the references or not given at all. Furthermore that the sample may include some references of the repository website with a purpose other than data citation.

Research Trends found 178,909 publications of the period 1996-2014 that referenced a data repository, which are 0.3% of the 55 million publication records in Scopus in 2014. 113,618 of the publications were articles, 37,410 conference papers, and 19,334 reviews. Between 2009 and 2013 an average annual growth (CAGR) of 19% was identified, 24% for articles, 7% for conference papers 7%, and 16% for reviews. The 178,909 publications received 1,879,964 citations by others, which represents a rather high average of 105 citations per publication. Based on an analysis of the publication titles Research Trends found that most of the publications addressed health-related topics.

The Research Trends study addressed papers with underlying data but did not investigate the question if there is a citation advantage for such papers. However, their results together with the results reported by Piwowar and colleagues (2007, 2013) suggest two assumptions: The amount of

[362] DataBib, http://databib.org; the catalog has been incorporated in re3data, http://www.re3data.org; it is no longer available on the DataBib website.

papers with underlying data deposited in a repository and the link to the data included in the list of references appears to be still rather low. Areas of research where a citation advantage may currently be identified are those which avail of state-of-the-art data repositories and depositing of data is mandatory. Some fields of the biomedical sciences fit these criteria, some fields of other sciences may fit too, but hardly fields of the humanities, or archaeology specifically. One exception may be comparing archaeological papers with data deposited in the repository of the Archaeological Data Service to other papers that did not make their underlying data available.

9.2.4 Citation of data

Concerning data citation we distinguish between two different forms with regard to the cited source of data: data that can be attributed to data producers (i.e. individual or groups of researchers), and databases which integrate data from different producers and are usually attributed to the organisation which curates and publishes the database. In the context of the current debate about data sharing of researchers our main focus are data producers.

We acknowledge however the tremendous service curators of research data databases provide to the research community. Some databases are heavily used and cited. To give but one example, a citation analysis of three databases curated by the US National Oceanographic Data Center (Belter 2014) showed that they are highly cited in the research literature. If the databases were counted as journal articles in Web of Science (WoS), two would have citation counts higher than 99% of all articles in Oceanography in the WoS Science Citation Index from any single publication year from 1995 to the present. Based on a different method, which involved full-text searches of websites of major publishers and Google Scholar each of the three databases would be ranked in the top 1% for citation counts of all articles published in Oceanography during the same year, two even in in the top 0.1%.

The most relevant source for citations of individual or groups of researchers (data producers) currently is Thomson Reuters' Data Citation Index (DCI). The DCI was launched in October 2012 and aims to provide *"a single point of access to quality research data from repositories across disciplines and around the world"*.[363]

According to a study by Torres-Salinas *et al*. (2013) the DCI in the first half of 2013 contained 2,623,528 data records of which 92% were produced between 2000 and 2013. Concerning discipline coverage 80% were natural sciences, 18% social sciences, and 2% arts & humanities data records. 75% of all records in the DCI database came from the Gene Expression Omnibus, UniProt Knowledgebase, PANGAEA and U.S. Census Bureau TIGER/Line Shapefiles. Concerning the other 25% there were 29 repositories that contained at least 4000 records and 64 that contained at least 100 records. The DCI covers three different types or records: data sets, data studies, and data repositories. Torres-Salinas *et al*. identified 2,475,534 data sets (94%), 159,280 data studies (6%), which describe studies or experiments (6%), and 96 data repositories, holding data sets and/or studies which are often indexed as a single record because indexing at a more granular level is not possible. The arts & humanities were present with 51,444 entities, 44,588 of which data sets, 6847 data studies and 9 repositories; there was a larger presence of data studies (13%) than in the DCI sample overall (6%).

Peters *et al*. (2015) found that across the Data Citation Index (DCI) sample of about 4 million data records from the period 1960-2014 about 85% were not cited, although, they identified a growing trend in citing data published since 2007.[364] The researchers looked for data items which received

[363] Thomson Reuters: Data Citation Index (DCI), http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/data-citation-index.html

[364] Interestingly, the Figshare repository accounted for almost 25% of the DCI database. Figshare is operated by the Digital Science brand of Macmillan Publishers. It is a "low barrier repository" which encourages

two or more citations of which there were 10,934 items, 907 with a DOI and 8027 an URL only. This sample included 21 items in the Archaeology Data Service which together had 75 citations in the DCI. Furthermore the researchers looked for such data items which also had at least one altmetrics score (e.g. tweet, Facebook like or comment, Mendeley reader, etc.) in the leading altmetrics service PlumX[365]: only 301 such data items were found.

Some subject areas attracted fewer citations than altmetrics scores, for example, 10 items in the Archaeology Data Service (with DOIs) received 47 citations compared to 139 scores, 4 cell biology items with 13 citations and 383 scores or 26 genetics & heredity items with 492 citations and 654 scores (the two cases with URLs only). According to the authors this demonstrates that altmetrics can complement impact analyses of data sharing and citation. However they emphasise that research data citations need to be studied in more detail, for example with regard to their rationale, relation to data types and structures, and data sharing, usage and citation activities.

9.2.5 Altmetrics (alternative metrics)

The novel research field of altmetrics (alternative metrics) develops and applies non-traditional measures of the dissemination, reach, use and impact of scholarly work based on online activities and interactions. This includes measures that are directly related to research outputs (i.e. views and downloads from publishers' websites or research repositories) as well as communication on the scholarly and social Web[366]. In general altmetrics address the issue that the evaluation of researchers based on their publications and citations by other scholars nourishes a limited, one-dimensional view of relevance which does not consider activities such as education and engagement with the public.

Altmetrics fuelled by increasing competition

While some advocates present altmetrics as a breakthrough towards measuring Web-based "open science" practices, the fuel that is driving the adoption of altmetrics is the increasing pressure for recognition and impact at all levels (cf. Taylor 2013). Research funders wish to see impact of their grants, university and research institutes worry about their ranking in national and internal rankings[367], journal publishers and editors are competing for authors and readership, and scholars are of course keen to demonstrate that their work is recognised, in the scholarly record as well as in other channels of communication. There is thirst for scholarly recognition in general, somewhat unfriendly called "narcissism" in one study of altmetrics (Wouters & Costas 2012). However arguably more important is that recognition in different arenas of communication may in some way or other translate into advantages in the fierce competition for jobs and research grants.

While a new field for media analysts and certainly well meant, using information on Facebook, Twitter *et al*. for impact metrics and emphasising the "immediacy" such impact is a disgrace to scholarship. Social media posts on Twitter or Facebook have a "shelf life" of about 3 hours (Bitly 2011), a lifespan that is *"nasty, brutish and short"* (Cervieri 2011). In comparison, a lot of literature cited in the humanities has been published many more years ago, in Greek and Latin philology on

researchers to archive smaller datasets and content. The repository assigns a DOI to each object to ensure citability and the data/content is shared under Creative Commons licenses. Figshare's social media component allows for promotion of the data/content and generates metrics based on uploads, sharing, etc. However, Peters *et al*. (2015) found that no item from Figshare was cited at least twice in the DCI.

[365] PlumX, http://www.plumanalytics.com

[366] Overviews of the many altmetrics indicators and sources are provided by Buschman & Michalek (2013) and Konkiel (2014); the largest set of is covered by Plum Analytics, http://www.plumanalytics.com/metrics.html

[367] For instance, the Snowball Metrics group of research-intensive universities in the UK, United States, Australia and New Zealand has included some altmetrics in their "Recipe Book" (Snowball Metrics 2014: 97-100).

average 25 years, not including the original texts (Crawford 2013), while in natural history taxonomists still reference scholarly work published since the 18th century (e.g. Carolus Linnaeus' *Systema Naturae*, 1735). The average "shelf life" of scientific content of course is more in the range of a few years. It varies according to discipline, is relatively short in the medical sciences whereas in the humanities references to older literature can be found more often (cf. *Section 9.2.1* above).

Some insights and critical aspects of altmetrics

What is particularly important about altmetrics is that they reflect the growing importance of the Web as the environment in which scholarly work is communicated, accessed, read, referenced and discussed, including by other participants than researchers. *"Indeed, it cannot be stressed highly enough that altmetrics are about more than just the numbers: the greater context and content of web activity is also hugely meaningful"* (Mounce 2013: 16).

The key to understanding and applying altmetrics appropriately is the concept of usage which covers many activities (e.g. viewing, downloading, bookmarking, commenting, recommending, etc.), whereas citation only means that a research publication has been referenced in another publication. As a matter of fact, usage events greatly exceed the number of citations. This also, and maybe even more so, can be expected for data resources.

Altmetrics are not alternatives but provide additional, potentially complementary measures to citation-based metrics of scholarly impact. Therefore it can be argued, despite many claims to the contrary, that altmetrics may have little effect on established frameworks of research evaluation. Adoption by the research community will also be low as most scholars see citations as the core currency of prestige and regard social media with suspicion.

In any case, altmetrics should be coupled with on-the-ground, empirical work by social sciences and humanities researchers on the scholarly and social impact or research communication; "*without this work – and without the mutual engagement of the humanities and altmetrics – the analytical part of altmetrics will only ever be a limited proxy for social impact"* (Taylor 2013). Social impact can mean many things, for example producing attention for current research topics, products, activities and events (e.g. in the case of Twitter feeds).

It should be noted that the main sources of altmetrics data (e.g. Facebook, Mendeley, Twitter and others) as well as altmetrics service providers, are private, for profit enterprises. The leading providers are PlumX/Plum Analytics[368] (owned by EBSCO Information Services) and Altmetrics.com[369] (own by Macmillan Publishers via their Digital Science[370] brand). While they mainly provide services to publishers, research libraries and other institutions, ImpactStory[371] offers researchers a "live CV" to *"share the full story of your research impact"*. ImpactStory has been developed based on research grants (NSF, JISC and Sloan Foundation) and is run by a not-for-profit organisation that currently charges $60/year for their service. Academic researchers who conduct studies on altmetrics topics often use data of one or more of these service providers.

Non-commercial organisations and interest groups currently mainly centre on conceptualizing data-level metrics and consultations with stakeholders, for example, the National Information Standards Organization (NISO) - Alternative Assessment Metrics Initiative[372], the joint Research Data Alliance / World Data System - Data Publishing Bibliometrics Working Group[373], and the Consortia Advancing

[368] Plum Analytics, http://www.plumanalytics.com

[369] Altmetric.com, http://www.altmetric.com

[370] Digital Science, http://www.digital-science.com/products/

[371] ImpactStory, https://impactstory.org

[372] NISO Alternative Assessment Metrics Initiative, http://www.niso.org/topics/tl/altmetrics_initiative/

[373] RDA/WDS Publishing Data Bibliometrics WG, https://rd-alliance.org/node/44532

Standards in Research Administration Information (CASRAI) Dataset Level Metrics Subject Group[374]. One exception is the Making Data Count project[375] in which the California Digital Library, Data Observation Network for Earth (DataONE) and Public Library of Science (PLOS) define and implement a practical suite of data metrics in the existing PLOS article-level metric tool.

Some results for humanities and archaeology

Costas *et al*. (2014) analysed the presence of different altmetrics provided by Altmetric.com for about 50,000 publications of different disciplines covered in Web of Science. The study found a relatively low but growing presence and density of social media activities around the publications. The biomedical & health sciences and social sciences & humanities publications showed the highest presence of at least one altmetrics score, 22.83% and 22.50% respectively, followed by life & earth sciences (15.94%) and natural sciences & engineering (8.98%). Only the social sciences & humanities had a (slightly) higher density of altmetrics per publication than citations (1.04 versus 0.9).

Fausto *et al*. (2012) conducted an analysis of blogging based on data of Research Blogging[376], an aggregator of scientific blog citations of peer-reviewed publications. In 2012, Research Blogging covered 1,236 weblogs with about 27,000 citations. The largest shares of citations were in the category Health & Medicine and Psychology, 15% and 13% respectively. The category Anthropology, which may include archaeological publications, reached 4%.

Björn Hellqvist (2010) studied citation practices in the humanities and emphasises that the relevance of citation analysis will be limited if their specific stylistic, epistemological and organisational background is not taken into account. More practically oriented, Björn Hammarfelt (2014) analysed the altmetric coverage of humanities 310 papers and 54 books published by Swedish researchers during 2012. Not surprisingly he found that Mendeley had the highest coverage of the papers (60%), followed by Twitter (20%), and that books were also often tweeted by publishers and authors. Few of the publications however were mentioned on weblogs and Facebook.

A favourite among the altmetrics studies is the impact of Twitter on views and citations of research papers of which some show a slight effect in this regard (e.g. Eysenbach 2011; Shuai *et al*. 2012; Thelwall *et al*. 2013 [covering also 19 other social web services]). More insights offer studies about how and why scholars use Twitter (e.g. Priem & Costello 2010). Survey based insights concerning archaeologists' use of Twitter, and what they think about this source of information, are provided by Lorna Richardson (2012, 2014 [survey results 2011-2013]). In general, users appreciate accurate tweets with links to further information by acknowledged institutions and individuals with real-life reputation and first-hand experience; for many it allows to follow events they cannot attend and/or stay up-to-date during periods of unemployment (cf. Richardson 2013b).

At the ARIADNE Open Access Session at the Annual Meeting of the European Association of Archaeologists (EAA) in Istanbul (3 September 2014), Doug Rocks-Macqueen[377] reported that Open Access Archaeology[378] since 2012 uses Internet bots to collect information about open access papers from various sources and automatically posts it on Twitter and Facebook, one post every two hours. On Twitter this promotion of OA papers has 5062 "followers" of whom about halve seem to be

[374] CASRAI: Dataset Level Metrics Subject Group, http://casrai.org/standards/subject-groups/dataset-level-metrics

[375] Making Data Count, http://mdc.lagotto.io

[376] Research Blogging, http://researchblogging.org

[377] Doug's Archaeology, https://dougsarchaeology.wordpress.com

[378] Open Access Archaeology, http://www.openaccessarchaeology.org

active. The effect is 100-200 clicks a day, i.e. people looking up one or more of the sources and probably accessing the OA papers.[379]

Peters *et al*. (2015) report that of 301 data items in Thomson Reuters' Data Citation Index with two or more citations and at least one altmetrics score 10 were deposited with a DOI in the Archaeology Data Service. These items received 47 citations in publications but had 139 altmetics scores. In the REWARD[380] project seven archaeologists prepared and deposited data in UCL's institutional repository and published a data paper in the Journal of Open Archaeology Data (JOAD). Within one year of publication the seven papers attracted 2629 views and 485 downloads. Most of the papers scored well in terms of ImpactStory altmetrics but within one year only one data paper was cited in a research article (Hole 2013).

9.2.6 (Alt)metrics for data archives

Importance of data archive usage metrics

Citations of both papers and data are generally rather low. For instance, article-level metrics for 63,771 papers in Public Library of Science (PLOS) open access journals in November 2012 show that they received about 124 million HTML page views, 27.7 million PDFs of papers were downloaded (22% of the views), but only 375,000 CrossRef citations of the papers were found, 0.3% of the views or 1.35% of the downloads (Pattinson 2012: 4). Thus a lot of consultation of papers, which may influence research activities, leaves no traces in citations. Peters *et al*. (2015) found that 85% of the about 4 million data records of the period 1960-2014 in Thomson Reuters' Data Citation Index were not cited.

The figures make clear the importance of metrics that are not based on citations but other forms of usage of publications and data. Such usage is especially relevant for data archives that lack means to track their impact through data citation in research publications. Through presenting other usage metrics/figures they can demonstrate relevance to the research community and funders, plan archive development, and mobilise new data contributions (Konkiel & Scherer 2013).

Rich and solid figures of data usage in various forms can support the sustainability of archives based on their recognition by the research community and funders. For potential new data contributors the figures can demonstrate that the shared data is perceived as valuable and that taking the time to share the data is worthwhile. Beside quantitative figures of growth exemplary stories of individual researchers and institutions that have benefitted from data sharing will be helpful to make the case and encourage others.

Digital archives should in the first place focus on indicators and evidence that are directly related to the research output they curate and make accessible (e.g. searches, views, downloads). Systems log-analysis and web analytics can allow correlating user activities with the effectiveness of the information system and relevance of the data being served. Furthermore user requirements and satisfaction surveys can be conducted on a regular basis or following the introduction of new services. Other proposed indicators, especially social media altmetrics, are currently relevant mainly as illustrative evidence of usage and may be convincing only for large and heavily used archives.

Particularly important are indicators that can expose the value that data curation adds to datasets such as organising the data in ways that ease discovery, access and re-use. This includes, but is not limited to, their work with researchers who deposit data, quality checks, work on metadata,

[379] Open Access Archaeology also provides "digests" of OA papers (currently at no. 697) on their Tumblr website http://openaccessarchaeology.tumblr.com

[380] REWARD - Researchers using Existing Workflows to Archive Research Data (UK, 10/2011-03/2012), http://www.ucl.ac.uk/reward

controlled vocabulary, indexing and interlinking. Indicators for value added are also important for raising the status of data curators in the research system.

Curators of the data archive of the US National Center for Atmospheric Research (NCAR) emphasise that current approaches to demonstrate the value of data centres *"have a difficult time capturing the nuanced work of data curators, including how shifts or changes in services impact end user consumption. This leaves the services and infrastructures, such as those developed by staff at the RDA [Research Data Archive], invisible to promotion or tenure awards at an individual level, and often ignored or overlooked by federal funding at an institutional level"* (Weber *et al*. 2013).

It should be noted that it is of course not possible to evaluate data re-use based on the mentioned usage indicators, only that the data has been searched, data records viewed and some or all of the data downloaded. What the users do with the downloaded data remains largely outside of what a digital archive can evaluate directly. One exception is if data users deposit a new dataset which combines derived and new data again in the same repository.

What an archive may attempt is to collect information about published research which demonstrates that data from the archive has been used. It is expected that increasing use of unique and persistent identifiers will ease this task in so far as archives assign such identifiers to deposited data, researchers cite re-used data with the DOI, and services are available that allow easy identification of such citations in the literature.

Current situation and examples of archive altmetrics

Many digital archives currently do not present usage metrics such as visits, searches, items viewed and downloaded. A survey of 35 data repositories of different disciplines found that 17 did not display any usage information (i.e. statistics/charts of views or downloads), 11 presented aggregated general metrics at the repository-level, and 7 some form of data-level metrics (Costas *et al*. 2013: 23). The Making Data Count project[381] surveyed data managers on metrics their repositories track and expose (Kratz & Strasser 2015b). A majority of the 71 respondents reported tracking landing page views and downloads, about 65% and 85%, respectively. Citations to the repository as a whole and/or of individual datasets were tracked by 20% of the repositories. Of the repositories that each metric only about one third display the results to site visitors. We may assume that in many cases the repository usage is rather low and therefore available figures are not shown. In general an archive that is used by many to look for relevant data will more likely present usage figures to demonstrate success and attract more users.

Below we highlight examples of data archives that have developed data usage indexes and other methods to prove their case:

US National Center for Atmospheric Research (NCAR)

The National Center for Atmospheric Research (NCAR) has developed a Data Usage Index for their Research Data Archive (Weber *et al*. 2013). The overall aim of the developers is a suite of metrics that can expose the value that data curators add to a dataset. Their conceptual framework acknowledges specificities of data curation systems and products which impact on how users search and access data and, consequently, the usage metrics a digital archive can present. One study applied the NCAR Data Usage Index on three of the most heavily used datasets which represent the diversity of the archive holdings (Weber & Thomer 2013). Each of the datasets is assigned one unique identifying number but composed of numerous files, which users often subset or download in various combinations.

The study results challenged assumptions about the two ways user can access data, via the archive website interface or advanced programmatic, API-based user access. Concerning the capability to

[381] Making Data Count, http://mdc.lagotto.io

sub-set, i.e. more selective data downloading, heavy use of course reduces the overall amount of data downloaded from an archive. The researchers warn that archives that do not allow for sub-setting but tout high download figures *"may actually be confusing the inefficiency of their architecture, with their impact on a user community".*

Global Biodiversity Information Facility (GBIF)

The Global Biodiversity Information Facility (GBIF)[382] is acknowledged as the core integrating database of worldwide biodiversity data, 579 million species observations (occurrences) of over 1.6 million species from 15,234 datasets of 770 data publishers. Nevertheless in recent years GBIF felt the need to demonstrate their relevance to secure continued support from stakeholders (especially data publishers) and long-term funding of their operation. One concern was that GBIF Secretariat knew little about the access and use of GBIF-mediated data for research purposes.

In 2009, the creation of a GBIF Data Usage Index was suggested and the developed index published in 2011 (Chavan & Ingwersen 2009; Ingwersen & Chavan 2011). The index consists of 14 quantitative indicators that capture different ways that GBIF-mediated data are searched and accessed, mostly based on a combination of web-analytics and log-analysis from the GBIF servers. Beside information of the GBIF-mediated data (data publishers, species covered, types of datasets, etc.) the index covers user interest and usage event data such as records searched, viewed and downloaded with regard to frequency and density, among others criteria.

Since 2008 the GBIF Secretariat also documents research publications that use GBIF-mediated data, a cumbersome activity, but until April 2015 resulting in 1409 identified publications (Schigel 2015: 15). The GBIF website features use cases from these publications (currently 24 cases). Moreover, in 2015 the GBIF Science Committee revamped their Ebbe Nielsen Prize for outstanding contributions to biodiversity informatics (awarded since 2001) as the GBIF Ebbe Nielsen Challenge[383]. The challenge seeks to encourage innovative uses of GBIF-mediated data.

Interuniversity Consortium for Political and Social Research (ICPSR)

ICPSR is an international consortium of more than 700 academic institutions and research organisations united by the mission to advance and expand social and behavioural research, and to act as global leader in data stewardship. ICPSR maintains a data archive[384] of more than 500,000 files of research, and has a Bibliography of Data-Related Literature of over 60,000 works using data held in the archive. The archive website captures usage data such as unique sessions of users and the names of ICPSR member institutions where data downloads occur. The bibliography more broadly tracks data usage and impact in the literature, including academic publications as well as government and agency reports, magazine and newspaper articles and audio-visual products.

Item records for publications link back to related datasets and the archive encourages citation of both data and related publications. In practice the curators of the ICPSR bibliography must invest much effort to identify data usage because explicit references are difficult to find in the literature or academic databases. Since 2008 the ICPSR archive assigns DOIs, both at the collection and the study level, which will improve the chance to identify data usage. The ICPSR collaborates with Thomson Reuters' Data Citation Index in this regard.

This summary is based on Konkiel (2013) and Moss (2013). The ICPSR archive and bibliography have been used in studies on data availability over time (Pienta *et al*. 2008), publication metrics (Pienta *et al*. 2011), and data re-use and citation (Konkiel 2013).

[382] GBIF, http://www.gbif.org

[383] GBIF Ebbe Nielson challenge, http://gbif.challengepost.com

[384] ICPSR data archive, http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp

Archaeology Data Service (ADS)

Usage figures for the website of the Archaeology Data Service (ADS) are available from surveys conducted in 2012 and reported in Beagrie & Houghton (2013a: 33-34). ADS has a broad user group which goes well beyond academia: 38% of users are conducting academic research, 19% private research, and 8% commercial research; 11% are heritage management users; 17% use ADS for general interest enquiries, 6% to support educational activities, and 1% for family history research. In 2012, the ADS website had 170,757 visits, served a total of 1,075,669 page views (not counting "bounces"), and 173,818 downloads of material, not including downloads via search engines (which are several times higher). 44% of website visits in 2012 were return visits, and over 11,000 visitors consulted the website more than 51 times. Concerning page views, 333,813 were Archsearch records or map views, 184,464 content of project archives, 293,826 of reference resources, 165,965 journals and series, and 91,626 grey literature, i.e. unpublished archaeological reports.

During the 23 month period from February 2013 to December 2014, the ADS website had 693,622 unique visitors who carried out a total of 4,620,070 actions including 4,244,893 page views and 324,492 downloads. During this period ADS modified the website's direct access filter functionality, allowing the website metrics to take account of users finding PDFs and images directly via Google. This has resulted in a dramatic difference between the average monthly users in the first eight months of the 23 month period (February - September 2013) and the figure for later months, 14,000 versus 39,000.[385]

The ADS presents detailed weekly access statistics, for example, between the 5th and 11th of October 2015 the website had 12,659 visits, over 50,000 page views, and about 3500 downloads[386]. Moreover the website shows archive-level usage figures for each of the deposited project archives, theses, journals & series and bibliographies (but not "grey literature", i.e. fieldwork reports). This includes visits to the archive, page views within the archive, file downloads, and totals. The figures cover the period since July 2011 or since the online release of the project archive thereafter.[387]

In October 2014 the Archaeology Data Service, in collaboration with *Internet Archaeology*, introduced a Digital Data Re-use Award. The award is intended to highlight the wide range of research carried out that re-uses data hosted at the ADS and raise awareness of the research potential of data re-use in archaeology and beyond (ADS 2014 and ADS 2015b).

9.2.7 Making Data Count survey results on data metrics

The Making Data Count project[388] in November/December 2014 conducted a pair of online surveys for researchers and data managers on data sharing behaviours, data discovery and use, and metrics (Kratz & Strasser 2015b). 73 data repository managers and 247 researchers contributed. Most responses were received from the United States followed by the United Kingdom, data managers 72% and 11%, researchers 57% and 14%, respectively. Data managers represented primarily academic (64%) and government-run (22%) repositories. Most of the researchers were academics (78%). More than half (53%) were biologists, but also researchers in environmental and social sciences were well represented (17% and 10%, respectively). Among the respondents were principal investigators (42%), postdocs (21%) and graduate students (19%). Below we mainly summarise the survey results concerning data metrics which are based on responses from 71 data managers and 247 researchers:

[385] ARIADNE D5.2 Initial Report on the Assessment of Online Access, January 2015, p.19.

[386] ADS Access Statistics, http://archaeologydataservice.ac.uk/about/accessStatistics

[387] For example see: Roman Amphorae: a digital resource, University of Southampton, 2005 (updated 2014), http://archaeologydataservice.ac.uk/archives/view/amphora_ahrb_2005/stats.cfm

[388] Making Data Count, http://mdc.lagotto.io

- *Citations:* Were by far the most relevant metric to both researchers and data managers, 85% of researchers and 61% of data managers ranked citations as the most interesting metric. Unfortunately citations are most challenging to capture; about 20% of the data managers reported that their repository collects citations.
- *Downloads:* Were the second-choice metric for researchers, while data managers – who were given a longer list to rank – put them in the middle, although 85% of repositories track them.
- *Page views:* Majorities of both researchers and data managers ranked page views as the least relevant metric; obviously reading information about a dataset on a landing page is not considered as a meaning usage metric for a dataset.
- *Social media:* The article does not present figures on social media but the authors state: *"We should not worry excessively about capturing social media (Twitter, Facebook, etc.) activity around data yet, because there is not much to capture. Only 9% of researchers said they would 'definitely' use social media to look for a dataset"*.

In summary the authors consider: *"In the short-term, page views and can be de-emphasized because of low status and lack of data-related activity respectively. While challenging, citations should be emphasized and collected as best as possible. Downloads should be emphasized as, at present, a happy medium: both reasonably valuable and reasonably easy to measure"*.

9.2.8 Academic credit for research software

Christine Borgman notes that data seldom stand alone, separable from various technical requirements for data re-use: "*Rarely are data self-describing, and rarely can they be interpreted outside their original context without extensive documentation. Interpreting scientific data often requires access to papers, protocols, analytical tools, instruments, software, workflows, and other components of research practice – and access to the people with whom those data originated. Sharing data may have little practical benefit if the associated hardware, software, protocols, and other technologies are proprietary, unavailable, or obsolete and if the people associated with the origins of the data cannot be consulted*" (Borgman 2015: 16).

Unfortunately, however, little research on such technical requirements has been conducted as yet in order to inform the discussion about open data and data re-use. One area that is discussed intensively is research software, specifically in fields of computational sciences.

In this study we will not address the many topics and issues concerning the role of research software in "open science". The intent of this section is mainly to highlight that besides open data it is important to consider also the means that are necessary for re-using data effectively. Sustained and readily available research software is a prime example. Like data also research software poses issues of appropriate academic recognition and reward.

Research software and reproducible research

Research software means software that is used for applications which support scientific purposes. Two main categories can be distinguishes: a) Scientific software that is produced to address specific scientific questions; b) Software that is created or adapted to support broader needs and different tasks of research communities: examples are utility software packages, software of scientific workflow systems or scientific gateways. Katz *et al*. (2014) discuss issues of different types of software with regard to collaborative development and sustainability.

In this section we mainly address scientific software that is developed for specific research problems. Some characteristics of such software are:

- The development requires scientific background and therefore is mainly carried out by domain scientists or teams of scientists and developers;

- the software pertains to specific domains, hence tends to be in a niche and have a small developer and user community, maybe only one or a few research groups; some software may however become part of an utility software package or evolve and support different research tasks;
- long-term sustainability/availability of the software is a challenge because of the small developer base, dependence on project funding, and shifts in research interests.

Scientific software that is well-maintained, accessible, documented and referencable is a key requirement for reproducible and extensible research. For extension the software should ideally be available open source. For reproduction and verification of reported research results a lot a information about the specific installation, control parameters etc. must be documented as well as the data used for the research made available.

In reality even the software as such is often not easily available (Veretnik *et al*. 2008) and the source code *"is commonly withheld, effectively rendering these programs 'black boxes' in the research work flow"* (Morin *et al*. 2012). An example can illustrate the difficulty and maybe exaggerated expectations concerning reproducible research: The EURASIP Journal of Advances in Signal Processing considered making reproducibility an important factor for accepting submitted papers. Therefore a special issue of papers was planned which could prove that their research results are reproducible, based on a number of criteria such as availability of software code, data, etc. 34 papers where submitted but only four published because the others *"were immediately rejected as they did not show any sign of reproducibility at all"* (Rupp *et al*. 2011).

A fair share of recognition for research software

Developers of research software demand a fair share of recognition for their contribution to the progress of the scientific enterprise (Jackson 2012; Todorov 2012). Similar to the situation with data, this is about a re-distribution of academic credit away from scholarly publications to shared means that enable new or enhanced ways of carrying out research. This necessary shift and the requirements for it are mostly discussed in the computational sciences, referring to the objective that research should be reproducible and extensible (Ince *et al*. 2012; Stodden *et al*. 2013a/b; Stodden & Miguez 2014).

While the request of credit for sharing research software is legitimate, at the same there is a growing amount of immature and abandoned software produced by academic developers that fill up open source software sharing platforms (e.g. GitHub, SourceForge and others). Obviously in many cases the goal has not been a working software tool but a number of publications around the tool development. As Hafer & Kirkpatrick (2009) note, *"Academic computer science has an odd relationship with software: Publishing papers about software is considered a distinctly stronger contribution than publishing the software".*

The imbalance in the academic recognition of the different scholarly contributions (publications, data, software) impacts on their curation, and software is a particularly difficult case in this regard (Wilson 2014). As for datasets the maintenance and sustained further development of relevant research software is arguably better served by dedicated organisations than individual developers.

There is some value in a quasi-Darwinian process of survival of the fittest, i.e. software that is recognised as useful and worth to be maintained. At that point an organisation with a broad or domain-specific scope could register, review and curate the software as well as make it readily available. Open source developer communities in a way fulfil such functions and are a model which works for some individual or suites of software.

In case of successful software development, sharing and application for research purposes there is the question of how to credit the software as a distinct scholarly contribution. This must be done by other researchers who used the tool and cite it in a standardised way in their publications. An editorial of *Nature Genetics* (2014) addresses requirements for giving credit and notes, *"Two needs stand out. First, code should have permanent identifiers such as the Digital Object Identifiers used by publishers, and authors of code should receive attribution for their programs as well as for their publications. Second, data sets and the code to handle data should be stored together, as metadata can cover both and the repositories then become attractors for communities, sometimes even evolving into environments where stored code can be run by third parties on stored data*".

The Software Sustainability Institute offers a detailed guide on how to describe and cite software (Jackson 2012). The requirements for software citation are generally the same as for data. Unfortunately little is known about a potential citation advantage for papers for which the special research software that was used for producing the results is available; one example is Vandewalle (2012) who provides some evidence for the field of signal and image processing.

9.3 ARIADNE data portal – Results of the lead user survey

The results of the ARIADNE online survey presented in the project Deliverable 2.1 (April 2014) allow a good general understanding of the needs of the archaeological research community with regard to e-infrastructure and services. For instance, the majority of the about 500 respondents wanted a portal that makes it more convenient to search for existing archaeological data that is stored in different archives/repositories (79%), and provides innovative or more powerful data discovery mechanisms (63%); a summary of some further results is given in *Section 7.3.1*.

However, the online survey was aimed to produce a broad overview of the current situation with regard to archaeological research data including, among other topics, data production, deposit and sharing/publication. Therefore the survey did not cover specific requirements which could guide the development of the ARIADNE data service portal. Such requirements were the focus of the "lead user" survey of existing portals, which is reported in the ARIADNE Second Report on Users' Needs (D2.2, February 2015).

The survey has been conducted by 23 archaeologists and data managers involved in the ARIADNE project who often use online resources for their work. They evaluated various archaeological websites (giving access to content/data of more than one institution or project) and some existing data portals of other domains. The survey participants looked for good practices and gave recommendations for services of the ARIADNE data portal.

The report in this chapter presents the method and results of the evaluation in April 2015 of the 34 survey recommendations by 28 experts of 21 partners. The evaluation has been carried out to focus the development of the services of the ARIADNE data portal on the most relevant services in the short to medium term. The evaluation results are taken account of in the focus area "Providing e-infrastructure services" (*Chapter 7*).

9.3.1 Brief description of the portals survey

The main objective of the survey was to provide more detailed evidence about user requirements as well as recommendations for the development of the ARIADNE data portal.

Participants

23 researchers and data managers involved in the ARIADNE project participated in the survey. These were "lead users" in that they make intensive use of searchable archives and other websites and therefore have a good understanding of the state-of-the-art and potential solutions that might serve their requirements even better (cf. the lead user concept developed by Von Hippel, 1986).

Of the 23 participants 18 were archaeologists with a very diverse background in terms of domains and geographic focus of research. 80% of the archaeologists had at least 10 years of professional experience. Besides using online resources since years many had also created and/or used data and digital tools in fields such as GIS/geospatial data, laboratory data, computational applications (e.g. statistical analysis), 3D models, various databases, metadata and conceptual knowledge.

Portals surveyed

A portal was defined as a website that provides access to content/data of more than one organisation or project, including digital archives which curate third-party data. 25 portals were evaluated, most of which were "international" in that they provide access to content/data from research not only in one country. More specifically,

- 15 entities focused only or to a large extent on archaeological content/data: websites of digital archives (ADS, Arachne, DANS, MAPPA, OpenContext, tDAR), scientific databases (ceraDAT, CHARISMA, PhytCore), content/data federations (CLAROS, Fasti Online, Pleiades/Pelagios, research programmes (Mapping Death, Portal to the Past), and one community website (Bone Commons).
- 5 entities were state-of-the art portals of other domains: CIARD-RING (a registry of food & agriculture research information services and repositories), EUROSTAT, GBIF - Global Biodiversity Information Facility, Global Change Master Directory (earth & environmental data), and SEDAC - Socioeconomic Data and Applications Center.
- Two e-research environments: Morphbank (analysis and sharing of biological research images) and Scratchpads (biodiversity/natural history research, with a focus on taxonomy),
- Three academic/professional networking and content sharing platforms: Academia.edu, Mendeley and ResearchGate, which are used also by archaeologists.

Survey participants' ideas and suggestions

The survey participants each evaluated 1 or more (some the same) portal and highlighted useful features which could serve as "good practices" when creating a portal for archaeological research data. The survey template also included a module "Ideas and suggestions" in which the participants were asked to suggest *"important and innovative features and functions which you would like to see in online archives and portals for researchers"*.

In total 127 suggestions were given, mostly short statements but also extended description of the expected advantage of a suggested portal feature, good practice examples, and potential improvements of current generation portals. The largest part (84) was suggested features (services, tools or other) of an online portal which would be most helpful for the lead users' research activities. The further suggestions (43) were various examples of good practice or addressed issues in current generation portals, potential improvements and other ideas and recommendations.

9.3.2 Survey recommendations for the data portal

From the lead users' various ideas suggestions the survey organisers at Salzburg Research derived and summarised 12 themes with in total 34 recommendations for the overall approach, design, services and specific features of the ARIADNE data portal. For each theme background, discussion and quotes of lead user statements are given the survey report. The list below presents the sets of recommendations:

1 - Implement a good overview and navigation of resources:

- Design a highly functional as well as attractive portal, e.g. with regard to an overview of searchable content/data and portal navigation. [1.1]

- Provide statistical information on the quantity and distribution of the data (e.g. per type of data, provider, country/area, period, etc.). [1.2]
- Enable a good understanding of the data resources and how they can be searched, accessed and (re-)used. [1.3]

2 - Ensure richness and added value of information:

- Seek to integrate many resources from affiliated institutions as well as from other initiatives within and beyond the archaeology sector (e.g. remote sensing or molecular biology data). [2.1]
- Emphasise the European/international dimension of the initial ARIADNE resources. [2.2]
- Create added value through linking to data and publication resources not held within the ARIADNE Registry. [2.3]
- Actively promote a culture of open data sharing in the sector, through making users aware of the importance of sharing (“give and take”) and point to good practice guides and community repositories (including archives in the ARIADNE federation). [2.4]

3 - Help users understand and use specific terminology:

- Provide aids for users not familiar with specific terminology/vocabularies (e.g. look up or dynamically invoked glossary or scope notes of thesauri). [3.1]
- Consider terminology support within search features as well as information pages for data resources, where possible in multi-lingual form. [3.2]
- Promote the provision of terminologies according to standards (e.g. ISO 25964) and in (semantic) web-based formats as required for machines and humans. [3.3]

4 - Integrate and link information resources:

- Make users aware that the portal does not store or create databases of primary data but operates based on the metadata of, and semantic relations between, data/content collections and items. [4.1]
- Provide integrated access to data and publications (i.e. include metadata of document archives and publishers). [4.2]
- Specifically support the inclusion and linking of information (metadata) from archaeological grey literature, which may be produced through Natural Language Processing (NLP). [4.3]

5 - Follow and promote Linked Data principles:

- Deploy Linked Open Data (LOD) to integrate information within the portal and to link to external resources which follow LOD principles (e.g. HTTP URLs). [5.1]
- Demonstrate advantages of Linked Data to encourage further uptake of LOD principles by archaeological institutions and projects. [5.2]
- Provide an LOD triple-store so that also external application developers can exploit resources of the portal for added value services (e.g. interlinking of databases). [5.3]

6 - Provide effective data search and filter functionality:

- Investigate the adequate set and implementation of search options directly with members of the user community (various option will be suggested which should be scrutinized). [6.1]
- Seek regular feedback on implemented solutions by the wider community. [6.2]

7 - Visualize data resources: Maps, timelines, and more:

- Offer search & filter functionality based on maps as well as date-ranges (timelines) & named periods. Both features will be required to allow users to select potentially relevant datasets. [7.1]
- Consider also visualizations based on Linked Data (e.g. density and web of links of the Linked Data graph around concepts). [7.2]

8 - Provide data preview and license information:

- Implement a data preview mechanism that enables portal users to check if discovered data resources are actually relevant for the study purpose. [8.1]
- Make users aware of available (or missing) license information of the data provider, e.g. to identify restrictions which impede re-use. [8.2]
- Enable portal users to filter available data according to allowed usage, from Public Domain to fully (c)-restricted. [8.3]

9 - Support different download options and external applications:

- Support download of data/metadata, including single item and bulk download as well as export/download in different open formats (for data that is not directly accessible, refer to the download page of the data repository). [9.1]
- Provide interfaces to allow external applications to exploit available data, metadata and conceptual knowledge resources (e.g. a well-documented API, OAI-PMH target, SPARQL endpoint). [9.2]
- Consider also RSS feeds for researchers and websites of research communities in particular subjects or geographic regions (e.g. alerts on new datasets in digital archives). [9.3]

10 - Provide personalized information services:

- Implement personalized information services which meet clearly evaluated needs of large segments of users. [10.1]
- Prioritise features which support users' control of specific searches, results and use of data. [10.2]
- Consider notifications (alerts), e.g. e-mail lists or RSS feeds, as good candidates for personalized services. [10.3]

11 - Enable linking and exchange of professional information:

- Enable linking of existing profiles of researchers on institutional websites and professional networking platforms to portal web pages (e.g. "MyData" pages). [11.1]
- Investigate how the portal might enable effective expert discussion. [11.2]

12 - Support online research work (e-research):

- Support integrated access as required for studying various research resources online (e.g. linking and comparing content). [12.1]
- Provide or link to tools which enable researchers to extract and combine data (e.g. images from different databases, numeric data to produce a derived dataset). [12.2]
- Provide or link to tools for data processing and analysis (e.g. statistical analysis, image data processing and analysis). [12.3]

9.3.3 Evaluation of the recommendations

To support decision making concerning the overall approach, design, services and specific features of the ARIADNE data portal, the recommendations of the portals survey also have been evaluated by the project consortium. This evaluation has been organised by Salzburg Research to provide consolidated results to the Technical Meeting in Leiden (19/20 February 2015) and subsequent work on the data portal. The results were not included in the project deliverable D2.2 but presented and discussed in Leiden as well as the Steering Committee Meeting and Technical Meeting (30 and 31 March 2015) in Siena.

The template based evaluation was conducted to assess the relevance and time-horizon of the recommendations for the data portal. Technical and data experts of 21 ARIADNE partners participated in the evaluation and provided in total 28 filled evaluation templates. Technical partners were invited to provide two independent evaluations (total 16 templates), data partners were asked to fill one template (total 12). The rationale behind this decision was that technical partners have a better expertise to consider how difficult it may be to implement certain features.

The evaluation template listed the 34 recommendations and the participants had to evaluate each recommendation with regard to relevance for the data portal (scores: high - 5 points, medium - 3, low = 1, not relevant = 0). Furthermore the time horizon was considered: “Within ARIADNE” meant that the suggested goal, service, feature etc. might be realized within the formal duration of the project (until January 2017), and for “Beyond ARIADNE” timespans of 1-2 years, 3-5 years and over 5 years could be chosen.

The table below presents the results across the 28 templates. To allow a good overview the text of the recommendations here is reduced to the key message (the full text of the numbered recommendations is given in the section above). The table includes the average relevance scores and evaluation with regard to the time-horizons (in %); the percentages of the options 3-5 years and over 5 years are summed up as over 5 years was either not chosen (20 cases) or only by a few respondents.

		Average relevance score (N=28)	Time horizon		
			Within ARIADNE (until Jan. 2017)	Beyond ARIADNE (1-2 y.)	Beyond ARIADNE (3+ y.)
1	**Implement a good overview and navigation of resources**	(4,1)			
1.1	Design a highly functional as well as attractive portal	**4,5**	**96%**	4%	0%
1.2	Provide statistical information on the quantity and distribution of the data	3,4	**75%**	18%	7%
1.3	Enable a good understanding of the data resources	**4,3**	**79%**	18%	4%
2	**Ensure richness and added value of information**	(3,4)			
2.1	Seek to integrate many resources from affiliated institutions as well as from other initiatives	**3,8**	21%	**50%**	29%
2.2	Emphasise the European/international dimension of the initial ARIADNE resources	2,9	**64%**	29%	8%
2.3	Create added value through linking to external data and publication resources	3,1	14%	**71%**	15%
2.4	Actively promote a culture of open data sharing in the sector	3,6	**75%**	7%	18%
3	**Help users understand and use specific terminology**	(3,6)			
3.1	Provide aids for users not familiar with specific terminology	3,6	**68%**	32%	0%
3.2	Consider terminology support within search features	3,6	**68%**	29%	4%
3.3	Promote the provision of terminologies according to standards	3,6	**61%**	29%	11%
4	**Integrate and link information resources**	(3,7)			
4.1	Make users aware that the portal does not store and create databases of primary data	**3,8**	**96%**	4%	0%
4.2	Provide integrated access to data and publications	**4,0**	**68%**	25%	7%
4.3	Specifically support the inclusion and linking of metadata from grey literature	3,4	39%	**43%**	18%
5	**Follow and promote Linked Data principles**	(3,9)			

5.1	Deploy Linked Open Data (LOD) to integrate information within the portal	**4,4**	**86%**	7%	7%
5.2	Demonstrate advantages of Linked Data to encourage further uptake of LOD principles	3,4	**61%**	25%	15%
5.3	Provide an LOD triple-store	**3,8**	**54%**	39%	7%
6	**Provide effective data search and filter functionality**	(3,8)			
6.1	Investigate search options directly with members of the user community	**4,1**	**75%**	25%	0%
6.2	Seek regular feedback on implemented solutions by the wider community	3,4	**61%**	32%	7%
7	**Visualize data resources: Maps, timelines, and more**	(3,8)			
7.1	Clear candidates for search & filter functionality: maps and date-ranges	**4,8**	**75%**	21%	0%
7.2	Consider also visualizations based on Linked Data	2,8	25%	**50%**	21%
8	**Provide data preview and license information**	(3,8)			
8.1	Implement a data preview mechanism	**4,0**	**68%**	21%	11%
8.2	Make users aware of available (or missing) license information of the data provider	**3,9**	**82%**	14%	4%
8.3	Enable portal users to filter available data according to allowed usage	3,6	**75%**	21%	4%
9	**Support different download options and external applications**	(3,4)			
9.1	Support download of data / metadata, including single item and bulk download	**4,0**	**79%**	21%	0%
9.2	Provide interfaces to allow external applications exploit available data	**4,0**	**75%**	21%	4%
9.3	Consider also RSS feeds for researchers and websites of research communities	2,1	25%	32%	**40%**
10	**Provide personalized information services**	(2,4)			
10.1	Implement personalized information services	2,4	29%	**57%**	11%
10.2	Prioritise features which support users' control of specific searches	3,1	39%	**54%**	4%
10.3	Consider notifications (alerts), e.g. e-mail lists or RSS feeds	1,9	21%	**39%**	36%
11	**Enable linking and exchange of professional information**	(2,1)			

11.1	Enable linking of existing profiles of researchers on institutional websites	2,6	32%	**43%**	25%
11.2	Investigate how the portal might enable effective expert discussion	1,7	11%	**57%**	32%
12	**Support online research work (e-research)**	(3,6)			
12.1	Support integrated access as required for studying various research resources online	**3,8**	36%	**54%**	11%
12.2	Provide or link to tools which enable researchers to extract and combine data	**3,7**	29%	**64%**	7%
12.3	Provide or link to tools for data processing and analysis	3,4	21%	**64%**	14%

General evaluation of the results

On a general level the results can be summarised as follows:

Some clear expectations:

- A highly functional as well as attractive portal
- Good overview of existing/available data resources
- Powerful search mechanisms, especially cross-searching repositories
- Search based on geo-location (maps) and date-ranges/chronologies (highest score)

Middle-high level:

- Terminology support
- Linking of / integrated access to data and publications
- Data preview mechanisms and licensing information
- Various data export and download options
- Interfaces to allow external applications data exploitation
- Support of online research work (e-research), but not within ARIADNE

Not appreciated

- Personalized portal services (e.g. alerts on possibly relevant new data)
- Linking of online professional information (e.g. researchers' profiles)
- Support of expert networking and discussion on the portal

Surprises

- High relevance attached to deploying Linked Open Data to integrate information within the portal and to link to external resources
- Relatively high scores for providing interfaces to allow external applications exploit available data, metadata and conceptual knowledge

Indeed, the portal should not be an "island", the wider information ecosystem and how to enable added valued beyond the portal should be considered.

Not a surprise is the relatively high, but arguable not high enough, importance assigned to promoting open data sharing via the portal (average score of 3.6). It seems that several evaluators could not envisage how a portal might serve effectively as a tool for data mobilisation. The main recommendations in this regard were to make users aware of a "give and take" obligation, point to guides to good practice in "open data", and suggest community repositories (including archives in the ARIADNE federation) where data can be deposited and made accessible.

The "Top 10" requirements

The "Top 10" requirements we highlight based on the following criteria: >60% of the evaluators attributed 4/5 of 5 possible points and considered the fulfilment of the requirement ≥68% "Within ARIADNE". Below the "Top 10" requirements are listed according to a logical implementation & usage sequence (*Imp = Importance, W.A. = Within ARIADNE*):

- Design a highly functional as well as attractive portal, e.g. with regard to overview of searchable content/data and portal navigation. – *Imp: 82%; 96% W.A.*

- Deploy Linked Open Data (LOD) to integrate information within the portal and to link to external resources which follow LOD principles (e.g. HTTP URIs and RDF). – *Imp: 79%; 86% W.A.*
- Enable a good understanding of the data resources and how they can be searched, accessed and (re-) used. – *Imp: 71%; 79% W.A.*
- Investigate the adequate set and implementation of search options directly with members of the user community (various option will be suggested which should be scrutinized). – Imp: 61%; 75% W.A.
- Offer search & filter functionality based on maps as well as date-ranges (timelines) & named periods. – *Imp: 93%; 75% W.A.*
- Provide integrated access to data and publications (i.e. include metadata of document archives and publishers). – *Imp: 64%; 68% W.A.*
- Implement data preview mechanisms that enable portal users to check if discovered data resources are actually relevant for the study purpose. – *Imp: 64%; 68% W.A.*
- Make users aware of available (or missing) license information of the data providers, e.g. to identify restrictions which impede re-use. – *Imp: 61%; 82% W.A.*
- Support download of data / metadata, including single item and bulk download as well as export/download in different open formats (for not directly accessible data refer to the download page of the data repository). – *Imp: 68%; 79% W.A.*
- Provide interfaces to allow external applications exploit available data, metadata and conceptual knowledge (e.g. well-documented API, OAI-PMH target, SPARQL endpoint). – *Imp: 61%; 75% W.A.*

Detailed discussion of the results

Theme 1: Implement a good overview and navigation of resources

The results confirm the goal to provide an attractive, highly functional portal that allows a good overview of existing archaeological data resources, is easy to navigate and understand with regard to how resources can be searched and accessed. Statistical information about the quantity and distribution of the data (e.g. per type of data, provider, country/area, period, etc.) appears to be considered less important. However statistical overviews were liked by evaluators of data partners (score 4.1, 12 templates) while not appreciated by technology partners (score 2.9, 16 templates). In general the figures show a high confidence of the partners that the basic features (cross-archive search, data discovery and access) will be realized within the formal duration of the project (until January 2017).

Theme 2: Ensure richness and added value of information

Here the evaluation shows doubts about the possibility to incorporate many data resources in addition to what the project partners aim to provide. While this goal is confirmed with a rather high score (3.8), the majority thought that it will require efforts beyond the project lifespan. But the participants noted that the portal should promote a culture of open data sharing in the sector (e.g. “give & take” principle, guides to good practice, community repositories). 75% thought that this is a priority of the project.

Theme 3: Help users understand and use specific terminology

The good score (3.6) of each of the three recommendations demonstrates the importance that the partners assign to the use of common terminology, which should be supported by the portal. Over two thirds of the participants thought that the suggested features (terminology aids, support within search) could be implemented in the project. About 60% expect that the project will promote the provision of archaeological thesauri and other vocabularies according to state-of-the-art standards (e.g. ISO 25964, Linked Data).

Theme 4: Integrate and link information resources

The participants thought that it is important to make users of the portal aware that it operates on top of underlying archives (i.e. does not store primary data) and uses metadata of, and semantic relations between, data/content collections and items (score 3.8., 96% within the project). Even more (score 4.0) considered that the portal should provide integrated access to data and publications (i.e. include metadata of document archives and publishers), but fewer (68%) expected that this will be realized in the project. Less importance was assigned to the specialist topic of using Natural Language Processing (NLP) to enable the inclusion and linking of archaeological grey literature.

Theme 5: Follow and promote Linked Data principles

Very well received was the suggestion to deploy Linked Open Data (LOD) for integrating information within the portal and linking external resources that follow LOD principles (score 4.4, 86% within the project). Also high is the score (3.8) for providing an LOD triple-store so that external application developers can use ARIADNE resources to interlink databases and create added value services. Somewhat less confident were the participants with regard to encouraging a wider uptake of LOD principles among archaeological institutions and projects (score 3.4), although about 60% expected that the project will promote this (e.g. through demonstration advantages of Linked Data).

Theme 6: Provide effective data search and filter functionality

The recommendations under this theme emphasise the need to involve the user community in the development of optimal portal functionality, especially search features. The high score of the suggestion to work directly with community members (4.1) confirms that the most project partners are aware of the importance of user-focused development for a successful portal. 75% expected such project activities and about 60% that regular feedback on implemented solutions by the wider community could be collected during the project.

Theme 7: Visualize data resources: Maps, timelines, and more

With a score of 4.8 appreciated most is the suggestion to offer search and filter functionality based on maps as well as date-ranges (timelines) and named cultural periods. 75% of the participants expected that this will be implemented by the project. Rather low is the score for visualizations based on Linked Data (2.8) which, however, is a topic for specialists.

Theme 8: Provide data preview and license information

The recommendations under this theme concern features that enable portal users to check if discovered data resources are relevant for their purposes. As this allows for effective data search and selection the suggestions were received very well. The scores are 4.0 for a data preview mechanism and 3.9 for information about data licenses (e.g. restrictions on re-use), and the expectation that these features might be implemented by the project were 68% and 82% respectively. Filtering data according to allowed usage, from Public Domain to fully (c)-restricted, still scored at 3.6 and with an expectation of 75%.

Theme 9: Support different download options and external applications

Highly appreciated would be different data/metadata export and download options (e.g. export in different formats, single item and bulk download), if possible on the portal. Furthermore interfaces for external applications to collect and use available data, metadata and conceptual knowledge resources. Those suggestions each got a score of 4.0 and most participants thought that they could be realized in the project, 79% and 75% respectively. Not considered as relevant was to provide RSS feeds for websites of research groups or individual researchers.

Theme 10: Provide personalized information services

Very low scored the suggestion to provide personalized information services, e.g. notification (alerts) about possibly relevant new data (2.4). There is also not a high average score (3.1) for features that would allow researchers control of searches or an individual space on the portal (e.g. “saved searches”, “MyData”/“OurData” pages). But such features were liked by evaluators of data partners (score 4.1, 12 templates) while not appreciated by technology partners (score 2.3, 16 templates). In general we assume that personalized portal services are not seen a priority of the project.

Theme 11: Enable linking and exchange of professional information

Overall least appreciated were suggestions to enable expert discussion on the portal or allow researchers to interlink their profile on an institutional website or professional networking platform with content on the portal. The latter suggestion was liked considerably less by evaluators of technical partners (score 2.1) than data partners (score 3.3, but still relatively low).

Theme 12: Support online research work (e-research)

The suggestions concerning support for research tasks on the portal got rather high scores, for instance 3.7 to provide or link to tools that enable extraction and combination of data, and 3.4 to provide or link to tools for data processing and analysis. But implementation of such support was not seen as a priority of the project.

10 References and other sources

The list of sources below includes all referenced literature and a selection of addition literature on topics covered in this report. Furthermore selected organisations, projects, data resources/services and standards mentioned in the report have been added.

3D-COFORM (EU, FP7-ICT, 12/2008-11/2012, created tools and a repository for 3D replicas of cultural heritage and archaeological artefacts), http://www.3d-coform.eu

3D-ICONS (EU, ICT-PSP, 02/ 2012-01/2015; produced and made available 3D models of archaeological and historical architecture for Europeana), http://3dicons-project.eu

Aalbersberg I.J., Heeman F., Koers H. & Zudilova-Seinstra E. (2012): Elsevier's article of the future enhancing the user experience and integrating data through applications. In: Insights - The UKSG Journal, 25(1): 33–43, http://insights.uksg.org/article/view/2048-7754.25.1.33/206

Academia.edu (academic information sharing and networking platform), https://www.academia.edu

ACCORD - Archaeology Community Co-Production of Research Data (UK, AHCR-funded Connected Communities project, 2013-2015), https://accordproject.wordpress.com

ACLS - American Council of Learned Societies (2006): Our Cultural Commonwealth. Prepared by the Commission on Cyberinfrastructure for the Humanities & Social Sciences. New York: ACLS, http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf

Acord S.K. & Harley D. (2012): Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0. In: New Media & Society, 15(3): 379–397, http://nms.sagepub.com/content/early/2012/12/20/1461444812465140.full.pdf+html

Adamick J. & Reznik-Zellen R. (2010a): Representation and Recognition of Subject Repositories. In: D-Lib Magazine, 16(9/10), http://www.dlib.org/dlib/september10/adamick/09adamick.html

Adamick J. & Reznik-Zellen R. (2010b): Trends in Large-Scale Subject Repositories. In: D-Lib Magazine, 16(11/12), http://www.dlib.org/dlib/november10/adamick/11adamick.html

Adams I.F., Miller E.L. & Storer M.W. (2011): Analysis of Workload Behavior in Scientific and Historical Long-Term Data Repositories. Storage Systems Research Center, Technical Report UCSC-SSRC-11-01, March 2011. http://www.ssrc.ucsc.edu/Papers/ssrctr-11-01.pdf

Adema J. & Ferwerda E. (2014): Publication Practices in Motion. The Benefits of Open Access Publishing for the Humanities, pp. 131-146, in: Dávidházi, Péter (ed., 2014): New Publication Cultures in the Humanities: Exploring the Paradigm Shift. Amsterdam University Press, https://oapen.org/download?type=document&docid=515678

Adema, Janneke (2010a): JALC User Needs: External Evaluation Report. SURFshare project Enriched publications in Dutch Archaeology. Leiden University, 10 January 2010, https://openreflections.files.wordpress.com/2008/10/wp8-external-evaluation-report.pdf

Adema, Janneke (2010b): Enriched publications in Dutch archaeology. In: Open Knowledge Blog, 20 September 2010, http://blog.okfn.org/2010/09/20/enriched-publications-in-dutch-archaeology/

ADS - Archaeology Data Service & Digital Antiquity (2013): Caring for Digital Data in Archaeology - A Guide to Good Practice. Oxbow Books

ADS - Archaeology Data Service & Digital Antiquity: Guides to Good Practice, online: http://guides.archaeologydataservice.ac.uk

ADS - Archaeology Data Service (2014): Digital Data Re-use Award. ADS Sound Bytes weblog, 7 October 2014, http://archaeologydataservice.ac.uk/blog/2014/10/digital-data-re-use-award/

ADS - Archaeology Data Service (2015a): ADS 3D Viewer. In: Sound Bytes from the ADS, weblog, 17 March 2015, http://archaeologydataservice.ac.uk/blog/2015/03/ads-3d-viewer/

ADS - Archaeology Data Service (2015b): We have a winner! Digital Data Reuse Award 2015. In: ADS Sound Bytes weblog, 12 October 2015, http://archaeologydataservice.ac.uk/blog/2015/10/we-have-a-winner-digital-data-reuse-award-2015/

ADS - Archaeology Data Service (n.d.): Guidance on the Deposition of Sensitive Digital Data, http://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy

ADS: Access Statistics, http://archaeologydataservice.ac.uk/about/accessStatistics

ADS: ADS-easy (system for submission of small-volume data sets), http://archaeologydataservice.ac.uk/easy/

ADS: Archaeological Site Index to Radiocarbon Dates from Great Britain and Ireland. Council for British Archaeology, 2000 (updated 2012), about 22,000 dates, http://archaeologydataservice.ac.uk/archives/view/c14_cba/overview.cfm

ADS: DataTrain - Open Access Post-Graduate Teaching Materials in Managing Research Data in Archaeology, http://archaeologydataservice.ac.uk/learning/DataTrain

ADS: Guidance on the Deposition of Sensitive Digital Data, http://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy

ADS: Library of Unpublished Fieldwork Reports, http://archaeologydataservice.ac.uk/archives/view/greylit/

ADS: Linked Open Data, http://data.archaeologydataservice.ac.uk

ADS: What is the ADS, and what do we do?, http://archaeologydataservice.ac.uk/about/background

AHDS - Arts and Humanities Data Service (2006): E-Science Scoping Study, documentation, http://www.ahds.ac.uk/e-science/e-science-scoping-study.htm

Ainsworth P. & Meredith M. (2009): e-Science for Medievalists: Options, Challenges, Solutions and Opportunities. Digital Humanities Quarterly 3.4, http://www.digitalhumanities.org/dhq/vol/3/4/000071/000071.html

Aitchison, Kenneth (2009a): Archaeology and the global financial crisis. In: Antiquity 83(319), http://www.antiquity.ac.uk/projgall/aitchison319/

Aitchison, Kenneth (2009b): After the 'Gold Rush': Global Archaeology in 2009. In: World Archaeology 41(4): 659-671, http://www.tandfonline.com/doi/abs/10.1080/00438240903363772

Aitchison, Kenneth (2011): The economic crisis and the coming second wave - how the economic situation will affect archaeological practice in universities and government. Archaeological Review from Cambridge, 26.1: 79-98, http://www.landward.eu/Aitchison%20ARC%202011.pdf

Aitchison, Kenneth (2014): Cultural Heritage in Times of Economic Crisis, pp. 1867-70, in: Encyclopedia of Global Archaeology. Springer, http://link.springer.com/referenceworkentry/10.1007%2F978-1-4419-0465-2_1177

Alcala F., Alcocer A., Alves F. *et al*. (2008): VENUS (Virtual ExploratioN of Underwater Sites) two years of interdisciplinary collaboration, pp. 250-258, in: Proceedings of VSMM'08 Conference on Virtual Systems and Multimedia Dedicated to Digital Heritage, Limassol, Cyprus, https://hal.archives-ouvertes.fr/hal-00361198

Allan, Robert (2009): Virtual Research Environments: From Portals to Science Gateways. Oxford: Chandos Publishing

ALLEA - European Federation of Academies of Sciences and Humanities, http://www.allea.org

ALLEA E-Humanities Working Group (2015): Going Digital: Creating Change in the Humanities. ALL European Academies (ALLEA), Berlin, http://dri.ie/sites/default/files/files/Going-Digital-digital-version.pdf

Allison, Penelope (2008): Dealing with legacy data – an introduction. In: Internet Archaeology 24, http://intarch.ac.uk/journal/issue24/introduction.html

Almansa-Sánchez J. & Richardson L. (2014): Working Group on Public Archaeology - Report Istanbul 2014. In: The European Archaeologist, Issue 42, Autumn 2014, 15-16, http://e-a-a.org/TEA/TEA42.pdf

Alsheikh-Ali A.A., Qureshi W., Al-Mallah M.H. & Ioannidis J.P. (2011): Public availability of published research data in high-impact journals". In: PloS one, 6(9) e24357, http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0024357#pone-0024357-g001

Altman M. & Crosas M. (2013): The Evolution of Data Citation: From Principles to Implementation. In: IASSIST Quarterly 37, Spring 2013: 62-70, http://www.iassistdata.org/downloads/iqvol371_4_altman.pdf

Altman M. & King G. (2007): A proposed standard for the scholarly citation of quantitative data. In: D-Lib Magazine, 13(3/4), http://www.dlib.org/dlib/march07/altman/03altman.html

Altmetric.com (altmetrics service provider, own by Macmillan Publishers via their Digital Science brand), http://www.altmetric.com

Altmetrics.org (research community website), http://altmetrics.org/manifesto/

Alvaro E., Brooks H., Ham M., Poegel S. & Rosencrans S. (2011): E-science librarianship: Field undefined. In: Issues in Science and Technology Librarianship, 66(11), http://www.istl.org/11-summer/article1.html

Ammon, Ulrich (2010): The hegemony of English, pp. 154-155, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Amorim R.C., Castro J.A, Rocha da Silva J. & Ribeiro C. (2015): A Comparative Study of Platforms for Research Data Management: Interoperability, Metadata Capabilities and Integration Potential, pp. 101-111, in: Rocha A. *et al*. (eds.): New Contributions in Information Systems and Technologies, Advances in Intelligent Systems and Computing, Vol. 353; pre-print, Vol. 353, http://dendro.fe.up.pt/pdf/papers/worldcist2015.pdf

Anagnostou P., Capocasa M., Milia N., Sanna E., Luzi D., Destro Bisol G. (2014): When data sharing gets close to 100%: what ancient human DNA studies can teach the Open Science movement. arXiv, 15 Jul 2014 (v.2), http://arxiv.org/abs/1407.3682

Ancient Charm project (EU, FP6, NEST - New and Emerging Science and Technology, 01/2006-12/2009): Analysis by neutron resonant capture imaging and other emerging neutron techniques: new cultural heritage and archaeological research methods. Final report, http://ec.europa.eu/research/fp6/nest/pdf/projects/ancient_charm.pdf

Ancient Charm project / Katalin T. Biró (Hungarian National Museum): Report on cultural heritage motivations of neutron-based imaging techniques. Project deliverable D2, http://www.ace.hu/acharm/D2-WP1.pdf

Ancient Lives (University of Oxford; invites non-experts to transcribe Egyptian papyri online; part of the Zooniverse network of projects), http://ancientlives.org

Anderson B. & Carlson S. (2006): Entangled Data: Knowledge & Community making in E (Social) Science. University of Essex, Chimera Working Paper: 2006-15, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.916&rep=rep1&type=pdf

Anderson S. & Blanke T. (2012): Taking the Long View: From e-Science Humanities to Humanities Digital Ecosystems. In: Historical Social Research 37.3: 147-164, http://nbn-resolving.de/urn:nbn:de:0168-ssoar-378350

Anderson S., Blanke T. & Dunn S. (2010): Methodological commons: arts and humanities e-Science fundamentals, in: Watson, P. *et al*. (eds.): e-Science: past, present and future, I. Phil. Trans. R. Soc. A, 368: 3779-96, http://rsta.royalsocietypublishing.org/content/368/1925/3779.full.pdf

Anderson, Sheila (2007): The Arts and Humanities and e-Science: Scoping Survey Report and Findings. Commissioned by the UK Arts and Humanities e-Science Initiative. King's College London, Department Centre for e-Research, December 2007, http://rae2007.cch.kcl.ac.uk/scoping-sa/e-science_report_final.pdf

Andrade M. & Baptista A.A. (2015): The Use of Application Profiles and Metadata Schemas by Digital Repositories: Findings from a Survey. DC-2015- International Conference on Dublin Core and Metadata Appliations, São Paulo, Brazil, http://dcevents.dublincore.org/IntConf/dc-2015/paper/download/362/373

Andrén, Sabina (2010): A transdisciplinary, participatory and action-oriented research approach – Sounds nice but what do you mean? Working paper, Human Ecology Division, Lund University, January 2010, http://lup.lub.lu.se/record/1744256/file/1744257.pdf

Andreoli-Versbach P. & Mueller-Langer F. (2013): Open access to data: An ideal professed but not practised. RatSWD Working Paper Series 215, February 2013, http://www.ratswd.de/dl/RatSWD_WP_215.pdf

Anichini F., Fabiani F., Gattiglia G. & Gualandi M. (2012): MAPPA. Methodologies applied to archaeological potential predictivity. Rome, https://www.academia.edu/2156935/MAPPA._Methodologies_Applied_to_Archaeological_Potential_Predictivity

Anichini F., Gattiglia G., Di Donato F. *et al*. (2012): Opening the Past. Archaeological Open Data. Pisa 9/06/2012. MapPapers 3-11: 51-95, http://mappaproject.arch.unipi.it/wp-content/uploads/2011/08/Pre_atti_online3.pdf

Antell K., Foote J.B., Turner J. & Shults B. (2013): Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. College and Research Libraries, http://crl.acrl.org/content/early/2013/04/05/crl13-464

APARSEN - Alliance for Permanent Access to the Records of Science Network (EU, FP7-ICT, 01/2011-12/2014), http://www.alliancepermanentaccess.org/index.php/aparsen/

APARSEN (2012): Report on peer review of research data in scholarly communication. Project deliverable D33.1A, 30-04-2012, http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/04/APARSEN-REP-D33_1A-01-1_0.pdf

APARSEN (2013a): Business Preparedness Report APARSEN project deliverable, D36.1, February 2013, http://aparsen.digitalpreservation.eu/pub/Main/ApanDeliverables/APARSEN-REP-D36_1-01-0_3.pdf

APARSEN (2013b): Report on Cost Parameters for Digital Repositories. APARSEN project deliverable D32.1, February 2013, http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2013/03/APARSEN-REP-D32_1-01-1_0.pdf

APARSEN (2013c): Report on Testing of Cost Models and Further Analysis of Cost Parameters. APARSEN project deliverable, D32.2, June 2013, http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2014/06/APARSEN-REP-D32_2-01-1_0_incURN.pdf

ARACHNE IDAI.objects database (central objects-database of the German Archaeological Institute and the Cologne Digital Archaeology Laboratory), http://arachne.uni-koeln.de

Archaeolandscapes / Arcland (EU, Culture Programme, 09/2010-09/2015; outreach and training programme focused on aerial, remote sensing and surveying techniques), http://www.arcland.eu

Archaeological Networks (information on the application of network analysis in archaeology; Tom Brughmans *et al*.) https://archaeologicalnetworks.wordpress.com

Archaeological Reference Resources Project (UK, commissioned by Historic England), http://www.archaeologicalreferenceresources.uk

Archaeology in Contemporary Europe (EU, Culture Programme, 2008-2012), project documentation at: http://archaeologydataservice.ac.uk/archives/view/ace_2011/index.cfm

Archaeotools - Data mining, facetted classification and E-archaeology (UK, e-Science Research Grant, 2007-2009), http://archaeologydataservice.ac.uk/research/archaeotools

Archambault E. & Larivière V. (2010): The limits of bibliometrics for the analysis of the social sciences and humanities literature, pp. 272-275, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Archäologische Informationen, Band 38: Fokus: Open Access & Open Data. Early view (27.10.2015), http://www.dguf.de/index.php?id=9

ARCHES - Archaeological Resources in Cultural Heritage / Perrin, Kathy *et al*. (eds., 2014): The Standard and Guide to Best Practice in Archaeological Archiving in Europe. EAC Guidelines volume 1 (Namur), online, http://archaeologydataservice.ac.uk/arches/Wiki.jsp?page=Main

ARCHES - Archaeological Resources in Cultural Heritage: a European Standard (EU Culture Programme, 06/2012-05/2014), project website, http://archaeologydataservice.ac.uk/arches/

ARCHES - Heritage Inventory & Management System (open source open-source geospatial software system, developed jointly by the Getty Conservation Institute and World Monuments Fund), http://archesproject.org

Archival Resource Key (ARK), https://wiki.ucop.edu/display/Curation/ARK

ArchLab - Consortium of Swedish Archaeological Research Laboratories, http://archlab.se

ARENA - Archaeological Records of Europe - Networked Access search portal (EU, Culture Programme, 2001-2004; 2009-2010 in the context of DARIAH); http://ads.ahds.ac.uk/arena/search/

ARIADNE - D12.1: Use Requirements, July 2014, http://www.ariadne-infrastructure.eu/Resources

ARIADNE - D12.3: Initial Infrastructure Implementation Report, June 2015, http://www.ariadne-infrastructure.eu/Resources

ARIADNE - D16.2: First Report on Natural Language Processing, May 2015, http://www.ariadne-infrastructure.eu/Resources

ARIADNE - D2.1: First Report on Users' Needs, April 2014, http://www.ariadne-infrastructure.eu/Resources/D2.1-First-report-on-users-needs

ARIADNE - D2.2: Second Report on Users' Needs, February 2015, http://www.ariadne-infrastructure.eu/content/view/full/1188

ARIADNE - D3.1: Initial Report on Standards and on the Project Registry, November 2013, http://www.ariadne-infrastructure.eu/Resources

ARIADNE - D3.2: Report on Project Standards, November 2013, http://www.ariadne-infrastructure.eu/Resources

ARIADNE - D3.3: Report on Data Sharing Policies, January 2014, http://www.ariadne-infrastructure.eu/Resources

ARIADNE - D5.2: Initial Report on the Assessment of Online Access, January 2015, http://www.ariadne-infrastructure.eu/Resources

ARIADNE - Dataset Catalogue Model (ACDM) support website, http://support.ariadne-infrastructure.eu

ARIADNE - Excavation and Monument Data SIG (2015): Survey of Excavation Data Management. Fentress E. *et al*., April 2014, http://www.ariadne-infrastructure.eu/Community/Special-Interest-Groups/Excavation-and-Monument-Data

ARIADNE - Expert Forum: Digital Futures of Archaeological Practice 2020-2025 (Athens, 2-3 July 2015), http://summerschool.dcu.gr/?page_id=19#expert-forum

ARIADNE - Landscape Services: http://landscape.ariadne-infrastructure.eu

ARIADNE - Registry, http://registry.ariadne-infrastructure.eu

ARIADNE - Special Interest Groups (SIGs), http://www.ariadne-infrastructure.eu/Community/Special-Interest-Groups/

ARIADNE - Visual Media Server, http://ariadne1.isti.cnr.it

ARL - Association of Research Libraries (2006): To Stand the Test of Time. Long-term Stewardship of Digital Data Sets in Science and Engineering. Arlington (VA), USA, September 2006, http://www.arl.org/storage/documents/publications/digital-data-report-2006.pdf

Armbruster C. & Romary L. (2010): Comparing Repository Types. In: International Journal of Digital Library Systems, 1(4): 61-73; manuscript: http://arxiv.org/ftp/arxiv/papers/1003/1003.4187.pdf

Armbruster, Chris (2008): Access, Usage and Citation Metrics: What Function for Digital Libraries and Repositories in Research Evaluation? Social Science Research Network Working Paper Series. Berlin: SSRN, http://dx.doi.org/10.2139/ssrn.1088453

Aronova E., Baker K.S. & Oreskes N. (2010): Big Science and Big Data in Biology: From the International Geophysical Year through the International Biological Program to the Long Term Ecological Research (LTER) Network, 1957–Present. In: Historical Studies in the Natural Sciences, 40(2): 183–224, http://interoperability.ucsd.edu/docs/10AronovaBakerOreskes_HNS.pdf

Aroyo L., Hyvönen E. & van Ossenbruggen J. (eds., 2007): Cultural Heritage on the Semantic Web. Co-located with the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference, Busan, Korea, proceedings of Workshop 9, http://www.cs.vu.nl/~laroyo/CH-SW/ISWC-wp9-proceedings.pdf

Artefacts - Online Encyclopedia of Archaeological Small Finds (France, collaborative database and search portal for archaeological finds in Europe), http://artefacts.mom.fr

Arxiv.org (international ePrints archive, Cornell University Library), http://arxiv.org

Asher T. & Guilhot N. (2010): The collapsing space between universities and think-tanks, pp. 341-345 in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Association of European-level Research Infrastructure Facilities (ERF-AISBL), http://www.erf-aisbl.eu

Association of Research Libraries (ARL): Contemporary User Studies, http://www.arl.org/rtl/plan/userstudies2009.shtml

Association of Research Libraries (ARL): E-Science Working Group, http://www.arl.org/rtl/escience/

AthenaPlus / McKenna G. *et al*. (2014): Survey and description of existing mapping models to LIDO. Project deliverable D2.2, August 2014, http://www.athenaplus.eu/getFile.php?id=422

Athienou Archaeological Project (Cyprus), http://sites.davidson.edu/aap/

Atici L., Kansa S., Lev-Tov J. & Kansa E. (2012): Other people's data: A demonstration of the imperative of publishing primary data. In: Journal of Archaeological Method and Theory, 1(3): 1-19; self-archived preprint, http://escholarship.org/uc/item/1nt1v9n2

Attwood T.K., Kell D.B., McDermott P. *et al*. (2010): Utopia documents: linking scholarly literature with research data. In: Bioinformatics, 26(18): i568–i574, http://bioinformatics.oxfordjournals.org/content/26/18/i568

Auckland, Mary (2012): Re-Skilling for Research: An Investigation into the Role and Skills of Subject and Liaison Librarians Required to Effectively Support the Evolving Information Needs of Researchers. London: Research Libraries UK, http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf

Auslander L., Bentley A., Halevi L., Sibum H.O. & Witmore C. (2009): AHR conversation: Historians and the study of material culture. In: The American Historical Review, 114(5): 1355-1404, https://steinhardt.nyu.edu/scmsAdmin/uploads/005/083/AHR%20conversations.pdf

Austin T. & Mitcham J. (2007): Preservation and Management Strategies for Exceptionally Large Data Formats: 'Big Data'. English Heritage / Archaeology Data Service, 28 September 2007, http://ads.ahds.ac.uk/project/bigdata/final_report/bigdata_final_report_1.3.pdf

Australian National Data Service (ANDS): Data Management Planning, http://ands.org.au/resource/data-management-planning.html

Averett E.W. *et al*. (2015): Mobilizing the Past for a Digital Future: the Potential of Digital Archaeology. Workshop at Wentworth Institute of Technology, Boston/MA, USA, 27–28 February 2015, http://uwm.edu/mobilizing-the-past/wp-content/uploads/sites/180/2015/02/NEHbrochure_final_small.pdf

Averett E.W., Counts D. & Gordon J. (2014): Mobilizing the Past for a Digital Future: The Potential of Digital Archaeology. National Endowment for the Humanities, Office of Digital Humanities – Digital Humanities Start-Up Grants, Level 1, application, http://www.neh.gov/files/grants/creighton_university_the_potential_of_digital_archaeology.pdf

AVROSS (2007): Accelerating Transition to Virtual Research Organisation in Social Science. Final Report, http://web.fhnw.ch/plattformen/avross/papers-and-prensentations/final-report/Final%20report,%20full%20draft,%20v2-2.pdf

AWOL - The Ancient World Online: Alphabetical List of Open Access Journals in Ancient Studies, http://ancientworldonline.blogspot.co.at/2012/07/alphabetical-list-of-open-access.html

Babeu, Alison (2011): Rome Wasn't Digitized in a Day: Building a Cyberinfrastructure for Digital Classicists. Washington, DC: Council on Library and Information Resources, http://www.clir.org/pubs/reports/pub150/pub150.pdf

Bailey, Charles W. (2010): Transforming Scholarly Publishing through Open Access: A Bibliography. Houston: Digital Scholarship, online: http://digital-scholarship.org/tsp/w/tsp.html

Bailey, Charles W. (2012): Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works. Houston: Digital Scholarship, online: http://digital-scholarship.org/dcbw/dcbw.htm

Bailey, Charles W. (2013): Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works - 2012 Supplement. Houston: Digital Scholarship, 18 March 2013, http://digital-scholarship.org/dcbw/s1/dcbw-s1.htm

Bailey, Charles W. (2014): Research Data Curation Bibliography. Version 4, 23 June 2014. Houston: Digital Scholarship, http://digital-scholarship.org/rdcb/rdcb.htm

Bajcsy, Peter (2010): Needs and Opportunities in Humanities, Arts and Social Sciences. Presentation at the Grand Challenges in Data-Intensive Discovery conference, San Diego, 26-28 October 2010, http://www.sdsc.edu/Events/gcdid2010/docs/gcdid_presentations/bajcsy.pdf

Baker K. & Millerand F. (2010): Infrastructuring ecology: challenges in achieving data sharing. In: Parker J., Vermeulen N. & Penders B. (eds.): Collaboration in the New Life Sciences. Ashgate; http://interoperability.ucsd.edu/docs/10BakerMillerand_InfrastructuringEcology_old.pdf

Bakker M. & Wicherts J.M. (2011): The (mis)reporting of statistical results in psychology journals. In: Behavior Research Methods, 43(3): 666-678, http://link.springer.com/article/10.3758%2Fs13428-011-0089-5

Balkestein M. & Tjalsma H. (2007): The ADA approach: retro-archiving data in an academic environment. In: Archival Science, 7(1): 89-105, http://link.springer.com/article/10.1007%2Fs10502-007-9053-7

Ball A. & Duke M. (2015): How to Cite Datasets and Link to Publications. DCC How-to Guide. Edinburgh: Digital Curation Centre, http://www.dcc.ac.uk/resources/how-guides/cite-datasets

Ball, Alex (2012): Review of Data Management Lifecycle Models. University of Bath, UK (unpublished), http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf

Ball, Alex (2013): Making Citation Work: A British Library DataCite Workshop. In: Ariadne, Issue 71, http://www.ariadne.ac.uk/issue71/datacite-2013-rpt

Ball, Alex (2014): How to License Research Data. DCC How-to Guides. Edinburgh: Digital Curation Centre, http://www.dcc.ac.uk/resources/how-guides/license-research-data#x1-6000

Balstad, Roberta (2010): The interdisciplinary challenges of climate change research, pp. 210-212, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Bardi A. & Manghi P. (2014): Enhanced Publications: Data Models and Information Systems. LIBER Quarterly 23(4), https://liber.library.uu.nl/index.php/lq/article/view/8445/9825

Barjak F., Lane J. *et al*. (2009): Case Studies of e-Infrastructure Adoption. In: Social Science Computer Review, 27(4): 583-600, http://ssc.sagepub.com/content/27/4/583.full.pdf

Barrett T., Troup D.B., Wilhite S.E. *et al*. (2011): NCBI GEO: archive for functional genomics data sets—10 years on. In: Nucleic Acids Research, 39: D1005–D1010, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3013736/

Bartling S. & Friesike S. (eds., 2014): Opening Science. The Evolving Guide on How the Web is Changing Research, Collaboration and Scholarly Publishing. Accessible as a SpringerOpen and printed book as well an online, dynamic version: http://www.openingscience.org/get-the-book/

Barton, Michael (2013): Stories of the past or science of the future? Archaeology and computational social science, pp. 151-178, in: A. Bevan & M. W. Lake (eds.): Computational Approaches to Archaeological Spaces. Walnut Creek: Left Coast Press, https://www.academia.edu/3651038/Stories_of_the_past_or_science_of_the_future_Archaeology_and_computational_social_science

BASE - Bielefeld Academic Search Engine, http://www.base-search.net

Beagrie - Charles Beagrie Ltd. & JISC (2011): KRDS Factsheet, July 2011, http://www.beagrie.com/static/resource/KRDS_Factsheet_0711.pdf

Beagrie - Charles Beagrie Ltd.: Keeping Research Data Safe (KRDS, UK, JISC funded projects, 2007-2011), http://beagrie.com/krds/

Beagrie N. & Houghton J. (2013a): The value and impact of the Archaeology Data Service. A study and methods for enhancing sustainability. Final Report, Charles Beagrie Ltd and Victoria University, September 2013, http://repository.jisc.ac.uk/5509/1/ADSReport_final.pdf

Beagrie N. & Houghton J. (2013b): The Value and Impact of the British Atmospheric Data Centre. Final Report, Charles Beagrie Ltd and Victoria University, 30 September 2013, http://repository.jisc.ac.uk/5382/1/BADCReport_Final.pdf

Beagrie N., Beagrie R. & Rowlands I. (2009): Research data preservation and access: the views of researchers. In: Ariadne, Issue 60, http://www.ariadne.ac.uk/issue60/beagrie-et-al

Beagrie N., Chruszcz J. & Lavoie B. (2008): Keeping research data safe. A cost model and guidance for UK universities. A study funded by JISC. Final Report, April 2008, http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf

Beagrie N., Lavoie B. & Woollard M. (2010): Keeping Research Data Safe 2. Final Report, Charles Beagrie Ltd., April 2010, http://www.webarchive.org.uk/wayback/archive/20140615221405/http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf

Beale G. & Reilly P. (2014): Virtual archaeology in a material world. New technologies enabling novel perspectives. Presentation at CAA 2014, Paris, 25 April 2014, http://eprints.soton.ac.uk/364454/1/VA%20in%20material%20world%20CAA14%20Paris.pdf

Beale N. & Beale G. (2012): The Potential of Open Models for Public Archaeology. Digital Futures 2012: The Third Annual Digital Economy All Hands Conference, Aberdeen, GB, 23-25 October 2012, http://eprints.soton.ac.uk/355830/1/de2012_submission_17.pdf

Beale, Nicole (2012): How community archaeology can make use of open data to achieve further its objectives. In: World Archaeology 44(4), 612-633, https://canvas.brown.edu/courses/917715/files/43868895/

Bechhofer S., Buchan I., De Roure D. *et al*. (2011): Why linked data is not enough for scientists, pp. 300-307, in: E-Science'10 - Proceedings of the IEEE Sixth International Conference on e-Science, Brisbane, Australia, 7-10 December 2010, http://eprints.soton.ac.uk/271587/5/research-objects-final.pdf

Bechhofer S., De Roure D., Gamble M., Goble C. & Buchan I. (2010): Research objects: towards exchange and reuse of digital knowledge. FWCS 2010 - The Future of the Web for Collaborative Science, Raleigh, USA, 26 April 2010, http://precedings.nature.com/documents/4626/version/1/files/npre20104626-1.pdf

Beck A. & Harrison D. (2014): Building an archaeological project repository II: Where are the research data repositories? In: Open Knowledge Blog, 17 April 2014, http://blog.okfn.org/2014/04/17/building-an-archaeological-project-repository-ii-where-are-the-research-data-repositories/

Beck A. & Neylon C. (2012): A vision for Open Archaeology. In: World Archaeology, 44(4): 479-497, http://www.tandfonline.com/toc/rwar20/44/4#.VEPIOskucyg

Beck, Anthony (2010): Dig the new breed, Part II – open archaeology and ethics. In: Open Knowledge Blog, 11 June 2010, http://blog.okfn.org/2010/06/11/dig-the-new-breed-open-archaeology-and-ethics/

Beck, Anthony (2013): Opening access to heritage resources: risk, opportunity or paradigm shift?, pp. 40-47, in: Archäologie und Informationssysteme. Arbeitshefte zur Denkmalpflege in Niedersachsen 42, http://www.landesarchaeologen.de/fileadmin/Dokumente/News/NI_Archaeologie_und_Informationssysteme_Arbeitsheft42.pdf#page=4

Beer, David (2013): Social media's politics of circulation have profound implications for how academic knowledge is discovered and produced. In: The Impact Blog, 29 July 2013, http://blogs.lse.ac.uk/impactofsocialsciences/2013/07/29/academic-knowledge-and-the-politics-of-circulation/

Bell G., Hey T. & Szalay A. (2009): Computer science: Beyond the data deluge. Science 323 (5919): 1297-1298, http://www.uvm.edu/~pdodds/files/papers/others/2009/bell2009a.pdf

Belmont Forum, E-Infrastructures and Data Management Collaborative Research Action (2014): Interim Report, version 1.1. Prepared by Lee Allison and Robert Gurney, 22 September 2014, http://www.bfe-inf.org/sites/default/files/doc-repository/Interim%20Report%20-%20E-Infrastructures%20%26%20Data%20Management%20CRA,%20version%201.1.pdf

Belter, Christopher W. (2014): Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets. In: PLoS ONE 9(3): e92590, http://dx.doi.org/10.1371/journal.pone.0092590

Benardou A. & Garnett V. (2015): Engaging with Europeana. Europeana Research weblog, http://research.europeana.eu/work-with-us/engaging-with-europeana

Benardou A., Constantopoulos P., Dallas C. & Gavrilis D. (2010): Understanding the Information Requirements of Arts and Humanities Scholarship. In: The International Journal of Digital Curation, 5(1): 18-33, http://www.ijdc.net/index.php/ijdc/article/viewFile/144/206

Benardou, Agiatis (2011a): Classical studies facing digital research infrastructures: from practice to requirements. In: The Digital Classicist, http://www.digitalclassicist.org/wip/wip2011-05ab.html

Benardou, Agiatis (2011b): Classical studies facing digital research infrastructures: from practice to requirements. Digital Classicist & Institute of Classical Studies Seminar, London, July 1st, 2011, http://www.digitalclassicist.org/wip/wip2011-05ab.pdf

Benardou, Agiatis (2015): "Archaeologists like us": One ring to rule us all? In: Research.Europeana weblog, 13 July 2015, http://research.europeana.eu/blogpost/archaeologists-like-us-one-ring-to-rule-us-all

Berggren A., Dell'Unto N., Forte M., Haddow S., Hodder I. *et al*. (2015): Revisiting reflexive archaeology at Çatalhöyük: integrating digital and 3D technologies at the trowel's edge. In: Antiquity, 89(344): 433-448, http://dx.doi.org/10.15184/aqy.2014.43, http://static1.squarespace.com/static/53568703e4b0feb619b78a93/t/55358638e4b076e61bf5f940/1429571838459/Antiquity+Digital+Techniques+Catalhoyuk+2015.pdf

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003), http://openaccess.mpg.de/Berlin-Declaration

Berman F., Lavoie B. *et al*. (2010): Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information. Final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, February 2010, http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Berman, Francine (2008): Got Data? A Guide to Preservation in the Digital Age. In: Communications of the ACM, 51(12), http://www.itpb.ucla.edu/documents/2009/p50-berman.pdf

Bernal, Isabel (2013): Open Access and the Changing Landscape of Research Impact Indicators: New Roles for Repositories [DIGITAL.CSIC]. In: Publications – Open Access Scholarly Publishing Journal, 2013-1: 56-77; http://digital.csic.es/bitstream/10261/79872/3/publications-01-00056_Bernal.pdf

Berry, David (2011): The computational turn: Thinking about the digital humanities. In: Culture Machine, Volume 12, http://www.culturemachine.net/index.php/cm/article/viewArticle/440

Bertrand L., Cotte M., Stampanoni M. *et al*. (2012): Development and trends in synchrotron studies of ancient and historical materials. In: Physics Reports, 519(2): 51-96, http://www.sciencedirect.com/science/article/pii/S0370157312000865

Bertrand L., Dillmann P. & Reiche I. (eds., 2015): Synchrotron radiation and neutrons in art and archaeology. In: Journal of Analytical Atomic Spectrometry, Issue 30 [a collection of 28 papers], http://pubs.rsc.org/en/journals/articlecollectionlanding?sercode=ja&themeid=20aaf6bd-1ff5-4bbb-ab8d-294f76ac4b70

Bertrand L., Thoury M. & Anheim E. (2013): Ancient materials specificities for their synchrotron examination and insights into their epistemological implications. In: Journal of Cultural Heritage, 14(4): 277-289, http://www.sciencedirect.com/science/article/pii/S1296207412001549

Bevan, Andrew (2012): Value, authority and the Open Society: some implications for digital and online archaeology, pp. 1-14, in: C. Bonacchi (ed.): Archaeology and Digital Communication: Towards Strategies of Public Engagement. London: Archetype (manuscript), http://discovery.ucl.ac.uk/1328917/1/1328917_Bevan12b_postprint.pdf

BIBO - Bibliographic Ontology (provides concepts and properties for describing citations and bibliographic references), http://bibliontology.com

Biehl P.F., Comer D.C., Prescott C. & Soderland H.A. (eds., 2014): Identity and Heritage. Contemporary Challenges in a Globalized World. Springer

Binding C. & Tudhope D. (2015): Mapping between linked data vocabularies in ARIADNE. Presentation at the 14th European Networked Knowledge Organization Systems (NKOS) Workshop; TPDL-2015 Conference, Poznań, Poland, 18 September 2015, https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2015/content/NKOS2015-presentation-tudhope.pdf

Binding, Ceri (2014): 5 star data – achieving the 5th star. Paper presented at the 13th European Networked Knowledge Organization Systems (NKOS) Workshop at DL2014 (joint JCDL/TPDL conferences), City University London, 11-12 September 2014, extended abstract, http://www.iskouk.org/sites/default/files/NKOS2014-abstract-binding_0.pdf

Bio2RDF: Linked Data for the Life Sciences, http://bio2rdf.org

BioSharing portal (bio-sciences), https://www.biosharing.org

BioVel - Biodiversity Virtual e-Laboratory (EU, FP7, 09/2011-12/2014), https://www.biovel.eu

BIOVIA (scientific workflow management system), http://accelrys.com

Birney, Ewan *et al*. (2009): Toronto International Data Release Workshop Authors: Prepublication data sharing. In: Nature, Vol. 461, 10 September 2009, pp. 168-170, http://www.nature.com/nature/journal/v461/n7261/full/461168a.html#a1

Birnholtz J.P. & Bietz M. (2003): Data at Work. Supporting sharing in science and engineering, pp. 339-348, in: Proceedings of the ACM SIGGROUP Conference on Supporting Group Work, Sanibel Island, FL, USA, preprint, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.79.4098&rep=rep1&type=pdf

Bishop, Libby (2013): Reply to: Natasha Mauthner and Odette Parry's 'Open access digital data sharing: principles, policies and practices'." In: Social Epistemology, review and reply collective, 2(8): 71-78, http://wp.me/p1Bfg0-U5

Bitly (2011): You just shared a link. How long will people pay attention? Bitly weblog, 6 September 2011, http://blog.bitly.com/post/9887686919/you-just-shared-a-link-how-long-will-people-pay

Bizer, Christian (2010): Data Linking, pp. 34-43, in: GRDI2020 - Global Research Data Infrastructures: Towards a 10-year vision for global research data infrastructures. GRDI2020 Consortium, http://www.grdi2020.eu/Repository/FileScaricati/9a85ca56-c548-47e4-8b0e-86c3534ad21d.pdf

Björk, Bo-Christer (2013): Open Access Subject Repositories – An Overview. In: Journal of the American Society for Information Science and Technology, 65(4): 698–706, preprint 2013, preprint, http://www.openaccesspublishing.org/repositories/Subject_Repositories.pdf

Blackwell C. & Crane G. (2009): Conclusion: Cyberinfrastructure, the Scaife Digital Library and Classics in a Digital age. In: Digital Humanities Quarterly 3.1, Winter 2009, http://www.digitalhumanities.org/dhq/vol/3/1/000035/000035.html

Blanke T. & Hedges M. (2013): Scholarly primitives: Building institutional infrastructure for humanities e-Science. In: Future Generation Computer Systems, 29(2): 654-661, https://kclpure.kcl.ac.uk/portal/files/6253556/BlankeHedges_Humanities_eScience_FGCS_1.pdf

Blanke T., Bryant M., Hedges M., Aschenbrenner A. & Priddy M. (2011): Preparing DARIAH. Seventh IEEE International Conference on eScience, Stockholm, https://de.dariah.eu/fileadmin/publications_presentations/Blanke_Bryant_Hedges_Aschenbrenner_Priddy_Preparing_DARIAH.pdf

Blanke T., Candela L., Hedges M., Priddy M. & Simeoni F. (2010): Deploying general-purpose virtual research environments for humanities research. In: Royal Society, Philosophical Transactions A, Vol. 368, Issue 1925 (Theme Issue: e-Science: past, present and future I), http://rsta.royalsocietypublishing.org/content/368/1925/3813

Blower J.D., Santokhee A., Milsted A.J. & Frey J.G. (2010): BlogMyData: a Virtual Research Environment for collaborative visualization of environmental data. In: All Hands Meeting 2010, Cardiff, GB, 13-16 September 2010, http://eprints.soton.ac.uk/164533/

Blumauer, Andreas (2013): The LOD cloud is dead, long live the trusted LOD cloud (7 June 2013), http://blog.semantic-web.at/2013/06/07/the-lod-cloud-is-dead-long-live-the-trusted-lod-cloud/

Boast R. & Biehl P. (2011): Archaeological Knowledge Production and Dissemination in the Digital Age, pp. 119-156, in: Kansa E. *et al*. (eds.): Archaeology 2.0: New Approaches to Communication and Collaboration. Cotsen Institute of Archaeology, UC Los Angeles, http://escholarship.org/uc/item/1r6137tb

Boast R. & Enote J. (2013): Virtual Repatriation: It is Neither Virtual nor Repatriation, pp. 103-116, in: Biehl P. & Prescott (eds.): Heritage in the Context of Globalization (SpringerBriefs in Archaeology 8), http://www.springer.com/us/book/9781461460763

Bodard G. & Mahony S. (eds., 2008): 'Though much is taken, much abides': Recovering antiquity through innovative digital methodologies. In: Digital Medievalist, 4 (special issue), http://www.digitalmedievalist.org/journal/4/DCintro/

Bodard G. & Mahony S. (eds., 2010): Digital Research in the Study of Classical Antiquity. Ashgate, http://www.ashgate.com/isbn/9780754677734

Bøgvad-Kejser U., Nielsen A.B. & Thirifays A. (2011): Cost model for digital preservation: Cost of digital migration. In: International Journal of Digital Curation, 6(1): 255-267, http://www.ijdc.net/index.php/ijdc/article/viewFile/177/246

Boland K., Ritze D., Eckert K. & Mathiak B. (2012): Identifying References to Datasets in Publications, pp. 150-161, in: Theory and Practice of Digital Libraries, LNCS Vol. 7489; pre-print, http://publications.wim.uni-mannheim.de/informatik/lski/Ritze2012TPDL.pdf

Bollen J., Van de Sompel H., Hagberg A. & Chute R. (2009): A Principal Component Analysis of 39 Scientific Impact Measures. In: PLoS ONE, 4(6): e6022, http://dx.doi.org/10.1371/journal.pone.0006022

Bollier, David (2011): The Promise and Peril of Big Data. The Aspen Institute Communications and Society Program: Washington DC, 2011. http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

Bollier, David (2011): The Promise and Peril of Big Data. The Aspen Institute Communications and Society Program: Washington DC, 2011.

http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

Bonacchi C., Bevan A., Pett A. *et al*. (2014): Crowd-sourced Archaeological Research: The MicroPasts Project. In: Archaeology International, No. 17: pp. 61-68, http://www.ai-journal.com/articles/10.5334/ai.1705/

Bonacchi, Chiara (ed., 2012): Archaeology and Digital Communication. Towards Strategies of Public Engagement. London: Archetype Publications

Bone D. & Burns P. (2011): An overview of content archiving services in scholarly publishing [LOCKSS, CLOCKSS, Portico, JSTOR). Supplement no. 1, http://allenpress.com/system/files/pdfs/library/archivingwhitepaper.pdf

BoneCommons ("a place for zooarchaeologists worldwide to communicate and share their work"; sponsored by the International Council for Archaeozoology (ICAZ), http://alexandriaarchive.org/bonecommons/

Borer E.T., Seabloom E.W., Jones M.B. & Schildhauer M. (2009): Some simple guidelines for effective data management. In: Bulletin of the Ecological Society of America, 90(2): 205-214, https://www.nceas.ucsb.edu/files/news/ESAdatamng09.pdf

Borgman C.L., Darch P.T., Sands A.E., Wallis J.C. & Traweek S. (2014): The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management. In: ACM/IEEE Joint Conference on Digital Libraries (JCDL 2014) and International Conference on Theory and Practice in Digital Libraries (TPDL 2014). London: IEEE Computer Society, http://works.bepress.com/cgi/viewcontent.cgi?article=1394&context=borgman

Borgman, Christine L. (2007): Scholarship in the Digital Age: Information, Infrastructure, and the Internet. Cambridge, MA: The MIT Press

Borgman, Christine L. (2009): The Digital Future is Now: A Call to Action for the Humanities. In: Digital Humanities Quarterly, 3(4), http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Borgman, Christine L. (2010): Research Data: Who will share what, with whom, when, and why? Fifth China – North America Library Conference 2010, Beijing, 8-12 September 2010, http://works.bepress.com/borgman/238

Borgman, Christine L. (2012): The Conundrum of Sharing Research Data. In: Journal of the American Society for Information Science and Technology, 63(6): 1059-1078, http://dx.doi.org/10.1002/asi.22634

Borgman, Christine L. (2015): If Data Sharing is the Answer, What is the Question?, pp. 15-16, in: ERCIM News 100: Special theme: Scientific Data Sharing and Re-use, January 2015, http://ercim-news.ercim.eu/images/stories/EN100/EN100-web.pdf

Bornman, Lutz (2014): Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. arXiv:1406.7091v2 [cs.DL], 10 September 2014, http://arxiv.org/abs/1406.7091

Bos N., Zimmerman A., Olson J. *et al*. (2007): From shared databases to communities of practice: A taxonomy of collaboratories. In: Journal of Computer-Mediated Communication, 12(2), http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00343.x/full

Bowker, Geoffrey C. (2000): Biodiversity Datadiversity. In: Social Studies of Science, 30(5): 643 – 683, http://www.ics.uci.edu/~gbowker/biodivdatadiv.pdf

Bowker, Geoffrey C. (2005): Memory Practices in the Sciences. Cambridge: MIT Press

Boyd D. & Crawford K. (2011): Six provocations for Big Data. Paper to be presented at 'A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society', Oxford Internet

Institute, 21 September 2011; manuscript, http://softwarestudies.com/cultural_analytics/Six_Provocations_for_Big_Data.pdf

Bradley R., Haselgrove C., Vander Linden M. & Webley L. (eds., 2011): Development-led Archaeology in Northwest Europe. Oxford: Oxbow Books

Bradley, John (2012): No job for techies: Technical contributions to research in digital humanities, pp. 11-26, in: Deegan M. & McCarty W. (eds.): Collaborative Research in the Digital Humanities. Farnham: Ashgate (DH2009 conference script online: https://www.yumpu.com/en/document/view/5530294/script-no-job-for-techies-kings-college-london)

Brembs B. & Colomb J. (2015): Sub-strains of Drosophila Canton-S differ markedly in their locomotor behavior [v2; ref status: indexed, http://f1000r.es/57i] F1000Research 3:176, http://f1000research.com/articles/3-176/v2#f4

Brembs B., Button K. & Munafò M. (2013): Deep impact: unintended consequences of journal rank. In: Front. Hum. Neurosci., 7:291, published online: 24 June 2013. Also in arXiv.org > cs > arXiv:1301.3748, v1, 16 Jan 2013 http://arxiv.org/abs/1301.3748

Breure L., Voorbij H. & Hoogerwerf M. (2011): Rich internet publications: “Show what you tell”. In: Journal of Digital Information, 12(1), http://journals.tdl.org/jodi/index.php/jodi/article/view/1606/1738

British Atmospheric Data Centre (BADC), the UK Natural Environment Research Council's (NERC) designated data centre for the atmospheric sciences, http://badc.nerc.ac.uk

British Museum: Semantic Web Collection Online, http://collection.britishmuseum.org

Brockman W.S., Neumann L., Palmer C.L. & Tidline T. J. (2001): Scholarly work in the humanities and the evolving information environment. Washington, DC: Council on Library and Information Resources, http://www.clir.org/pubs/reports/pub104/pub104.pdf

Brown A.G., Basell L.S. & Butzer K.W. (eds., 2011): Geoarchaeology, Climate Change, and Sustainability. GSA Special Papers 476. Boulder: Geological Society of America, http://specialpapers.gsapubs.org/content/476

Brown S. & Swan A. (2007): Researchers’ use of academic libraries and their services. A report commissioned by the Research Information Network and the Consortium of Research Libraries, April 2007, http://www.rin.ac.uk/system/files/attachments/Researchers-libraries-services-report.pdf

Brown S., Ross R., Gerrard D. *et al*. (2006): RePAH - A User Requirements Analysis for Portals in the Arts and Humanities. Final Report, http://repah.dmu.ac.uk/report/pdfs/RePAHReport-Complete.pdf

Bruce T.R. & Hillmann D. (2013): Metadata Quality in a Linked Data Context. VoxPopuLII weblog, 24 January 2013, https://blog.law.cornell.edu/voxpop/2013/01/24/metadata-quality-in-a-linked-data-context/

Brughmans, Tom (2010). Connecting the dots: towards archaeological network analysis. In: Oxford Journal of Archaeology, 29(3): 277-303, http://core.ac.uk/download/pdf/270448.pdf

Brughmans, Tom (2013): Thinking through networks: A Review of Formal Network Methods in Archaeology. Journal of Archaeological Method and Theory, 20(4): 623-662, https://www.academia.edu/1646219/Brughmans_T._2013._Thinking_Through_Networks_A_Review_of_Formal_Network_Methods_in_Archaeology._Journal_of_Archaeological_Method_and_Theory_20_623_662

Brughmans, Tom (2014): Evaluating network science in archaeology. A Roman Archaeology perspective. University of Southampton, Faculty of Humanities, Doctoral Thesis, http://eprints.soton.ac.uk/371700/

Büchler M., Franzini E., Franzini G. & Moritz M. (2014): Scaling Historical Text Re-use. 2nd Workshop on Big Humanities Data (27 October 2014), held in conjunction with the IEEE International Conference on Big Data (IEEE BigData 2014), Washington DC, USA, https://bighumanities.files.wordpress.com/2014/10/buchler.pdf

Buddenbohm S., Enke H., Hofmann M. et al. (2015): Success Criteria for the Development and Sustainable Operation of Virtual Research Environments. In: D-Lib Magazine, 21(9/10), http://www.dlib.org/dlib/september15/buddenbohm/09buddenbohm.html

Burdick A., Drucker J., Lunenfeld P. *et al*. (2012): Digital Humanities. MIT Press, 2012

Burns C.S., Lana A. & Budd J.M. (2013): Institutional Repositories: Exploration of Costs and Value. In: D-Lib Magazine, 19(1/2), http://www.dlib.org/dlib/january13/burns/01burns.html

Burtenshaw, Paul (2010): A Reply to "What is Public Archaeology?". In: Present Pasts, 1(1), http://www.presentpasts.info/articles/10.5334/pp.8/

Buschman M. & Michalek A. (2013): Are Alternative Metrics Still Alternative? In: ASIS&T Bulletin, 39(4): 35-39, https://asis.org/Bulletin/Apr-13/Bulletin_AprMay13_Final.pdf

C14.sk (C14 radiocarbon data of Central Europe; Comenius University in Bratislava, Department of Archaeology, http://www.c14.sk

CAA - Computer Applications and Quantitative Methods in Archaeology: Online Proceedings, http://proceedings.caaconference.org

Calhoun, Craig (2010): North American social science: trends in and beyond disciplines, pp. 219-224, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Calhoun, Karen (2013a): Digital Library Collections: Repositories. Prepublication copy submitted to Facet Publishing, 16 September 2013, http://ecommons.library.cornell.edu/bitstream/1813/39196/2/04chapter4%20repos%20preprint.pdf

Calhoun, Karen (2013b): The Prospects of Open Access Repositories. Prepublication copy submitted to Facet Publishing, 16 September 2013, http://ecommons.cornell.edu/bitstream/1813/39203/2/08chapter8%20repo%20prospects%20preprint.pdf

Calhoun, Karen (2014): Exploring Digital Libraries: Foundations, Practice, Prospects. London: Facet Publishing & Chicago: ALA Neal-Schuman, http://www.facetpublishing.co.uk/title.php?id=048200&category_code=#table-of-contents-tab / sample: http://www.alastore.ala.org/pdf/Calhoun_sample.pdf

CALIPSO – Way for Light (EU, FP7, coordinates the European synchrotron and free electron laser (FEL) facilities, including the three ESFRI roadmap projects, European XFEL, EuroFEL and the European Synchrotron User Organisation (ESRF) Upgrade Programme, towards an integrated network, http://www.calipso.wayforlight.eu

Callaghan S., Tedds J., Kunze J. *et al*. (2014): Guidelines on recommending data repositories as partners in publishing research data. In: International Journal of Data Curation, 9(1): 152–163. http://www.ijdc.net/index.php/ijdc/article/view/9.1.152

Callaghan, Sarah (2013): Data journals – as soon-to-be-obsolete stepping stone to something better? In: CitingBytes weblog, 4 April 2013, http://citingbytes.blogspot.co.at/2013/01/data-journals-as-soon-to-be-obsolete.html

Candela L., Castelli D., Manghi P. & Tani A. (2015): Data Journals: A Survey. In: Journal of the Association for Information and Science Technology, 01- 2015, http://dx.doi.org/10.1002/asi.23358, pre-print, http://figshare.com/articles/Data_Journals_A_Survey/1358039

Candela, Leonardo (2010): Data Use – Virtual Research Environments, pp. 91-98, in: GRDI2020 - Global Research Data Infrastructures: Towards a 10-year vision for global research data infrastructures. GRDI2020 Consortium, http://www.grdi2020.eu/Repository/FileScaricati/9a85ca56-c548-47e4-8b0e-86c3534ad21d.pdf

Caraher, Bill (2012): Archaeological Data and Small Projects: A Draft, https://mediterraneanworld.wordpress.com/2012/12/31/archaeological-data-and-small-projects-a-draft/

Caraher, Bill (2012): Small Archaeological Projects and the Social Context of their Data, weblog, 17 December 2012, https://mediterraneanworld.wordpress.com/2012/12/17/small-archaeological-projects-and-the-social-context-of-their-data/

Caraher, Bill (2013): iPads in the Field and Reflections on Archaeology's Digital Future [Pyla-Koutsopetria Project]. In: Archaeology of the Mediterranean World weblog, 23 May 2013, https://mediterraneanworld.wordpress.com/2013/05/23/ipads-in-the-field-and-reflections-on-archaeologys-digital-future/

Caraher, William (2013): Slow Archaeology. In: North Dakota Quarterly. North Dakota Quarterly, Vol. 80.2: 80: 43-52, https://www.academia.edu/14327107/Slow_Archaeology

Caraher, William (2015, in preparation): Slow Archaeology: Technology, Efficiency, and Archaeological Work. In: Walcek-Averett E., Gordon J. & Counts D. (eds.): Mobilizing the Past. Grand Forks, ND: The Digital Press at the University of North Dakota

CARARE - Connecting Archaeology and Architecture in Europeana (EU, ICT-PSP, 02/2010-01/2013), http://www.carare.eu

CARARE: The CARARE metadata schema (v1.1, 2011-11; v2.0, 2013-02), http://www.carare.eu/bul/Support/CARARE-metadata-schema

Carlson S. & Anderson B. (2007): What *are* data? The many kinds of data and their implications for data re-use. In: Journal of Computer-Mediated Communication, 12(2): 635-651, http://onlinelibrary.wiley.com/enhanced/doi/10.1111/j.1083-6101.2007.00342.x/

Carroll, Michael W. (2015): Sharing Research Data and Intellectual Property Law: A Primer. In: PLoS Biology, 13(8): e1002235, August 2015, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4551669/

Carusi A. & Reimer T. (2010): Virtual Research Environment - Collaborative Landscape Study. A JISC funded project. Oxford e-Research Centre and Centre for e-Research, King's College London, http://www.jisc.ac.uk/media/documents/publications/vrelandscapereport.pdf

Caruso G., Briguglio L., Matthews B., Tona C. & Albani M. (2013): Modelling Data Value in Digital Preservation. iPres2013, http://purl.pt/24107/1/iPres2013_PDF/Modelling%20Data%20Value%20in%20Digital%20Preservation.pdf

Carver M.O.H. & Klápště J. (eds., 2011): The Archaeology of Medieval Europe. Volume 2: Twelfth to Sixteenth Centuries. Aarhus University Press

CASRAI - Consortia Advancing Standards in Research Administration Information, http://casrai.org

CASRAI: Dataset Level Metrics Subject Group, http://casrai.org/standards/subject-groups/dataset-level-metrics

CASRAI: Dictionary, http://dictionary.casrai.org

Castelli D., Manghi P. & Thanos C. (2013): A vision towards scientific communication infrastructures: on bridging the realms of research digital libraries and scientific data centre. In: International Journal on Digital Libraries, 13(3/4): 155–169, http://wiki.lib.sun.ac.za/images/9/9d/A-vision-for-osc-infrastructure.pdf

Castro J. A., Ribeiro C., Rocha da Silva J. (2013): Designing an Application Profile Using Qualified Dublin Core: A Case Study with Fracture Mechanics Datasets, pp. 47-52, in: DC-2013 - Proc. International Conference on Dublin Core and Metadata Applications, Lisbon, http://dcevents.dublincore.org/IntConf/dc-2013/paper/viewFile/180/76

Çatalhöyük - Excavations of a Neolithic Anatolian Höyük, http://www.catalhoyuk.com

Caves, C.M. (2014): High-impact-factor syndrome. In: American Physical Society - APS News, 23(10), http://www.aps.org/publications/apsnews/201411/backpage.cfm

Cayless H., Roueché C., Elliott T. & Bodard G. (2009): Epigraphy in 2017. In: Digital Humanities Quarterly, Winter 2009 (Special Issue: Changing the Center of Gravity: Transforming Classical Studies Through Cyberinfrastructure), http://www.digitalhumanities.org/dhq/vol/3/1/000030.html

CCSDS - Consultative Committee for Space Data Systems (2012): Reference Model for an Open Archival Information System (OAIS). Recommended Practice CCSDS 650.0-M-2, June 2012, http://public.ccsds.org/publications/archive/650x0m2.pdf

CCSDS (2011): Audit and certification of trustworthy digital repositories. NASA, Washington, D.C. http://public.ccsds.org/publications/archive/652x0m1.pdf

CENDARI - Collaborative European Digital Archive Infrastructure (EU, FP7, 02/2012-01/2016), http://www.cendari.eu

Center for Research Libraries (2007): Ten Principles [10 principles digital repositories should exhibit in order to be trusted], http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re

Center for Research Libraries (2007): TRAC - Trustworthy repositories audit & certification: criteria and checklist. Chicago & Dublin (USA); OCLC and CRL, http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Central European Research Infrastructure Consortium (CERIC-ERIC), http://www.ceric-eric.eu

Centre for Past Climate Change, School of Archaeology, Geography and Environmental Sciences (SAGES) at University of Reading, https://www.reading.ac.uk/sages/research/CPCC/

Centre for Science and Technology Studies (CWTS), Leiden University, http://www.cwts.nl

Centre for the Future of the Museum (2014): TrendsWatch 2014. American Alliance of Museums, http://www.aam-us.org/resources/center-for-the-future-of-museums/projects-and-reports/trendswatch

Centre SMAArt - Scientific Methodologies applied to Archaeology and Art (University of Perugia, Italy), http://www.smaart.it

Cervieri, Michael (2011): The lifespan of online content is nasty, brutish and short (21 June 2011), http://www.archive-org-2013.com/org/s/2013-12-18_3365322_2/The-lifespan-of-online-content-is-nasty-brutish-and-short-ScribeMedia-org/

CESSDA - Consortium of European Social Science Data Archives (ESFRI research infrastructure), http://cessda.net

Chao, Tiffany C. (2015): Mapping methods metadata for research data. In: International Journal of Digital Curation, 10(1): 82-94, http://www.ijdc.net/index.php/ijdc/article/view/10.1.82

Chapman R. & Wylie A. (eds., 2014): Material Evidence: Learning from Archaeological Practice Paperback. Routledge

CHARISMA - Cultural Heritage Advanced Research Infrastructures: Synergy for a Multidisciplinary Approach to Conservation/Restoration (EU, FP7, Integrating Activity, 10/2009-09/2013), http://www.charismaproject.eu

CHARISMA: Art database portal, http://archives-charisma-portal.eu

CHARM - Cultural Heritage Abstract Reference Model (C. Gonzalez-Perez *et al*., Institute of Heritage Sciences [Incipit], Spain), http://www.charminfo.org

Chase A.F., Chase D.Z., Fisher C.T. *et al*. (2012): Geospatial revolution and remote sensing LiDAR in Mesoamerican archaeology. In: PNAS - Proceedings of the National Academy of Sciences, 109(3): 12916-21, http://www.pnas.org/content/109/32/12916.full.pdf

Chavan V. & Ingwersen P. (2009): Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. In: BMC Bioinformatics, 10(Suppl 14):S2, http://www.biomedcentral.com/1471-2105/10/S14/S2

Chavan V. & Penev L. (2011): The data paper: a mechanism to incentivize data publishing in biodiversity science. In: BMC Bioinformatics, 12(Suppl 15):S2, http://www.biomedcentral.com/1471-2105/12/S15/S2

Chepstow-Lusty A.J., Frogley M.R., Bauer B.S. *et al*. (2009): Putting the rise of the Inca Empire within a climatic and land management context. In: Climate of the Past, 5: 375–388, http://www.clim-past.net/5/375/2009/cp-5-375-2009.pdf

Childs T. & Kinsey K. (2003): Cost of curating archaeological collections. A study of repository fees in 2002 and 1997/98. US National Park Service, Washington, http://www.nps.gov/archeology/tools/feesstud.htm

Childs, S. Terry (2006): Archaeological Collections: Valuing and Managing an Emerging Frontier, pp. 204-210, in: Getty Conservation Institute / Agnew N. & Bridgland J. (2006): Of the Past, For the Future: Integrating Archaeology and Conservation. Proceedings of the Conservation Theme at the World Archaeological Congress, June 2003. Los Angeles, USA, http://www.getty.edu/conservation/publications_resources/pdf_publications/of_past_for_future.html

Christian, Gideon E. (2009): Building a sustainable framework for open access to research data through information and communications technologies. International Development Research Centre, Canada, December 2009, http://idl-bnc.idrc.ca/dspace/bitstream/10625/41336/1/129183.pdf

Chue Hong N., Hole B. & Moore S. (2013): Software Papers: improving the reusability and sustainability of scientific software. figshare, http://dx.doi.org/10.6084/m9.figshare.795303

CIDOC - Conceptual Reference Model (CIDOC-CRM), http://www.cidoc-crm.org

CIDOC-CRM - Definition of the CIDOC Conceptual Reference Model. Version 6.1, February 2015, http://www.cidoc-crm.org/docs/cidoc_crm_version_6.1.pdf

CIDOC-CRM - ISO 21127:2006, revised 2014: Information and documentation - A reference ontology for the interchange of cultural heritage information, http://www.iso.org/iso/catalogue_detail?csnumber=57832

CIDOC-CRM extension: CRMarcheo: the Excavation Model, http://www.ics.forth.gr/isl/index_main.php?l=e&c=711

CIDOC-CRM extension: CRMdig: a model for provenance metadata, http://www.ics.forth.gr/isl/index_main.php?l=e&c=656

CIDOC-CRM extension: CRMgeo: a spatiotemporal model, http://www.ics.forth.gr/isl/index_main.php?l=e&c=661

CIDOC-CRM extension: CRMinf: an Argumentation model, http://www.ics.forth.gr/isl/index_main.php?l=e&c=713

CIDOC-CRM extension: CRMsci: the Scientific Observation Model, http://www.ics.forth.gr/isl/index_main.php?l=e&c=663

CIDOC-CRM extensions (overview), http://www.ics.forth.gr/isl/CRMext/

Ciuchini, Paolo (2010): Archaeological policies and practices across Europe: national developments and international trends, http://www.archaeojobs.com/2010/07/archaeological-policies-and-practices.html / https://www.academia.edu/273719/Archaeological_policies_and_practices_across_Europe_national_developments_and_international_trends

Civic Epistemologies (EU, FP7, 08/2014-03/2016, promotes participation of citizens in humanities and cultural heritage research), http://www.civic-epistemologies.eu

CLARIN - Common Language Resources and Technology Initiative (ERIC), http://www.clarin.eu

CLAROS – The World of Art on the Semantic Web (hosted by University of Oxford's e-Research Centre), http://www.clarosnet.org

Climate Change and Heritage Management: Measuring and Monitoring the Impacts of Future Climate and Environmental Change on the Historic Environment and Cultural Resources. EAA-2015 coonference session, abstracts and videos. Doug's Archaeology (Doug Rocks-Macqueen) weblog, https://dougsarchaeology.wordpress.com/2015/11/03/

Climate for Culture (FP7-ENV, 2009-2014), project for the protection cultural heritage against impacts of climate change, http://www.climateforculture.eu

Climate of the Past - An interactive open-access journal of the European Geosciences Union, http://www.climate-of-the-past.net

CLIR - Council on Library and Information Resources (2009): Working Together or Apart: Promoting the Next Generation of Digital Scholarship. Washington: CLIR, http://www.clir.org/pubs/reports/pub145/pub145.pdf

CLIR - Council on Library and Information Resources (2013): Research Data Management. Principles, Practices, and Prospects. DataRes project report. Washington: CLIR, http://www.clir.org/pubs/reports/pub160

CLOCKSS - Controlled LOCKSS (service for long-term preservation of digital content), http://www.clockss.org

COAR - Confederation of Open Access Repositories (2013): Incentives, Integration, and Mediation: Sustainable Practices for Populating Repositories, June 2013, http://www.coar-repositories.org/files/Sustainable-best-practices_final.pdf

COAR - Confederation of Open Access Repositories (2015): COAR Roadmap - Future Directions for Repository Interoperability, February 2015, https://www.coar-repositories.org/files/Roadmap_final_formatted_20150203.pdf

COAR - Confederation of Open Access Repositories (launched in October 2009, about 100 members and partners), http://www.coar-repositories.org

COAR - Confederation of Open Access Repositories report (2012): The Current State of Open Access Repository Interoperability, October 2012, https://www.coar-repositories.org/files/COAR-Current-State-of-Open-Access-Repository-Interoperability-26-10-2012.pdf

Cochran, Angela (2014): Interview with Thomson Reuters: InCites Platform Offers New Analytics and Transparency. In: Scholarlykitchen weblog, 8 August 2014, http://scholarlykitchen.sspnet.org/2014/08/08/qa-with-thomson-reuters-incites-platform-offers-new-analytics-and-transparency/

Cochrane G., Cook C.E. & Birney E. (2012): The future of DNA sequence archiving. GigaScience 1:2, http://www.gigasciencejournal.com/content/1/1/2

CODATA - Committee on Data for Science and Technology of the International Council for Science (ICSU), http://www.icsu.org

CODATA - ICSTI Task Group on Data Citation Standards and Practices (2013): Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data. In: Data Science Journal 12, 13 September 2013, https://www.jstage.jst.go.jp/article/dsj/12/0/12_OSOM13-043/_article

Cole, Roland J. (2012): Some observations on the practice of open data as opposed to its promise. In: Journal of Community Informatics, 8(2), http://ci-journal.net/index.php/ciej/article/view/920

Collar A., Coward F., Brughmans T. & Mills B.J. (2015): Networks in Archaeology: Phenomena, Abstraction, Representation. In: Journal of Archaeological Method and Theory, 22(1): 1-32, http://link.springer.com/article/10.1007/s10816-014-9235-6

Colley, Sarah (2013): Social media and archaeological communication: an Australian survey. In: Archäologische Informationen, Bd. 36, https://journals.ub.uni-heidelberg.de/index.php/arch-inf/article/view/15385

Collins, Sandra (2014): Service provider perspective: Digital Repository of Ireland. Presentation at 4C/DPC Conference, London, 17-18 November 2014, http://4cproject.eu/component/docman/doc_download/74-4c-conference-day-1-session-3-collins?Itemid=

Comer D.C., Soderland H.A., Biehl P.F. & Prescott C. (2014): The Future Challenges of Heritage and Identity in a Globalized World, pp. 161-167, in: Identity and Heritage. Contemporary Challenges in a Globalized World. Springer

Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age (2009): Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. Washington: National Academy of Sciences, http://www.nap.edu/catalog/12615.html

Connected Island (2012): Introducing 'A Connected Island?', Iza Romanowska and Tom Brughmans, . 21 January 2013, http://blog.soton.ac.uk/connectedisland/

Connected Island (2013): Citation Network Analysis. In: Archaeological Networks weblog, 18 February 2013, https://archaeologicalnetworks.wordpress.com/2013/02/18/connected-island-citation-network-analysis/

Constantopoulos P., Dallas C., Androutsopoulos I. *et al.* (2009): DCC&U: An Extended Digital Curation Lifecycle Model. In: The International Journal of Digital Curation, 4(1): 34-45, http://www.ijdc.net/index.php/ijdc/article/viewFile/100/75

Cooper A. & Green C. (2015): Embracing the complexities of "Big Data" in archaeology: the case of the English Landscape and Identities Project. In: Journal of Archaeological Method and Theory, 22(1), http://link.springer.com/article/10.1007%2Fs10816-015-9240-4

Cooper J. & Sheets P. (eds., 2012): Surviving Sudden Environmental Change: Answers from Archaeology. Boulder: University Press of Colorado, http://www.upcolorado.com/university-press-of-colorado/item/download/363_92a5da8833cdd3f69f5bb4dcc8b2d7ef

Copernicus (2013): Free and open access to Sentinel satellite data (15 November 2013, http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Free_access_to_Copernicus_Sentinel_satellite_data

CoPoRI - Communication and Policy development for pan-European Research Infrastructures (FP7, 10/2011-03/2014, supports ESFRI), http://www.copori.eu

Corns A., Kennedy L. & Stular B. (2014): Impressions from the ARIADNE Community, pp. 83-89, in: ARIADNE – The Way Forward to Digital Archaeology in Europe. ARIADNE publication, November 2014, http://www.ariadne-infrastructure.eu/content/download/4569/26666/version/2/file/Ariadne+Booklet.pdf

Corrall S., Kennan M.A. & Afzal W. (2013): Bibliometrics and research data management: emerging trends in library research support services. In: Library Trends, 61(3), http://muse.jhu.edu/login?auth=0&type=summary&url=/journals/library_trends/v061/61.3.corrall02.html

Costa S., Beck A., Bevan A. & Ogden J. (2012): Defining and advocating Open Data in archaeology, pp. 449-456, in: Proceedings of the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), University of Southampton, 26-30 March 2012, http://dare.uva.nl/cgi/arno/show.cgi?fid=516092

Costa, Stefano (2014): Open licenses for archaeological data matter: the case of AustArch, 29 July 2014, http://archaeology.okfn.org/tag/radiocarbon/

Costas R., Meijer I., Zahedi Z. & Wouters P. (2013): The Value of Research Data. Metrics for datasets from a cultural and technical point of view. Center for Science and Technology Studies, Leiden University, http://www.knowledge-exchange.info/Default.aspx?ID=586

Costas R., Zahedi Z. & Wouters P. (2014): Do 'altmetrics' correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. Centre for Science and Technology Studies (CWTS), Leiden University, https://openaccess.leidenuniv.nl/handle/1887/23041

Costello M.J., Michener W.K., Gahegan M., Zhang Z.Q. & Bourne P.E. (2013): Biodiversity data should be published, cited, and peer reviewed. In: Trends in Ecology and Evolution, 28(8): 454–461, http://www.sciencedirect.com/science/article/pii/S0169534713001092

Costs4C - Collaboration to Clarify the Cost of Curation (2014a): Evaluation of Cost Models and Needs & Gaps Analysis, 4C project deliverable, June 2014 http://www.4cproject.eu/d3-1

Costs4C - Collaboration to Clarify the Cost of Curation (2014b): Cost Concept Model and Gateway Specification, 4C project deliverable D3.2, August 2014, http://www.4cproject.eu/d3-2-ccm-grs

Costs4C - Collaboration to Clarify the Costs of Curation (EU, FP7-ICT, 02/2013-01/2015), http://www.4cproject.eu

Council of Europe (1985): Convention for the Protection of the Architectural Heritage of Europe, Granada, 3.X.1985, http://conventions.coe.int/treaty/en/Treaties/Html/121.htm

Council of Europe (1992): European Convention on the Protection of the Archaeological Heritage (Revised), Valetta, 16.I.1992, http://conventions.coe.int/Treaty/en/Treaties/Html/143.htm

Council of Europe (2000): European Landscape Convention, Florence, 20.X.2000, http://conventions.coe.int/treaty/en/Treaties/Html/176.htm

Council of Europe: European Convention on the Protection of the Archaeological Heritage (Revised), Valetta, 16.I.1992, http://conventions.coe.int/Treaty/en/Treaties/Html/143.htm

Council of Europe: Framework Convention on the Value of Cultural Heritage for Society, Faro, 27.X.2005, http://conventions.coe.int/Treaty/en/Treaties/Html/199.htm

Council of the European Union (2009): The Future of ICT research, innovation and infrastructures: Adoption of Council Conclusions. Council of the European Union Document No. 16128/9, http://europa.eu/documentation/official-docs/index_en.htm

Council of the European Union (2014): Conclusions on cultural heritage as a strategic resource for a sustainable Europe. Brussels, 20 May 2014, https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/educ/142705.pdf

Council of Europe: Cultural heritage, http://www.coe.int/t/dg4/cultureheritage/

Cox A.M, Verbaan E. & Sen B. (2012): Upskilling liaison librarians for research data management. In: Ariadne, Issue 70, http://www.ariadne.ac.uk/issue70/cox-et-al

Cox A.M. & Pinfield S. (2013): Research data management and libraries: Current activities and future priorities. In: Journal of Librarianship and Information Science, June 2013, http://eprints.whiterose.ac.uk/76107/7/WRRO_76107.pdf

Cragin M., Palmer C., Carlson J. & Witt M. (2010): Data sharing, small science and institutional repositories. In: Transactions of the Royal Society, A-368: 4023-4038. http://rsta.royalsocietypublishing.org/content/368/1926/4023.full.pdf

Crane G., Babeu A. & Bamman D. (2007): eScience and the humanities. In: International Journal on Digital Libraries, 7(1): 117–122, http://www.perseus.tufts.edu/~amahoney/jdlsciencecomm.pdf

Crawford, Gregory A. (2013): A Citation Analysis of the Classical Philology Literature: Implications for Collection Development. In: Evidence Based Library and Information Practice, 8(2), http://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/18598/15201

Creative Commons (2009): Defining “Noncommercial”: A Study of How the Online Population Understands “Noncommercial Use”, September 2009, https://wiki.creativecommons.org/wiki/Defining_Noncommercial

Creative Commons (2013): 4.0/NonCommercial, https://wiki.creativecommons.org/wiki/4.0/NonCommercial

Creative Commons (CC) licenses, https://creativecommons.org/licenses/?lang=en

Creative Commons: Data, https://wiki.creativecommons.org/wiki/Data

CRediT - Contributor Roles Taxonomy project (develops an open standard for expressing roles intrinsic to research), http://credit.casrai.org

Cross May S. & Crosby V. (2004): Holy Grail or Poison Chalice? Challenges in Implementing Digital Excavation Recording. Proceedings of the CAA 2004 conference, Prato, Italy, 13-17 April 2004, pp. 49-54, http://proceedings.caaconference.org/files/2004/06_May_Crosby_CAA_2004.pdf

Crowston K. & Qin J. (2011): A Capability Maturity Model for Scientific Data Management: Evidence from the Literature. In: Proceedings of the American Society for Information Science and Technology, 48(1): 1–9, http://crowston.syr.edu/system/files/110718%20CMM%20ASISTpaper.pdf

Cultural Heritage Partners (2012): Federal Laws and Regulations Requiring Curation of Digital Archaeological Documents and Data, 25 October 2012, http://www.digitalantiquity.org/wp-uploads/2013/05/2013-CHP-Legal-Analysis-of-Fed-Req-for-Curation-of-Dig-Arch-Docs-Data-.pdf

Curation Costs Exchange (CCEx, established by Costs4C), http://www.curationexchange.org

Curcin V. & Ghanem M. (2008): Scientific workflow systems - can one size fit all? Proceedings of the IEEE, CIBEC'08 - Cairo International Biomedical Engineering Conference, http://www.doc.ic.ac.uk/~vc100/papers/Scientific_workflow_systems.pdf

Curtis G., Hammond M. & Oppenheim C. (2013): Access to Citation Data: Cost-benefit and Risk Review and Forward Look. JISC, September 2013, http://repository.jisc.ac.uk/5371/1/Access-to-Citation-data-report-final.pdf

Curtis+Cartwright Consulting (2010): Using the Cloud for Research. Final Report. Guildford, UK, http://www.jisc.ac.uk/whatwedo/programmes/researchinfrastructure/usingcloudcomp.aspx

DAACS - Digital Archaeological Archive of Comparative Slavery (Robert H. Smith ICJS, Department of Archaeology, Monticello, USA), http://www.daacs.org

DADAISM - Digging into Archaeological Data and Image Search Metadata (Digging Into Data Challenge, round 3, UK, AHRC/ESRC, 06/2014-12/2015), http://dadaism-did.org

Dallas, Costis (2009): From Artefact Typologies to Cultural Heritage Ontologies: Or, an Account Of the Lasting Impact of Archaeological Computing. In: Archeologia e Calcolatori, 20: 205-221, http://soi.cnr.it/archcalc/indice/PDF20/17_Dallas.pdf

Dallas, Costis (2015): Curating Archaeological Knowledge in the Digital Continuum: from Practice to Infrastructure. In: Open Archaeology, 1(1), 176–207, http://www.degruyter.com/view/j/opar.2014.1.issue-1/opar-2015-0011/opar-2015-0011.xml?format=INT

Dallmeier-Tiessen S., Darby R., Gitmans K. *et al.* (2014): Enabling Sharing and Reuse of Scientific Data. In: New Review of Information Networking, 19(1): 16-43, http://www.tandfonline.com/doi/full/10.1080/13614576.2014.883936

Danish National Archives / Danish Royal Library, http://www.costmodelfordigitalpreservation.dk

DANS - Data Archiving and Networked Services, Netherlands, http://www.dans.knaw.nl/en

DANS - E-Depot for Dutch Archaeology (EDNA), http://www.edna.nl

DARIAH - Digital Research Infrastructure for the Arts and Humanities (ERIC), https://www.dariah.eu

DART - Detection of Archaeological residues using Remote sensing Techniques (UK, Science and Heritage programme, AHRC and EPSRC, 10/2010-09/2013), DARTPortal, https://dartportal.leeds.ac.uk

DASISH - Data Service Infrastructure for the Social Sciences and Humanities (EU, FP7, 01/ 2012-12/2014), http://dasish.eu

Data Citation Index (Thomson Reuters; launched in October 2012), http://wokinfo.com/products_tools/multidisciplinary/dci/

Data Conservancy (NSF DataNet funded project), http://dataconservancy.org

Data Curation Profiles (Purdue University) website, http://docs.lib.purdue.edu/datacurationprofiles/

Data Documentation Initiative (DDI), http://www.ddialliance.org

Data Seal of Approval, http://www.datasealofapproval.org

DataCite, http://www.datacite.org

DataCite: Metadata Schema for the Publication and Citation of Research Data, https://schema.datacite.org

DataNet Federation Consortium (USA, NSF DataNet funded project), http://datafed.org

DataONE - Data Observation Network for Earth (USA, NSF DataNet funded project, focused on environmental and earth science data), http://www.dataone.org

DataONE: Best practices (data management), https://www.dataone.org/all-best-practices

Data without Boundaries - DwB project (EU, FP7, 12/2011-04/2015; quantitative social and economic data services), http://www.dwbproject.org

David P.A. & Spence M. (2003): Towards Institutional Infrastructures for e-Science: The Scope of the Challenge. Final Report to the Joint Information System Committee of the UK Research Councils, Oxford Internet Institute, Research Report No. 2, http://www.oii.ox.ac.uk/resources/publications/OIIRR_E-Science_0903.pdf

Dávidházi, Péter (ed., 2014): New Publication Cultures in the Humanities: Exploring the Paradigm Shift. Amsterdam University Press, https://oapen.org/download?type=document&docid=515678

Davidson, Osha G (2014): Data drives everything (but the bridges need a lot of work). In: Earthzine, 11 June 2014, http://earthzine.org/2014/06/11/data-drives-everything-but-the-bridges-need-a-lot-of-work/

Davies M.I.J. & M'Mbogori F.N. (eds., 2013): Humans and the Environment: New Archaeological Perspectives for the 21st Century. Oxford: Oxford University Press.

Davies T., Perini F. & Alonso J.M. (2013): Researching the emerging impacts of open data - ODDC conceptual framework. International Development Research Centre & WWW Foundation 'Exploring the Emerging Impacts of Open Data in Developing Countries' (ODDC) research programme, July 2013, http://www.opendataresearch.org/sites/default/files/posts/Researching%20the%20emerging%20impacts%20of%20open%20data.pdf

Davis, Philip M. (2010): Access, Readership, Citations: A randomized controlled trial of scientific journal publishing. PhD thesis, Cornell University, USA, http://ecommons.cornell.edu/bitstream/1813/17788/1/Davis,%20Philip.pdf

Davis, Philip M. (2011): Open access, readership, citations: a randomized controlled trial of scientific journal publishing. In: The FASEB Journal, 25(7): 2129-34, http://www.fasebj.org/content/25/7/2129.full

DCCD - Digital Collaboratory for Cultural Dendrochronology, http://dendro.dans.knaw.nl

DCH-RP - Digital Cultural Heritage Roadmap for Preservation - Open Science Infrastructure for Digital Cultural Heritage in 2020 (EU, FP7, 10/2012-09/2014, http://www.dch-rp.eu

De Roure, David (2010): Replacing the Paper: The Twelve Rs of the e-Research Record. Scilogs E-Research blogs, 27 November 2010, http://www.scilogs.com/eresearch/replacing-the-paper-the-twelve-rs-of-the-e-research-record/

De Roure, David (2012): More Rs than Pirates. In: Scilogs E-Research blogs, 10 June 2012, http://www.scilogs.com/eresearch/more-rs-than-pirates/

De Roure, David (2013): Towards Computational Research Objects, pp. 16-19, in: Proceeding of DPRMA '13 - 1st International Workshop on Digital Preservation of Research Methods and Artefacts, Indianapolis, 25-26 July 2013. New York: ACM, https://dl.dropboxusercontent.com/u/15772302/ComputationalResearchObjects.pdf

Dearing, J. A. (2006): Climate-human-environment interactions: resolving our past. Climate of the Past, 2: 187–203, http://www.clim-past.net/2/187/2006/cp-2-187-2006.pdf

Decker E. & Volkmann A. (2015): RetroDig - Standardised work flows for retro-digitalisation: a case study of the excavation records of the roman fort of Heidelberg-Neuenheim, http://www.asia-europe.uni-heidelberg.de/en/research/heidelberg-research-architecture/projects/retrodig.html

Deegan M. & McCarty W. (eds., 2012): Collaborative Research in the Digital Humanities. Ashgate

Deelman E., Gannon D., Shields M. & Taylor I. (2009): Workflows and e-Science: An overview of workflow system features and capabilities. In: Future Generation Computer Systems, 25(5): 528-540, http://www.sciencedirect.com/science/article/pii/S0167739X08000861

Degryse P. & Shortland A.J. (2013): Nourishing archaeology and science. In: PNAS - Proceedings of the National Academy of Sciences of the United States of America, 110(51), https://lirias.kuleuven.be/bitstream/123456789/437071/1/PNAS-2013-Degryse-20352-3.pdf

Demoule J.-P. (2010): The crisis – economic, ideological and archaeological, pp. 13-17, in: Schlanger N. & Aitchison K. (eds.): Archaeology and the Global Crisis. Multiple Impacts, Possible Solutions. Tervuren: Culture lab Éditions, http://digital.csic.es/bitstream/10261/27633/1/2010_25Archaeology-and-the-crisis.pdf

Destro-Bisol G., Anagnostou P., Capocasa M. *et al*. (2014): Perspectives on open science and scientific data sharing: an interdisciplinary workshop. In: Journal of Anthropological Sciences, Vol. 92: 1-22, http://www.isita-org.com/jass/Contents/2014vol92/Destro/Destro.pdf

Dextre, Stella (2014): Workshop report: Vocabularies and the Potential for Linkage (VPL link) Workshop. Jointly organized by ISKO UK, DCMI and BCS IRSG, London, 23 June 2014, https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2014/content/VPLinkWorkshopReport2014-07-09.pdf

Di Cosmo, Nicola (2011): DNA, History, and Archaeology. In: IAS - The Institute Letter, Spring 2011 (Institute for Advanced Study, New Jersey, USA), article online; https://www.ias.edu/about/publications/ias-letter/articles/2011-spring/dna-history-archaeology-di-cosmo

DigCurV - Digital Curator Vocational Education Europe project (EU, Leonardo da Vinci, 01/2011-06/2013), http://www.digcur-education.org

Digital Antiquity (see also tDAR), http://www.digitalantiquity.org

Digital Classicist, http://www.digitalclassicist.org

Digital Curation Centre - DCC (UK; offers a wide range of data management resources), http://www.dcc.ac.uk

Digital Karnak (University of California at Los Angeles), http://dlib.etc.ucla.edu/projects/Karnak/

Digital Repository of Ireland (national repository for social and cultural data), http://dri.ie

Digital Science (owned by Macmillan Publishers; among other services provide the Figshare repository), http://www.digital-science.com/products/

DigitalPreservationEurope (2008): Repository planning checklist and guidance DPE-D3.2. Glasgow: HATII, http://www.digitalpreservationeurope.eu/platter.pdf

DigVentures - Archaeology in your hands (crowd-funding platform), http://www.digventures.com

Dijk E. & Doorn P. (2014): Providing access to research data, publications and current research information at Data Archiving and Networked Services at DANS, pp. 12-19, in: Proceedings of the Conference on Grey Literature and Repositories, Prague: National Library of Technology, http://nrgl.techlib.cz/index.php/Proceedings_2014 (or: http://repozitar.techlib.cz/record/809/files/idr-809_4.pdf

DIPIR - Dissemination Information Packages for Information Reuse project (USA, funded by the Institute of Museum and Library Services, http://dipir.org

Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32013L0037

Directive 2014/52/EU of the European Parliament and of the Council of 16 April 2014 amending Directive 2011/92/EU on the assessment of the effects of certain public and private projects on the environment, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0052

Directive 96/9/EC of the European Parliament and of the Council on the Legal Protection of Databases, http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31996L0009

Directory of Open Access Books – DOAB (OAPEN Foundation), http://www.doabooks.org

DISCO - Discovering the Archaeologists of Europe (2009): Transnational Report. Prepared by Kenneth Aitchison / Institute for Archaeologists, January 2009, http://www.discovering-archaeologists.eu/DISCO_Transnational_Report.pdf

DISCO - Discovering the Archaeologists of Europe (2014): Transnational Report. York Archaeological Trust, November 2014, http://www.discovering-archaeologists.eu/national_reports/2014/transnational_report.pdf

DISCO - Discovering the Archaeologists of Europe project (EU, Leonardo da Vinci II, 2006-2008; Lifelong Learning Programme, 2012-2014), http://www.discovering-archaeologists.eu

DM2E - Digital Manuscripts to Europeana project (EU, ICT-PSP, 02/2012-01/2015), http://dm2e.eu

DOAP - Description of a Project (defines terms for describing software projects, particularly those that are Open Source), http://trac.usefulinc.com/doap

Doerr M. & Hiebel G. (2013): CRMgeo: Linking the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement. Technical report: ICS-FORTH/TR-435, April 2013, http://www.cidoc-crm.org/docs/Technical%20Report435-CRMgeo.pdf

Doerr M. & Oldman D. (2013): The Costs of Cultural Heritage Data Services: The CIDOC CRM or Aggregator formats? Dominic Oldman weblog, 13 June 2013, http://www.oldman.me.uk/blog/costsofculturalheritage/

Doerr M., Bekiari C., Kritsotaki A., Hiebel G. & Theodoridou M. (2014b): Modelling Scientific Activities: Proposal for a global schema for integrating metadata about scientific observation. Paper presented at the CIDOC 2014 Conference, 6th-11th Sept. 2014, Dresden/Germany, http://www.cidoc2014.de/images/sampledata/cidoc/papers/E-2_Bekiari_paper.pdf

Doerr M., de Jong G., Konsolaki K., Norton B., Oldman D., Theodoridou M. & Wikman T. (2014a): The SYNERGY Reference Model of Data Provision and Aggregation. Draft, June 2014, http://www.cidoc-crm.org/docs/SRM_v0.1.pdf

Doerr M., Hiebel G. and Eide Ø. (2013): CRMgeo: Linking the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement. Technical Report: ICS-FORTH, http://www.ics.forth.gr/tech-reports/2013/2013.TR435_CRMgeo_CIDOC_CRM_GeoSPARQL.pdf

Doerr M., Hiebel G. and Eide Ø. (2013): CRMgeo: Linking the CIDOC CRM to GeoSPARQL through a Spatiotemporal Refinement. Technical Report: ICS-FORTH, http://www.ics.forth.gr/tech-reports/2013/2013.TR435_CRMgeo_CIDOC_CRM_GeoSPARQL.pdf

Doerr M., Kritsotaki A. & Boutsika A. (2011): Factual argumentation - a core model for assertions making. In: Journal on Computing and Cultural Heritage, 3(3), Article No.8, http://dl.acm.org/citation.cfm?id=1921615

DOI - Digital Object Identifier, http://www.doi.info

Dombrowski Q. & Perkins J. (2014): TaDiRAH: Building Capacity for Integrated Access. In: dh+lib weblog, 21 May 2014, http://acrl.ala.org/dh/2014/05/21/tadirah-building-capacity-integrated-access/

Doorn, Peter (2013): Cost and benefits of digital preservation in a federated data infrastructure: Who pays what and who gets what in The Netherlands? EUDAT Conference: Cost and Funding Models, Rome, 28-30 October 2013, http://eudat.eu/sites/default/files/DOORN%20EUDAT%2030OCT2013.pdf

Dormans S. & Kok J.(2010): An alternative approach to large historical databases. Exploring best practices with collaboratories. In: Historical Methods, 43(3): 97-107, http://www.tandfonline.com/doi/abs/10.1080/01615440.2010.496443

Doug's Archaeology (Doug Rocks-Macqueen) weblog, https://dougsarchaeology.wordpress.com

DRAMBORA - Digital Repository Audit Method Based on Risk Assessment (EU-funded project, Digital Curation Center & DigitalPreservationEurope, 2006-2009), http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora

DRIVER - Digital Repository Infrastructure Vision for European Research (DRIVER II: 12/2007-11/2009), http://www.driver-support.eu

DRYAD (bio-sciences data repository, since 2008, sponsored by the National Science Foundation, USA), http://datadryad.org

DRYAD: Joint Data Archiving Policy (JDAP) statement (adopted by many bio-sciences journals), http://datadryad.org/pages/jdap

DRYAD: Metadata Application Profile, http://wiki.datadryad.org/Metadata_Profile

Dubinsky, Ellen (2014): A current snapshot of institutional repositories: Growth rate, disciplinary content and faculty contributions. In: Journal of Librarianship and Scholarly Communication, 2(3): eP1167, http://www.jlsc-pub.org/jlsc/vol2/iss3/3/

Dublin Core Collection Application Profile (DCCAP), http://dublincore.org/groups/collections/collection-application-profile/

Ducke, Benjamin (2012): Natives of a connected world: free and open source software in archaeology. In: World Archaeology, 44(4): 571-579, http://www.tandfonline.com/toc/rwar20/44/4#.VSen25Mufh

Duin D., King D. & van den Besselaar P. (2012): Identifying Audiences of E-Infrastructures - Tools for Measuring Impact. PLOS ONE, 7(12): e50943, http://dx.doi.org/10.1371/journal.pone.0050943

Dunn S. & Hedges M. (2012): Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research. Centre for e-Research, Department of Digital Humanities, King's College London, November 2012, http://crowds.cerch.kcl.ac.uk/wp-content/uploads/2012/12/Crowdsourcing-connected-communities.pdf

Dunn, Stuart (2011): Poor Relatives or Favorite Uncles? Cyberinfrastructure and Web 2.0: A Critical Comparison for Archaeological Research, pp. 95-117, in: Kansa E. *et al*. (eds.): Archaeology 2.0: New Approaches to Communication and Collaboration. Cotsen Institute of Archaeology, UC Los Angeles, http://escholarship.org/uc/item/1r6137tb

Dunn, Stuart (2012): CAA - The Digital Humanities and Archaeology Venn, 1 April 2012, http://stuartdunn.wordpress.com/2012/04/01/caa1-the-digital-humanities-and-archaeology-venn-diagram/

Duranti L. & Shaffer E. (eds., 2012): Memory of the World in the Digital Age: Digitization and Preservation. Conference proceedings. UNESCO, http://ciscra.org/docs/UNESCO_MOW2012_Proceedings_FINAL_ENG_Compressed.pdf

Duşa A., Nelle D., Stock G. & Wagner G. (eds., 2014): Facing the Future: European Research Infrastructures for the Humanities and Social Sciences. Berlin: Scivero, online: http://www.allea.org/Content/ALLEA/2014_06_04-FACING_THE_FUTURE.pdf

Dutch Dataverse Network (managed by DANS - Data Archiving and Networked Services, https://dataverse.nl

Dutton W.H. & Jeffreys P. (eds.): World Wide Research: Reshaping the Sciences and Humanities. Cambridge: MIT Press

Dutton W.H. & Meyer E.T. (2009): Experience with New Tools and Infrastructures of Research: An exploratory study of distance from, and attitudes toward, e-Research. In: Prometheus, 27(3): 223-238, http://www.tandfonline.com/doi/abs/10.1080/08109020903127802

Dutton, William H. (2013): The social shaping of digital research. In: International Journal of Social Research Methodology, 16(3): 177-195, http://www.researchgate.net/publication/256014826_The_Social_Shaping_of_Digital_Research

EAGLE - Europeana Network of Ancient Greek and Latin Epigraphy project (EU, ICT-PSP, 04/2013-03/2016), http://www.eagle-network.eu

Eaker, Chris (2012): Architectural History / Epigraphy - University of Tennessee Knoxville. Data Curation Profiles Directory: Vol. 4, Article 1, http://docs.lib.purdue.edu/dcp/vol4/iss1/1/

Earl, Graeme (2012): The Big Digital Archaeology Digital Humanities Venn Diagram Show, 28 March 2012, https://storify.com/graemeearl/the-big-digital-archaeology-digital-humanities-ven

Earth Science Information Partners (ESIP): Data Management Short Course for Scientists, http://commons.esipfed.org/datamanagementshortcourse

EarthCube (2011-, USA, NSF-funded geosciences e-infrastructure), http://earthcube.org

EarthCube (2012a): EarthCube Roadmap for Creating the Semantic/Ontologic Infrastructure for the Geosciences. Prepared by the Semantic/Ontology Community Group. July 2012 http://earthcube.org/document/2012/semantic-ontologic-infrastructure-roadmap

EarthCube (2012b): EarthCube Roadmap. Prepared by the Cross-Domain Interoperability Test Bed Group, August 2012, http://earthcube.org/document/2012/earthcube-cross-domain-interoperability-roadmap

EarthCube (2014): EarthCube: Past, Present, and Future. Edited by Y. Gil, M Chan, B. Gomez & B. Caron. EarthCube Project Report EC-2014-3, December 2014, http://earthcube.org/sites/default/files/doc-repository/EC-PPF-FINAL.pdf

Eastern Vani Survey project (Georgia), http://www.lsa.umich.edu/kelsey/fieldwork/currentfieldwork/vanirepublicofgeorgia

EC & ESF - European Commission and European Science Foundation (2007): Trends in European Research Infrastructures. Analysis of data from the 2006/07 survey, July 2007, http://ec.europa.eu/research/infrastructures/pdf/survey-report-july-2007_en.pdf

Edgeworth M., Benjamin J., Clarke B. *et al*. (2014): Archaeology the Anthropocene - Forum. In: Journal of Contemporary Archaeology 1: 73-132, https://www.equinoxpub.com/journals/index.php/JCA/article/view/24264/pdf

Edgeworth, Matt (2014a): The Relationship Between Archaeological Stratigraphy and Artificial Ground and its Significance in the Anthropocene, pp. 91-108, in: Waters *et al*. (eds.): A Stratigraphical Basis for the Anthropocene. GSL Special Publications 395. Bath, UK: The Geological Society Publishing House, http://sp.lyellcollection.org/content/395/1.toc

Edgeworth, Matt (2014b): From Spade-Work to Screen-Work: New Forms of Archaeological Discovery in Digital Space, pp. 40-58, in: Carusi A., Hoel A.S., Webmoor T. & Woolgar S. (eds): Visualization in the Age of Computerization. Routledge; manuscript, https://www.academia.edu/6566988/From_Spadework_to_Screenwork_new_forms_of_archaeological_discovery_in_digital_space

EDISON - Education for Data Intensive Science to Open New science frontiers (H2020 project, 08/2015-08/2017), http://edison-project.eu

Edwards P., Pignotti E., Mellish C. *et al*. (2014): Lessons Learnt from the Deployment of a Semantic Virtual Research Environment. In: Web Semantics Journal, Vol. 27, http://www.websemanticsjournal.org/index.php/ps/article/view/377/400

Edwards P.N., Jackson S.J., Bowker G.C. & Knobel C.P. (2007) : Understanding infrastructure: Dynamics, tensions, and design. Workshop report, NSF Office of Cyberinfrastructure, http://cohesion.rice.edu/Conferences/Hewlett/emplibrary/UI_Final_Report.pdf

Edwards P.N., Mayernik M., Batcheller A.L., Bowker G.C. & Borgman C.L. (2011): Science friction: data, metadata, and collaboration. In: Social Studies of Science, 41(5): 667-690, http://pne.people.si.umich.edu/PDF/EdwardsEtAl2011ScienceFriction.pdf

Edwards, Rachel (2013): Archaeological archives and museums. Society of Museum Archaeologists, http://www.socmusarch.org.uk/docs/Archaeological-archives-and-museums-2012.pdf

EEF - European E-infrastructure Forum (2010): ESFRI project requirements for Pan-European e-infrastructure resources and facilities, 28 April 2010, http://www.osiris-online.eu/RI%20Reports/eef-report.pdf

EFFLA - European Forum of Forward Looking Activities (2014): Science 2.0: the deep unbundling. Final Report, 1 March 2014, http://europa.eu/espas/orbis/document/science-20-deep-unbundling

EGI Foundation (EGI.eu; coordinates a pan-European platform of distributed computing infrastructures), http://www.egi.eu

EGI.eu - European Grid Infrastructure (2014): Open Science Commons, 10 November 2014, http://www.egi.eu/news-and-media/publications/OpenScienceCommons_v2.pdf

EGI-Engage project (EU, H2020, 03/2015-08/2017; promotes open science commons and supports research communities with technical and data solutions), https://www.egi.eu/about/egi-engage/

EGI-InSPIRE (2011): Integration of clouds and virtualization into the European production infrastructure. Project deliverable 2.6, March 2011, https://documents.egi.eu/document/258/

EHRI - European Holocaust Research Infrastructure (EU, FP7, 10/2010- 03/2015; Horizon2020, 05/2015-04/2019; develops a portal and tools for the Holocaust research community), http://www.ehri-project.eu

Einav L. & Levin J. (2013): The Data Revolution and Economic Analysis. In: Innovation Policy and the Economy, 14(1): 1-24, http://web.stanford.edu/~jdlevin/Papers/BigData.pdf

e-InfraNet (2013): “Open” as the default modus operandi for research and higher education. Prepared by L. van der Vaart *et al*. e-InfraNet (ERA-NET) project, https://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2013/e-InfraNet-Open-as-the-Default-Modus-Operandi-for-Research-and-Higher-Education.pdf

e-IRG - e-Infrastructures Reflection Group (2005): e-Infrastructures Roadmap. Prepared by M. Leenaars, M. Heikkurinen, P. Louridas & F. Karayannis. Published by e-IRGSP, The Hague, December 2005, http://isoc.nl/michiel/eirg-roadmap.pdf

e-IRG - e-Infrastructures Reflection Group (2012): Cloud Computing for research and science: a holistic overview, policy, and recommendations. e-IRG Taskforce, http://www.e-irg.eu/images/stories/dissemination/e-irg_cloud_computing_paper_v.final.pdf

e-IRG - e-Infrastructures Reflection Group (2013): White Paper 2013. The Hague, The Netherlands, http://e-irg.eu/white-papers

e-IRG - e-Infrastructures Reflection Group (2015): Best Practices for the Use of E-Infrastructures by Large-scale Research Infrastructures, February 2015, http://e-irg.eu/guidelines

EIROforum (international research centres), http://www.eiroforum.org

Eiteljorg, Harrison II (2004): Computing for Archeologists, pp.20-30, in: Schreibman R. *et al.* (eds.): A Companion to Digital Humanities. Oxford: Blackwell (a scan of the volume is available at: http://chw.zinbun.kyoto-u.ac.jp/~chris/kg/mat/digital-humanities-big.pdf)

eIUS (2009a), University of Oxford & NCeSS: eIUS – e-Infrastructure Use Cases and Service Usage Models, website, http://projects.oucs.ox.ac.uk/eius/

eIUS (2009b): Archaeology Experience Report, July 2009, http://projects.oucs.ox.ac.uk/eius/outputs/Archaeology_ExpRep-pub.pdf eIUS (2009c): Archaeology : eIUS e-Infrastructure Case Study, video, http://www.youtube.com/watch?v=LxZcI0ikKV0

eIUS (2009d): Use Case 11 – Archaeology, http://projects.oucs.ox.ac.uk/eius/outputs/Archaeology_FinalUseCase-1.pdf

Elizabeth Alton (2014): What Trends Are Shaping The Future of Museums?, 5 June 2014, http://entertainmentdesigner.com/news/museum-design-news/what-trends-are-shaping-the-future-of-museums/

Elliott T., Heath S. & Muccigrosso J. (2012): Report on the Linked Ancient World Data Institute. In: ISQ - Information Standards Quarterly, 24(2/3): 43-45, http://www.niso.org/apps/group_public/download.php/9404/CR_Elliott-etal_LAWDI_isqv24no2-3.pdf

Elliott T., Heath S. & Muccigrosso J. (eds., 2014): Current Practice in Linked Open Data for the Ancient World. Institute for the Study of the Ancient World, New York University. ISAW Papers 7, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Ellis S. & Wallrodt J. (2011): iPads at Pompeii, http://classics.uc.edu/pompeii/index.php/news/1-latest/142-ipads2010.html

Ellrott K., Zmasek C.M., Weekes D. *et al*. (2010): TOPSAN: a dynamic web database for structural genomics. In: Nucleic Acids Research, Vol. 39, Database Issue, D494–D496, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3013775/

Elsevier (2013): Elsevier Announces its Scopus Book Titles Expansion Program. Amsterdam, 8 October 2013, http://www.elsevier.com/about/press-releases/science-and-technology/elsevier-announces-its-scopus-book-titles-expansion-program

Elsevier (2014): Scopus content: Book Expansion project update, 17 June 2014, http://blog.scopus.com/posts/scopus-content-book-expansion-project-update

Elsevier (2015): Scopus Content update: 75,000 book titles and counting, 30 March 2015, http://blog.scopus.com/posts/scopus-content-update-75000-book-titles-and-counting

Elsevier: Article of the Future project website, http://www.articleofthefuture.com

Elsevier: Data in Brief FAQ, http://www.journals.elsevier.com/data-in-brief/policies-and-guidelines/data-in-brief-faq/

Elsevier: Executable Papers Challenge website, http://www.executablepapers.com

eLTER - European Long-Term Ecosystem and Socio-Ecological Research Infrastructure (EU, H2020, 06/2015-05/2019), http://www.lter-europe.net/projects/eLTER

EMTO (Early Modern Thought Online), http://emto-nanopub.referata.com/wiki/EMTO_Nanopub

Encyclopedia of Egyptology (UCLA), http://uee.ucla.edu; see also: https://escholarship.org/uc/nelc_uee

Endangered Archives Programme, http://eap.bl.uk, programme weblog: http://britishlibrary.typepad.co.uk/endangeredarchives/

Engage - Engaging research with e-Infrastructure (UK, JISC-funded project), http://www.engage.ac.uk

Engels T.C.E., Ossenblok T.L.B. & Spruyt E.H.J. (2012): Changing Publication Patterns in the Social Sciences and Humanities, 2000-2009'. In: Scientometrics, 93/2: 373–90, https://www.ecoom.be/sites/ecoom.be/files/downloads/Engels%20et%20al%20changing%20pub%20patterns%20SSH%20Scientometrics%202012.pdf

Enriquez V., Walker-Judson S., Walker N.M. *et al*. (2010): Data citation in the wild. Presented at International Digital Curation Conference 2010, Chicago/USA, Nature Precedings, http://precedings.nature.com/documents/5452/version/1

ENUMERATE (2014): Survey Report on Digitisation in European Cultural Heritage Institutions 2014. Prepared by Natasha Stroeker and René Vogels, Panteia (NL), January 2014, http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2014.pdf

Epigraphic Database Heidelberg, http://www.epigraphische-datenbank-heidelberg.de

EPIRIA (2014): Evaluation of pertinence and impact of the EU research infrastructure activity in FP7. A study for the European Commission by the Technopolis Group and Empirica, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=publications

EPOCH - Excellence in Processing Open Cultural Heritage (EU, FP6-IST, Network of Excellence, 2004-2008), http://www.epoch-net.org

eResearch2020 (2010): The Role of e-Infrastructures in the Creation of Global Virtual Research. Study for the European Commission, DG Information Society and Media by Empirica GmbH *et al*., http://www.eresearch2020.eu

eResearch2020 (2010a): The Role of e-Infrastructures in the Creation of Global Virtual Research Communities. Empirica *et al*. Final Report, Bonn and Brussels, February 2010, http://www.eresearch2020.eu/eResearch2020%20Final%20Report.pdf

eResearch2020 (2010b): eResearch 2020 - The Role of e-Infrastructures in the Creation of Global Virtual Research Communities. Study brochure, Empirica, Bonn, http://www.eresearch2020.eu/eResearch%20Brochure%20EN.pdf

ERF-AISBL - Association of European-level Research Infrastructure Facilities, http://www.erf-aisbl.eu

ERF-AISBL (2014): Executive Summary of the Workshop on Big Data and Open Data, Royal Museums of Art and History, Brussels, 7-8 May 2014, http://www.elettra.eu/Conferences/2014/BDOD/

Eric T. Meyer E.T. & Schroeder R. (2009): Untangling the web of e-Research: Towards a sociology of online knowledge. In: Journal of Informetrics, 3(3): 246-260 (authors' manuscript), https://www.researchgate.net/publication/222432865_Untangling_the_web_of_e-Research_Towards_a_sociology_of_online_knowledge

Eschenfelder K. & Johnson A. (2011): The Limits of sharing: Controlled data collections. In: Proceedings of the American Society for Information Science and Technology, 48(1): 1–10, http://www.asis.org/asist2011/proceedings/submissions/62_FINAL_SUBMISSION.pdf

e-SciDR (2008): Towards a European e-Infrastructure for e-Science Digital Repositories. A report for the European Commission; prepared by The Digital Archiving Consultancy Ltd, UK, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/e-scidr.pdf

ESF - European Science Foundation & EUROHORCS (2011): Basic Requirements for Research Infrastructures in Europe. ESF Member Organisation Forum on Research Infrastructures. March 2011 (endorsed by the EUROHORCs on 14 April 2011), http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/CEO_Unit/MO_FORA/MOFORUM_ResearchInfra/General/Basic_requirements_V1_new_ESF_logo.pdf&t=1413646980&hash=2c2ccdfe1c00ec5440f9c63dd93946131e57a982

ESF - European Science Foundation (2010): The Future of Knowledge. Mapping Interfaces. Interdisciplinary Workshop Report. ESF: October 2010, http://www.esf.org/fileadmin/Public_documents/Publications/Future_of_Knowledge.pdf

ESF - European Science Foundation (2011): Research Infrastructures in the Digital Humanities. Science Policy Briefing 42. Strasbourg: ESF, http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf

ESF - European Science Foundation (2013): Research Infrastructures in the European Research Area. A report by the ESF Member Organisation Forum on Research Infrastructures. Strasbourg, March 2013, http://www.esf.org/fileadmin/Public_documents/Publications/mof_research_infrastructures.pdf

ESFRI - European Strategy Forum on Research Infrastructures (2006): European Roadmap for Research Infrastructures. Report 2006. Luxembourg: Publications Office of the European Union, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap

ESFRI - European Strategy Forum on Research Infrastructures (2011): Strategy Report on Research Infrastructures. Roadmap 2010. Luxembourg: Publications Office of the European Union, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap

ESFRI - European Strategy Forum on Research Infrastructures, http://ec.europa.eu/research/esfri/

EU-ARTECH - Access Research and Technology for the Conservation of the European Cultural Heritage (EU, FP6, Integrated Infrastructures Initiative, 06/2004-05/2009), http://www.eu-artech.org

EUDAT - European Data Infrastructure (EU, FP7-Infrastructures, 10/2011-09/2014, developed several data services; continued under H2020, 03/2015-02/2018), http://www.eudat.eu

e-Uptake - Enabling Uptake of e-Infrastructure Services project, http://www.engage.ac.uk/e-uptake

EUROCORR - European Correspondence to Jacob Burckhardt (06/2010-05/2015, funded by the European Research Council), http://burckhardtsource.org

EuroCRIS - Current Research Information Systems: The European Organisation for International Research Information, http://www.eurocris.org

EuroCRIS: CERIF - Common European Research Information Format (European standard for Current Research Information), http://eurocris.org/cerif/main-features-cerif

European Archaeological Council (2015): Amersfoort Agenda – Setting the agenda for the future of archaeological heritage management in Europe, pp. 15-24, in: Setting the Agenda – Giving New Meaning to the European Archaeological Heritage. Edited by P.A.C. Schut, D. Scharff & L.C. de Wit. EAC Occasional Paper no. 10, Budapest: Archaeolingua, http://european-archaeological-council.org/files/amersfoort_agenda_english.pdf

European Archaeological Council (EAC), http://european-archaeological-council.org

European Association of Archaeologists (EAA), http://e-a-a.org

European Association of National Research Facilities, http://www.europeanresearchfacilities.eu

European Association of Science Editors (EASE): Impact factor statement, http://www.ease.org.uk/publications/impact-factor-statement

European Commission & European Science Foundation (2007): Trends in European Research Infrastructures. Analysis of data from the 2006/07 survey. July 2007, http://ec.europa.eu/research/infrastructures/pdf/survey-report-july-2007_en.pdf

European Commission (2012): Online survey on scientific information in the digital age, Brussels, 2012, http://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-scientific-information-digital-age_en.pdf

European Commission (2013-12a): Guidelines on data management in Horizon 2020 research projects. Version 1.0, 11 December 2013: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission (2013-12b): Commission launches pilot to open up publicly funded research data, Brussels, 16 December 2013, http://europa.eu/rapid/press-release_IP-13-1257_en.htm

European Commission (2014): Cultural heritage. Digitisation, online accessibility and digital preservation. Report on the Implementation of Commission Recommendation 2011/711/EU, 2011-2013, http://ec.europa.eu/information_society/newsroom/image/recommendation-2011-2013_progress%20report-final-clean-shared%20with%20eac-ga%20approved-22-09-2014-final_6953.pdf

European Commission, Communication (2011): Open data. An engine for innovation, growth and transparent governance (12.12.2011), http://ec.europa.eu/information_society/policy/psi/docs/pdfs/opendata2012/open_data_communication/en.pdf

European Commission, Communication (2012): Towards better access to scientific information: Boosting the benefits of public investments in research (17.7.2012), http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf

European Commission, Communication (2014): Towards an integrated approach to cultural heritage for Europe. COM/2014/0477 final, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2014:477:FIN

European Commission, Culture (2014): Mapping of Cultural Heritage actions in European Union policies, programmes and activities, http://ec.europa.eu/culture/library/reports/2014-heritage-mapping_en.pdf

European Commission, DG Connect (2013a): Digital science in Horizon 2020. March 2013, http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=2124

European Commission, DG Connect (2013b): Research Data e-Infrastructures: Framework for Action in H2020, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/framework-for-action-in-h2020_en.pdf

European Commission, DG Connect (2014): Public consultation on 'Science 2.0': Science in Transition, http://ec.europa.eu/research/consultations/science-2.0/consultation_en.htm, background document, http://ec.europa.eu/research/consultations/science-2.0/background.pdf)

European Commission, DG Enterprise and Industry (2014): Innovation Union Scoreboard 2014. Brussels, http://ec.europa.eu/enterprise/policies/innovation/files/ius/ius-2014_en.pdf

European Commission, DG Research & Innovation (2015a): Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 2.0, 30 October 2015, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

European Commission, DG Research & Innovation (2015b): Guidelines on Data Management in Horizon 2020. Version 2.0, 30 October 2015, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission, Directorate General: Communications Networks, Content and Technology, Units: Digital Science and eInfrastructure, http://ec.europa.eu/dgs/connect/en/who-we-are

European Commission, eInfrastructures unit (2013): Research Data e-Infrastructures: Framework for Action in H2020, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/framework-for-action-in-h2020_en.pdf

European Commission: European Environmental Impact Assessment Directive, http://ec.europa.eu/environment/eia/eia-legalcontext.htm

European Correspondence to Jacob Burckhardt (EUROCORR, 06/2010-05/2015, funded by the European Research Council), http://burckhardtsource.org

European E-Infrastructure Forum (2010): ESFRI project requirements for Pan-European e-infrastructure resources and facilities, Geneva, 28. April 2010, http://www.osiris-online.eu/RI%20Reports/eef-report.pdf

European E-Infrastructure Forum (EEF), http://www.einfrastructure-forum.eu

European Expert Group Report (2010): Cost Control and Management Issues of Global Research Infrastructures. October 2010, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=publications

European Marine Board (2014): Land beneath the Waves: Submerged Landscapes and Sea Level Change. A joint geoscience-humanities strategy for European Continental Shelf Prehistoric Research. Prepared by N.C. Flemming, M.N Çağatay, F.L Chiocci *et al*. Position Paper 21 of the European Marine Board, Ostend, Belgium, October 2014, http://www.splashcos.org/sites/splashcos.org/files/Land%20Beneath%20the%20Waves-244%20%281%29_0.pdf

European Science Foundation (2011): Research Infrastructures in the Digital Humanities. Science Policy Briefing 42. Strasbourg, http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf

European Union - Europe 2020, http://ec.europa.eu/europe2020/

European Union Open Data Portal, http://open-data.europa.eu

Europeana (2014): Europeana for Research. Final Policy Recommendations. Athens, June 2014, http://pro.europeana.eu/documents/858566/2691fd56-c0c0-4b4f-98dd-fa1eecd4a8dc

EUROPEANA (gateway to digital content of archives, libraries and museum across Europe), http://www.europeana.eu

Europeana Cloud (2015): Content priorities for Humanities and Social Sciences research. Project deliverable 1.4, January 2015, http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.4%20Content%20priorities%20for%20Humanities%20and%20Social%20Sciences%20research%20communities.pdf

Europeana Cloud (eCloud) project (EU, ICT-PSP, 02/2013-01/2016), http://pro.europeana.eu/web/europeana-cloud

Europeana Foundation (2011): Strategic Plan, 2011-2015. National Library of the Netherlands, January 2011, http://www.pro.europeana.eu/c/document_library/get_file?uuid=c4f19464-7504-44db-ac1e-3ddb78c922d7&groupId=10602

Europeana Research - Liberating Cultural Heritage for Use in Research, http://research.europeana.eu

EuroRis-Net Observatory of Research Infrastructures' Network, http://observatory.euroris-net.eu

EUTEMA (2014): NGC-R - Next Generation Computing Roadmap. A study prepared for the European Commission, DG Communications Networks, Content & Technology. eutema GmbH (Austria) in co-operation with Optimat, EPCC and 451 Research. Final Report, http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=6636

Executive Office of the President. Office of Science and Technology Policy, Washington, D.C, 20502: Memorandum for the Heads of Executive Departments and Agencies, Subject: Increasing Access to the Results of Federally Funded Scientific Research, 22 February 2013, http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Eysenbach, Gunther (2011): Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. In: Journal of Medical Internet Research, 13(4), http://www.jmir.org/2011/4/e123/

Fagan, Brian (1995): Archaeology's dirty secret. In: Archaeology, 48(4): 14–17

FAIMS - Federated Archaeological Information Management System (Australia), https://www.fedarch.org

FAIMS - Federated Archaeological Information Management System: Mobile Platform, https://www.fedarch.org/wordpress/faims-mobile-platform/

Fairclough, Graham (2014): 'I wanted to be an archaeologist'. A very 21th century crisis of confidence. Presentation at the EAC Annual Meeting 2014 / 15th EAC Heritage Management Symposium: Setting the Agenda: Giving new meaning to the European archaeological heritage, Amersfoort, Netherlands 20-22 March 2014; version 28/06/2014, http://www.european-archaeological-council.org/files/eac_2014_graham_fairclough.pdf

Fairclough, Graham (2015): 'I wanted to be an archaeologist' – A 21st-century crisis of confidence?, pp. 37-44 in: European Archaeological Council (EAC): Setting the Agenda – Giving New Meaning to the European Archaeological Heritage. Edited by P.A.C. Schut, D. Scharff & L.C. de Wit. EAC Occasional Paper no. 10, Budapest: Archaeolingua

Faniel I., Kansa E., Whitcher-Kansa S., Barrera-Gomez J. & Yakel E. (2013): The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, 295-304 (preprint), http://www.oclc.org/content/dam/research/publications/library/2013/faniel-archae-data.pdf

Faniel I.M. & Jacobsen T.E. (2010a): Reusing scientific data: A research framework. CSCW 2010 Workshop on "The Changing Dynamics of Scientific Collaboration", CSCW 2010, Savannah, GA, 6-10 February 2010, https://www.sci.utah.edu/images/Workshops/cscw2010/faniel_2.pdf

Faniel I.M. & Jacobsen T.E. (2010b): Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. In: Computer Supported Cooperative Work (CSCW), 19(3-4), 355–375, http://dx.doi.org/10.1007/s10606-010-9117-8

Faniel I.M. & Zimmerman A. (2010): Beyond the Data Deluge: A Research Agenda for Large-Scale Data Sharing and Reuse. 6th International Digital Curation Conference, Chicago, December 6-8, 2010, http://scientificdatasharing.com/wp-content/uploads/2011/01/Faniel_Zimmerman.pdf

Farnel S. & Shiri A. (2014): Metadata for Research Data: Current Practices and Trends, pp. 37-42, in: DC-2014 – International Conference on Dublin Core and Metadata Appliations. Papers, Project Reports and Posters for DC-2014 in Austin, Texas, 8-11 October 2014, http://dcpapers.dublincore.org/pubs/article/viewFile/3714/1937

Fasti Online (International Association of Classical Archaeology), http://www.fastionline.org

Fausto S., Machado F.A., Bento L.F. J. *et al*. (2012): Research Blogging: Indexing and Registering the Change in Science 2.0. In: PLOS ONE, 10(4): e0124184, http://dx.doi.org/10.1371/journal.pone.0050109

Fecher B. & Friesike S. (2014): Open Science: One Term, Five Schools of Thought. In: Bartling S. & Friesike S. (eds.): Opening Science, http://book.openingscience.org/basics_background/open_science_one_term_five_schools_of_thought.html

Fecher B., Friesike S. & Hebing M. (2015a): What Drives Academic Data Sharing? In: PLOS ONE, 10(2), e0118053, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118053

Fecher B., Friesike S., Hebing M., Linek S. & Sauermann A. (2015b): A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing. DIW Berlin, Discussion Paper, No. 1454, http://www.hiig.de/wp-content/uploads/2015/02/dp1454.pdf

Fecher, Benedikt (2014): Data sharing angst – An insight to an ongoing research on data sharing in academia. Alexander von Humboldt Institute for Internet and Society, weblog, 27 January 2014, http://www.hiig.de/en/data-sharing-angst-an-insight-to-an-ongoing-research-on-data-sharing-in-academia/

FECYT - Spanish Foundation for Science and Technology (2012): The preservation and reuse of scientific data in Spain. Report of the RECOLECTA Working Group on The depositing and management of data in Open Access. December 2012. Madrid: FECYT http://recolecta.fecyt.es/sites/default/files/contenido/documentos/Scientific_Data_in_Spain_Report_EN.pdf

Fedrigo A., Broström A., Brink K. *et al*. (2015): Neutron and X-ray imaging applied to archaeological barley seeds from Southern Scandinavia, a comparative study. Presented at TECHNART 2015 - Catania, Italy, 27-30 April 2015 (expected to be published in the Microchemical Journal).

Ferwerda, Eelco (2014): Open access monograph business models. In: UKSG Insights, supplement 1, April 2014, http://insights.uksg.org/articles/10.1629/2048-7754.46/

Field expedition: Mongolia (invites non-experts to identify potential archaeological sites in Mongolia on satellite images), http://exploration.nationalgeographic.com/mongolia

Fishare (online repository), http://figshare.com/about

Fisher C.T., Hill J.B., Feinman G.M. (eds., 2009): The Archaeology of Environmental Change: Socionatural Legacies of Degradation and Resilience. University of Arizona Press

Flatman, Joe (2010): The Economics of Public Archaeology: A Reply to "What is Public Archaeology?". In: Present Pasts, 1(1), http://www.presentpasts.info/articles/10.5334/pp.9/

FORCE11 - Future of Research Communication and e-Scholarship, https://www.force11.org

FORCE11 (2014): Joint Declaration of Data Citation Principles, https://www.force11.org/datacitation/

FORCE11 (n.d.): Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0 (FAIR princples), https://www.force11.org/node/6062

Ford, Emily (2013): Defining and characterizing open peer review: a review of the literature. In: Journal of Scholarly Publishing, 44(4): 311-326, http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1000&context=ulib_fac

Fossilworks (provides search and analysis tools for the Paleobiology Database), http://fossilworks.org

FOSTER - Facilitate Open Science Training for European Research (EU, FP7 project, 02/2014-01/2016), http://www.fosteropenscience.eu

Foulonneau M. & André F. (2008): Investigative Study of Standards for Digital Repositories and Related Services. Amsterdam University Press, http://dare.uva.nl/document/93727

Foulonneau, Muriel (2012): Cultural Heritage Collections: From Content Curation to Semantic Services and the Semantic Web. In: Uncommon Culture, Vol. 3, no. 5/6, pp. 53-60, http://uncommonculture.org/ojs/index.php/UC/article/view/4716/3679

Frank R.D., Yakel E. & Faniel I.M. (2015): Destruction/Reconstruction: Preservation of Archaeological & Zoological Research Data. In: Archival Science, January 2015, http://deepblue.lib.umich.edu/handle/2027.42/111298

Franzoni C. & Sauermann H. (2014): Crowd science: The organization of scientific research in open collaborative projects. In: Research Policy, 43(1): 1-20, http://www.sciencedirect.com/science/article/pii/S0048733313001212

Free Your Metadata project (iMinds / Ghent University and MaSTIC / Université Libre de Bruxelles) http://freeyourmetadata.org

Freeman, Mark A. (2015): „Not for Casual Readers:" An Evaluation of Digital Data from Virginia Archaeological Websites. Master's Thesis, University of Tennessee, http://trace.tennessee.edu/cgi/viewcontent.cgi?article=4785&context=utk_gradthes

Frenken K., Hoekman J. & Hardeman S. (2010): The globalization of research collaboration, pp. 144-148, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Friedlander, Amy *et al*. (2009): Working together or apart: Promoting the next generation of digital scholarship. Report of a workshop co-sponsored by the Council on Library and Information Resources and the National Endowment for the Humanities. Washington, DC: CLIR, http://www.clir.org/pubs/reports/pub145/pub145.pdf

Frodeman R., Thompson-Klein R. & Mitcham C. (eds., 2010): The Oxford Handbook of Interdisciplinarity. Oxford UP

Fry J., Lockyer S., Oppenheim C., Houghton J. & Rasmussen B. (2008): Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes. JISC project report, November 2008, http://ie-repository.jisc.ac.uk/279/

Fulford M., O'Riordan E.J., Clarke A. & Rains M. (2010): Silchester Roman Town: developing virtual research practice 1997-2008, pp. 15-34, in: Bodard G. & Mahony S. (eds.): Digital Research in the Study of Classical Antiquity. Farnham: Ashgate

Funtowicz S. & Ravetz J. (2003): Post-Normal Science. In: International Society for Ecological Economics: Internet Encyclopaedia of Ecological Economics, February 2003, http://leopold.asu.edu/sustainability/sites/default/files/Norton,%20Post%20Normal%20Science,%20Funtowicz_1.pdf

G8 Open Data Charter (2013) – Government of the United Kingdom, The Cabinet Office (2013): Policy paper: G8 Open Data Charter and Technical Annex, 18 June 2013, https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

Gaimster D. & Majewski T. (eds., 2009): International handbook of historical archaeology. New York: Springer, http://mayrasandy.files.wordpress.com/2012/08/36192861-hist-or-archaeology.pdf

Galle, Jillian (2012): Will today's graduate training in Historical Archaeology predict the future of digital research archives? In: Society for Historical Archaeology weblog, 29 May 2012, http://education8314.rssing.com/browser.php?indx=45442997&item=11

Gargouri Y., Hajjem C., Larivière V. *et al*. (2010): Self-selected or mandated, open access increases citation impact for higher quality research. In: PLOS One, 5(10), e13636, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013636

Garoufallou E. & Papatheodorou, C. (2014): A critical introduction to metadata for e-science and e-research. In: International Journal of Metadata, Semantics and Ontologies, 9(1): 1-4, http://www.inderscience.com/storage/f118410395762121.pdf

Garth B. & Reilly P. (2014): Additive Archaeology: Towards a virtual archaeology reprinted? Paper presented at CAA Greece, Rythymnon, Crete, 8 March 2014, http://eprints.soton.ac.uk/364455/1/Additive%20Archaeology%20CAA14%20GR%20P%20Reilly.pdf

Gartner (2015): Gartner's 2015 Hype Cycle for Emerging Technologies, press release, 18 August 2015 http://www.gartner.com/newsroom/id/3114217

Gattiglia, Gabriele (2014): Think Big about Data: Archaeology and the Big Data Challenge. Presentation at EAA 2014 conference, Istanbul, 13 September 2014, http://www.ariadne-infrastructure.eu/content/download/3645/21009/version/1/file/EAA2014_OpenAccess_session_13_Gabriele.Gattiglia_MAPPA_13092014.pdf

Gattiglia, Gabriele (2015): Think big about data: Archaeology and the Big Data challenge. In: Archäologische Informationen, Band 38: Fokus: Open Access & Open Data, early view: http://www.dguf.de/fileadmin/AI/ArchInf-EV_Gattiglia.pdf

Gaugne R., Gouranton V., Dumont G., Chauaut A. & Arnaldi B. (2014): Immersia, an open immersive infrastructure: doing archaeology in virtual reality. In: Archeologia e Calcolatori, suppl. 5: 1-10, https://hal.archives-ouvertes.fr/hal-01003383/file/Archeologia_Immersia.pdf

GBIF - Global Biodiversity Information Facility (2013): Global Biodiversity Informatics Outlook: Delivering Biodiversity Knowledge in the Information Age. Copenhagen: GBIF Secretariat 2013, http://www.gbif.org/orc/?doc_id=5353

GBIF - Global Biodiversity Information Facility, http://www.gbif.org

GCMD – Global Change Master Directory (NASA; provides access to more than 34,000 Earth & Environmental dataset and service descriptions; including datasets relevant to archaeology), http://gcmd.nasa.gov

Geertz, Clifford (1973): Thick description: toward an interpretive theory of culture, in: The interpretation of cultures: selected essays. New-York: Basic Books, 1973, pp. 3-30, available online at: http://hypergeertz.jku.at/GeertzTexts/Thick_Description.htm

GenBank (includes about 10,000 records of ancient DNA of humans, animals, plants, bacteria, etc.), http://www.ncbi.nlm.nih.gov/genbank/

Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo/

Geonames, http://www.geonames.org

Geser G. & Niccolucci F. (2012): Virtual museums, digital reference collections and e-science environments. In: Uncommon Culture, Vol. 3, no. 5/6, pp. 12-37, http://uncommonculture.org/ojs/index.php/UC/article/view/4714/3677

Geser, Guntram (2014a): From User Needs to the Innovation Plan, pp. 9-24, in: ARIADNE – The Way Forward to Digital Archaeology in Europe. ARIADNE publication, November 2014, http://www.ariadne-infrastructure.eu/content/download/4569/26666/version/2/file/Ariadne+Booklet.pdf

Geser, Guntram (2014b): Open Access and Open Data in Archaeology – Report on the ARIADNE Session at EAA 2014. Available at: http://www.ariadne-infrastructure.eu/Media/Files/EAA-2014-Open-Access-Session-Report

GESIS (Leibniz Institute for the Social Sciences): Datorium, https://datorium.gesis.org

Gestrich, Nikolas (2011): Putting a Price on the Past: The Ethics and Economics of Archaeology in the Marketplace - A Reply to "What is Public Archaeology". In: Present Pasts, 3(2), http://www.presentpasts.info/articles/10.5334/pp.8/

Getty Research Institute: Getty Vocabularies as Linked Open Data, http://www.getty.edu/research/tools/vocabularies/lod/

Gezelter, Dan (2011): Advice to junior faculty who want to do get promoted doing Open Science. In: Open Science weblog, 23 September 2011, http://www.openscience.org/blog/?p=463

Gibney E. & Van Noorden R. (2013): Scientists losing data at a rapid rate. In: Nature, 19.12.2013, http://www.nature.com/news/scientists-losing-data-at-a-rapid-rate-1.14416

Giguet-Covex C., Pansu J., Arnaud F. *et al*. (2014): Long livestock farming history and human landscape shaping revealed by lake sediment DNA. In: Nature Communications 5: 3211, http://dx.doi.org/10.1038/ncomms4211 (research data: http://dx.doi.org/10.5061/dryad.h11h7)

Gilissen, Valentijn (2014): Dutch archaeological data depositing, processing, archiving and accessing at DANS: A repository with ten years of history, setting its sails to the future. Presentation at Deutscher Archäologiekongress 2014, Berlin, http://www.ianus-fdz.de/attachments/download/756/01_Gilissen_DANS.pdf

Gimenez-Toledo E., Tejada-Artigas C & Manana-Rodriguez J. (2013): Evaluation of Scientific Books Publishers in Social Sciences and Humanities: Results of a Survey. In: Research Evaluation, 22/1: 64-77, http://rev.oxfordjournals.org/content/22/1/64.abstract

Gingras Y. & Mosbah-Natanson S. (2010): Where are social sciences produced?, pp. 149-153, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Gold, Anna (2007a): Cyberinfrastructure, data and libraries. Part 1: A cyberinfrastructure primer for librarians. In: D-Lib Magazine, 13(9/10), http://www.dlib.org/dlib/september07/gold/09gold-pt1.html

Gold, Anna (2007b): Cyberinfrastructure, data and libraries. Part 2: Libraries and the data challenge: roles and actions for libraries. In: D-Lib Magazine, 13(9/10), http://www.dlib.org/dlib/september07/gold/09gold-pt2.html

Golden P. & Shaw R. (2015): Period assertion as nanopublication. The PeriodO period gazetteer. In: Semantics, Analytics, Visualisation: Enhancing Scholarly Data. Workshop Co-Located with WWW'15 -24th International World Wide Web Conference, Florence, Italy. http://cs.unibo.it/save-sd/2015/papers/html/golden-savesd2015.html

Gonçalves-Curty R. & Qin J. (2014): Towards a model for research data reuse behaviour. In: Proceedings of ASIST 2014, Seattle, 1-4 November 2014, https://www.asis.org/asist2014/proceedings/submissions/posters/194poster.pdf

Gonzalez-Perez C. & Hug C. (2013): Crafting Archaeological Methodologies: Suggesting Situational Method Engineering for the Humanities and Social Sciences, pp. 807-820, in: Archaeology in the Digital Era - Volume II, CAA 2012, Southampton, 26-30 March 2012, Amsterdam Univ.Press, https://www.academia.edu/7783958/Archaeology_in_the_Digital_Era._Papers_from_the_40th_Annual_Conference_of_Computer_Applications_and_Quantitative_Methods_in_Archaeology_CAA_._Printed_Book_and_E-Proceedings

Gonzalez-Perez C. & Martín-Rodilla P. (2014): Integration of Archaeological Datasets Through the Gradual Refinement of Models, 193-204, in: Proceedings of CAA 2014, Paris, 22-25 April 2014, http://www.archaeopress.com/ArchaeopressShop/Public/download.asp?id={5CACE285-4C48-41AE-809E-E98B65C9E4CD}

Goodale N., Bailey D., Fondak T. & Nauman A. (2013): iTrowel: Mobel Devices as Transformative Technology in Archaeological Field Research. In: SAA Archaeological Record, 13.3: 18-22, http://www.saa.org/portals/0/saa/publications/thesaaarchrec/may2013.pdf

Goodman A., Pepe A., Blocker A.W. *et al*. (2014): Ten Simple Rules for the Care and Feeding of Scientific Data. In: PLOS Computational Biology, 10(4): e1003542, http://dx.doi.org/10.1371/journal.pcbi.1003542

Görögh E., Kędzierska E., Kavalchuk N., Stepniak J., Dzivak J., Pejšová P. & Vyčítalová H. (2014): Enhanced publications in V4 countries, in: Conference on Grey Literature and Repositories: proceedings 2014: the Value of Grey Literature in Repositories. Prague: National Library of Technology, http://repozitar.techlib.cz/record/807/files/idr-807_4.pdf

Gorraiz J., Gumpenberger C. & Schlögl C. (2012): Differences and similarities in usage versus citation behaviours observed for five subject areas. Elsevier Bibliometric Research Program, http://ebrp.elsevier.com/pdf/2012_Proposal3b_Gorraiz.pdf

Gorraiz J., Gumpenberger C. & Schlögl C. (2014): Usage versus citation behaviours in four subject areas. In: Scientometrics, 101-2: 1077-1095, http://dx.doi.org/10.1007/s11192-014-1271-1

Gorraiz J., Purnell P.J. & Glänzel W. (2013): Opportunities for and limitations of the Book Citation Index. In: ASIS&T Journal, 64(7): 1388-98, http://dx.doi.org/10.1002/asi.22875

Gothic Past - Visual Archive of Gothic Architecture and Sculpture in Ireland (Trinity College, Ireland), http://www.gothicpast.com

Gow A. & Molloy L. (2014): Ahead of the CurV: Digital Curator Vocational Education. In: C. Mills, M. Pidd & E. Ward (eds.): Proceedings of the Digital Humanities Congress 2012. Studies in the

Digital Humanities. Sheffield: HRI Online Publications, http://www.hrionline.ac.uk/openbook/chapter/dhc2012-gow

Graham-Campbell J. & Valor M. (eds., 2007): The Archaeology of Medieval Europe. Volume 1. The Eighth to Twelfth Centuries AD. Aarhus University Press

Gray J. & Szalay A. (2004): Where the Rubber Meets the Sky: Bridging the Gap between Databases and Science. Technical Report, MSR-TR-2004-110, http://arxiv.org/ftp/cs/papers/0502/0502011.pdf

Gray J., Szalay A.S., Thakar A.R. & Stoughton C. (2002): Online scientific data curation, publication, and archiving, pp. 103-107, in: Proceedings of SPIE, Vol. 4846: Virtual Observatories, http://arxiv.org/ftp/cs/papers/0208/0208012.pdf

Gray, Jim (2007): Jim Gray on eScience. A Transformed Scientific Method, pp. xvii-xxxi, in: The Fourth Paradigm. Data-Intensive Scientific Discovery. Edited by T. Hey, S. Tansley and K. Tolle. Redmond, WA: Microsoft Research 2009, http://research.microsoft.com/en-us/collaboration/fourthparadigm/

GRDI2020 - Global Research Data Infrastructures (2010): Towards a 10-year vision for global research data infrastructures. Technological & Organisational Aspects of a Global Research Data Infrastructure: A view from the experts. GRDI2020 Consortium, http://www.grdi2020.eu/Repository/FileScaricati/9a85ca56-c548-47e4-8b0e-86c3534ad21d.pdf

GRDI2020 - Towards a 10-year Vision for Global Research Data Infrastructures project (EU, FP7, 01/2010-12/2012), http://www.grdi2020.eu

GRDI2020 (2012): Global Research Data Infrastructures: The Big Data Challenges. Final Roadmap Report, February 2012, http://www.grdi2020.eu/Repository/FileScaricati/e2b03611-e58f-4242-946a-5b21f17d2947.pdf

Green A.G. & Gutmann M.P. (2007): Building partnerships among social science researchers, institution-based repositories, and domain specific data archives. In: OCLC Systems and Services: International Digital Library Perspectives, 23: 35-53; manuscript, http://deepblue.lib.umich.edu/handle/2027.42/41214

Green D. & Roy M.D. (2008): Things to Do While Waiting for the Future to Happen: Building Cyberinfrastructure for the Liberal Arts. In: EDUCAUSE Review, 43(4), http://www.educause.edu/ero/article/things-do-while-waiting-future-happen-building-cyberinfrastructure-liberal-arts

Green, Toby (2010): We need publishing standards for datasets and data tables. OECD Publishing White Paper, revised version February 2010, http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2010/wp.8.e.pdf

Greenberg J., Swauger S. & Feinstein E. (2013): Metadata Capital in a Data Repository [DRYAD]. DC-2013 – Proc. International Conference on Dublin Core and Metadata Applications, Lisbon, 2-6 September 2013, http://dcevents.dublincore.org/IntConf/dc-2013/paper/viewFile/189/172

Greenberg, Jane (2014): Metadata Capital via a linked data HIVE. Advances in Classification Research Online, 24(1): 95-61, https://journals.lib.washington.edu/index.php/acro/article/viewFile/14680/12320

Gregory I.N. & Ell P.S. (2008): Historical GIS: Technologies, Methodologies, and Scholarship. Cambridge University Press (Cambridge Studies in Historical Geography)

GridTalk (2010): GridBriefing 13: New projects, new requirements: Future needs of the ESFRI projects, June 2010, http://www.gridtalk.org/Documents/GridBriefing-Future_needs_ESFRI_projects.pdf

Griffin R.E. / CODATA Task Group 'Data At Risk' (2015): When are Old Data New Data? GeoResJ vol. 6, June 2015, 92–97, http://www.sciencedirect.com/science/article/pii/S2214242815000121

Grima, Reuben (2010): A Reply to "What is Public Archaeology?" In: Present Pasts, 1(1), http://www.presentpasts.info/articles/10.5334/pp.10/

Groen W.J.M., Márquez-Grant N. & Janaway R. (2015): Forensic Archaeology: A Global Perspective. John Wiley & Sons

Grootveld M. & van Egmond J. (2012): Peer-Reviewed Open Research Data: Results of a Pilot. In: International Journal of Digital Curation, 7(2): 81-91, http://ijdc.net/index.php/ijdc/article/view/221/290

Grossmann, Alexander (2014): Where did our Peer Review Mojo go? In: ScienceOpen weblog, 29 September 2014, http://blog.scienceopen.com/2014/09/peer-review-mojo/

Guermandi M.P. & Salas Rossenbach K. (2013): Twenty years after Malta: preventive archaeology in Europe and in Italy. Istituto per i Beni Artistici, Culturali e Naturali della Regione Emilia Romagna, Bologna, Italy, http://online.ibc.regione.emilia-romagna.it/I/libri/pdf/twenty_years_after_malta_26_09_2013web.pdf

Guetzkow J., Lamont M. & Mallard G. (2004): What is Originality in the Humanities and the Social Sciences? In: American Sociological Review, 69/2: 190–212, http://www.gregoiremallard.com/site/Articles_files/OriginalityPaperFromProquest.pdf

Guibault L. & Wiebe A. (eds., 2013): Safe to be open. Study on the protection of research data and recommendations for access and usage. OpenAIRE / Universitätsverlag Göttingen, http://www.ivir.nl/publicaties/download/1129

Guibault, Lucie (2013): Licensing Research Data under Open Access Conditions. Chapter to be published in: D. Beldiman (ed.): Information and Knowledge: 21st Century Challenges in Intellectual Property and Knowledge Governance: Cheltenham: Elgar, http://www.ivir.nl/publicaties/download/759

Guldi J. & Armitage D. (2014): The History Manifesto. Cambridge UP, http://historymanifesto.cambridge.org/files/6114/1227/7857/historymanifesto.pdf

Gunn, William (2014): Gauging openness, measuring impact. In: Research Trends, Issue 37, http://www.researchtrends.com/issue-37-june-2014/gauging-openness-measuring-impact/

Guttmann-Bond, Erika (2010): Sustainability out of the past. How archaeology can save the planet. In: World Archaeology, 42(3), 355-366; available at https://www.academia.edu/1074449/Sustainability_out_of_the_past_how_archaeology_can_save_the_planet

Haak Marcial L. & Hemminger B.M. (2010): Scientific data repositories on the Web: An initial survey. In: Journal of the American Society for Information Science and Technology, 61(10):2029–48, http://onlinelibrary.wiley.com/doi/10.1002/asi.21339/full

Hackett, Edward J. (2010): Peer review and social science research funding, pp. 264-266, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Hafer L.W. & Kirkpatrick A.E. (2009): Assessing open source software as a scholarly contribution. In: Communications of the ACM, 52(12), 126-129, http://dl.acm.org/citation.cfm?id=1610285&CFID=500091250&CFTOKEN=70398928

Hagedorn G., Mietchen D., Morris R.A. *et al.* (2011): Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. In: Smith V. & Penev L. (eds.): e-Infrastructures for data publishing in biodiversity science. ZooKeys 150: 127-149, http://zookeys.pensoft.net/articles.php?id=3036

Hahnel, Mark (2015a): 2015 - The year of open data mandates. In: Figshare weblog, 7 January 2015, http://figshare.com/blog/2015_The_year_of_open_data_mandates/143

Hahnel, Mark (2015b): What's the point of open academic data? Who will it affect? In: Figshare weblog, 12 January 2015, http://figshare.com/blog/What_s_the_point_of_open_academic_data_Who_will_it_affect_/144

Halevi G. & Moed H.F. (2012): The evolution of big data as a research and scientific topic: Overview of the literature. Research Trends 30, September 2012, http://www.researchtrends.com/issue-30-september-2012/the-evolution-of-big-data-as-a-research-and-scientific-topic-overview-of-the-literature/

Halevi G. & Moed H.F. (2012): The evolution of big data as a research and scientific topic: Overview of the literature. Research Trends 30, September 2012, http://www.researchtrends.com/issue-30-september-2012/the-evolution-of-big-data-as-a-research-and-scientific-topic-overview-of-the-literature/

Hall M. & Silliman, S.W. (2006): Introduction: Archaeology of the Modern World. In: Historical Archaeology. Oxford: Blackwell, http://www.faculty.umb.edu/stephen_silliman/Articles/Introduction%20%28Hall%20and%20Silliman%29.pdf

Hall, Neil (2014): The Kardashian index: a measure of discrepant social media profile for scientists. In: Genome Biology, 15: 424, http://rd.springer.com/article/10.1186%2Fs13059-014-0424-0

Hames, Irene *et al*. (2012): Report on the International Workshop on Contributorship and Scholarly Attribution. Harvard University, Institute for Quantitative Social Science (IQSS), 16 May 2012, http://projects.iq.harvard.edu/attribution_workshop

Hamilakis Y. & Anagnostopoulos A. (2009): What is Archaeological Ethnography? In: Public archaeology: archaeological ethnographies, 8(2/3): 65–87, https://www.academia.edu/240591/What_is_archaeological_ethnography

Hamilton S., Hill J.D., Morris E., Woodward A. (1992): Guidelines for the Analysis and Publication of Prehistoric Ceramics. Prehistoric Ceramics Research Group, Occasional Paper 2. Oxford

Hamilton, Sue (2014): Under-Representation in Contemporary Archaeology. In: PIA. Papers from the Institute of Archaeology / University College London, Vol. 24, http://www.pia-journal.co.uk/article/view/pia.469/648

Hammarfelt, Björn (2014): Using altmetrics for assessing research impact in the humanities. In: Scientometrics, 101(2): 1419-1430 (pre-print), http://bada.hb.se/bitstream/2320/13511/1/Using%20altmetrics%20for%20assessing%20research%20impact%20in%20the%20humanities.pdf

Handke C., Guibault L. & Vallbé J.J. (2015): Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research, 7 June 2015. Available from Social Science Research Network (SSRN), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2608513

Harding, Anthony (2008): A conversation with Colin Renfrew. In: European Journal of Archaeology, 11(2/3): 143-170, https://www.academia.edu/1638249/A_conversation_with_Colin_Renfrew_Professor_Lord_Renfrew_of_Kaimsthorn_

Hardman, Catherine (2014): Repository perspective: Archaeology Data Service. Presentation at 4C/DPC Conference, London, 17-18 November 2014, http://4cproject.eu/component/docman/doc_download/73-4c-conference-day-1-session-3-hardman?Itemid=

Harley, Diane (2007): Use and Users of Digital Resources. A survey explored scholars' attitudes about educational technology environments in the humanities. In: Educause Quarterly, 30(4), https://net.educause.edu/ir/library/pdf/eqm0742.pdf

Harley, Diane *et al*. (2010a): Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines. – Executive Summary and Full Report. University of California Berkeley, Center for Studies in Higher Education, eScholarship, http://escholarship.org/uc/cshe_fsc

Harley, Diane *et al*. (2010b): Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines. – Archaeology Case Study. University of California Berkeley, http://escholarship.org/uc/item/15x7385g#page-37

Harnad S. & Brody T. (2004): Comparing the impact of open access (OA) vs. non-OA articles in the same journals. D-lib Magazine, 10(6), http://www.dlib.org/dlib/june04/harnad/06harnad.html

Harpring, Patricia (2014): Linked Open Data in the Cultural Heritage World: Issues for Information Creators and Users. CLIR – Council on Library and Information Resources weblog, 20 March 2014, http://connect.clir.org/blogs/patricia-harpring/2014/03/20/linked-open-data-in-the-cultural-heritage-world-issues-for-information-creators-and-users

Harris, Trevor M. (2012): Interfacing archaeology and the world of citizen sensors: exploring the impact of neogeography and volunteered geographic information on an authenticated archaeology. In: World Archaeology, 44(4), 580-591, http://www.tandfonline.com/toc/rwar20/44/4#.VSen25Mufh

Hartigan, John (2015): Cultural evolution as dialectic. In: Savage Minds weblog, 21 April 2015, http://savageminds.org/2015/04/21/cultural-evolution-as-dialectic/#more-16800

Harvey D. & Jim Perry (eds., 2015): The Future of Heritage as Climates Change: Loss, Adaptation and Creativity (Key Issues in Cultural Heritage). Routledge

Harzing A.W., Alakangas S. & Adams D. (2014): hIa: An individual annual h-index to accommodate disciplinary and career length differences. In: Scientometrics, 99(3): 811-821 (authors' manuscript), http://www.harzing.com/download/hia.pdf

Harzing website: Publish or Perish (citation metrics), http://www.harzing.com/pop.htm

Hasnain A., Syeda Sana e Zainab, Kamdar M.R. *et al*. (2015): A roadmap for navigating the life sciences linked Open Data Cloud, pp. 97-112, in: Semantic Technology: JIST 2014 - 4th Joint International Conference, Chiang Mai, Thailand, 9-11 November 2014. Lecture Notes in Computer Science 8943, http://maulik-kamdar.com/wp-content/uploads/2014/10/JIST2014.pdf

Haustein S. & Siebenlist T. (2011): Applying social bookmarking data to evaluate journal usage. In: Journal of Informetrics, 5(3): 446-457, http://dx.doi.org/10.1016/j.joi.2011.04.002

Havlicek I. & Wilson (2012): Zagora goes digital. Zagora Dig Blog, 7 November 2012, http://www.powerhousemuseum.com/zagora/2012/11/07/zagora-goes-digital/#more-2845

Heaton, Janet (2004): Reworking Qualitative Data. SAGE Research Methods; chapter "What is Secondary Analysis?" available online at: http://www.uk.sagepub.com/gray3e/study/chapter20/book%20chapters/What_is_Secondary_Analysis.pdf

Hedges M., Blanke T. & Marciano R. (2013): IEEE Big Data 2013, Workshop 2: "Big Data and the Humanities", Santa Clara, USA, 8 October 2013, http://bighumanities.net/events/ieee-bigdata-2013/workshop-program/

Hedges M., Blanke T. & Marciano R. (2014): IEEE Big Data 2014, workshop "Big Humanities Data", Maryland, USA, 27 October 2014, http://bighumanities.net/events/ieee-bigdata-oct-2014/big-humanities-data-workshop-program/

Hedges M., Blanke T., Fabiane S. *et al*. (2012): Sheer curation of experiments: data, process, provenance. In: Journal of Digital Information, 13(1), https://journals.tdl.org/jodi/index.php/jodi/article/view/5883

Hedges, Mark (2009): Grid-enabling Humanities Datasets. In: Digital Humanities Quarterly, 3(4), http://digitalhumanities.org/dhq/vol/3/4/000078/000078.html

Hedstrom M., Niu J., & Marz K. (2008): Incentives for data producers to create "archive/ready" data: Implications for archives and records management. In: Proceedings from the Society of American Archivists Research Forum. San Francisco: SAA, http://files.archivists.org/conference/2008/researchforum/M-HedstromJ-Niu-SAA-ResearchPaper-2008.pdf

HeiDATA (Heidelberg University, data repository, launched in June 2014), https://heidata.uni-heidelberg.de/dvn/

Heidorn, P. Bryan (2008): Shedding Light on the Dark Data in the Long Tail of Science. In: Library Trends, 57(2): Institutional Repositories: Current State and Future. Edited by S. Sheeves & M. Cragin, http://hdl.handle.net/2142/9127

Hellqvist, Björn (2010): Referencing in the humanities and its implications for citation analysis. In: Journal of the American Society for Information Science and Technology, 61(2): 310-318, http://onlinelibrary.wiley.com/doi/10.1002/asi.21256/abstract

Henneken, Edwin (2015): RDAP Review: Unlocking and Sharing Data in Astronomy. In: ASIS&T Bulletin, April/May 2015, https://www.asis.org/Bulletin/Apr-15/AprMay15_RDAP_Henneken.html

Henson, Don (2011): Does archaeology matter?, pp. 123-130, in: Moshenska G. & Dhanjal S. (eds.): Community archaeology. Themes, methods and practices. Oxford University Press (manuscript), https://www.academia.edu/1162187/Does_Archaeology_Matter

Henty, Margaret (2008): Developing the capability and skills to support e-research. In: Ariadne, Issue 55, http://www.ariadne.ac.uk/issue55/henty/

HERA-NET (2006): The HERA Survey on Infrastructural Research Facilities and Practices for the Humanities in Europe. S. Kaur-Pedersen and G.M. Kladakis, Danish Agency for Science, Technology and Innovation, September 2006, http://www.ucm.es/info/eurohum/docs/heraonhumanities.pdf

Heritage Council of Ireland: Significant Unpublished Irish Archaeological Excavations 1930-1997, website, http://heritagecouncil.ie/unpublished_excavations/section1.html

Heritage Data - Linked Data Vocabularies for Cultural Heritage, http://www.heritagedata.org

Heritage Data: Vocabulary Matching Tool, http://heritagedata.org/vocabularyMatchingTool/

HeritageToGather - Creating 3D models from photographs (2013-2015, funded under the UK Humanities Research Council's 'Connected Communities' programme), http://heritagetogether.org

Hettrick, Simon (2014): It's impossible to conduct research without software, say 7 out of 10 UK researchers. In: Software Sustainability Institute weblog, http://www.software.ac.uk/blog/2014-12-04-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers

Hettrick, Simon *et al*. (2014): UK Research Software Survey 2014 [dataset], 4 December 2014, https://zenodo.org/record/14809#.VWef6kYjHEQ

Heuvel C. van den (2006): Modeling historical evidence in digital maps: A preliminary sketch. In: e-Perimetron, 1(2): 113-126, http://www.e-perimetron.org/Vol_1_2/Heuvel/van%20den%20Heuvel.pdf

Hey T. & Hey J. (2006): E-science and its implications for the library community. In: Library Hi Tech, 24(4): 515–528; manuscript, http://eprints.rclis.org/9202/1/heyhey_final_web.pdf

Hey T., Tansley S. & Tolle K. (eds., 2009): The Fourth Paradigm. Data-Intensive Scientific Discovery. Redmond, WA: Microsoft Research, http://research.microsoft.com/en-us/collaboration/fourthparadigm/

Hicks D. & Wang J. (2011): Coverage and overlap of the new social sciences and humanities journal lists. In: Journal of the American Society for Information Science and Technology, 62(2): 284-294; preprint, http://works.bepress.com/diana_hicks/22/

Hicks, Diana (2005): The four literatures of social science. Chapter 22, in: H. Moed, W. Glänzel & U. Schmoch (eds.): Handbook of Quantitative Science and Technology Research. Kluwer Academic; manuscript: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.10.500&rep=rep1&type=pdf

Hine, Christine (2006): Computerization movements and scientific disciplines: The reflexive potential of new technologies, pp. 26-47, in: New infrastructures for knowledge production. Understanding e-science. London: Idea Group

Hine, Christine (2008): Systematics as Cyberscience: Computers, Change and Continuity in Science. Cambridge, MA: MIT Press.

Hinrichs E. & Krauwer S. (2014): The CLARIN Research Infrastructure: Resources and Tools for E-Humanities Scholars, pp. 1525–31, in: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, May 2014, http://www.clarin.nl/sites/default/files/415_Paper.pdf

Hirsch Hadorn, Gertrude *et al*. (2008): Handbook of Transdisciplinary Research. Dordrecht: Springer

Hoang D.T., Kaur J. & Menczer F. (2010): Crowdsourcing Scholarly Data. In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, Raleigh/NC, USA, 26-27 April 2010, http://journal.webscience.org/321/

Hodder I., Karlsson H. & Olsen B. (2008): 40 Years of Theoretical Engagement: A Conversation with Ian Hodder In: Norwegian Archaeological Review, 41(1): 26-42, https://hodder.squarespace.com/s/40-years-of-theoretical-engagement.pdf

Hodder I., Schaffer J., Yazicioğlu G.B. & Marshall M.E. (n.d.): To The Trowel's Edge: An Interview with Ian Hodder, http://ucexchange.uchicago.edu/interviews/hodder.html

Hodson, Simon (2012): Institutional Data Repositories and the Curation Hierarchy, 6 August 2012, http://researchdata.jiscinvolve.org/wp/2012/08/06/institutional-data-repositories-and-the-curation-hierarchy-reflections-on-the-dcc-icpsr-workshop-at-or2012-and-the-royal-societys-science-as-an-open-enterprise-report/

Hodson, Simon (2013): Supporting Research Data Management in Universities: the Jisc Managing Research Data Programme. Presentation at Cerif 4 Datasets Workshop, Glasgow, 12 July 2013, https://cerif4datasets.files.wordpress.com/2013/07/jisc-background-to-the-rdm-programme.pptx

Hofmann, Daniela (2015): What have genetics ever done for us? The implications of aDNA data for interpreting identity in Early Neolithic Central Europe. In: European Journal of Archaeology, 18(3): 454-476, https://www.academia.edu/9658297/What_Have_Genetics_Ever_Done_for_Us_The_Implications_of_aDNA_Data_for_Interpreting_Identity_in_Early_Neolithic_Central_Europe

Hole B., Lin L., McCann P. & Wheatley P. (2010): LIFE3: A predictive costing tool for digital collections. 7th International Conference on Preservation of Digital Objects (iPRES2010), Vienna, 19-24 September 2010, http://www.life.ac.uk/3/docs/Ipres2010_life3_submitted.pdf

Hole, Brian (2012): A Call for Open Scholarship in Archaeology. In: Bonacchi, Chiara (ed.): Archaeologists and Digital Communication: Towards Strategies of Public Engagement. London: Archetype; preprint, http://discovery.ucl.ac.uk/1326267/1/1326267.pdf

Hole, Brian (2013): REWARD Project, JISC Final Report. University College London, 27 February 2013, http://www.ucl.ac.uk/reward/documents/reward_final_report_1-1.pdf

Hole, Frank (2007): Agricultural sustainability in the semi-arid Near East. In: Climate of the Past, 3: 193–203, www.clim-past.net/3/193/2007/cp-3-193-2007.pdf

Holmgren, Sverker (2014): The Development of the European e-Infrastructure Landscape. Implications for ESFRI projects? Workshop: Launch of the 2015-16 ESFRI Roadmap Update, Trieste, 25 September 2014, http://www.copori.eu/_media/9_e-irg-ESFRI-launch-event-140925.pdf

Holtorf, Cornelius (2007): Archaeology is a Brand! The Meaning of Archaeology in Contemporary Popular Culture. Oxford: Archaeopress

Hoogerwerf M., Lösch M., Schirrwagen J. *et al*. (2013): Linking Data and Publications: Towards a Cross-Disciplinary Approach [OpenAIRE]. In: International Journal of Digital Curation, 8(1): 244-254, http://www.ijdc.net/index.php/ijdc/article/view/257

Howard, Jennifer (2013): Rise of 'Altmetrics' Revives Questions About How to Measure Impact of Research. In: The Chronicle, 3 June 2013, online, http://chronicle.com/article/Rise-of-Altmetrics-Revives/139557/

Hu, Di (2011): Advancing Theory? Landscape Archaeology and Geographical Information Systems. Papers from the Institute of Archaeology, University College London, 21, http://pia-journal.co.uk/article/view/pia.381/446

Huang M.H. & Chang Y.W. (2008): Characteristics of research output in social sciences and Humanities: from a research evaluation perspective. In: Journal of the American Society for Information Science and Technology 59(11): 1819-28, http://onlinelibrary.wiley.com/doi/10.1002/asi.20885/abstract

Hubblesite (2011): Hubble racks up 10,000 science papers. News release STScI-2011-40, 6 December 2011, http://hubblesite.org/newscenter/archive/releases/2011/40/image/a/

Hudson M.J., Aoyama M., Hoover K.C. & Uchiyama J. (2012): Prospects and Challenges for an Archaeology of Global Climate Change. In: Wiley Interdisciplinary Reviews: Climate Change 3: 313-328, http://onlinelibrary.wiley.com/doi/10.1002/wcc.174/abstract

Hudson, Mark J. (2013): Navigating Disciplinary Challenges to Global Sustainability Science: An Archaeological Model. In: Documenta Praehistorica, XL: 219-226, http://revije.ff.uni-lj.si/DocumentaPraehistorica/article/view/40.17/811

Hudson, Richard L. (2012, rapporteur): Open Infrastructures for Open Science. Horizon 2020 consultation report, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/open-infrastructure-for-open-science.pdf

Hug C., Salinesi C., Deneckère R. & Lamassé S. (2011): Process Modelling for Humanities: Tracing and Analysing Scientific Processes, pp. 245-255, in: Proceedings of CAA 2011, Beijing, China. Amsterdam University Press, https://hal-paris1.archives-ouvertes.fr/file/index/docid/662699/filename/CAA2011_Hug_et_al.pdf

Hug S.E. & Ochsner M. (2014): A Framework to Explore and Develop Criteria for Assessing Research Quality in the Humanities. In: International Journal of Education Law and Policy, 10(1), http://www.psh.ethz.ch/research/publications/ijelp_inpress.pdf

Hugenholtz, Bernt (2005): Abuse of Database Right. Sole-source information banks under the EU Database Directive, pp. 203-219, in: Lévêque F. & Shelanski H. (eds.): Antitrust, patents and copyright: EU and US perspectives. Cheltenham: Elgar; pre-print, http://www.ivir.nl/publicaties/download/953

Huggett J., Lock G. & Reilly P. (2014): What do you want from Digital Archaeology? Roundtable at CAA-2014 Conference, Paris, 25 April 2014, http://caa2014.sciencesconf.org/resource/page/id/30;

Huggett J., Lock G. & Reilly P. (2015): Challenging Digital Archaeology – the discussion continues. Roundtable at CAA-2015, Siena, 2 April 2015, http://2015.caaconference.org/program/roundtables/rt1/

Huggett, Jeremy (2012a): Disciplinary Issues: Challenging the Research and Practice of Computer Applications in Archaeology, pp. 13-24, in: Archaeology in the Digital Era. Papers from the CAA 2012, Southampton, 26-29 March 2012. Amsterdam University Press, http://dare.uva.nl/cgi/arno/show.cgi?fid=516092

Huggett, Jeremy (2012b): Core or Periphery? - Digital Humanities from an Archaeological Perspective. In: Thaller M. (ed.): The Cologne Dialogue on Digital Humanities 2012 – Controversies around the Digital Humanities (Special Issue of Historical Social Research, Vol. 37/3), http://www.cceh.uni-koeln.de/files/Huggett_final.pdf

Huggett, Jeremy (2012c): What lies beneath: lifting the lid on archaeological computing, pp. 204-214, in: Chrysanthi A., Murrietta Flores P. & Papadopoulos C. (eds.): Thinking Beyond the Tool: Archaeological Computing and the Interpretative Process. Archaeopress, http://eprints.gla.ac.uk/61333/1/61333.pdf

Huggett, Jeremy (2012d): Lost in information? Ways of knowing and modes of representation in e-archaeology. In: World Archaeology, 44(4): 538-552, http://www.tandfonline.com/toc/rwar20/44/4#.VSen25Mufh

Huggett, Jeremy (2014): Promise and Paradox: Accessing Open Data in Archaeology. In: C. Mills, M. Pidd & E. Ward (eds.): Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities. Sheffield: HRI Online Publications, http://www.hrionline.ac.uk/openbook/chapter/dhc2012-huggett

Huggett, Jeremy (2015a): Challenging Digital Archaeology. In: Open Archaeology, 1(1): 79-85, http://www.degruyter.com/view/j/opar.2014.1.issue-1/opar-2015-0003/opar-2015-0003.xml?format=INT

Huggett, Jeremy (2015b): A Manifesto for an Introspective Digital Archaeology. In: Open Archaeology, 1(1): 86-95, http://www.degruyter.com/view/j/opar.2014.1.issue-1/opar-2015-0002/opar-2015-0002.xml?format=INT

Huggett, Jeremy (2015c): Big Data Analytics. In: Introspective Digital Archaeology weblog, 8 February 2015, https://introspectivedigitalarchaeology.wordpress.com/2015/02/08/big-data-analytics/

Huggett, S. & Taylor M. (2014): Elsevier expands metrics perspectives with launch of new altmetrics pilots, 1 March 2014, http://www.elsevier.com/authors-update/story/publishing-trends/elsevier-expands-metrics-perspectives-with-launch-of-new-altmetrics-pilots

Huggett, Sarah (2014): A quick look at references to research data repositories. In: Research Trends, Issue 38, http://www.researchtrends.com/issue-38-september-2014/a-quick-look-at-references-to-research-data-repositories/

Hughes, Ryan (2012): Taking Survey Digital: Implementing a Paperless Workflow on the Eastern Vani [Georgia] Survey. Paper read at the Redford Conference in Archaeology, Tacoma, WA, 25-27 October, http://archaeology.pugetsound.edu/RedfordConference2012/

Huijboom N. & Van den Broek T. (2011): Open data: an international comparison of strategies. In: European Journal of ePractice, Issue 12, http://share-psi.eu/papers/TNO-epractice.pdf

Hull D., Pettifer S.R. & Kell D.B. (2008): Defrosting the Digital Library: Bibliographic Tools for the Next Generation Web. In: PLoS Computational Biology, 4(10): e1000204, http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000204

Humanities at Scale / HaS-DARIAH (EU, H2020, 09/2015-08/2017; aims to develop a DARIAH platform and digital humanities user community), http://dariah.eu/about/humanities-at-scale.html

Huvila, Isto (2006): The Ecology of Information Work. A Case Study of Bridging Archaeological Work and Virtual Reality Based Knowledge Organisation. ÅBO Akademi University Press, http://www.doria.fi/bitstream/handle/10024/4153/TMP.objres.83.pdf

Hyams, Elspeth (2008): Data librarianship: a gap in the market. In: Update magazine, 7(6), http://www.era.lib.ed.ac.uk/handle/1842/2499

Hypotheses (a publication platform for academic weblogs in the humanities and social sciences), http://en.hypotheses.org

Hyvönen E., Mäkelä E., Salminen M. *et al*. (2005): MuseumFinland - Finnish Museums on the Semantic Web. Journal of Web Semantics, 3(2):224–241, http://www.seco.tkk.fi/publications/2005/hyvonen-makela-et-al-museumfinland-finnish-2005.pdf

Hyvönen, Eero (2012): Publishing and Using Cultural Heritage Linked Data on the SemanticWeb. Palo Alto, CA, USA: Morgan & Claypool

IANUS - Research Data Centre for Archaeology and Classical Studies (Germany; coord. German Archaeological Institute; funded by the German Research Foundation), http://www.ianus-fdz.de

IANUS - Schäfer F., Heinrich M. & Jahn S. (2014): Stakeholderanalyse 2013 zu Forschungsdaten in den Altertumswissenschaften. Teil 1: Ergebnisse [Version 1.0] Hrsg. IANUS, http://www.ianus-fdz.de/projects/ergebnisse/wiki

IANUS - Schäfer F., Heinrich M. & Jahn S. (2014): Stakeholderanalyse 2013 zu Forschungsdaten in den Altertumswissenschaften. Teil 1: Ergebnisse, Version 1.0, 21. Januar 2014, http://www.ianus-fdz.de/attachments/download/572/Teil1-Ergebnisse_2014-02-26_FINAL.pdf (http://www.ianus-fdz.de/projects/ap3-community/wiki/Stakeholderanalyse

ICONCLASS (classification system for cultural content), http://www.iconclass.org

ICONCLASS as Linked Open Data, http://www.iconclass.org/help/lod

ICPSR - Interuniversity Consortium for Political and Social Research: Data archive, http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp

ICSU - International Council for Science, http://www.icsu.org

ICSU - World Data System (WDS), https://www.icsu-wds.org

ICSU - World Data System: Data portal, https://www.icsu-wds.org/services/data-portal

IHOPE - Integrated History and future of People on Earth, http://ihopenet.org

IHOPE (2015): Global Environmental Change Threats to Heritage and Long Term Observing Networks of the Past, 10 June 2015, http://ihopenet.org/global-environmental-change-threats-to-heritage-and-long-term-observing-networks-of-the-pas/

ILTER - International Long Term Ecological Research Network, http://www.ilternet.edu

ImpactStory (altmetrics service provider), https://impactstory.org

Imperial College (2010): Research information management: Developing tools to inform the management of research and translating existing good practice. Imperial College London & Elsevier, sponsored by JISC, http://www.jisc.ac.uk/media/documents/programmes/RIM/RIMTNT_FinalReport.pdf

Ince D.C., Hatton L. & Cumming J.G. (2012): The Case for Open Computer Programs. In: Nature 482 (7386): 485–488, http://dx.doi.org/10.1038/nature10836

INCREMENTAL project (UK, JISC Managing Research Data programme, 10/2009-03/2011), http://www.lib.cam.ac.uk/preservation/incremental/

Ingraham, Thomas (2015): First living scientific figure: articles can now keep pace with scientific discovery. In: F1000 weblog, 22 April 2015, http://blog.f1000research.com/2015/04/22/first-living-scientific-figure-articles-can-now-keep-pace-with-scientific-discovery/

Ingwersen P. & Chavan V. (2011): Indicators for the Data Usage Index (DUI): an incentive for publishing primary biodiversity data through global information infrastructure. In: BMC Bioinformatics, 12(Suppl 15):S3 (15 December 2011), http://www.biomedcentral.com/1471-2105/12/S15/S3

INNET - Innovative Networking in Infrastructure for Endangered Languages (EU, FP7, 10/2011-09/2014, developed a network of language archives), http://www.innet-project.eu

INSPIRE Data Specification on Protected Sites – Technical Guidelines, Version 3.2, 17.04.2014, http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PS_v3.2.pdf

INSPIRE - Infrastructure for Spatial Information in the European Community (based on the EU Directive 2007/2/EC and subsequent regulations; the INSPIRE spatial data themes includes protected sites), http://inspire.ec.europa.eu/theme/ps/

Intellectual Property Issues in Cultural Heritage (IPinCH), http://www.sfu.ca/ipinch/

Intellectual Property Issues in Cultural Heritage (IPinCH, 2008-2014) project, http://en.wikipedia.org/wiki/Intellectual_Property_Issues_in_Cultural_Heritage_project

International Association for Classical Archaeology (AIAC), http://www.aiac.org

International Conferences on Research Infrastructure, first conference in Copenhagen, 21- 23 March 2012, http://www.icri2012.dk, second conference in Athens, 2-4 April 2014, http://www.icri2014.eu

International Tree-Ring Data Bank (NOAA / NCEI Paleoclimatology Program & World Data Center for Paleoclimatology), http://www.ncdc.noaa.gov/paleo/treering.html

International Union of Geological Sciences (IUGS), International Commission on Stratigraphy (ICS), Subcommission on Quaternary Stratigraphy: Working Group on the 'Anthropocene', http://quaternary.stratigraphy.org/workinggroups/anthropocene/

Internet Archaeology (e-journal, since 1996), http://intarch.ac.uk

Internet Archaeology: Data Papers, http://intarch.ac.uk/authors/data-papers.html

Ioannidis J.P.A. (2014): How to make more published research true. PLoS Medicine 11(10): e1001747, http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001747

Ioannidis, J.P.A. (2006): Concentration of the Most-Cited Papers in the Scientific Literature: Analysis of Journal Ecosystems. In: PLoS ONE, 1: e5. http://dx.plos.org/10.1371/journal.pone.0000005

IPANEMA - Ancient Materials Research Platform (Synchrotron SOLEIL, France), http://ipanema.cnrs.fr

IPERION CH - Integrated Platform for the European Research Infrastructure on Cultural Heritage (H2020, 2015-2019; aims to establish a distributed European research infrastructure for restoration and conservation of cultural heritage), http://www.iperionch.eu

iRODS - Integrated Rule-Oriented Data System, http://irods.org

ISA - Investigation/Study/Assay tools (managed by the Oxford e-Research Centre), http://www.isa-tools.org

Isaksen, Leif (2011): Archaeology and the Semantic Web. Thesis, University of Southampton, School of Electronics and Computer Science, December 2011, http://eprints.soton.ac.uk/206421/1/Thesis.pdf

ISO 14721:2012: Space data and information transfer systems – Open archival information system (OAIS) Reference model, http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

ISO 16363:2012: Space data and information transfer systems – Audit and certification of trustworthy digital repositories, http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

ISO/IEC 11179: Information Technology - Metadata registries standards, http://metadata-standards.org/11179/

ISO/IEC 24744:2007,2014: Software Engineering - Metamodel for Development Methodologies, http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=62644

Istituto Nazionale di Fisica Nucleare (INFN): Cultural Heritage Network (CHNet), http://infnbeniculturali.net

IWGDD (2009): Harnessing the Power of Digital Data for Science and Society. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. Washington/DC, January 2009, http://www.nitrd.gov/About/Harnessing_Power_Web.pdf

IWGSC - National Science and Technology Council, Committee on Science, Interagency Working Group on Scientific Collections (2009): Scientific Collections: Mission-Critical Infrastructure for Federal Science Agencies. Washington DC, http://www.whitehouse.gov/sites/default/files/sci-collections-report-2009-rev2.pdf

Jackson, Mike (2012): How to cite and describe software. Software Sustainability Institute website, 4 September 2012, http://software.ac.uk/so-exactly-what-software-did-you-use

Jacobs C.A. & Worley S.J. (2009): Data curation in climate and weather: Transforming our ability to improve predictions through global knowledge sharing. In: International Journal of Digital Curation, 4(2): 68-79, http://www.ijdc.net/index.php/ijdc/article/view/119/122

Jahnke L., Ashercover A. & Keralis S. (2012): The Problem of Data. Washington: Council on Library and Information Resources, August 2012, http://www.clir.org/pubs/reports/pub154/pub154.pdf

Jankowski N.W., Scharnhorst A., Tatum C. & Tatum Z. (2013): Enhancing scholarly publications: Developing hybrid monographs in the humanities and social sciences. In: Scholarly and Research Communication, 4(1), http://src-online.ca/index.php/src/article/viewFile/40/123

Jankowski, Nicholas W. (ed., 2009): E-Research: Transformation in Scholarly Practice. New York: Routledge

Janssens K., Alfeld M., Van der Snickt G. *et al*. (2013): The Use of Synchrotron Radiation for the Characterization of Artists' Pigments and Paintings. In: Annual Review of Analytical Chemistry, Vol. 6: 399-425, http://www.annualreviews.org/doi/abs/10.1146/annurev-anchem-062012-092702

Jantz R.C. & Wilson M.C. (2008): Institutional repositories: Faculty deposits, marketing, and the reform of scholarly communication. Journal of Academic Librarianship, 34(3), 186-195; https://rucore.libraries.rutgers.edu/rutgers-lib/37278/

Jeffrey, Stuart (2012): A new Digital Dark Age? Collaborative web tools, social media and long-term preservation. In: World Archaeology, 44(4): 553-570, http://www.tandfonline.com/toc/rwar20/44/4#.VSen25Mufh

JISC & OAPEN (2013): Open Access Monographs in the Humanities and Social Sciences conference, 1 and 2 July 2013, The British Library, London, https://www.jisc-collections.ac.uk/Documents/Reports/OA%20Monographs%20in%20HSS%20Conference%20Report%20Oct%202013.pdf

Jobmann A., Hoffmann C.P., Künne S., Peters I., Schmitz J. & Wollnik-Korn G. (2014): Altmetrics for large, multidisciplinary research groups: Comparison of current tools. Bibliometrie - Praxis und Forschung, 3, http://www.bibliometrie-pf.de/article/viewFile/205/258

Johnson K.M. & Ouimet W.B. (2014): Rediscovering the lost archaeological landscape of southern New England using airborne light detection and ranging (LiDAR). In: Journal of Archaeological Science, 43: 9–20, http://stonewall.uconn.edu/wp-content/uploads/sites/534/2014/03/Johnson-and-Ouimet-2014-Rediscovering-the-lost-archaeological-landscape-of-southern-New-England-using-airborne-LiDAR.pdf

Johnson L., Adams Becker S., Freeman A. (2013): The NMC Horizon Report: 2013 Museum Edition. Austin, Texas: The New Media Consortium, http://www.nmc.org/pdf/2013-horizon-report-museum-EN.pdf

Johnston, Lisa (2010): User-needs assessment of the research cyberinfrastructure for the 21st century (June 21, 2010). International Association of Scientific and Technological University Libraries, 31st Annual Conference, http://docs.lib.purdue.edu/iatul2010/conf/day1/5

Joint Task Force on Library Support for E-Science (2007): Agenda for Developing E-Science in Research Libraries. Association of Research Libraries, November 2007, http://old.arl.org/bm~doc/ARL_EScience_final.pdf

Jones, Elisabeth *et al*. (2008): E-science talking points for ARL Deans and Directors. Association of Research Libraries, 24 October 2008, http://www.arl.org/storage/documents/publications/e-science-talking-points.pdf

Jones, Sarah (2012): Developments in research funder data policy [in the UK]. International Journal of Digital Curation, 7(1): 114–125, http://www.ijdc.net/index.php/ijdc/article/view/209/278

Jöns, Hauke (2014): Discover the submerged prehistory of Europe – aims, methods and outcomes of the European Splashcos-network. EAC Annual Meeting 2014 / 15th EAC Heritage Management Symposium: Setting the Agenda: Giving new meaning to the European archaeological heritage, Amersfoort, Netherlands 20-22 March 2014, http://www.european-archaeological-council.org/files/eac_2014_hauke_j__ns_comp.ppt

Jöns, Hauke (2015): Discover the submerged prehistory of Europe - Scientific background, aims, methods, outcomes and perspectives of the European SPLASHCOS network, pp. 119-136, in: European Archaeological Council (EAC): Setting the Agenda – Giving New Meaning to the European Archaeological Heritage. Edited by P.A.C. Schut, D. Scharff & L.C. de Wit. EAC Occasional Paper no. 10, Budapest: Archaeolingua

Journal of Intercultural and Interdisciplinary Archaeology, http://www.jiia.it

Journal of Open Archaeology Data (since 2012), http://openarchaeologydata.metajnl.com

Journal of Open Research Software, http://openresearchsoftware.metajnl.com

JPI-CH - Joint Programming Initiative on Cultural Heritage and Global Change (2014): Strategic Research Agenda. Rome: MIBACT, http://www.jpi-culturalheritage.eu/2014/02/strategic-research-agenda-sra/

JPI-CH - Joint Programming Initiative on Cultural Heritage and Global Change (since 10/2011 supported by projects under the EU FP7 Environment and ERA-Net programmes), http://www.jpi-culturalheritage.eu

Kahn, Michael (2010): Measure for measure: quantifying the social sciences, pp. 359-385, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Kameník J., Havránek V., Smolík J., Velemínský P., Lynnerup N., Bruzek J. & Vellev J. (2013): Was he murdered or was he not? – Part I: Analyses of mercury in the remains of Tycho Brahe. In: Archaeometry, 55 (6): 1187-1195, http://onlinelibrary.wiley.com/doi/10.1111/j.1475-4754.2012.00729.x/abstract

Kansa E. & Whitcher-Kansa S. (2009): Open Context: Developing Common Solutions for Data Sharing. In: CSA Newsletter, Vol. XXI, No. 3, http://csanet.org/newsletter/winter09/nlw0902.html

Kansa E. & Whitcher-Kansa S. (2011): Guidelines for Web-based data publication in archaeology. A working document to inform archaeologists about sharing data, from the field to the Web. The Alexandria Archive Institute / Open Context, January 2011, http://ux.opencontext.org/blog/good-practice/

Kansa E. & Whitcher-Kansa S. (2013): We all know that a 14 is a sheep: data publication and professionalism in archaeological communication. In: Journal of Eastern Mediterranean Archaeology and Heritage Studies, 1(1): 88–97; preprint available at: https://escholarship.org/uc/item/9m48q1ff

Kansa E., Whitcher-Kansa S. & Arbuckle B. (2014): Publishing and pushing: mixing models for communicating research data in archaeology. In: International Journal of Digital Curation 9(1), http://dx.doi.org/10.2218/ijdc.v9i1.301

Kansa E., Whitcher-Kansa S. & Goldstein L. (2013): On Ethics, Sustainability, and Open Access in Archaeology, pp. 15-22, in: The SAA Archaeological Record, 13(4), September 2013, http://onlinedigeditions.com/publication/repo26/16146/175171/66647a9807469ad281a26db155f4b85cc6271664.pdf#page=17

Kansa E., Whitcher-Kansa S. & Watrall E. (eds., 2011): Archaeology 2.0: New Approaches to Communication and Collaboration. Cotsen Institute of Archaeology, UC Los Angeles, http://escholarship.org/uc/item/1r6137tb

Kansa, Eric (2011): Introduction: New directions for the digital past, pp. 1-25, in: Kansa E. *et al*. (eds.): Archaeology 2.0: New Approaches to Communication and Collaboration. Cotsen Institute of Archaeology, UC Los Angeles, http://escholarship.org/uc/item/1r6137tb

Kansa, Eric (2012): Openness and archaeology's information ecosystem. In: World Archaeology, 44(4): 498–520; preprint, http://escholarship.org/uc/item/4bt04063

Kansa, Eric (2014a): Open Context and Linked Data. ISAW Papers 7.10, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/kansa/

Kansa, Eric (2014b): The Need to Humanize Open Science, pp. 31–58, in: Moore, S. A. (ed.): Issues in Open Research Data. London: Ubiquity, http://dx.doi.org/10.5334/ban

Kansa, Eric (2014c): It's the Neoliberalism, Stupid: Why Instrumentalist Arguments for Open Access, Open Data, and Open Science Are Not Enough. In: Impact of Social Sciences, http://blogs.lse.ac.uk/impactofsocialsciences/2014/01/27/its-the-neoliberalism-stupid-kansa/

Kansa, Eric (2015): Contextualizing Digital Data as Scholarship in Eastern Mediterranean Archaeology. In: CHS Research Bulletin, 3(2), http://nrs.harvard.edu/urn-3:hlnc.essay:KansaE.Contextualizing_Digital_Data_as_Scholarship.2015

Kansa, Eric (2015, in preparation): Click Here to Save Archaeology. In: Walcek-Averett E., Gordon J. & Counts D. (eds.): Mobilizing the Past. Grand Forks, ND: The Digital Press at the University of North Dakota, working draft, https://ekansa-pubs.github.io/click-here-to-save-archaeology/

Karasti H. & Baker K. (2008): Digital Data Practices and the Long Term Ecological Research Program Growing Global. In: International Journal of Digital Curation, 3(2): 42-58, http://www.ijdc.net/index.php/ijdc/article/viewFile/86/57

Karasti H., Baker K.S. & Millerand F. (2010): Infrastructure Time: Long-term Matters in Collaborative Development. In: Computer Supported Cooperative Work, 19: 377-415, http://cce.lternet.edu/docs/bibliography/163ccelter.pdf

Katz, Daniel S. *et al*. (2014): Summary of the First Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE1). In: Journal of Open Research Software, 2(1): e6: 1-21, http://dx.doi.org/10.5334/jors.an

Kaye J., Vertesi J., Avery S. *et al*. (2006): To have and to hold: exploring the personal archive, pp. 275-284, in: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '06), New York: ACM Press; pre-print, http://alumni.media.mit.edu/~jofish/writing/tohaveandtohold.pdf

Kelleher M. & Hoogland E. (eds., 2011): Changing Publication Culture in the Humanities: Young Researchers Forum ESF Humanities Spring 2011, Strasbourg: Ireg, 2012, http://www.esf.org/fileadmin/Public_documents/Publications/Changing_Publication_Cultures_Humanities.pdf

Kennan M.A. & Kingsley D.A. (2009): The state of the nation: a snapshot of Australian institutional repositories. In: First Monday, 14(2), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2282/2092

Kepler (scientific workflow management system), https://kepler-project.org/users/projects-using-kepler

Key Perspectives (2010): Data Dimensions: Disciplinary Differences in Research Data Sharing, Reuse and Long Term Viability. Digital Curation Center, http://www.dcc.ac.uk/sites/default/files/SCARP%20SYNTHESIS_FINAL.pdf

Kilbride, William (2006): Grand challenges: Grand opportunities? Archaeology, the historic environment sector and the E-science programme. UK AHDS E-science Scoping Study, expert seminar, 24 July 2006, http://www.ahds.ac.uk/e-science/documents/Archaeology-grand-challenges.pdf

Killick, David (2015): Using evidence from natural sciences in archaeology, pp. 159-172, in: Chapman R. & Wylie A. (eds.): Material Evidence. Learning from archaeological practice. London & New York: Routledge.

Kim Y. & Stanton J.M. (2013): Institutional and individual influences on scientists' data sharing behaviors: A multilevel analysis. In: Proceedings of ASIS&T 2013, Montreal, 1-6 November 2013, http://www.asis.org/asist2013/proceedings/submissions/papers/123paper.pdf

Kindling M., Schirmbacher P. & Simukovic E. (2013): Forschungsdatenmanagement an Hochschulen: das Beispiel der Humboldt-Universität zu Berlin. In: LIBREAS. Library Ideas 23, http://libreas.eu/ausgabe23/07kindling/

King J.C., Harley D., Earl-Novell S. *et al*. (2006): Scholarly communication: Academic values and sustainable models. Center for the Studies in Higher Education. Paper CSHE-16-06, July 27 2006 [case study on anthropology: 48-63], http://repositories.cdlib.org/cshe/CSHE-16-06

Kintigh K. & Altschul J.H. (2010): Sustaining the Digital Archaeological Record. In: Heritage Management, 3(2): 264-274, http://www.digitalantiquity.org/wp-uploads/2011/01/20110127-Kintigh-Altschul-Forum-on-Sustaining-the-Digital-Archaeological-Record-from-Heritage-Management.pdf

Kintigh K., Altschul J., Kinzig A.P., Limp F. *et al*. (2015): Cultural Dynamics, Deep Time, and Data: Planning Cyberinfrastructure Investments for Archaeology. In: Advances in Archaeological Practice, 3(1): 1-15, https://www.cni.org/wp-content/uploads/2015/02/archaeology-CI.pdf

Kintigh K.W., Altschul J.H., Beaudry M.C. *et al*. (2014a): Grand challenges for archaeology. In: American Antiquity, 79(1): 5-24, http://www2.lawrence.edu/fast/PEREGRIP/Publications/2014AmAntiqGrandChallenges.pdf

Kintigh K.W., Altschul J.H., Beaudry M.C. *et al*. (2014b): Grand challenges for archaeology. In: PNAS - Proceedings of the National Academy of Sciences, 111(3), http://www.pnas.org/content/111/3/879.full

Kintigh K.W., McManamon F.P. & Spielmann K.A. (2010): Synthesis and Cyberinfrastructure for SBE Research. White Paper in Response to: SBE 2020: Future Research in the Social, Behavioral & Economic Sciences. Arizona State University, School of Human Evolution & Social Change & Center for Digital Antiquity, http://www.tdar.org/wp-uploads/2010/12/201010115-Cyberinfrastructures-for-SBE-Research-final-w-abstract.docx

Kintigh, Keith (2006): The challenge of archaeological data integration. Paper presented at the meeting of the Union Internationale des Sciences Préhistoriques et Protohistoriques: Session Technology and Methodology for Archaeological Practice, Lisbon, September 2006, http://archaeoinformatics.org/articles/Kintigh2006UISPP.pdf

Kirschenbaum, Matthew (2014): What is “Digital Humanities", and why are they saying such terrible things about It? In: differences 25.1 (Special issue: In the Shadows of the Digital Humanities), http://differences.dukejournals.org/content/25/1.toc

Klein, JulieT (2010): The Metaphorics of Mapping Interdisciplinary Knowledge, pp. 17-24, in: European Science Foundation: The Future of Knowledge. Mapping Interfaces. Interdisciplinary Workshop Report. ESF, October 2010, http://www.esf.org/fileadmin/Public_documents/Publications/Future_of_Knowledge.pdf

Kleiner B., Renschler I., Wernli B., Farago P. & Joye D. (eds., 2013): Understanding Research Infrastructures in the Social Sciences. Commissioned by the Swiss Foundation for Research in Social Sciences. Zürich: SEISMO

Klimpel, Paul (2012): Free knowledge based on Creative Commons Licenses. Consequences, risks, and side-effects of the license module Non-Commercial – NC. Wikimedia Germany, iRights.info & Creative Commons Germany, May 2012, http://openglam.org/files/2013/01/iRights_CC-NC_Guide_English.pdf

Klump, Jens (2011): Criteria for the trustworthiness of data centres. In: D-Lib Magazine, 17(1-2), http://www.dlib.org/dlib/january11/klump/01klump.html

Knappett, Carl (2011): An Archaeology of Interaction. Networks Perspectives on Material Culture and Society. Oxford University Press

Knappett, Carl (2011): Introduction: Why Networks?, pp. 3-15, in: An Archaeology of Interaction: Network Perspectives on Material Culture and Society. Oxford University Press, https://www.academia.edu/2570224/Network_Analysis_in_Archaeology

Knappett, Carl (ed., 2013): Network Analysis in Archaeology. New Approaches to Regional Interaction. Oxford University Press

Knowledge Exchange (2013): Making Data Count: Research data availability and research assessment, Workshop, Berlin, http://www.knowledge-exchange.info/Default.aspx?ID=577

Knowledge Unlatched (initiative for open access books), http://www.knowledgeunlatched.org

Kohler T.A. & Van der Leeuw S.E. (eds., 2007): Model-Based Archaeology of Socionatural Systems (SAR, Santa Fe, NM)

Kolbmann, Wibke (2014): A Digital Registry for Archaeological Find Spots and Excavation Documentation in IANUS. Humboldt-Universität zu Berlin, Institut für Bibliotheks- und

Informationswissenschaft, http://edoc.hu-berlin.de/series/berliner-handreichungen/2014-359/PDF/359.pdf

Kollias S. & Cousins J. (eds., 2008): Semantic Interoperability in the European Digital Library. Proceedings of the First International Workshop, SIEDL 2008, Tenerife, June 2, 2008, pp. 14-35, http://image.ntua.gr/swamm2006/SIEDLproceedings.pdf

Kominko, Maja (ed., 2015): From Dust to Digital: Ten Years of the Endangered Archives Programme. OpenBook Publishers, http://www.openbookpublishers.com/htmlreader/978-1-78374-062-8/Contents.xhtml

Konkiel S. & Scherer D. (2013): New Opportunities for Repositories in the Age of Altmetrics. In: ASIS&T Bulletin, 39(4): https://asis.org/Bulletin/Apr-13/Bulletin_AprMay13_Final.pdf

Konkiel, Stacy (2013): Tracking citations and altmetrics for research data: Challenges and opportunities. In: Bulletin of the American Society for Information Science and Technology, August/September 2013, https://www.asis.org/Bulletin/Aug-13/AugSep13_Konkiel.html

Konkiel, Stacy (2014): Tracking the impacts of data – beyond citations. e-Science Community weblog, 16 October 2014, http://esciencecommunity.umassmed.edu/2014/10/16/data-metrics/

Korn N. & Oppenheim C. (2011): Licensing Open Data: A Practical Guide. May 2011, http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf

Kozinets, Robert V. (2009): Netnography. Doing Ethnographic Research Online. Sage Publications Ltd.

Kratz J. & Strasser C. (2014): Data publication consensus and controversies. F1000Research 3: 94, http://dx.doi.org/10.12688/f1000research.4518

Kratz J. & Strasser C. (2015a): Researcher Perspectives on Publication and Peer Review of Data. In: PLoS ONE 10(4): e0123377, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0117619

Kratz J. & Strasser C. (2015b): Making data count. In: Scientific Data 2, http://www.nature.com/articles/sdata201539

Kratz, John (2013): Data citation developments. In: DataPub weblog, 11 October 2013, http://datapub.cdlib.org/2013/10/11/data-citation-developments/

Kratz, John (2014): Fifteen ideas about data validation (and peer review). In: DataPub weblog, 8 May 2014, http://datapub.cdlib.org/2014/05/08/fifteen-ideas-about-data-validation-and-peer-review/

Kratz, John (2015): Make data rain. In: DataPub weblog, 8 January 2015, http://datapub.cdlib.org/2015/01/08/make-data-rain/

Kreutzer, Till (2014): Open Content - A Practical Guide to Using Creative Commons Licences. German Commission for UNESCO, HBZ - Library Service Centre & Wikimedia Deutschland, https://www.unesco.de/fileadmin/medien/Dokumente/Kommunikation/Open_Content_A_Practical_Guide_to_Using_Open_Content_Licences_web.pdf

Kriegel H.-P., Kröger P., van der Meijden C.H., Obermaier H., Peters J. & Renz M. (2010): Towards Archaeo-informatics: Scientific Data Management for Archaeobiology. In: Scientific and Statistical Database Management. Proceeding of SSDBM 2010. Springer, LNCS, Vol.6187, 169-177, http://fdb.vetmed.uni-muenchen.de/upload/dateien/1_10.pdf

Kriesberg A., Frank R.D., Faniel I.M. & Yakel E. (2013): The Role of Data Reuse in the Apprenticeship Process. ASIS&T Proceedings, 50(1), http://deepblue.lib.umich.edu/handle/2027.42/106839

Kristiansen K., Šmejda L. & Turek J. (eds., 2015): Paradigm Found: Archaeological Theory: Present, Past and Future. Essays in Honour of Evžen Neustupný. Oxbow Books

Kristiansen, Kristian (2001): Borders of Ignorance: Research Communities and Language, pp.38–43, in: Kobylinski Z. (ed.): Quo Vadis Archaeologia?WhitherEuropean Archaeology in the 21st Century. European Science Foundation / Polish Academy of Sciences, Warsaw; https://www.academia.edu/326277/Borders_of_Ignorance_Research_Communities_and_Language

Kristiansen, Kristian (2009): Contract archaeology in Europe: an experiment in diversity. In: World Archaeology, 41(4): 641-648, http://www.tandfonline.com/doi/abs/10.1080/00438240903371486?journalCode=rwar20

Kristiansen, Kristian (2009): Contract archaeology in Europe: an experiment in diversity. In: World Archaeology, 41(4): 641-648, http://www.tandfonline.com/doi/abs/10.1080/00438240903371486?journalCode=rwar20

Kruse F. & Thestrup J.B. (2014): Research libraries' new role in research data management, current trends and visions in Denmark. In: LIBER Quarterly - The Journal of the Association of European Research Libraries, 23(4), http://liber.library.uu.nl/index.php/lq/article/view/9173/9761

Kunze J., Cruse P., Hu R., Abrams S. *et al*. (2011): Practices, Trends, and Recommendations in Technical Appendix Usage for Selected Data-Intensive Disciplines. California Digital Library, http://www.cdlib.org/services/uc3/docs/dax.pdf

Kurtz M.J. & Bollen J. (2010): Usage bibliometrics. In: Annual review of information science and technology, 44(1): 1-64, arXiv:1102.2891v1 [cs.DL], http://arxiv.org/ftp/arxiv/papers/1102/1102.2891.pdf

Kvalheim V. & Kvamme T. (2014): Policies for Sharing Research Data in Social Sciences and Humanities. A survey about research funders' data policies. International Federation of Data Organisations (IFDO), http://www.ada.edu.au/documents/ifdo-report-on-policies-for-data-sharing

Kwok, Roberta (2013): Research impact: Altmetrics make their mark. In: Nature, 500 (7463): 491-493, http://www.nature.com/naturejobs/science/articles/10.1038/nj7463-491a

Labrador, Angela M. (2012): Ontologies of the future and interfaces for all: archaeological databases for the twenty-first century. In: Archaeologies, 8(3): 236–249. http://link.springer.com/article/10.1007/s11759-012-9203-2

Lagerlöf, Agneta (ed., 2013): Who cares? Perspectives on Public Awareness, Participation and Protection in Archaeological Heritage Management. European Archaeological Council, EAC Occasional Paper 8. Budapest: Archaeolingua, http://european-archaeological-council.org/publications

Lagoze C. & Van de Sompel H. (2007): Open Archives Initiative – Object Reuse and Exchange. Compound Information Objects: The OAI-ORE Perspective, 28 May 2007, http://www.openarchives.org/ore/documents/CompoundObjects-200705.html

Lagoze C., Edwards P., Sandvig C. & Plantin J-P (2015): Should I Stay or Should I Go? Alternative Infrastructures in Scholarly Publishing. In: International Journal of Communication, 9: 1052-71, http://ijoc.org/index.php/ijoc/article/view/2929/1354

Lagoze, Carl (2010): Lost Identity: The Assimilation of Digital Libraries into the Web. PhD thesis, Information Science, Cornell University, http://www.cs.cornell.edu/lagoze/dissertation/dissertation.html

Lagoze, Carl (2014): Big Data, data integrity, and the fracturing of the control zone. In: Big Data & Society, July-December 2014: 1-11, http://bds.sagepub.com/content/spbds/1/2/2053951714558281.full.pdf

Lake, Mark (ed., 2012): Special issue "Open Archaeology", World Archaeology, 44(4), http://www.tandfonline.com/toc/rwar20/44/4

Lake, Mark (2012): Open archaeology [introduction to the special issue]. World Archaeology, 44(4): 471-478, http://dx.doi.org/10.1080/00438243.2012.748521

LAMS - Laboratoire d'archéologie moléculaire et structural, http://www.umr-lams.fr/?lang=fr

Lane, Paul J. (2015): Archaeology in the age of the Anthropocene: A critical assessment of its scope and societal contributions. In: Journal of Field Archaeology, 40(5): 485-498, http://www.maneyonline.com/doi/abs/10.1179/2042458215Y.0000000022?af=R

Lange, Guus (ed., 2004): Reference Collections. Foundation for Future Archaeology. In: Proceedings of the international conference on the European electronic Reference Collection, 12-13 May 2004, ROB, Amersfoort, The Netherlands, http://www.cultureelerfgoed.nl/sites/default/files/publications/reference-collections-foundation_for_future_archaeology.pdf

LAPSI - The European Thematic Network on Legal Aspects of Public Sector Information (01/2010-12/2014), http://www.lapsi-project.eu; http://ec.europa.eu/digital-agenda/en/news/legal-aspects-public-sector-information-lapsi-thematic-network-outputs

Larivière V., Archambault E. & Gingras Y. (2008): Long-Term Variations in the Aging of Scientific Literature: From Exponential Growth to Steady-State Science (1900–2004). In: Journal of the American Society for Information Science and Technology, 59(2): 288-296, http://www.cirst.uqam.ca/portals/0/docs/pub/jasist_aging2008.pdf

Larivière V., Gingras Y., & Archambault É. (2008): The decline in the concentration of citations, 1900-2007. In: Journal of the American Society for Information Science and Technology, 60: 858-862, http://dx.doi.org/10.1002/asi.21011

Last House on the Hill (Centre for Digital Archaeology, University of California, Berkeley), http://lasthouseonthehill.org

Lauriault T.P., Craig B.L., Taylor D.R. & Pulsifer P. (2007): Today's data are part of tomorrow's research: archival issues in the sciences. In: Archivaria, 64: 123-179. http://journals.sfu.ca/archivar/index.php/archivaria/article/viewFile/13156/14404

LaValle S., Lesser E., Shockley R., Hopkins M.S. & Kruschwitz N. (2011): Big data, analytics and the path from insights to value. In: MIT Sloan Management Review, 52(2): 21-32, http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value

Lavoie B. & Dempsey L. (2004): Thirteen ways of looking at…digital preservation. In: D-Lib Magazine, 10 (7/8), http://www.dlib.org/dlib/july04/lavoie/07lavoie.html

Lavoie, Brian (2014): Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition). DPC Technology Watch Report 14-02 October 2014. Digital Preservation Coalition, http://dx.doi.org/10.7207/TWR14-02

Law M. & Morgan C. (2014): The Archaeology of Digital Abandonment: Online Sustainability and Archaeological Sites. In: Present Pasts, 6(1), http://www.presentpasts.info/articles/10.5334/pp.58/

Lawrence B., Jones C., Matthews B., Pepler S. & Callaghan S. (2011): Citation and peer review of data: Moving towards formal data publication. In: International Journal of Digital Curation, 6(2): 4-37, http://www.ijdc.net/index.php/ijdc/article/viewFile/181/265

Lazorchak, Butch (2012): A Digital Asset Sustainability and Preservation Cost Bibliography. Library of Congress, The Signal: Digital Preservation, 26 June 2012, http://blogs.loc.gov/digitalpreservation/2012/06/a-digital-asset-sustainability-and-preservation-cost-bibliography/

Leathem C. & Adrian D. (2015): Survey and Analysis of Basic Social Science and Humanities Research at the Science Academies and Related Research Organisations of Europe (SASSH). Union of the German Academies of Sciences and Humanities in cooperation with All European Academies

(ALLEA), http://www.irishhumanities.com/assets/Uploads/ALLEA-2015-ProjectReport-SASSH.pdf

LeClere, Felicia (2010): Too many researchers are reluctant to share their data. In: The Chronicle of Higher Education, 3 August 2010, http://chronicle.com/article/Too-Many-Researchers-Are/123749

Lectorium - Open Repository of Historical Sciences, https://lectorium.edu.pl

Lee C.A. & Tibbo H.R. (2007): Digital Curation and Trusted Repositories: Steps Toward Success. In: Journal of Digital Information, 8(2), https://journals.tdl.org/jodi/index.php/jodi/article/view/5883/5890

Lee C.P., Bietz M.J. & Thayer A. (2010): Research-Driven Stakeholders in Cyberinfrastructure Use and Development. Proceedings of the International Symposium on Collaborative Technologies and Systems, Chicago, 17-21 May 2010, pp. 163–172, https://depts.washington.edu/csclab/wordpress/wp-content/uploads/Lee-Stakeholders-CTS-2010.pdf

Lee C.P., Dourish P. & Mark G. (2006): The Human Infrastructure of Cyberinfrastructure. CSCW 2006 conference, Banff, Canada, 4-8.11.2005, http://www.ics.uci.edu/~gmark/CSCW06.pdf

Lehmann E.H. & Mannes D. (2012): Wood investigations by means of radiation transmission techniques. In: Journal of Cultural Heritage, 13(3), Supplement, S35–S43, http://www.sciencedirect.com/science/article/pii/S1296207412000660

LERU (2013): League of European Research Universities (LERU) Roadmap for Research Data. Advice Paper, No. 14, December 2013, http://www.uzh.ch/research/LERU_Roadmap_for_Research_data.pdf

Levi, Alma S. (2013): Humanities 'big data': myths, challenges, and lessons, pp. 33-36, in: Proc. IEEE International Conference on Big Data 2013, Silicon Valley, USA, 6-9 October 2013; pre-print, https://bighumanities.files.wordpress.com/2013/09/2_5_levi_paper.pdf

Lewis S., de Castro P. & Jones R. (2012): SWORD: Facilitating Deposit Scenarios. In: D-Lib Magazine, 18(1/2), http://www.dlib.org/dlib/january12/lewis/01lewis.html

Lewis, John A. (2014a): Institutional Research Data Management Technical Infrastructures. University of Sheffield, 19 November 2014, http://figshare.com/authors/joalewis/96626

Lewis, John A. (2014b): A Review of Options for the Development of Research Data Management Technical Infrastructure at the University of Sheffield. University of Sheffield, University Library, 25 November 2014, figshare, http://files.figshare.com/1806064/RDMTI_v2.5.pdf

Lewis, Martin (2010): Libraries and the management of research data, pp. 145-168, in: McKnight S. (ed.): Envisioning future academic library services: initiatives, ideas and challenges. London: Facet; manuscript, http://eprints.whiterose.ac.uk/11171/1/LEWIS_Chapter_v10.pdf

Lewis, Michael (2013): The PAS - a rather British solution.The mandatory reporting and voluntary recordingof archaeological objects in England and Wales. In: https://www.academia.edu/3841961/The_PAS_-_A_Rather_British_Solution_the_mandatory_reporting_and_voluntary_recording_of_archaeological_objects_in_England_and_Wales

Leydesdorff L., Hammarfelt B. & Salah A. (2011): The structure of the Arts & Humanities Citation Index: A mapping on the basis of aggregated citations among 1,157 journals. In: Journal of the American Society for Information Science and Technology, 62(12), 2414-26, http://arxiv.org/ftp/arxiv/papers/1102/1102.1934.pdf

LIBER - Ligue des Bibliothèques Européennes de Recherche / Association of European Research Libraries (over 400 members), http://libereurope.eu

LIBER - Ligue des Bibliothèques Européennes de Recherche / Association of European Research Libraries, Birte Christensen-Dalsgaard *et al*. (2012): Ten recommendations for libraries to get started with research data management. Final report of the Working group on E-Science / Research Data Management, July 2012, http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf

LIBER – Scholarly Communication and Research Infrastructures Steering Committee (2014): 11 Case Studies Released on Research Data Management in Libraries, June 2014, http://libereurope.eu/committee/scholarly-research/research-data-management-case-studies/

LIDO - Lightweight Information Describing Objects (museum metadata specification), http://network.icom.museum/cidoc/working-groups/lido/what-is-lido/

LIFE - Life Cycle Information for E-Literature (UK, JISC, three phases, 2006-2010), http://www.life.ac.uk

Limp, W. Fredrick (2011): Web 2.0 and Beyond, or On the Web, Nobody Knows You're an Archaeologist, pp. 265-280, in: Kansa E. *et al*. (eds.): Archaeology 2.0: New Approaches to Communication and Collaboration. Cotsen Institute of Archaeology, UC Los Angeles, http://escholarship.org/uc/item/1r6137tb

Lin A.Y-M., Huynh A., Lanckriet G. & Barrington L. (2014): Crowdsourcing the Unknown: The Satellite Search for Genghis Khan. In: PLoS ONE, 9(12): e114046, http://dx.doi.org/10.1371/journal.pone.0114046

Lin J. & Fenner M. (2013): The Many Faces of Article-Level Metrics. In: ASIS&T Bulletin, 39(4): 27-30, https://asis.org/Bulletin/Apr-13/Bulletin_AprMay13_Final.pdf

Lin J. & Strasser C. (2014): Recommendations for the Role of Publishers in Access to Data. In: PLoS Biology, 12(10): e1001975, http://dx.doi.org/10.1371/journal.pbio.1001975

Linked Data - Connect Distributed Data across the Web, http://linkeddata.org

Linked Heritage (EU, CIP-PSP, 04/2011-09/2013; employed Linked Data technology), http://www.linkedheritage.eu

Linked Open Data (LOD) cloud diagram, http://lod-cloud.net

Lisée C., Larivière V. & Archambault E. (2008): Conference proceedings as a source of scientific information: A bibliometric analysis. In: Journal of the American Society for Information Science and Technology (JASIST), 59(11): 1776-84, http://www.science-metrix.com/pdf/JASIST_Proceedings.pdf

LOCKSS - Lots of Copies Keep Stuff Safe (service for long-term preservation of digital content) http://www.lockss.org

Lossau, Norbert (2012): An Overview of Research Infrastructures in Europe — and Recommendations to LIBER. In: LIBER Quarterly - The Journal of the Association of European Research Libraries, 21(3-4), http://liber.library.uu.nl/index.php/lq/article/view/8028/8386

Lourdi I., Papatheodorou C. & Doerr M. (2009): Semantic Integration of Collection Description: Combining CIDOC/CRM and Dublin Core Collections Application Profile. In: D-Lib, 15(7-8), http://www.dlib.org/dlib/july09/papatheodorou/07papatheodorou.html

Lozny, Ludomir R. (2011): Archaeology in the age of globalization: local meanings, global interest, pp. 21-49, in: Lozny, L.R. (ed): Comparative archaeologies: a sociological view of the science of the past. New York: Springer

Lozny, Ludomir R. (ed., 2011): Comparative archaeologies: a sociological view of the science of the past. New York: Springer

LTER Long Term Ecological Research Europe, http://www.lter-europe.net

LTER Long Term Ecological Research Europe: Ecological Research Site and Dataset Registry (DEIMS), http://data.lter-europe.net/deims/

Luckman, S. (2004): More than the sum of its parts: The humanities and communicating the "hidden work" of research, pp. 82–90, in: Kenway J., Bullen E. & Robb S. (eds.): Innovation and Tradition. The Arts, Humanities, and the Knowledge Economy. New York: Peter Lang

Luzi D., Ruggieri R., Biagioni S. & Schiano E. (2013): Data sharing in environmental sciences: A survey of CNR researchers. In: International Journal of Grey Literature, 9: 69-81, http://www.opengrey.eu/item/display/10068/700182

Lynch, Clifford (2013): The Next Generation of Challenges in the Curation of Scholarly Data, pp. 395-408, in: Ray, Joyce M. (ed.): Research Data Management: Practical Strategies for Information Professionals. Purdue University Press, http://www.cni.org/wp-content/uploads/2013/10/Research-Data-Mgt-Ch19-Lynch-Oct-29-2013.pdf

Lyon L., Rusbridge C., Neilson C. & Whyte A. (2010): Disciplinary approaches to sharing, curation, reuse and preservation. DCC-SCARP JISC Final Report, March 2010, http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf

Lyon, Liz (2009): Open Science at Web-Scale: Optimising Participation and Predictive Potential. Consultative Report. Joint Information Systems Committee (JISC), UK, UKOLN, Version V1.0, November 2009, http://opus.bath.ac.uk/30056/1/open-science-report-6nov09-final-sentojisc.pdf

Lyon, Liz (2012): The informatics transform: Re-engineering libraries of the data decade. In: International Journal of Digital Curation, 7(1): 126-138, http://ijdc.net/index.php/ijdc/article/view/210

MacColl J. & Jubb M. (2011): Supporting Research: Environments, Administration and Libraries. Dublin (USA): OCLC Research. http://www.oclc.org/research/publications/library/2011/2011-10.pdf

MacEachern, Scott (2013): Genetics and archaeology, pp.65-76, in: Mitchell P. & Lane J.L. (eds.). The Oxford Handbook of African Archaeology. Oxford University Press; available at http://www.researchgate.net/publication/258849996_Genetics_and_archaeology

Madella M., Alexandre A. & Ball T. (2005): International Code for Phytolith Nomenclature 1.0. In: Annals of Botany 96(2): 253-60, http://aob.oxfordjournals.org/content/96/2/253.full.pdf

Madsen, Torsten (2004): Classification and archaeological knowledge bases, pp. 35-42, in: Lange, Guus (ed.), Reference Collections. Foundation for Future Archaeology. ROB, Amersfoort, Netherlands, http://www.cultureelerfgoed.nl/sites/default/files/publications/reference-collections-foundation_for_future_archaeology.pdf

Making Data Count – Data Level Metrics Project of DataONE, PLOS and UC3 - University of California Curation Center (USA, NSF-funded, 10/2014-09/2015), http://articlemetrics.github.io/MDC/

Making Data Count project (2015): Unit 1 Report, http://files.figshare.com/1939585/Make_Data_Count_Unit_1_Final_Report.pdf

Malaverri J.E.G., Santanchè A. & Medeiros C.B. (2014): A provenance-based approach to evaluate data quality in eScience. In: International Journal of Metadata, Semantics and Ontologies, 9(1): 15–28, http://www.inderscience.com/info/inarticle.php?artid=59127

Manghi P., Bolikowski L., Houssos N. & Schirrwagen J. (2014): Report on the First Workshop on Linking and Contextualizing Publications and Datasets. In: SIGMOD Record, 43(2): 57-60, http://www.sigmod.org/publications/sigmod-record/1406/pdfs/11.reports.Manghi.pdf

Manning, Sturt W. (2015): Radiocarbon dating in Archaeology. History, progress and present status, pp. 128-158, in: Chapman R. & Wylie A. (eds.): Material Evidence. Learning from archaeological practice. London & New York: Routledge

MANTRA - Research Data Management Training (University of Edinburgh, UK), http://datalib.edina.ac.uk/mantra/

Manyika J., Chui M., Farrell D. *et al*. (2013): Open data: Unlocking innovation and performance with liquid information. McKinsey, October 2013, http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information

MAPPA Open Data (data repository at the University of Pisa, Italy), http://mappaproject.arch.unipi.it/?lang=en

Mapping Death - Boundaries and Territories in Ireland, 1st to 8th centuries AD (Irish National Strategic Archaeological Research (INSTAR) project, 2008-2010, http://www.mappingdeath.ie

Marchant R., Cleef A., Harrison S.P. *et al*. (2009): Pollen-based biome reconstructions for Latin America at 0, 6000 and 18,000 radiocarbon years ago. In: Climate of the Past, 5: 725-767, http://www.clim-past.net/5/725/2009/cp-5-725-2009.pdf

Mařík, Jan (2013): Amateurs and professional archaeologists: Legal models for their cooperation in the Czech Republic, pp. 105-108, in: Lagerlöf, Agneta (ed.): Who cares? Perspectives on Public Awareness, Participation and Protection in Archaeological Heritage Management. EAC Occasional Paper 8, Budapest; http://www.arup.cas.cz/wp-content/uploads/2010/05/detektory_EAC.pdf

Marker, Hans J. (2014): Report from the DASISH SSH Workshop, Gothenburg, 4–5th October 2013, pp. 207-215, in: Duşa A., Nelle D., Stock G. & Wagner G. (eds., 2014): Facing the Future: European Research Infrastructures for the Humanities and Social Sciences. Berlin: Scivero, online: http://www.allea.org/Content/ALLEA/2014_06_04-FACING_THE_FUTURE.pdf

Markey K., Young Rieh S., St. Jean B. *et al*. (2007): Census of Institutional Repositories in the United States: MIRACLE Project Research Findings. CLIR report 140, http://www.clir.org/pubs/abstract/reports/pub140

Markl V., Hoeren T., Krcmar H. *et al*. (2013): Innovationspotenzialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen (Big Data Management). Im Auftrag des Bundesministeriums für Wirtschaft und Technologie; TU Berlin, TU München & Westfälische Wilhelms-Universität Münster; November 2013, http://www.dima.tu-berlin.de/fileadmin/fg131/Publikation/BDM_Studie/StudieBiDaMa-online-v2.pdf

Marshall, Yvonne (2002): What is community archaeology? In: World Archaeology, 34(2): 211-219, http://faculty.ksu.edu.sa/archaeology/Publications/General/What%20is%20community%20archaeology.pdf

Marston J.M., d'Alpoim Guedes J. & Warinner C. (eds.): Method and Theory in Paleoethnobotany. University Press of Colorado

Marwick B., Fairbairn A. & O'Connor S. (2009): New Directions in Archaeological Science. ANU Press (terra australis 28), open access http://www.doabooks.org/doab?func=fulltext&rid=14949

Matthew, Alison (2015): Free mapping data will elevate flood risk knowledge. In: UK Environment Agency weblog, 16 June 2015, https://environmentagency.blog.gov.uk/2015/06/16/free-mapping-data-will-elevate-flood-risk-knowledge/

Mauthner N. & Parry O. (2013): Open access digital data sharing: principles, policies, and practices. In: Social Epistemology, 27(1): 47-67, http://www.tandfonline.com/doi/full/10.1080/02691728.2012.760663

Mayernik M.S. *et al*., (2007) Adding context to content: The CENS deployment center. UCLA Graduate School of Education & Information, Department of Information Studies, http://polaris.gseis.ucla.edu/cborgman/pubs/asist_2007_mwbp_fin.pdf

Mayernik M.S., Callaghan S., Leigh R., Tedds J. & Worley S. (2014): Peer Review of Datasets: When, Why, and How. In: Bulletin of the American Meteorological Society, February 2014, e-View, http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-13-00083.1

Mazel A., Galani A., Maxwell D. & Sharpe K. (2012): 'I want to be provoked': public involvement in the development of the Northumberland Rock Art on Mobile Phones project. In: World Archaeology, 44(4): 592-611, http://www.tandfonline.com/toc/rwar20/44/4#.VSen25Mufh

McCabe M.J. & Snyder C.M. (2014): Identifying the effect of Open Access on citations using a panel of science journals. In: Economic Inquiry, 52(4): 1284–1300, http://onlinelibrary.wiley.com/doi/10.1111/ecin.12064/abstract

McGann, James G. (2015): 2014 Global Go To Think Tanks Index Report. Think Tanks and Civil Societies Program, University of Pennsylvania, Philadelphia, 3-1-2015, http://repository.upenn.edu/cgi/viewcontent.cgi?article=1008&context=think_tanks

McKeague P., Corns A. & Shaw R. (2012): Developing a Spatial Data Infrastructure for Archaeological and Built Heritage. In: International Journal of Spatial Data Infrastructures Research, Volume 7: 38-65, http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/239

Mediterranean Archaeology and Archaeometry (open access journal), http://www.maajournal.com

Mehler, Natascha (2013): Breaking new Ground: Historical Archaeology in Central Europe, pp. 11-29, in: Historical Archaeology in Central Europe. Rockville: Society for Historical Archaeology, Special Publication 10, https://histarch.univie.ac.at/fileadmin/user_upload/ag_histarch/PDF-Dateien/Mehler/SHA_H-A_Central_Europe_Mehler.pdf

Meier zu Verl C. & Horstmann W. (eds., 2011): Studies on subject-specific requirements for open access infrastructure. Bielefeld: Universitätsbibliothek Bielefeld, https://pub.uni-bielefeld.de/luur/download?func=downloadFile&recordOId=2445229&fileOId=2479864

Meloni, Maurizio (2014): How biology became social, and what it means for social theory. In: The Sociological Review, 62(3): 593-614, http://onlinelibrary.wiley.com/doi/10.1111/1467-954X.12151/abstract

Mendeley (academic information sharing and networking platform), http://www.mendeley.com

MERIL - Mapping of the European Research Infrastructure Landscape (inventory of Research Infrastructures of all types and across all scientific domains), http://portal.meril.eu

METS - Metadata Encoding and Transmission Standard, http://www.loc.gov/standards/mets/

Meyer, W.J., Crumley, C.L. (2011): Historical ecology: Using what works to cross the divide, pp. 109-134, in: Moore T. & Armada L. (eds.): Atlantic Europe in the First Millennium BC: Crossing the Divide. Oxford University Press

MIBBI - Minimum Information for Biological and Biomedical Investigations https://www.biosharing.org/standards/

Michener W.K. & Jones M.B. (2012): Ecoinformatics: supporting ecology as a data-intensive science. In: Trends in Ecology & Evolution, 27(2): 85-93, http://www.cell.com/trends/ecology-evolution/pdf/S0169-5347%2811%2900339-9.pdf

MicroPasts project (University College London, Institute of Archaeology), http://micropasts.org

Microsoft Research (2006): Towards 2020 Science, http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/

MIDAS Heritage - The UK Historic Environment Data Standard. Version 1.1, October 2012, https://www.historicengland.org.uk/images-books/publications/midas-heritage/

Miles, Alistair (2007): Zoological Case Studies in Digital Curation – DCC SCARP / ImageStore. Weblog, 27 June 2007, https://alimanfoo.wordpress.com/category/sheer-curation/

Millerand F., Ribes D., Baker K.S. & Bowker G.C. (2012): Making an Issue out of a Standard [EML]: Storytelling Practices in a Scientific Community. In: Science, Technology, & Human Values, 38(1): 7-43, http://interoperability.ucsd.edu/docs/millerand12_STHV_storytelling.pdf

Milloy, Caren (2013): Innovative approaches to publishing open access monographs. JISC inform, Issue 47, http://www.jisc.ac.uk/inform/inform37/OpenAccessMonographs.html#.VV-db0YjHEQ

Mohammadi E. & Thelwall M. (2014): Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. 14th International Conference of the Society of Scientometrics and Informetrics Vienna, Austria, 15-19 July 2013, extended version of the paper, http://www.scit.wlv.ac.uk/~cm1993/papers/EhsanMendeleyAltmetrics.pdf

Möller, Erik (2007): Licenses/NC - The case for Free use: reasons not to use a Creative Commons -NC license (last modified on 9 December 2014). Freedomdefined.org, http://freedomdefined.org/Licenses/NC

Molloy J.C. (2011): The Open Knowledge Foundation: Open Data Means Better Science. In: PLoS Biology, 9(12): e1001195, http://dx.doi.org/10.1371/journal.pbio.1001195

Monastersky, Richard (2013). Publishing frontiers: The library reboot. In: Nature, 495 (7442), 430-432, http://www.nature.com/news/publishing-frontiers-the-library-reboot-1.12664

Mongeon P. & Paul A. (2014): The journal coverage of bibliometric databases: A comparison of Scopus and Web of Science. 77th ASIS&T annual meeting: SIGMET Workshop, Seattle, USA, 5 November 2014, http://www.asis.org/SIG/SIGMET/data/uploads/sigmet2014/mongeon.pdf, cf. http://de.slideshare.net/apaulhus/the-journal-coverage-of-bibliometric-databases-a-comparison-of-scopus-and-web-of-science)

Mons B., van Haagen H., Chichester C. *et al*. (2011): The value of data. In: Nature Genetics, 43: 281–283, http://www.nature.com/ng/journal/v43/n4/full/ng0411-281.html

Mooney H. & Newton M.P. (2012): The Anatomy of a Data Citation: Discovery, Reuse, and Credit. In: Journal of Librarianship and Scholarly Communication, 1(1): http://www.jlsc-pub.org/jlsc/vol1/iss1/6/

Moore R., Hardman C., Richards J. & Xia L. (2012): ADS easy: an Automated e-archiving System for Archaeology, pp. 299-306, in: Archaeology in the Digital Era. Papers from the CAA 2012, Southampton, 26-29 March 2012. Amsterdam University Press, http://dare.uva.nl/cgi/arno/show.cgi?fid=516092

Moore, Niamh (2007): (Re)Using Qualitative Data? In: Sociological Research Online, 12(3), http://www.socresonline.org.uk/12/3/1.html

Moore, Ray (2014): Accessioning is ADS-easy. Day of Archaeology weblog, 22 July 2014, http://www.dayofarchaeology.com/accessioning-is-ads-easy/

MoRe (Metadata & Object Repository) aggregator, http://more.dcu.gr

Morgan C. & Eve S. (2012): DIY and digital archaeology: what are you doing to participate? In: World Archaeology, 44(4): 521-537; manuscript: https://www.academia.edu/2365783/DIY_and_digital_archaeology_what_are_you_doing_to_participate

Morgan, Colleen L. (2012): Emancipatory Digital Archaeology. PhD diss., University of California, Berkeley, 2012, https://escholarship.org/uc/item/08s0v1fb#page-4 (or: https://www.academia.edu/2997156/Emancipatory_Digital_Archaeology)

Morin A., Urban J., Adams P. *et al*. (2012): Shining light into black boxes. In: Science 336(6078), 159–160, http://hkl.hms.harvard.edu/uploads/image/pdfs/sliz2012science.pdf

Morovati, Meredith (2015): What were the most downloaded data packages in 2014? In: DRYAD weblog, 16 January 2015, http://blog.datadryad.org/2015/01/16/what-were-the-most-downloaded-data-packages-in-2014/

Moshenska G. & Burtenshaw P. (2010): Response: the Values of Archaeology. In: Present Pasts, 1(1), http://www.presentpasts.info/articles/10.5334/pp.11/

Moshenska, Gabriel (2010): What is Public Archaeology? In: Present Pasts, 1(1), http://www.presentpasts.info/articles/10.5334/pp.7/

Moshenska, Gabriel (2013): Reflections on the 1943 'Conference on the Future of Archaeology'. In: Archaeology International [UCL], Volume 16: 128-139, http://www.ai-journal.com/articles/10.5334/ai.1606/

Moshenska, Gabriel (2013): Reflections on the 1943 'Conference on the Future of Archaeology'. In: Archaeology International, Volume 16: 128-139, http://www.ai-journal.com/articles/10.5334/ai.1606/

Moss, Elizabeth (2013): Viable Data Citation: Expanding the impact of social science research. Presentation at RDAP13 - Research Data Access & Preservation Summit, Panel on Data Citation and Altmetrics, Baltimore, USA, 5 April 2013, http://www.slideshare.net/asist_org/rdap13-moss

Motz C. & Carrier S.C. (2012): Paperless Recording at the Sangro Valley Project, pp. 25-30, in: Archaeology in the Digital Era: Papers from CAA 2012 conference, Southampton, 26-29 March 2012. Amsterdam University Press, http://dare.uva.nl/cgi/arno/show.cgi?fid=516092

Mounce, Ross (2013): Open access and altmetrics: distinct but complementary. In: ASIS&T Bulletin 39(4): 14-17, https://asis.org/Bulletin/Apr-13/Bulletin_AprMay13_Final.pdf

Mowers S., Humphrey C. & Perry C. (2013): Survey of Researchers' Needs and Practices regarding Research Data Management in Canada. Summary Report, 1 August 2013, http://gsg.uottawa.ca/data/open/aa-interim-survey-report/20130801-en.pdf

Mrozowski, S. A. (2009): Pulling the threads together: Issues of theory and practice in an archaeology of the modern world, in: Horning A. & Palmer M. (eds.): Crossing Paths or Sharing Tracks? Future directions in the archaeological study of post-1550 Britain and Ireland. Boydell Press, http://www.geos.ed.ac.uk/~nabo/meetings/glthec/groups/group_data/ecodynamics/CH26MROZOWSKI.pdf

Murray-Rust P., Neylon C., Pollock R. & Wilbanks J. (2010): Panton Principles - Principles for open data in science, 19 February 2010, http://pantonprinciples.org

Mutz R., Bornmann L. & Daniel H.-D. (2012): Types of research output profiles: A multilevel latent class analysis of the Austrian Science Fund's final project report data. In: Research Evaluation, 22(2): 118-133, http://rev.oxfordjournals.org/content/22/2/118.long

myExperiment (virtual environment for sharing of biological research workflows; UK, EPSRC, JISC and other grants, 2007/on-going), http://www.myexperiment.org

Naderi Beni A., Lahijani H., Mousavi Harami R. *et al*. (2013): Caspian sea-level changes during the last millennium: historical and geological evidence from the south Caspian Sea. In: Climate of the Past, 9: 1645–65, http://www.clim-past.net/9/1645/2013/cp-9-1645-2013.pdf

Nanopub.org: What is a Nanopublication, http://nanopub.org/wordpress/?page_id=65

National Center for Ecological Analysis and Synthesis (NCEAS), United States, https://www.nceas.ucsb.edu

National Evolutionary Synthesis Center (NESCent), United States, http://www.nescent.org

National Research Centre on Human Evolution (CENIEH), Spain, http://www.cenieh.es

Nature Genetics, editorial (2014): Credit for code. Nature Genetics, 46(1), http://www.nature.com/ng/journal/v46/n1/full/ng.2869.html

Nature, editors (2009): Data's shameful neglect. In: Nature 461, 145, 10 September 2009, http://www.nature.com/nature/journal/v461/n7261/full/461145a.html

Nawrotzki K. & Dougherty J. (2013): Writing History in the Digital Age. Ann Arbor: University of Michigan Press, online publication, http://dx.doi.org/10.3998/dh.12230987.0001.001

Naylor J. & Richards J. (2005): Third party data for first class research. In: Archeologia e Calcolatori, Vol. 16: 83-91, http://soi.cnr.it/archcalc/indice/PDF16/NAYLOR83-92.pdf

NEARCH – New ways of Engaging audiences, Activating societal relations and Renewing practices in Cultural Heritage (EU, Culture Programme, 2013-2018), http://www.nearch.eu

Nederhof, A.J. (2006): Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A review. In: Scientometrics, 66(1): 81-100, http://link.springer.com/article/10.1007%2Fs11192-006-0007-2

NeDIMAH - Network for Digital Methods in the Arts and Humanities (European Science Foundation, Research Networking Programme, 05/2011-05/2015), http://www.nedimah.eu

NeDiMAH ICT methods ontology (NEMO), http://nemo.dcu.gr; see also: http://www.nedimah.eu/workgroups/development-ict-methods-taxonomy

Nelson M.C., Kintigh K., Abbott D.R. & Anderies J.M. (2010): The cross-scale interplay between social and biophysical context and the vulnerability of irrigation-dependent societies: Archaeology's long-term perspective. In: Ecology and Society, 15(3), http://ltvtp.shesc.asu.edu/Nelson%20et%20al%202010.pdf

Nelson, Bryn (2009): Empty archives. Most researchers agree that open access to data is the scientific ideal, so what is stopping it happening? 10 September 2009. In: Nature, Vol. 461, http://www.nature.com/news/2009/090909/full/461160a.html

NEMO - NeDiMAH ICT methods ontology, http://nemo.dcu.gr

NERC Natural Environment Research Council (UK): Data Centres, http://www.nerc.ac.uk/research/sites/data/

Neustupný, Evžen (1998): Mainstreams and Minorities in Archaeology. Archaeologia Polona, Volume 35-36: 13-23, http://www.iaepan.edu.pl/archaeologia-polona/object/1

Niccolucci F. & Richards J.D. (2013): ARIADNE: Advanced Research Infrastructures for Archaeological Dataset Networking in Europe. In: The European Archaeologist, Issue 39, Summer 2013, http://e-a-a.org/TEA/archive/TEA_39_SUMMER_2013/rep1_39.pdf

Niccolucci, Franco (2014): The ARIADNE approach to Digital Cultural Heritage, pp. 179-186, in: Duşa A., Nelle D., Stock G. & Wagner G. (eds., 2014): Facing the Future: European Research Infrastructures for the Humanities and Social Sciences. Berlin: Scivero, online: http://www.allea.org/Content/ALLEA/2014_06_04-FACING_THE_FUTURE.pdf

Niccolucci, Franco (2015): Digital archaeological archives of Europe: opportunities and challenges, pp. 111-118, in: European Archaeological Council (EAC): Setting the Agenda – Giving New Meaning to the European Archaeological Heritage. Edited by P.A.C. Schut, D. Scharff & L.C. de Wit. EAC Occasional Paper no. 10, Budapest: Archaeolingua

Nicholas D., Rowlands I., Watkinson A. *et al*. (2012): Digital repositories ten years on: what do scientific researchers think of them and how do they use them? In: Learned Publishing 25: 195-206, http://ciber-research.eu/download/20120620-Digital_repositories_ten_years_on.pdf

Nicholas D., Watkinson A., Jamali H.R. *et al*. (2015): Peer review: still king in the digital age. In: Learned Publishing 28(1): 15-21, http://ciber-research.eu/download/20140120-Peer_review-Learned_Publishing_2015.pdf

NISO - National Information Standards Organization (USA): Alternative Assessment Metrics (Altmetrics) Initiative (07/2013-), http://www.niso.org/topics/tl/altmetrics_initiative/

NISO ISQ (2013): Information Standards Quarterly (ISQ), 25(2), Summer 2013, Topic: Altmetrics, http://www.niso.org/publications/isq/2013/v25no2

Nomisma.org (collaborative project to provide numismatic concepts and datasets based on Linked Open Data principles), http://nomisma.org

Nowakowski P., Ciepiela E., Harężlak D. *et al.* (2011): The Collage Authoring Environment. In: Procedia Computer Science, 4: 608–617, http://www.sciencedirect.com/science/article/pii/S1877050911001220

Nowotny H., Scott P. & Gibbons M. (2001): Re-thinking Science: Knowledge and the Public in an Age of Uncertainty. Cambridge: Polity

Nowotny, Helga (2010): Out of science – out of sync?, pp. 319-322, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Nowviskie, Bethany (2011): Where Credit Is Due: Preconditions for the Evaluation of Collaborative Digital Scholarship. In: Profession (Modern Language Association), Vol. 2011: 169–181, http://www.mlajournals.org/doi/pdf/10.1632/prof.2011.2011.1.169

Nowviskie, Bethany (2014): Digital humanities in the Anthropocene. In: Nowviskie.org weblog, 10 July 2014, http://nowviskie.org/2014/anthropocene/

NSF - National Science Foundation – Blue-Ribbon Advisory Panel on Cyberinfrastructure (2003): Revolutionizing Science and Engineering Through Cyberinfrastructure, https://arizona.openrepository.com/arizona/bitstream/10150/106224/1/report.pdf

NSF - National Science Foundation, Cyberinfrastructure Council (2007): Cyberinfrastructure Vision for 21st Century Discovery. Arlington, VA., USA, March 2007, http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

NSF - National Science Foundation, National Science Board (2005): Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. Washington/DC: NSF, http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf

O'Brien, Karen (2010): Responding to the global environmental change: social sciences of the world unite!, pp. 11-12, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

O'Neill B. & Shankar K. (2015): Hurling Against a Haystack: The Incentives and Challenges of Open Data in the Republic of Ireland. iConference 2015, Newport Beach, Calif., 24-27 March 2015, https://www.ideals.illinois.edu/bitstream/handle/2142/73437/155_ready.pdf?sequence=2

OAGR - Online Ancient Genome Repository (Australia), https://www.oagr.org.au

OAI-ORE - Open Archives Initiative - Object Reuse and Exchange (ORE), defines a RDF-based model to represent aggregations of resources as enumerations),http://www.openarchives.org/ore/

OAI-PMH - Open Archives Initiative - Protocol for Metadata Harvesting, http://www.openarchives.org/pmh/

OAPEN - Open Access Publishing in European Networks project (EU, eContentplus, 09/2008-02/2011), http://project.oapen.org

Ochsner M., Hug S.E. & Daniel H.-D. (2012): Four Types of Research in the Humanities: Setting the Stage for Research Quality Criteria in the Humanities. In: Research Evaluation, 22/2: 79-92, http://rev.oxfordjournals.org/content/early/2012/11/30/reseval.rvs039

Odyssee programme (projects on unpublished archaeological investigations, 2009-2013, funded by the Netherlands Organisation for Scientific Research), http://www.nwo.nl/onderzoek-en-resultaten/programmas/odyssee

OECD (2004): Declaration on Access to Research Data from Public Funding. Paris: OECD, http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157&Lang=en&Book=False

OECD (2007): Principles and Guidelines for Access to Research Data from Public Funding. Paris: OECD, http://www.oecd.org/dataoecd/9/61/38500813.pdf

OECD (2010): Establishing Large International Research Infrastructures: Issues and Options. December 2010, http://www.oecd.org/sti/scienceandtechnologypolicy/47027330.pdf

OECD (2015): Making Open Science a Reality. OECD, https://www.innovationpolicyplatform.org/sites/default/files/DSTI-STP-TIP%282014%299-REV2_0_0_0_0.pdf

OECD Global Science Forum (2008): Second Activity on Policy Issues Related to Scientific Research Collections. Final Report on Findings and Recommendations, Washington, 17 July 2008, http://www.oecd.org/science/sci-tech/42237442.pdf

Office of Science and Innovation (OSI), e-Infrastructure Working Group (2010): Developing the UK's e-infrastructure for science and innovation, http://www.nesc.ac.uk/documents/OSI/report.pdf

Old Weather (UK National Maritime Museum; invites non-experts to transcribe ship logbooks for historical research), http://www.oldweather.org

Oldman D., Doerr M., de Jong G., Norton B. & Wikman T. (2014): Realizing Lessons of the Last 20 Years: A Manifesto for Data Provisioning & Aggregation Services for the Digital Humanities (A Position Paper). In: D-Lib Magazine, 20(7/8), http://www.dlib.org/dlib/july14/oldman/07oldman.html

Oldman, Dominic (2012): The British Museum, CIDOC CRM and the Shaping of Knowledge. Dominic Oldman weblog, 4 September 2012, http://www.oldman.me.uk/blog/the-british-museum-cidoc-crm-and-the-shaping-of-knowledge

Oomen J. & Aroyo L. (2011): Crowdsourcing in the cultural heritage domain: opportunities and challenges, pp. 138-149, in: Proceedings of the 5th International Conference on Communities and Technologies (C&T'11), QUT, Brisbane, Australia, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.6933&rep=rep1&type=pdf

Open Access Archaeology (helps researchers discover open access resources in the field of Archaeology), http://www.openaccessarchaeology.org

Open Archaeology (journal, since 2014), http://www.degruyter.com/view/j/opar

OAI - Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH), http://www.openarchives.org/pmh/

Open Book Publishers (publish open access monographs and other books, since 2008), http://www.openbookpublishers.com

Open Citation (OpCit) project: Reference Linking and Citation Analysis for Open Archives. The effect of open access and downloads ('hits') on citation impact: a bibliography of studies [2004-2013], http://opcit.eprints.org/oacitation-biblio.html

Open Context (2009): Sharing Archaeological Data: Part 3 – What is your Dream Tool? In: Heritage Bytes weblog, 29 April 2009, http://ux.opencontext.org/blog/2009/04/29/sharing-archaeological-data-part-3-%E2%80%93-what-is-your-dream-tool/

Open Context (2011): Guidelines for Web-based Publication in Archaeology. Prepared by E. Kansa & S. Whitcher-Kansa, January 2011, http://ux.opencontext.org/blog/wp-content/uploads/2011/05/Guidelines_Jan2011.rtf

Open Context (archaeological data publication platform, Alexandria Archive Institute, USA), http://opencontext.org

Open Data Barometer (2013): 2013 Global Report. Open Data Institute and WWW Foundation, Report author: Tim Davies, 31 October 2013, http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf

Open Data Commons (ODC) licenses, http://opendatacommons.org/licenses/

Open Data Institute & Thomson Reuters (2014): Creating Value with Identifiers in an Open Data World, October 2014, http://site.thomsonreuters.com/site/data-identifiers/

Open Edition (France, humanities and social sciences information and publication portal / platforms: Calenda, Hypotheses, OpenEdition Books, Revues.org), http://www.openedition.org

Open Knowledge Foundation: Open Definition, Version 2.0, http://opendefinition.org/od/

Open Plaque (community project documenting commemorative plaques in the UK), http://openplaques.org

Open Pompei (project on transparency and open data in archaeological research), http://www.openpompei.it

OpenAIRE - Open Access Infrastructure for Research in Europe (12/2009-07/2018), http://www.openaire.eu

OpenAIRE (2014): OpenAIRE2020: A new horizon for open science, 2 December 2014, https://www.openaire.eu/news-events/openaire2020-press-release

OpenAIRE: What is the Open Research Data Pilot?, https://www.openaire.eu/h2020-oa-data-pilot

OpenDOAR - Directory of Open Access Repositories, http://www.opendoar.org

OpenMinTeD - Open Mining Infrastructure for Text & Data (EU, Horizon 2020, 06/2015-05/2018), http://openminted.eu

Opificio delle Pietre Dure (Florence, Italy), http://www.opificiodellepietredure.it

Opitz R. & Cowley D. (eds., 2012): Interpreting Archaeological Topography – Airborne Laser Scanning, 3D Data and Ground Observation. Occasional Publications of the Aerial Archaeology Research Group No. 5. Oxford: Oxbow Books (http://www.oxbowbooks.com/oxbow/interpreting-archaeological-topography.html)

Opportunities for Data Exchange / Kotarski R. *et al*. (2012): Report on best practices for citability of data and on evolving roles in scholarly communication. ODE project, 21 June 2012, http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode/outputs/

Opportunities for Data Exchange / Reilly S. *et al*. (2011): Report on integration of data and publications. ODE project, 17 October 2011, http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode/outputs/

ORCID (unique identifiers for researchers), http://orcid.org

OREA - Institute for Oriental and European Archaeology, Austrian Academy of Sciences, http://www.orea.oeaw.ac.at

Orlandi S., Santucci R., Casarosa V. & Liuzzo P.M. (2014): Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the first EAGLE International Conference. Rome: Sapienza Università Editrice, http://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf

Ortman S.G., Cabaniss A.F.H., Sturm J.O. & Bettencourt L.M.A. (2014): The Pre-History of Urban Scaling. In: PLoS ONE, 9(2): e87902, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0087902

Orton C. & Hughes M. (2013): Pottery in Archaeology, 2nd edition. Cambridge Manuals in Archaeology. Cambridge UP

Overbeck Laise, Kristen (2007): The Heritage Health Index [2004] Findings on Digital Collections. In: First Monday, 12(7), http://firstmonday.org/issues/issue12_7/laise/index.html

Owens, Carol (2009): When metadata worlds collide: The hunter/gatherer dichotomy. In: Journal of Digital Asset Management 5, 181-184, http://www.palgrave-journals.com/dam/journal/v5/n4/full/dam200914a.html

Oxford Archaeology Digital, http://oadigital.net

Page, Roderic (2015): Nanopublications and annotation: a role for the Biodiversity Data Journal? In: iPhylo weblog, 27 July 2015, http://iphylo.blogspot.co.at/2015/07/nanopublications-and-annotation-role.html

Palaiologk A., Economides A., Tjalsma H. & Sesink, L. (2012): An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS. In: International Journal on Digital Libraries, 12(4): 195-214, http://link.springer.com/article/10.1007/s00799-012-0092-1

Palaiologk A., Tjalsma H. & Sesink L. (2010): Costs of Digital Archiving: the case of DANS. Presentation at IASSIST 2010, Cornell University, Ithaca, USA, 2 June 2010, http://www.iassistdata.org/downloads/2010/2010_c3_sesink_etal.pdf

Palaiologk, Anna (2013): Cost and value analysis of digital data archiving. Presentation at the eFiscal final workshop, Amsterdam, 28 January 2013, http://efiscal.eu/files/presentations/amsterdam/DANS_AnnaPalaiologk_efiscal-final-workshop.pdf

Paleobiology Database, http://paleobiodb.org

Paleobot.org (collaborative web resource for archaeobotanical research), http://www.paleobot.org

Palmer C.L., Teffeau L.C. & Pirmann C.M. (2009): Scholarly information practices in the online environment. Dublin, OH: OCLC, http://www.oclc.org/research/activities/past/workagenda/researchinfo/support/practices.htm

Palmer C.L., Teffeau L.C. & Pirmann C.M. (2009): Scholarly information practices in the online environment. Dublin (USA): OCLC, http://www.oclc.org/research/activities/past/workagenda/researchinfo/support/practices.htm

Palmer C.L., Weber N.M., Muñoz T. & Renear A. (2013): Foundations of data curation: The pedagogy and practice of purposeful work with data. In: Archive Journal, 3, http://www.archivejournal.net/issue/3/archives-remixed/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/

Palmer C.L., Weber N.W. & Cragin M.H. (2011): The Analytic Potential of Scientific Data: Understanding Re-Use Value. In: Proceedings of ASIST 2011, Annual Meeting, New Orleans, http://www.asis.org/asist2011/proceedings/submissions/174_FINAL_SUBMISSION.pdf

Paludan-Müller, Carsten (2009): Europe - A constrained and fragmented space on the edge of the continental landmasses. Crossroad, battlefield and melting pot, pp. 75-83, in: Heritage and Beyond. Council of Europe, Strasbourg, https://book.coe.int/eur/en/cultural-heritage/4256-heritage-and-beyond.html

Paludan-Müller, Carsten (2013a): Caring about the past requires care for the present, pp. 87- 91, in: Lagerlöf, Agneta (ed.): Who cares? Perspectives on Public Awareness, Participation and Protection in Archaeological Heritage Management. EAC Occasional Paper No. 8. Namur, http://european-archaeological-council.org/publications

Paludan-Müller, Carsten (2013b): On rivers, mountains, seas and ideas – or What vast spaces and long lines mean to culture and history, pp. 617-622, in: Bergerbrant S. & Sabatini S. (eds.): Counterpoint: Essays in Archaeology and Heritage Studies in Honour of Professor Kristian Kristiansen. BAR Int. Se. 2508. Oxford: Archaeopress

Paludan-Müller, Carsten (2015): Opinion Editorial: The European Association of Archaeologists. Some thoughts for the future of our Association. In: The European Archaeologist, Issue 43: 5-9, http://e-a-a.org/TEA/TEA43.pdf

PaNdata Open Data Infrastructure (EU, FP7, 11/2011-10/2014; federation of user and data management systems of European photon and neutron facilities), http://pan-data.eu

PANGAEA - Data Publisher for Earth & Environmental Science (hosted by the Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, Germany), http://www.pangaea.de

Park J. & Tosaka Y. (2010): Metadata Creation Practices in Digital Repositories and Collections: Schemata, Selection Criteria, and Interoperability, pp. 104-116, in: Information Technologies and Libraries, September 2010, https://ejournals.bc.edu/ojs/index.php/ital/article/download/3136/2750

Parker B. & Cobb P. (eds., 2012): Kenan Tepe: Digital Data and Media (Released 2012-06-30). Open Context, http://opencontext.org/projects/3DE4CD9C-259E-4C14-9B03-8B10454BA66E

PARSE.Insight (2009): Insight into digital preservation of research output in Europe. Project deliverable D3.4: Survey Report, 9 December 2009, http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf

Parsons M.A. & Fox P.A. (2013): Is data publication the right metaphor? In: Data Science Journal, Vol.12; authors' manuscript, http://semanticommunity.info/@api/deki/files/29002/12_WDS-042.pdf

Parsons M.A., Duerr R. & Minster J.-B. (2010): Data citation and peer-review. Eos - Transactions American Geophysical Union, 91(34): 297-298, http://modb.oce.ulg.ac.be/wiki/upload/Alex/EOS_data_citation.pdf

PARTHENOS - Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies (EU, H2020, 05/2015-04/2019), http://www.parthenos-project.eu

PARTHENOS (2015): It's a start. Voices and impressions from the PARTHENOS kick-off meeting, PARTHENOS website, http://www.parthenos-project.eu/kick-off-article-tbu/

PASTEUR4OA - Open Access Policy Alignment Strategies for European Union Research project (EU, FP7, Science in Society), http://www.pasteur4oa.eu

PASTEUR4OA (2015): Report on policy recording exercise, including policy typology and effectiveness and list of further policymaker targets. Project deliverable 3.1, March 2015, http://www.pasteur4oa.eu/sites/pasteur4oa/files/deliverables/PASTEUR4OA%20Work%20Package%203%20Report%20final%2010%20March%202015.pdf

Pattinson, Damian (2012): PLOS Article-Level Metrics. http://septentrio.uit.no/index.php/SCS/article/viewFile/3045/2930

Peer L., Green A. & Stephenson E. (2014): Committing to Data Quality Review. In: International Journal of Digital Curation, 9(1): 263-291, http://www.ijdc.net/index.php/ijdc/article/view/9.1.263/358

Pegasus (scientific workflow management system), https://pegasus.isi.edu/applications

Pelagios: Enable Linked Ancient Geodata In Open Systems (UK, JISC-funded project), http://pelagios-project.blogspot.com

Pepe A., Goodman A., Muench A., Crosas M. & Erdmann C. (2014): How do astronomers share data? Reliability and persistence of datasets linked in AAS publications and a qualitative study of data practices among US astronomers. In: PLoS ONE 9(8): e104798, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104798

Perez-Sanz A., Li G., González-Sampériz P. & Harrison S.P. (2014): Evaluation of modern and mid-Holocene seasonal precipitation of the Mediterranean and northern Africa in the CMIP5 simulations. In: Climate of the Past, 10: 551.568, http://www.clim-past.net/10/551/2014/cp-10-551-2014.pdf

Peric B., Ochsner M., Hug S.E. & Daniel H.-D. (2013): Arts and Humanities Research Assessment Bibliography (AHRABi), http://www.psh.ethz.ch/crus/bibliography/ahrabi.pdf

PeriodO - Periods, Organized, http://perio.do

Perrin, T. M. (2012): The perceived contradiction between archaeology as an art and archaeology as a science: finding common ground. University of Leicester, 25 August 2012, https://www.academia.edu/5137592/Archaeology_as_an_Art_or_Archaeology_as_a_Science_Finding_Common_Ground

Perry S. & Beale N. (2015): The Social Web and Archaeology's Restructuring: Impact, Exploitation, Disciplinary Change. In: Open Archaeology, 1: 153.165, http://www.degruyter.com/view/j/opar.2014.1.issue-1/opar-2015-0009/opar-2015-0009.xml

Peter Kraker (2013): Visualizing Research Fields based on Scholarly Communication on the Web. Doctoral Thesis, University of Graz, Austria, December 2013, http://unipub.uni-graz.at/download/pdf/233911

Peters I., Kraker P., Lex E., Gumpenberger C. & Gorraiz J. (2015): Research Data Explored: Citations versus Altmetrics. In: arXiv:1501.03342 [cs.DL], 14 January 2015, http://arxiv.org/ftp/arxiv/papers/1501/1501.03342.pdf

Phelps C., Heidl R. & Wadhwa A. (2012): Knowledge, networks, and knowledge networks a review and research agenda. In: Journal of Management, 38(4): 1115-66, http://mackcenter.wpengine.netdna-cdn.com/wp-content/uploads/2013/07/Phelps-Heidl-Wadhwa_Knowledge-Networks-and-Knowledge-Networks.pdf

Pienta A., Alter G. & Lyle J. (2011): The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data. Proceedings of iPRES-2011, Singapore, 1-4 November 2011, https://fedora.phaidra.univie.ac.at/fedora/get/o:294251/bdef:Content/get

Pienta A., Gutmann M., Hoelter L., Lyle J. & Donakowski D. (2008): The LEADS database at ICPSR: Identifying important "at risk" social data. Roundtable paper presented at the American Sociological Association Annual Meeting 2008, Boston, MA, https://www.academia.edu/11149655/The_LEADS_Database_at_ICPSR_Identifying_Important_At_Risk_Social_Science_Data1

Pinfield S., Salter J., Bath P. *et al*. (2014): Open-access repositories worldwide, 2005-2012: Past growth, current characteristics and future possibilities. In: Journal of the American Society for Information Science and Technology, http://eprints.whiterose.ac.uk/76839/15/wrro_76839.pdf

Piscitelli, Matthew (2013): Using social networks to fundraise. Crowdfunding for the archaeologist, pp. 36-39, in: The SAA Archaeological Record, 13(4), September 2013, http://onlinedigeditions.com/publication/repo26/16146/175171/66647a9807469ad281a26db155f4b85cc6271664.pdf#page=17

Piwowar H., Vision T.J. & Whitlock M.C. (2011): Data archiving is a good investment; full text and details for the letter in Nature 473 (7347): 285, https://researchremix.wordpress.com/2011/05/19/nature-letter/

Piwowar H.A. & Chapman W.W. (2010): Public sharing of research datasets: a pilot study of associations. In: Journal of Informetrics, 4(2): 148.156, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3039489/

Piwowar H.A. & Vision T.J. (2013): Data reuse and the open data citation advantage. In: PeerJ, 1:e175, http://dx.doi.org/10.7717/peerj.175

Piwowar H.A., Day R.S. & Fridsma D.B. (2007): Sharing detailed research data is associated with increased citation rate. In: PLoS ONE, 2(3):e308, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308

Piwowar, Heather (2011): Who shares? Who doesn't? Factors associated with openly archiving raw research. In: PLOS ONE, 6(7): e18657, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0018657

Piwowar, Heather (2012): Making data Count: tracking data impact through the scholarly literature and beyond. Published in Github, 27 July 2012, https://github.com/tjv/datacitationstudy/blob/master/citationpractices_knit_.md

Piwowar, Heather (2013): Introduction: Special section: Altmetrics: What, Why and Where?. In: ASIS&T Bulletin, 39(4): 8-9, https://asis.org/Bulletin/Apr-13/Bulletin_AprMay13_Final.pdf

Plale B., Kouper I., Seiffert K. & Konkiel S. (2013): Repository of NSF-funded Publications and Related Datasets: “Back of Envelope” Cost Estimate for 15 years. Indiana University, March 2013, http://eprints.rclis.org/20072/4/Plale-2013-NSF-repository-estimate.pdf

Pleiades: a community-built gazetteer and graph of ancient places (hosted by the Institute for the Study of the Ancient World, New York University), http://pleiades.stoa.org

Plieninger T. & Bieling C. (eds., 2012): Resilience and the Cultural Landscape: Understanding and Managing Change in Human-Shaped Environments. Cambridge University Press

PlumX (2015): Plum Analytics: PlumX’s Current List of Metrics (the most comprehensive list of altmetrics indicators), http://www.plumanalytics.com/metrics.html

PlumX / Plum Analytics (altmetrics service provider), http://www.plumanalytics.com

Pollard A.M. & Bray P. (2007): A Bicycle Made for Two? The Integration of Scientific Techniques into Archaeological Interpretation. In: Annual Review of Anthropology, 36: 245-259, http://www.annualreviews.org/doi/full/10.1146/annurev.anthro.36.081406.094354

Pompeii Archaeological Research Project: Porta Stabia (PARP:PS, Italy), http://classics.uc.edu/pompeii/

Pontille D. & Torny D. (2010): The controversial policies of journal ratings: evaluating social sciences and humanities. In: Research Evaluation, 19(5): 347-360, https://halshs.archives-ouvertes.fr/halshs-00568746v2/document

Poole, Alex (2013): “Curate Thyself” and the DigCCurr Experts' Meeting: Communication, Collaboration, and Strategy in Digital Curation Education. In: D-Lib Magazine, 19(5/6), http://www.dlib.org/dlib/may13/poole/05poole.html

Poole, Alex (2014): How Has Your Science Data Grown? Digital Curation and the Human Factor: a Critical Literature Review. In: Archival Science, 15(2): 101-139; manuscript: https://alexhpoole.files.wordpress.com/2014/01/archival-science-how-has-your-science-data-grown-preprint1.pdf

Portable Antiquities Scheme (encourages the voluntary online recording of archaeological objects found by members of the public in England and Wales), http://finds.org.uk

Portal to the Past (online archive of 18 archaeological projects undertaken under the aegis of the Canadian Institute in Greece), http://www.portal.cig-icg.gr

Porter, Dot (2013): Medievalists and the Scholarly Digital Edition. In: Scholarly Editing - The Annual of the Association for Documentary Editing, Vol. 34, http://www.scholarlyediting.org/2013/essays/essay.porter.html

Portico (service for long-term preservation of digital content), http://www.portico.org

Poynder, Richard (2014): Interview with Kathleen Shearer, Executive Director of the Confederation of Open Access Repositories. In: Open and Shut? Richard Poynder weblog, 4 May 2014, http://poynder.blogspot.co.at/2014/05/interview-with-kathleen-shearer.html

PRACE - Partnership for Advanced Computing in Europe, http://www.prace-ri.eu

Prehistoric Ceramics Research Group (2011): The Study of Prehistoric Pottery: General Policies and Guidelines for Analysis and Publications. Occasional Paper No 1 and No 2. - 3rd Edition Revised 2011, http://www.pcrg.org.uk/Publications1-2.htm

PREPARDE - Peer Review for Publication & Accreditation of Research Data in the Earth Sciences (UK JISC funded international project, 2012-2013), http://www2.le.ac.uk/projects/preparde

Press, Gill (2015): Gartner: What's the most hyped technology? IoT, not big data. In: Tech Page One, 27 August 2015, http://www.techpageone.co.uk/technology-uk-en/gartner-whats-hyped-technology-iot-big-data/#.Vi1iKisWnEQ

Priem J. & Costello K.L. (2010): How and why scholars cite on Twitter. In: Proceedings of the American Society for Information Science and Technology, 47(1): 1-4, http://jasonpriem.org/self-archived/Priem_Costello_Twitter.pdf

Priem J. & Hemminger B.H. (2010): Scientometrics 2.0: New metrics of scholarly impact on the social Web. In: First Monday, 15(7), http://firstmonday.org/ojs/index.php/fm/article/view/2874

Priem J., Piwowar H. & Hemminger B.M. (2012): Altmetrics in the Wild: Using Social Media to Explore Scholarly Impact, http://arxiv.org/html/1203.4745v1 (article summary: http://altmetrics.org/altmetrics12/priem/)

Pritchard S.M., Carver L. & Anand S. (2005): Collaboration for knowledge management and campus informatics. EDUCAUSE – Centre for Applied Research. Research Bulletin, 2005, Issue 2, http://net.educause.edu/ir/library/pdf/ERB0502.pdf

Procter R., Voss A., Asgari-Targhi M. (2013): Fostering the human infrastructure of e-research. In: Information, Communication and Society, 16(10): 1668-91, http://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.715667?journalCode=rics20#.VbzcN_kjE2A

Procter R., Williams R., Stewart J. *et al*. (2010a): If you build it, will they come? How researchers perceive and use web 2.0. London: Research Network Information, http://wrap.warwick.ac.uk/56246/

Procter R., Williams R., Stewart S. *et al*. (2010b): Adoption and use of Web 2.0 in scholarly communications. In: Phil. Trans. R. Soc. A 2010, 368 : 4039-56, http://rsta.royalsocietypublishing.org/content/368/1926/4039.full.html#

Pryor G. & Donnelly M. (2009): Skilling up to do data: whose role, whose responsibility, whose career? In: International Journal of Digital Curation, 4(2): 158-170, http://ijdc.net/index.php/ijdc/article/view/126

Pryor, Graham (2009): Multi-scale Data Sharing in the Life Sciences: Some Lessons for Policy Makers. In: International Journal of Digital Curation, 4(3), http://ijdc.net/index.php/ijdc/article/view/135/178

Pryor, Graham (2012): DCC institutional engagements. Reflections on building capacity and capability for research data services. Presentation at the workshop on Institutional Repositories and Data - Roles and Responsibilities, Open Repositories 2012 conference, 9 July 2012, http://www.dcc.ac.uk/webfm_send/867

Puschmann C. & Burgess J. (2014): Metaphors of big data. In: International Journal of Communication, 8: 1690–1709, http://ijoc.org/index.php/ijoc/article/view/2169/1162

Pyla-Koutsopetria Archaeological Project (Cyprus), http://pylakoutsopetria.wordpress.com

Pyla-Koutsopetria Archaeological Project, data published through Open Context, http://opencontext.org/projects/3F6DCD13-A476-488E-ED10-47D25513FCB2

Qin J. & Li K. (2013): How Portable Are the Metadata Standards for Scientific Data? A Proposal for a Metadata Infrastructure, pp. 25-34, in: DC-2013 - Proc. International Conference on Dublin Core and Metadata Applications, Lisbon, http://dcevents.dublincore.org/IntConf/dc-2013/paper/viewFile/162/73

R4L - The Repository for the Laboratory (UK, JISC-funded project, completed in 2007), http://r4l.eprints.org

Rabinowitz, Adam (2014): It's about time: Historical Periodization and Linked Ancient World Data. ISAW Paper 7.22, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/rabinowitz/

Radiocarbon Context Database (University of Cologne, Institute for Prehistory; 2874 dates of 1274 Near Eastern sites), http://context-database.uni-koeln.de

Radiocarbon Palaeolithic Database Europe (Katholieke Universiteit Leuven, Center for Archaeological Sciences; about 13,000 dates), http://ees.kuleuven.be/geography/projects/14c-palaeolithic/

RADON - Radiocarbon Dates Online 2014 (University of Kiel, Germany, about 4300 dates for the Neolithic of Central Europe and Southern Scandinavia), http://radon.ufg.uni-kiel.de

Rains, Michael (2011): Creating a virtual research environment for archaeology, pp. 159-169, in: Kansa E. *et al*. (eds.): Archaeology 2.0: New Approaches to Communication and Collaboration. Cotsen Institute of Archaeology, UC Los Angeles, http://escholarship.org/uc/item/1r6137tb

Rains, Michael (2015): Integrated databased design and use into recording methodologies, pp. 79-91, in: Chapman R. & Wylie A. (eds.): Material evidence. Learning from archaeological practice. London: Routledge

Rathje W.L., Shanks M. & Witmore C. (eds., 2013): Archaeology in the Making: Conversations Through a Discipline. Routledge

RDA Europe - Research Data Alliance Europe project (EU, FP7-Infrastructures+H2020, 09/2012-03/2018), http://europe.rd-alliance.org

RDA Europe (2014): The Data Harvest: How sharing research data can yield knowledge, jobs and growth. A RDA Europe report, December 2014, http://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20Final.pdf

re3data - Registry of Research Data Repositories, http://www.re3data.org

RECODE - Policy Recommendations for Open Access to Research Data in Europe project (EU, FP7-SiS, 02/2013-01/2015, http://www.recodeproject.eu

RECODE (2014a): Infrastructure and technology challenges, Project deliverable 2.1, March 2014, http://recodeproject.eu/wp-content/uploads/2014/04/D2.1-Infrastructure-and-technology-challenges.pdf

RECODE (2014b): Legal and ethical issues in open access and data dissemination and preservation. Project deliverable D3.1, April 2014, http://recodeproject.eu/wp-content/uploads/2014/05/D3.1-legal-and-ethical-issues-FINAL.pdf

RECODE (2014c): Institutional barriers and good practice solutions. Project deliverable 4.1, September 2014, http://recodeproject.eu/wp-content/uploads/2014/09/RECODE-D4.1-Institutional-barriers-FINAL.pdf

RECODE (2015): Policy guidelines for open access and data dissemination and preservation. Project deliverable D5.1, February 2015, http://recodeproject.eu/wp-content/uploads/2015/02/RECODE-D5.1-POLICY-RECOMMENDATIONS-_FINAL.pdf

RECOLECTA - Recolector de Ciencia Abierta [Open Science Harvester], Spain, http://recolecta.fecyt.es/home?language=en

Reichman O.J., Jones M.B. & Schildauer M.P. (2011): Challenges and opportunities to open data in ecology. In: Science 331: 703-705, http://www.nefmc.org/research/cte_mtg_docs/110316/Challenges%20and%20Opportunities%20of%20Open%20Data%20in%20Ecology.pdf

Renear, Allen H. *et al*. (2008): Collection/Item Metadata Relationships. Berlin: Humboldt University, 2008, pp. 80-89, www.edoc.hu-berlin.de/conferences/dc-2008/renear-allen-80/PDF/renear.pdf

Renfrew, Colin (2001): Loot, legitimacy and ownership: the ethical crisis in archaeology. London: Duckworth

Repositories Support Project (2013): RSP Survey – Institutions: Summary Data, 21 March 2013, http://www.rsp.ac.uk/pmwiki/index.php?n=Institutions.Summary

Research Blogging (aggregator of research weblog posts on topics of various scientific fields), http://researchblogging.org

Research Councils UK (2010): Delivering the UK's E-Infrastructure for Research and Innovation. Report commissioned by the Department for Business Innovation and Skills. July 2010, http://www.rcuk.ac.uk/RCUK-prod/assets/documents/research/esci/e-Infrastructurereviewreport.pdf

Research Councils UK: Common Principles on Data Policy, http://www.rcuk.ac.uk/research/datapolicy/

Research Data Alliance (RDA) - "Research Data Sharing without barriers", https://rd-alliance.org

Research Data Alliance (RDA): Working and Interest Groups, https://www.rd-alliance.org/groups

Research Data Strategy Working Group (2011): Mapping the Data Landscape: Report of the 2011 Canadian Research Data Summit, December 2011, http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca

Research Trends (2013): New Perspectives on the Arts & Humanities. In: Research Trends, Issue 32, 15-19, http://www.researchtrends.com/wp-content/uploads/2013/02/Research_Trends_Issue32.pdf

ResearchGate (academic information sharing and networking platform), http://www.researchgate.net

Researchinformation.info (2015): Open-access books slowly on the rise, says PCG [Publishers Communication Group], 28 May 2015, http://www.researchinformation.info/news/news_story.php?news_id=1915

ResearchSpace - Creating the Cultural Heritage Knowledge Graph project (British Museum), http://www.researchspace.org

REWARD - Researchers using Existing Workflows to Archive Research Data project (UK, 10/2011-03/2012), http://www.ucl.ac.uk/reward

Richards J., Austin T. & Hardman C. (2010): Covering the Costs of Digital Curation. In: Heritage Management, 3(2): 255-263, http://www.maneyonline.com/doi/abs/10.1179/hma.2010.3.2.255

Richards J., Niven K. & Jeffrey S. (2013): Preserving our Digital Heritage: Information Systems for Data Management and Preservation, pp. 311-326, in: Ch'ng E. *et al*. (eds.): Visual Heritage in the Digital Age. Springer Series on Cultural Computing, http://radar.gsa.ac.uk/3080/1/Gaffney_%26_Chnng_Chapter_16.pdf

Richards, Julian (2004): Online Archives. In: Internet Archaeology 15, http://intarch.ac.uk/journal/issue15/richards_toc.html

Richards, Julian (2009): From anarchy to good practice: The evolution of standards in archaeological computing. In: Archeologia e Calcolatori 20: 27-35, http://eprints.whiterose.ac.uk/10707/

Richards, Julian (2012): Digital Infrastructures for Archaeological Research: A European Perspective. In: The CSA Newsletter, Vol. XXV, No. 2, http://csanet.org/newsletter/fall12/nlf1202.html

Richards, Julian (2014): Safeguarding our heritage for the future. Preserving valuable archaeological data saves time and money [Case study: Impact]. Arts and Humanities Research Council, July 2014, http://www.ahrc.ac.uk/What-We-Do/Build-the-evidence-base/Documents/Safeguarding-our-heritage.pdf

Richardson, Lorna (2012): Twitter and archaeology: an archaeological network in 140 characters or less, pp. 15-24, in: Bonacchi C. (ed.): Archaeology and Digital Communication. Towards Strategies of Public Engagement. London: Archetype Publications, 15-24; manuscript: https://www.academia.edu/2492383/Twitter_and_Archaeology_An_Archaeological_Network_in_140_Characters_or_Less

Richardson, Lorna (2013a): A Digital Public Archaeology? In: Papers from the Institute of Archaeology - University College London, 23(1): 1-12, http://www.pia-journal.co.uk/article/download/pia.431/550

Richardson, Lorna (2013b): Archaeology and Twitter survey 2013. Excel Spreadsheet (Dataset), http://discovery.ucl.ac.uk/1436372/

Richardson, Lorna (2014a): Online Survey Data from Twitter and Archaeology Surveys 2011-2013. Journal of Open Archaeology Data 3:e3, DOI: http://dx.doi.org/10.5334/joad.ae

Richardson, Lorna (2014b): Understanding Archaeological Authority in a Digital Context. In: Internet Archaeology, Issue 38, http://intarch.ac.uk/journal/issue38/richardson_index.html

Richardson, Matthew (2013): Multidisciplinarity: Mapping the multidisciplinarity of the Arts & Humanities. In: New Perspectives on the Arts & Humanities. Research Trends, Issue 32, 15-19, http://www.researchtrends.com/wp-content/uploads/2013/02/Research_Trends_Issue32.pdf

Ridge, Mia (ed., 2014): Crowdsourcing our Cultural Heritage. Ashgate (introduction available at: http://www.ashgate.com/pdf/SamplePages/Crowdsourcing-our-Cultural-Heritage-Intro.pdf)

Riding the Wave (2010): How Europe can gain from the rising tide of scientific data. Final report of the European High-level Expert Group on Scientific Data. A submission to the European Commission, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

RIN - Research Information Network & British Library (2009): Patterns of information use and exchange: case studies of researchers in the life sciences. November 2009, RIN: London, http://www.bl.uk/reshelp/experthelp/science/science@blevents/previousevents/rincasestudy/Patterns_information_use-REPORT_Nov09.pdf

RIN - Research Information Network & Key Perspectives (2008): To Share or not to Share: Publication and Quality Assurance of Research Data Outputs. Main Report and Annex. RIN: London, http://eprints.soton.ac.uk/266742/1/Published_report_-_main_-_final.pdf

RIN - Research Information Network (2010a): If you build it, will they come? How researchers perceive and use web 2.0. London: RIN, http://wrap.warwick.ac.uk/56246/1/WRAP_Procter_If you build it will they come.pdf

RIN - Research Information Network (2010b) E-infrastructure – taking forward the strategy. Synthesis of interviews with expert stakeholders. A Research Information Network report, September 2010, http://www.rin.ac.uk/system/files/attachments/E-infrastructure_V02_KL.pdf

RIN - Research Information Network (2011a): Reinventing Research? Information Practices in the Humanities. London: RIN, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1859267

RIN - Research Information Network (2011b): Data centres: their use, value and impact. RIN: London, September 2011 [JISC-funded study; the empirical research was undertaken by a team from Technopolis Group], http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/media/documents/publications/general/2011/datacentres.pdf

RIN & NESTA - National Endowment for Science, Technology, and the Arts (2010): Open to All? Case Studies of Openness in Research. Report prepared by the Digital Curation Centre. London: RIN, http://apo.org.au/files/Resource/nesta-rin_open_science_v01.pdf

Riris P. & Corteletti R. (2015): A New Record of Pre-Columbian Engravings in Urubici (SC), Brazil using Polynomial Texture Mapping. In: Internet Archaeology 38, http://intarch.ac.uk/journal/issue38/riris_index.html

RISE - Research Infrastructures Staff Exchange: a new scheme for staff mobility within European Research Infrastructures, proposed by the EIROforum and ERF, http://www.europeanresearchfacilities.eu/IMG/pdf/ERF_EIROforum_mobility_october2011_final.pdf

RItrain - Research Infrastructures Training Programme (EU, H2020, 9/2015-8/2019), http://ritrain.eu

Ritze D. & Boland K. (2013): Integration of Research Data and Research Data Links into Library Catalogues, pp. 35-40, in: DC-2013 - Proc. International Conference on Dublin Core and Metadata Applications, Lisbon, http://dcevents.dublincore.org/IntConf/dc-2013/paper/viewFile/156/74

ROAR - Registry of Open Access Repositories, http://roar.eprints.org

ROARMAP - Registry of Open Access Repositories Mandatory Archiving Policies, http://roarmap.eprints.org

Robey, David (2013): Infrastructure needs for the Digital Humanities. e-Research South website, http://www.eresearchsouth.ac.uk/news/infrastructure-needs-for-the-digital-humanities

Robinson, Nicholas A. (2013): Keynote: Sustaining Society in the Anthropocene Epoch. In: Denver Journal of International Law and Policy, Vol. 41: 467-506, http://digitalcommons.pace.edu/cgi/viewcontent.cgi?article=1925&context=lawfaculty

Robinson-García N., Torres-Salinas D., Zahedi Z. & Costas R. (2014): New data, new possibilities: Exploring the insides of Altmetric.com. In: arXiv:1408.0135 [cs.DL], 1 August 2014, http://arxiv.org/ftp/arxiv/papers/1408/1408.0135.pdf

Roche D.G., Lanfear R., Binning S.A. *et al.* (2014): Troubleshooting public data archiving: Suggestions to increase participation. In: PLOS Biology 12(1), e1001779, http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001779

Rockman M. & Flatman J. (eds., 2012): Archaeology in society: Its relevance in the modern world. New York: Springer

Rockman, Marcy (2012): The necessary role of archaeology in climate change mitigation and adaptation, pp. 193-215, in: Rockman M. & Flatman J. (eds., 2012): Archaeology in society: Its relevance in the modern world. New York: Springer, https://www.academia.edu/5551872/The_Necessary_Roles_of_Archaeology_in_Climate_Change_Mitigation_and_Adaptation

Rodrigo A., Alberts S., Cranston K. *et al*. (2013): Science Incubators: Synthesis Centers and Their Role in the Research Ecosystem. In: PLoS Biology, 11(1): e1001468, http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001468

Rogers, Everett M. (1962): Diffusion of Innovations. New York: Free Press

Romary, Laurent (2014): Sustainable data for sustainable infrastructures (pp. 187-201). In: Duşa A., Nelle D., Stock G. & Wagner G. (eds., 2014): Facing the Future: European Research Infrastructures for the Humanities and Social Sciences. Berlin: Scivero, online: http://www.allea.org/Content/ALLEA/2014_06_04-FACING_THE_FUTURE.pdf

Rosen A.M, Laparidou S. & Ramsey M.N. (2015): Special Issue: 'The Anthropocene in the Longue Durée'. The Holocene, 25(10), http://hol.sagepub.com/content/25/10.toc

Rosenblum, Leah F. (2009): If Malinowski had been a Blogger. In: LIBREAS. Library Ideas 14, http://libreas.eu/ausgabe14/002ros.htm

Rosenthal, David (2014): Costs: Why do we care?, 18 November 2014, http://blog.dshr.org/2014/11/talk-costs-why-do-we-care.html

Ross S., Sobotkova A., Ballsun-Stanton B, & Crook P. (2013): Creating eResearch Tools for Archaeologists: The Federated Archaeological Information Management Systems project. In: Australian Archaeology, No. 77, December 2013, https://www.academia.edu/5690498/Creating_eResearch_Tools_for_Archaeologists_The_Federated_Archaeological_Information_Management_Systems_project

Roth , Wolff-Michael (2005): Making Classifications (at) Work: Ordering Practices in Science. In: Social Studies of Science, 35: 581-621, http://web.uvic.ca/~mroth/PREPRINTS/MakingClassification104.pdf

Rouder, Jeffrey N. (2015): The What, Why, and How of Born-Open Data, http://pcl.missouri.edu/sites/default/files/p_9.pdf

Royal Netherlands Academy of Arts and Sciences (2011): Quality indicators for research in the humanities. Interim report by the Committee on Quality Indicators in the Humanities, https://www.knaw.nl/en/news/publications/quality-indicators-for-research-in-the-humanities/@@download/pdf_file/20111024.pdf

Ruggles, Clive (2011): Pushing back the frontiers or still running around the same circles? 'Interpretative archaeoastronomy' thirty years on. In: Proceedings of the International Astronomical Union, Vol. 7, Symposium S278: 1-18, http://dx.doi.org/10.1017/S1743921311012427

Ruiz, Sergio (2014): CrossRef and DataCite announce new initiative to accelerate the adoption of DOIs for data publication and citation. DataCite website, 10 November 2014, https://www.datacite.org/CrossRefDataCiteinitiative

Rupp M., Gini F., Pérez-Neira A. *et al*. (2011): Reproducible research in signal processing. In: EURASIP Journal on Advances in Signal Processing 93, http://asp.eurasipjournals.com/content/2011/1/93

Rusbridge, Chris (2007): Create, curate, re-use: the expanding life course of digital research. Educause Australasia 07 conference paper, https://www.era.lib.ed.ac.uk/handle/1842/1731

Rusbridge, Chris (2008): Tomorrow, and tomorrow, and tomorrow: poor players on the digital curation stage, pp. 207-217, in: Earnshaw R. & Vince J. (eds.): Digital Convergence – Libraries of the Future. Springer; manuscript available at: https://www.era.lib.ed.ac.uk/handle/1842/2150

Rutkin, Aviva (2015): Data archaeology helps builders avoid buried treasure. In: New Scientist 3006, January 2015, https://www.newscientist.com/article/mg22530062.400-data-archaeology-helps-builders-avoid-buried-treasure/

Rutner J. & Schonfeld R.C. (2012): Supporting the Changing Research Practices of Historians. Final Report from ITHAKA S+R. 10 December 2012, http://www.sr.ithaka.org/wp-content/uploads/2015/08/supporting-the-changing-research-practices-of-historians.pdf

Sabbioni C., Brimblecombe P. & Cassar M. (eds., 2010): The Atlas of Climate Change Impact on European Cultural Heritage: Scientific Analysis and Management Strategies. London: Anthem Press, http://www.anthempress.com/the-atlas-of-climate-change-impact-on-european-cultural-heritage-pb

Sakellariadi, Anastasia (2014): Review of “Integrating Archaeology: Science – Wish – Reality”. Edited by N. Schücker. Papers of the International Conference on the Social Role, Possibililties and Perspectives of Classical Studies. Römisch-Germanische Kommission des Deutschen Archäologischen Instituts, Frankfurt a.M., 2012. In: European Journal of Archaeology, 17(4): 755-759

Salo, Dorothea (2010): Retooling Libraries for the Data Challenge. In: Ariadne, Issue 64, http://www.ariadne.ac.uk/issue64/salo/

Salvemini F., Grazzi F., Fedrigo A. *et al*. (2013): Revealing the secrets of composite helmets of ancient Japanese tradition. In: European Physical Journal Plus, 128: 87, http://link.springer.com/article/10.1140%2Fepjp%2Fi2013-13087-y

Samson, Kurt (2008): Data Sharing: Making Headway in a Competitive Research Milieu. In: Annals of Neurology, 64(1): A13-6, July 2008, http://hinxtongroup.files.wordpress.com/2010/10/samson_ann-neurol_2008.pdf

Sánchez Solís, Barbara (2014): Factors for Enabling Sharing and Reuse of Research Data. A study of the National Open Access Desk (NOAD) Austria, March 2014, https://phaidra.univie.ac.at/detail_object/o:344008

Sands, Rob (2009): Artefact Studies, Artefact Data and Information Technology: A point in time. University College Dublin, School of Archaeology, November 2009, http://www.ucd.ie/archaeology/documentstore/hc_reports/add/17366_ucd_final_report_v1.2.pdf

Sandve G.K., Nekrutenko A., Taylor J. & Hovig E. (2013): Ten simple rules for reproducible computational research. In: PLoS Computational Biology, 9(10), e1003285, http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003285

Sandweiss D. & Kelley, A.R. (2012): The Archaeological Record as a Paleoclimatic and Paleoenvironmental Archive. In: Annual Review of Anthropology, Vol. 21: 371-409, http://www.annualreviews.org/eprint/CAFcG2ZZWdAg4paqE9IG/full/10.1146/annurev-anthro-092611-145941

Sangro Valley Project (Tornareccio, Abruzzo, Italy), http://www.sangro.org

SCARP - Disciplinary Approaches to Sharing, Curation, Reuse and Preservation (UK, JISC-funded project, coord. Digital Curation Centre, 01/2007-07/2009), http://www.dcc.ac.uk/projects/scarp

Scarre, Geoffrey (2014): The Ethics of Digging, in: Sandis, Constantine (ed.): Cultural Heritage Ethics: Between Theory and Practice. Open Book Publishers, October 2014, http://dx.doi.org/10.11647/OBP.0047.07

Scheliga K. & Friesike S. (2014): Putting open science into practice: A social dilemma? In: First Monday, 19(9), http://journals.uic.edu/ojs/index.php/fm/article/view/5381/4110

Schigel, Dmitry (2015): Open biodiversity information: international perspective. Helsinki, 22 May 2015, http://de.slideshare.net/Luomus/6-schigel-laji22-may

Schittek K., Mächtle B., Schäbitz F. *et al.* (2015): Holocene environmental changes in the highlands of the southern Peruvian Andes (14° S) and their impact on pre-Columbian cultures. In: Climate of the Past, 11: 27–44, http://www.clim-past.net/11/27/2015/cp-11-27-2015.pdf

Schlanger N. & Aitchison K. (eds., 2010): Archaeology and the Global Crisis. Multiple Impacts, Possible Solutions. Tervuren: Culture lab Éditions, http://digital.csic.es/bitstream/10261/27633/1/2010_25Archaeology-and-the-crisis.pdf

Schlanger, Nathan (2010): Postscript: on dead canaries, guinea-pigs and other Trojan horse, pp. 107-115, in: Schlanger N. & Aitchison K. (eds.): Archaeology and the Global Crisis. Multiple Impacts, Possible Solutions. Tervuren: Culture lab Éditions, http://digital.csic.es/bitstream/10261/27633/1/2010_25Archaeology-and-the-crisis.pdf

Schneider, Jeri (2006): Why we need a data citation standard: Lessons learned from compiling ICPSR's Bibliography of Data-Related Literature. In: ICPSR Bulletin, xxvi(2): 9-12, http://www.icpsr.umich.edu/files/ICPSR/org/publications/bulletin/2006-Q1.pdf

Schwarzman, Sasha (2010): Supplemental materials survey. In: Information Standards Quarterly 22(3): 23-26, http://www.niso.org/apps/group_public/download.php/4886/Schwarzman_SuppMatlsSurvey_ISQ_v22no3.pdf

Science Europe (association of European Research Funding Organisations [RFO] and Research Performing Organisations [RPO]), http://www.scienceeurope.org

Science magazine (2011): Science Staff introduction to the Special Issue 'Dealing with Data'. In: Science, Vol. 331, no. 6018, 11 February 2011: 692-693, http://www.sciencemag.org/content/331/6018/692.full

Science Metrix (2004): The Use of Bibliometrics in the Social Sciences and Humanities. Prepared by E. Archambault and E. Vignola Gagné for the Social Sciences and Humanities Research Council of Canada (SSHRCC). Final Report, August 2004, http://www.science-metrix.com/pdf/SM_2004_008_SSHRC_Bibliometrics_Social_Science.pdf

Science-Metrix (2014a): Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels, 1996–2013. Study for the European Commission, DG Research and Innovation. D1.8, April 2014, http://science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf

Science-Metrix (2014b): State-of-art analysis of OA strategies to peer-review publications. Study for the European Commission, DG Research and Innovation. D2.1, April 2014, http://science-metrix.com/files/science-metrix/publications/d_2.1_sm_ec_dg-rtd_oa_policies_in_the_era_update_v05p_0.pdf

Science-Metrix (2014c): State-of-art analysis of OA strategies to scientific data. Study for the European Commission, DG Research and Innovation. D2.2, April 2014, http://science-metrix.com/files/science-metrix/publications/d_2.2_sm_ec_dg-rtd_oa_data_in_the_era_update_v04p_0.pdf

Science-Metrix (2014d): Comparative analysis of the strengths and weaknesses of existing open access strategies. Study of the European Commission, DG Research and Innovation. D2.3, April 2014, http://science-metrix.com/files/science-metrix/publications/d_2.3_sm_ec_dg-rtd_oa_comparative_analysis_update_v04p_0.pdf

Science-Metrix (2014e): Evolution of Open Access Policies and Availability, 1996–2013. Study for the European Commission, DG Research and Innovation. D4.5, April 2014, http://science-metrix.com/files/science-metrix/publications/d_4.5_sm_ec_dg-rtd_oa_policies_proportion_oa_1996-2013_v05p_0.pdf

Science-Metrix (2014f): Evolution of Open Access Policies and Availability, 1996-2013 – Summary Report. Study for the European Commission, DG Research and Innovation. D4.6, April 2014, http://science-metrix.com/files/science-metrix/publications/d_4.6_sm_ec_dg-rtd_oa_policies_proportion_oa_1996-2013_v12p_0.pdf

Scully, Lorcan (2013): Making Remote Sensing Data Less Remote. Data Sharing and the ArcLand Spatial Data Infrastructure. University College Dublin, School of Archaeology [thesis], https://www.academia.edu/5400371/Making_Remote_Sensing_Data_Less_Remote_Data_Sharing_and_the_ArcLand_Spatial_Data_Infrastructure

SEAD - Strategic Environmental Archaeology Database (Umeå University, Environmental Archaeology Lab, Sweden), http://www.sead.se

SEAD - Sustainable Environment Actionable (Data NSF DataNet funded project, builds cyber-infrastructure for multi-disciplinary sustainability science), http://sead-data.net

Seddon A.W.R., Mackay A.W., Baker A.G. *et al*. (2014): Looking Forward Through the Past: Identification of 50 Priority Research Questions in Palaeoecology. In: Journal of Ecology 102: 256-267, http://onlinelibrary.wiley.com/enhanced/doi/10.1111/1365-2745.12195/

Sedransk N., Young L.J., Kelner K.L. *et al*. (2010): Make Research Data Public? – Not Always so Simple: A Dialogue for Statisticians and Science Editors. In: Statistical Science, 25(1): 41-50; http://projecteuclid.org/download/pdfview_1/euclid.ss/1280841732

SENESCHAL - Semantic Enrichment Enabling Sustainability of Archaeological Links (UK, AHRC-funded project, 03/2013-02/2014), http://hypermedia.research.southwales.ac.uk/kos/seneschal/, see also: http://www.heritagedata.org

SERSCIDA - Support for Establishment of National/Regional Social Sciences Data Archives (EU, FP7, Support action, 01/2012-06/2014), http://www.serscida.eu

Shanks M. & McGuire R.H. (1996): The Craft of Archaeology. In: American Antiquity, 61(1): 75-88, https://www.academia.edu/944074/The_Craft_of_Archaeology

Shanks M. & Witmore C. (2012): Archaeology 2.0? Review of Archaeology 2.0: New Approaches to Communication and Collaboration. In: Internet Archaeology, Issue 32, http://intarch.ac.uk/journal/issue32/shanks-witmore.html

Sharma S., Tim U.S., Wong J., Gadia S. & Sharma S. (2014): A Brief Review on Leading Big Data Models. In: Data Science Journal, Vol. 13: 138-157, http://datascience.codata.org/articles/abstract/10.2481/dsj.14-041/

Shearer K., Haigh S. & Whitehead M. (2015): Portage: Supporting Canadian innovation through shared expertise and stewardship of research data. IFLA World Library and Information Congress, Cape Town, South Africa, 15-21 August 2015, http://library.ifla.org/1263/1/198-shearer-en.pdf

Shearer, Kathleen (2014): Wagging the Long Tail - Current Metadata Practices for Long Tail Research Data. EUDAT meeting, 25 September 2014, http://www.iassistdata.org/downloads/2014/2014_3i_shearer.pdf

Shema H., Bar-Ilan J. & Thelwall M. (2012): Research Blogs and the Discussion of Scholarly Information. In: PLoS ONE, 7(5): e35869, http://dx.doi.org/10.1371/journal.pone.0035869

SHERPA/JULIET (information on research funders' open access policies/mandates), http://www.sherpa.ac.uk/juliet/

Shotton, David (2011): Why researchers don't publish data. In: Opencitations weblog, 4 August 2011, http://opencitations.wordpress.com/2011/08/04/why-researchers-don%E2%80%99t-publish-data/

Shuai X., Pepe A. & Bollen J. (2012): How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations. In: PLoS ONE 7(11): e47523, http://dx.doi.org/10.1371/journal.pone.0047523

Sieber J.E. & Trumbo B.E. (1995): (Not) giving credit where credit is due: Citation of data sets. Science and Engineering Ethics, 1(1): 11-20, http://link.springer.com/article/10.1007%2FBF02628694

Silchester Town Life project (University of Reading, UK), http://www.reading.ac.uk/silchester/

Simpson, Faye. A. (2009): The values of community archaeology: a comparative assessment between the UK and US. PhD thesis, University of Exeter, August 2009, https://ore.exeter.ac.uk/repository/handle/10036/86115

Sinclair, Anthony (2014): Archaeology Is...? Mapping Archaeological Research 2004-2013. Current Research in British Prehistory: a Liverpool symposium, 21-22 August 2014, Liverpool, https://www.researchgate.net/publication/266638775_Archaeology_Is..._Mapping_Archaeological_Research_2004-2013

Sinnott R.O. *et al*. (2005): Large-scale data sharing in the life sciences: Data standards, incentives, barriers and funding models (The Joint Data Standards Study). e-Science Technical Report Series, August 2005, http://www.nesc.ac.uk/technical_papers/UKeS-2006-02.pdf

Smith L. & Waterton E. (2009): Heritage, Communities and Archaeology. Bloomsbury

Smith M.E., Feinman G.M., Drennan R.D. *et al*. (2012): Archaeology as a social science. In: PNAS - Proceedings of the National Academy of Sciences: Current Issue, 109(20), http://www.pnas.org/content/109/20/7617.full.pdf

Smith, Michael E. (2009): Just How Useful is Archaeology for Scientists and Scholars in Other Disciplines? Paper for the symposium "Archaeology Beyond Archaeology" at the Annual Meeting of the Society for American Archaeology, Atlanta, USA, http://www.public.asu.edu/~mesmith9/1-CompleteSet/0-MES-09-SAA-IntroPaper.pdf

Smith, Michael E. (2015): Is archaeology relevant? Is "relevance" irrelevant? Publishing Archaeology weblog, 14 February 2015, http://publishingarchaeology.blogspot.co.at/2015/02/is-archaeology-relevant-is-relevance.html

Snijder, Ronald (2010): The profits of free books – an experiment to measure the impact of Open Access publishing. In: Learned Publishing, 23(4): 293-301 (website provides the content as well as data of the study), https://sites.google.com/site/theprofitsoffreebooks/

Snowball Metrics (2014): Snowball Metrics Recipe Book. 2nd edition, June 2014, http://www.snowballmetrics.com/wp-content/uploads/snowball-recipe-book_HR.pdf

Sobotkova, Adela (2013): The Use of Information Technology in Australian Archaeology: the FAIMS Digital Data Survey Report. FAIMS project, 25 November 2013, https://fedarch.org/documents/DigitalDataSurveyReport.pdf

Society for Phytolith Research, International Committee on Phytolith Taxonomy – ICPT, ICPN committee, http://www.phytolithsociety.org/international-committee-on-.html

Socio-Environmental Synthesis Center (SESYNC), United States, http://www.sesync.org

Software Sustainability Institute, http://www.software.ac.uk

SPARC Europe (European branch of the Scholarly Publishing and Academic Resources Coalition), http://sparceurope.org

SPARC Europe: The Open Access Citation Advantage Service, http://sparceurope.org/oaca/

Sparks, Sue (2005): JISC Disciplinary Differences Report. Rightscom Ltd, London, http://www.jisc.ac.uk/uploaded_documents/Disciplinary%20Differences%20and%20Needs.doc

Spielmann K. & Kintigh K. (2011): The digital archaeological record: The potentials of archaeozoological data integration through tDAR. In: The SAA Archaeological Record, January 2015: 22–25, http://digitaleditions.sheridan.com/publication/?i=58423&p=24

SPLASHCOS - Submerged Prehistoric Archaeology and Landscapes of the Continental Shelf (EU, COST Action, 2009-2013), http://www.splashcos.org

SPLASHCOS (2013): Marine Industry & Submerged Prehistoric Archaeology. Sharing Data, Developing Understanding and Delivering Best Practice. Prepared by Julie Satchell, Maritime Archaeology Trust, UK. COST Action TD0902, http://www.splashcos.org/sites/splashcos.org/files/SPLASHCOS_Marine_Industry_Guide_FINAL.pdf

STACHEM – Science and Technology for Archaeology and Cultural Heritage in the Eastern Mediterranean (EU FP7 project, 11/2008-04/2010), http://www.cyi.ac.cy/index.php/stachem-overview.html

STAR - Semantic Technologies for Archaeological Resources (UK, AHRC, funded project, 2007-2010), http://hypermedia.research.southwales.ac.uk/kos/star/

Steele, Colin (2003): Phoenix Rising: New Models for the Research Monograph? In: Learned Publishing, 16(2): 111-122, https://digitalcollections.anu.edu.au/bitstream/1885/40718/2/111-122.pdf

Steele, Colin (2008): Scholarly Monograph Publishing in the 21st Century: The Future More Than Ever Should Be an Open Book. In: JEP - the journal of electronic publishing, 11(2), http://dx.doi.org/10.3998/3336451.0011.201

STELLAR - Semantic Technologies Enhancing Links and Linked data for Archaeological Resources (UK, AHCR funded project, 2010-2011), http://hypermedia.research.southwales.ac.uk/kos/stellar/

STM Association (2007): Brussels Declaration, http://www.stm-assoc.org/public-affairs/resources/brussels-declaration/

Stodden V. & Miguez S. (2014): Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. In: Journal of Open Research Software, 2(1):e21, http://dx.doi.org/10.5334/jors.ay

Stodden V. Borwein J. & Bailey D.H. (2013a): 'Setting the Default to Reproducible' in Computational Science Research. In: Siam News, 3 June 2013, http://sinews.siam.org/DetailsPage/tabid/607/ArticleID/351/%E2%80%9CSetting-the-Default-to-Reproducible%E2%80%9D-in-Computational-Science-Research.aspx

Stodden V., Bailey D.H., Borwein J., LeVeque R.L., Rider W. & Stein W. (eds., 2013b): Setting the Default to Reproducible. Reproducibility in Computational and Experimental Mathematics. Developed collaboratively by the ICERM workshop participants, http://stodden.net/icerm_report.pdf

Stodden, Victoria (2010): The Scientific Method in Practice: Reproducibility in the Computational Sciences. MIT Sloan Research Paper, No. 4773-10, SSRN Electronic Journal http://papers.ssrn.com/abstract=1550193

Stott D., Boyd D.S., Beck A. & Cohn A.G. (2015): Airborne LiDAR for the Detection of Archaeological Vegetation Marks Using Biomass as a Proxy. In: Remote Sensing,7: 1594-1618, http://www.mdpi.com/2072-4292/7/2/1594

Strasser, Charly (2013): Closed Data… Excuses, Excuses. Data Pub, 24 April 2013, http://datapub.cdlib.org/closed-data-excuses-excuses/

Suarez A.V. & Tsutui N.D. (2004): The Value of Museum Collections for Research and Society. In: BioScience, 54(1), http://nature.berkeley.edu/tsutsuilab/SuarezTsutsui2004Biosci.pdf

SUDAMIH - Supporting Data Management Infrastructure in the Humanities (2010): Sudamih Researcher Requirements Report. J.A.J. Wilson & M. Patrick, Oxford University Computing Services, http://sudamih.oucs.ox.ac.uk/docs/Sudamih_FinalReport_v1.0.pdf

SUDAMIH - Supporting Data Management Infrastructure in the Humanities. UK JISC funded project, 10/2009- 03/2011, http://sudamih.oucs.ox.ac.uk

Sun N. & Li X.Q. (2012): The quantitative reconstruction of the palaeoclimate between 5200 and 4300 cal yr BP in the Tianshui Basin, NW China. In: Climate of the Past, 8: 625–636, http://www.clim-past.net/8/625/2012/cp-8-625-2012.pdf

Šupak Smolčić, Vesna (2013): Salami publication: definitions and examples. In: Biochemia Medica 23(3): 237-141, http://www.biochemia-medica.com/2013/23/237

SURF Foundation (2009): The European Repository Landscape 2008. Inventory of Digital Repositories for Research Output. M. van der Graaf and M. Vernooy-Gerritsen. Amsterdam University Press, http://dare.uva.nl/document/150724

SURF Foundation (2013): European Landscape Study of Research Data Management. For SIM4RDM - Support Infrastructure Models for Research Data Management. W. Mossink, M. Bijsterbosch and J. Nortier, May 2013, SURF, http://www.sim4rdm.eu/docs/project-outputs/sim4rdm-landscape-report

Sustainable Archaeology (Canada), http://sustainablearchaeology.org

Svensson, Patrik (2010): The Landscape of Digital Humanities. In: Digital Humanities Quarterly, 4.1, http://www.digitalhumanities.org/dhq/vol/4/1/000080/000080.html

Svensson, Patrik (2011): From Optical Fiber to Conceptual Cyberinfrastructure. In: Digital Humanities Quarterly, 5.1, http://digitalhumanities.org/dhq/vol/5/1/000090/000090.html

Swan A. & Brown S. (2008): Skills, Role and Career Structure of Data Scientists and Curators: Assessment of Current Practice and Future Needs. Report to JISC. Truro, UK: Key Perspectives, http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx

Swan A. & Chan L. (2010): Citation impact. In: Open Access Scholarly Information Sourcebook, http://www.openoasis.org/index.php?option=com_content&view=article&id=560&Itemid=391

Swan, Alma (2010): The Open Access citation advantage. Studies and results to date. University of Southampton, http://eprints.soton.ac.uk/268516/2/Citation_advantage_paper.pdf

Swauger S. & Vision T.J. (2015): What Factors Influence Where Researchers Deposit their Data? A Survey of Researchers Submitting to Data Repositories. In: International Journal of Digital Curation, 10(1), 68-81, http://www.ijdc.net/index.php/ijdc/article/view/10.1.68

Swedish National Data Service - SND (Sweden), http://snd.gu.se

SWORD - Simple Web service Offering Repository Deposit (a lightweight protocol for moving repository content from one location to another), http://swordapp.org

SYNERGY - The SYNERGY Reference Model of Data Provision and Aggregation. Current contributors: M. Doerr, G. de Jong, K. Konsolaki, B. Norton, D. Oldman, M. Theodoridou, T. Wikman. Draft, June 2014, http://www.cidoc-crm.org/docs/SRM_v0.1.pdf

TaDiRAH - Taxonomy of Digital Research Activities in the Humanities (v. 0.5.1, 05/2014), available on Github, https://github.com/dhtaxonomy/TaDiRAH

Talia, Domenico (2013): Workflow Systems for Science: Concepts and Tools. In: ISRN Software Engineering, Volume 2013, http://www.hindawi.com/journals/isrn/2013/404525/

Tattersall, Andy (2014): Comment, discuss, review: An essential guide to post-publication review sites. London School of Economics - Maximising the impact of academic research, weblog, 8

November 2014, http://blogs.lse.ac.uk/impactofsocialsciences/2014/11/08/comment-discuss-review-an-essential-guide/

Tayler, Mike (2013): Towards a common model of citation: some thoughts on merging altmetrics and bibliometrics. In: Research Trends, Issue 35, December 2013, http://www.researchtrends.com/issue-35-december-2013/towards-a-common-model-of-citation-some-thoughts-on-merging-altmetrics-and-bibliometrics/

Taylor & Francis (2014): Open Access Survey. June 2014. Taylor & Francis / Routledge, http://www.tandf.co.uk/journals/explore/open-access-survey-june2014.pdf

Taylor C.F., Field D., Sansone S.A. *et al*. (2008): Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. In: Nature Biotechnology 26: 889-896, http://www.nature.com/nbt/journal/v26/n8/full/nbt.1411.html

Taylor M., Zijlstra H. & Colledge L. (2014): Elsevier is expanding its use of altmetrics, 14 September 2014, http://www.elsevier.com/authors-update/story/impact-metrics/elsevier-is-expanding-its-use-of-altmetrics

Taylor, Mike (2013): Exploring the Boundaries: How Altmetrics Can Expand Our Vision of Scholarly Communication and Social Impact. In: Information Standards Quarterly, 25(2): 27-32, http://www.niso.org/apps/group_public/download.php/11274/IP_Taylor_Altmetrics_Social_Impact_isqv25no2.pdf

tDAR - The Digital Archaeological Record (Digital Antiquity consortium, USA), http://www.tdar.org

Tenopir C., Allard S., Douglass K. *et al*. (2011): Data Sharing by Scientists: Practices and Perceptions. In: PLoS ONE, 6(6): e21101, http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101

Tenopir C., Birch B. & Allard S. (2012): Academic libraries and research data services. Association of College & Research Libraries, http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Terra Populus - Integrated Data on Population and the Environment (TerraPop, NSF DataNet funded project), http://www.terrapop.org

Terras M., Nyhan J. & Vanhoutte E. (2013): Defining Digital Humanities. A Reader. Ashgate, http://www.ashgate.com/isbn/9781409469636

Terras M., Warwick C. & Fisher C. (2010): Integrating New Technologies into Established Systems: a case study from Roman Silchester. In: Proceedings of the CAA conference, Williamsburg, USA, 22-26 March 2009, http://discovery.ucl.ac.uk/1324501/1/1324501.pdf

Terras, Melissa (2010): Digital curiosities: resource creation via amateur digitization. In: Literary and Linguist Computing, 25(4): 425-438, http://discovery.ucl.ac.uk/171071/1/Terras_Amateur_Digitisation.pdf

Testa, Jim (2011): Globalization of Web of Science, 2005-2010. Web of Science, info, http://wokinfo.com/products_tools/multidisciplinary/webofscience/contentexp/expansionessay/

The Archaeologist 82 - Winter 2011 - Best practice in community archaeology, http://www.archaeologists.net/sites/default/files/ta82.pdf

The Archaeologist 86 - Winter 2012 - Adding value to development, http://www.archaeologists.net/sites/default/files/TA86-Winter2012-value.pdf

The Archaeologist 89 - Autumn 2013 - Making waves; archaeology and impact, http://www.archaeologists.net/sites/default/files/TA89-2013-Autumn.pdf

The Archaeologist 90, Winter 2013: Archaeological Archives, http://www.archaeologists.net/sites/default/files/TA90-Website.pdf

The Economist (2013): Unreliable research. Trouble at the lab: Scientists like to think of science as self-correcting. To an alarming degree, it is not. The Economist, 19 October 2013, http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble

The Royal Society (2012a): Science as an Open Enterprise. Final Report, June 2012, https://royalsociety.org/policy/projects/science-public-enterprise/Report/

The Royal Society (2012b): Science as an Open Enterprise: Costs of digital repositories, https://royalsociety.org/policy/projects/science-public-enterprise/digital-repositories/

The Royal Society (2012c): Science as an Open Enterprise: Case studies of data use, https://royalsociety.org/policy/projects/science-public-enterprise/case-studies/

Thelwall M. & Kousha K. (2014a): Academia.edu: Social network or Academic Network? In: Journal of the Association for Information Science and Technology, 65(4): 721-731; manuscript, http://www.scit.wlv.ac.uk/~cm1993/papers/Academia_preprint.pdf

Thelwall M. & Kousha K. (2014b): ResearchGate: Disseminating, communicating and measuring scholarship? In: Journal of the Association for Information Science and Technology, May 2014; manuscript, http://www.scit.wlv.ac.uk/~cm1993/papers/ResearchGate.pdf

Thelwall M., Haustein S., Larivière V. & Sugimoto C.R. (2013): Do altmetrics work? Twitter and ten other social web services. In: PLoS ONE, 8(5): e64841, http://dx.doi.org/10.1371/journal.pone.0064841

Theodoridou M., Tzitzikas Y., Doerr M., Marketakis Y. & Melessanakis V. (2010): Modeling and querying provenance by extending CIDOC CRM. In: Distributed and Parallel Databases, Vol. 27(2), 169-210, http://link.springer.com/article/10.1007%2Fs10619-009-7059-2

Thessen A.E. & Patterson D.J. (2011): Data issues in the life sciences. In: ZooKeys, 150: 15–51, http://www.pensoft.net/journals/zookeys/article/1766/data-issues-in-the-life-sciences

Thomson Reuters: Book Citation Index, http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/book-citation-index.html

THOR - Technical and Human infrastructure for Open Research project (EU, H2020, 06/2015-12/2017), http://project-thor.eu

Thwaites, Harold (2013): Digital Heritage: What Happens When We Digitize Everything?, pp. 327-348, in: Ch'ng E., Gaffney V. & Chapman H. (eds.): Visual Heritage in the Digital Age. Springer

Tijssen, Robert J.W. (2010): Discarding the ‘basic science/applied science’ dichotomy: A Knowledge Utilization Triangle classification system of research journals. Centre for Science and Technology Studies (CWTS), Leiden University, https://openaccess.leidenuniv.nl/handle/1887/15046

Timpson A., Colledge S., Crema W. et al. (2014): Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. In: Journal of Archaeological Science, vol. 52, 549–557, http://www.sciencedirect.com/science/article/pii/S0305440314002982

Tjalsma H. & Rombouts J. (2010): Selection of research data: guidelines for appraising and selecting research data. DANS Studies in Digital Archiving 6. Utrecht: SURF Foundation, http://www.dans.knaw.nl/nl/over/organisatie-beleid/publicaties/DANSselectionofresearchdata.pdf

Todorov, Ilian (2012): Is the Work of Scientific Software Engineers Recognised in Academia? In: Software Sustainability Institute weblog, http://software.ac.uk/blog/2012-04-23-work-scientific-software-engineers-recognised-academia

Torrence R., Rehren T. & Martinón-Torres M. (eds., 2015): Scoping the Future of Archaeological Science: Papers in Honour of Richard Klein. Journal of Archaeological Science, 56: 1-248, http://www.sciencedirect.com/science/journal/03054403/56

Torres-Salinas D., Martín-Martín A. & Fuente-Gutiérrez (2013): An introduction to the coverage of the Data Citation Index (Thomson-Reuters): Disciplines, document types and repositories. EC3 Working Papers 11, June 2013. arXiv:1306.6584v2 [cs.DL], http://arxiv.org/ftp/arxiv/papers/1306/1306.6584.pdf

Torsen M. & Anderson J. (2010): Intellectual Property and the Safeguarding of Traditional Cultures: Legal Issues and Practical Options for Museums, Libraries and Archives. World Intellectual Property Organization (WIPO), December 2010, http://www.wipo.int/edocs/pubdocs/en/tk/1023/wipo_pub_1023.pdf

Troll Covey, Denise (2011): Recruiting Content for the Institutional Repository: The Barriers Exceed the Benefits. In: Journal of Digital Information, 12(3), https://journals.tdl.org/jodi/index.php/jodi/article/view/2068/1757

Tsonis A.A., Swanson K.L., Sugihara G. & Tsonis P.A. (2010): Climate change and the demise of Minoan civilization. In: Climate of the Past, 6: 525–530, http://www.clim-past.net/6/525/2010/cp-6-525-2010.pdf

Tudhope D., Binding C., Jeffrey S., May K. & Vlachidis A. (2011): A STELLAR role for knowledge organisation systems in digital archaeology. In: ASIS&T Bulletin, 37(4): 15-18, http://www.asis.org/Bulletin/Apr-11/AprMay11_Tudhope_etAl.pdf

Tys, Dries (2015): Dare to choose: make research the product, pp. 69-74, in: European Archaeological Council (EAC): Setting the Agenda – Giving New Meaning to the European Archaeological Heritage. Edited by P.A.C. Schut, D. Scharff & L.C. de Wit. EAC Occasional Paper no. 10, Budapest: Archaeolingua

UCL & Emerald - University College London & Emerald Group Publishing Ltd (2010) Social media and research workflow (14 December 2010), http://ciber-research.eu/download/20101111-social-media-report.pdf

Udell, Jon (2015): When Open Access is the norm, how do scientists work together online? PLOS weblogs, http://blogs.plos.org/scicomm/2015/04/13/hello-world/

Uhlir, Paul F. (2010): Information gulags, intellectual straightjackets, and memory holes: three principles to guide the preservation of scientific data. In: Data Science Journal, 9, ES1-ES5, 7 October 2010, https://www.jstage.jst.go.jp/article/dsj/9/0/9_Essay-001-Uhlir/_pdf

Uhlir, Paul F. (ed., 2012): For attribution: Developing scientific data attribution and citation practices and standards: Summary of an international workshop. Washington, D.C.: National Academies Press, http://www.nap.edu/catalog.php?record_id=13564

UK Data Archive (social and economic research data), http://www.data-archive.ac.uk

UMAC Worldwide Database of University Museums & Collections (ICOM Committee for University Museums & Collections - UMAC), http://publicus.culture.hu-berlin.de/collections/

UNESCO (2015): UNESCO's Open Access (OA) Curriculum. Curriculum for Library Schools. Module 2: Open Access Infrastructure. Paris: UNESCO, http://unesdoc.unesco.org/images/0023/002322/232204E.pdf

Universal Numeric Fingerprint (UNF), http://guides.dataverse.org/en/latest/developers/unf/index.html

University of California Curation Center - UC3 (USA), http://www.cdlib.org/uc3/

University of California Curation Center (UC3): DMPTool [Data Management Planning Tool], https://dmptool.org

University of California Curation Center (UC3): Merrit repository (part of the California Digital Library), http://www.cdlib.org/uc3/merritt/

University of California Curation Center (UC3): TCP - Total Cost of Preservation (2011-), https://wiki.ucop.edu/display/Curation/Cost+Modeling

Unsworth, John M. (2000): Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this? King's College, London, http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html

UrCrowdsource (invites non-experts to transcribe excavation documents; part of the UrOnline project), http://urcrowdsource.org

UrOnline - Ur of the Chaldees: A Virtual Vision of Woolley's Excavations (University of Pennsylvania Museum and the British Museum), http://www.ur-online.org/about/

Van de Noort, Robert (2011): Conceptualising Climate Change Archaeology. In: Antiquity 85: 1039-48, pre-print, https://www.academia.edu/3663250/Conceptualising_climate_change_archaeology

Van de Noort, Robert (2013): Climate Change Archaeology: Building Resilience from Research in the World's Coastal Wetlands. Oxford: Oxford University Press, http://onlinelibrary.wiley.com/doi/10.1111/1095-9270.12125/abstract

Van de Sompel H. & Lagoze C. (2009): All aboard: toward a machine-friendly scholarly communication system, pp. 193-199, in: Hey T., Tansley S. & Tolle K. (eds.): Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, WA; Microsoft, http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_sompel_lagoze.pdf

Van de Sompel H., Lagoze C., Nelson M.L. *et al*. (2009a): Adding e-science assets to the data web. Linked Data on the Web (LDOW2009), 20 April 2009, Madrid, Spain, http://events.linkeddata.org/ldow2009/papers/ldow2009_paper8.pdf

Van de Sompel H., Pepe A., Mayernik M. & Borgman C. (2009b): From artefacts to aggregations: Modeling scientific life cycles on the semantic web. In: Journal of the American Society for Information Science and Technology, 61(3), http://arxiv.org/ftp/arxiv/papers/0906/0906.2549.pdf

Van der Graaf, Maurits (2007): DRIVER: Seven Items on a European Agenda for Digital Repositories. In: ARIADNE, Issue 52, http://www.ariadne.ac.uk/issue52/vandergraf/

Van der Leeuw S.E. & Redman C.L. (2002): Placing archaeology at the center of socio-natural studies. In: American Antiquity, 67(4): 597- 606, http://www.jstor.org/stable/i271767

Van der Linde S.J., van de Dries M.H., Schlanger N. & Slappendel C.G. (eds., 2012): Introduction: European Archaeology Abroad: Global Settings, Comparative Perspectives. Leiden: Sidestone, https://www.academia.edu/6903287/European_Archaeology_Abroad._Global_Settings_Comparative_Perspectives

Van Gorp P. & Mazanek S. (2011): SHARE: a web portal for creating and sharing executable research papers. In: Procedia Computer Science, 4, 589–597, http://www.sciencedirect.com/science/article/pii/S1877050911001207

Van Hooland S. & Verborgh R. (2014): Linked Data for Libraries, Archives, and Museums. How to clean, link and publish your metadata. Facet Publishing / website, http://book.freeyourmetadata.org

Van Hooland S., De Wilde M., Verborgh R., Steiner T. & Van de Walle R. (2014): Exploring Entity Recognition and Disambiguation for Cultural Heritage Collections. In: Digital Scholarship in the Humanities, 30(2), 2015: 262-279; preprint 2014, http://freeyourmetadata.org/publications/named-entity-recognition.pdf

Van Hooland S., Verborgh R. & van de Walle A. (2012): Joining the Linked Data Cloud in a Cost-Effective Manner. Information Standards Quarterly 24(2/3): 24-28, http://www.niso.org/apps/group_public/download.php/9423/IP_VanHooland-etal_%20LD-Cloud_isqv24no2-3.pdf

Van Hooland S., Verborgh R., De Wilde M. *et al*. (2013): Evaluating the success of vocabulary reconciliation for cultural heritage collections. In: ASIS&T Journal, 64(3): 464–479; preprint, http://freeyourmetadata.org/publications/freeyourmetadata.pdf

Van House, N. A. (2002): Trust and Epistemic Communities in Biodiversity Data Sharing, ACM Joint Conference on Digital Libraries, Portland, OR, 2002.

Van Noorden, Richard (2011): Science publishing: The trouble with retractions. In: Nature 478, October 2011, http://www.nature.com/news/2011/111005/full/478026a.html

Van Noorden, Richard (2013): Researchers opt to limit uses of open-access publications. Nature.com, news, 6 February 2013, http://www.nature.com/news/researchers-opt-to-limit-uses-of-open-access-publications-1.12384

Van Noorden, Richard (2014): Online collaboration: Scientists and the social network (results of a Nature survey), 13 August 2014, http://www.nature.com/news/online-collaboration-scientists-and-the-social-network-1.15711

Van Zundert, Joris (2012): If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities. In: Historical Social Research, 37(3): 165–86, http://www.cceh.uni-koeln.de/files/vanZundert_0.pdf

Vandenbroucke, Rosette (2015): Best Practices for the use of e-Infrastructures by large-scale research infrastructures. An e-IRG guidelines document, pp. 6-8, in: e-IRG Newsletter, Spring 2015, http://e-irg.eu/documents/10920/12770/e-IRG+Newsletter+2015-1+Online+version+%286+Mbyte%29.pdf

Vandewalle, Patrick (2012): Code Sharing Is Associated with Research Impact in Image Processing. In: Computing in Science & Engineering, 14(4): 42–47; available at http://infoscience.epfl.ch/record/206184/files/Vandewalle12.pdf

Velden, Theresa (2013): Explaining field differences in openness and sharing in scientific communities. CSCW 2013. Conference on Computer supported cooperative work, proceedings, pp. 445-458 ; preprint, http://www.cs.cornell.edu/~tvelden/pubs/2013-cscw_preprint.pdf

VENUS - Virtual Exploration of Underwater Sites project (EU, FP6-IST, 07/2006-06/2009), http://archaeologydataservice.ac.uk/research/venus

Veretnik S., Fink J.L. & Bourne P.E. (2008): Computational biology resources lack persistence and usability. In: PLoS Computational Biology, 4(7): e1000136, http://dx.doi.org/10.1371/journal.pcbi.1000136

Vernooy-Gerritsen M., Pronk G. & van der Graaf M. (2009): Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe. In: Ariadne, Issue 59, http://www.ariadne.ac.uk/issue59/vernooy-gerritsen-et-al

Vertesi, Janet (2010): Tweeting Spacecraft: Communicating Science in the Age of Web 2.0. In: CAP Journal - Communicating Astronomy with the Public, No.10, December 2010, pp. 30-33, http://www.capjournal.org/issues/10/10_30.pdf

ViBRANT - Virtual Biodiversity Research and Access Network for Taxonomy (EU, FP7, 12/2010-11/2013), http://vbrant.eu

Vierkant P. & Kindling M. (2014): Welche Institutionen betreiben Open-Access-Repositorien in Deutschland? In: LIBREAS. Library Ideas, 26, http://libreas.eu/ausgabe26/07vierkantkindling/

Vierkant, Paul (2013): 2012 Census of Open Access Repositories in Germany: Turning Perceived Knowledge Into Sound Understanding. In: D-Lib Magazine 19(11-12), http://www.dlib.org/dlib/november13/vierkant/11vierkant.html

Vines T. H., Albert A.Y., Andrew R.L *et al*. (2014): How Does the Availability of Research Data Change With Time Since Publication? In: Current Biology, 24(1): 94–97, http://www.cell.com/current-biology/abstract/S0960-9822%2813%2901400-0

Vines T.H., Andrew R.L., Bock D.G. *et al*. (2013): Mandated data archiving greatly improves access to research data. In: FASEB Journal, 27(4): 1304–08, http://www.fasebj.org/content/27/4/1304

Visual Computing Lab (CNR-ISTI, Pisa), http://vcg.isti.cnr.it

Vlaeminck S. & Wagner G.G. (2014): On the Role of Research Data Centres in the Management of Publication-related Research Data. Results of a survey among scientific infrastructure service providers in the field of social sciences. LIBER Quarterly 23(4), http://liber.library.uu.nl/index.php/lq/article/view/9356/9900

Volk C.J., Lucero Y. & Barnas K. (2014): Why is data sharing in collaborative natural resource efforts so hard and what can we do to improve it? In: Environmental Management, 53(5): 883-893, https://www.researchgate.net/publication/260611778_Why_is_Data_Sharing_in_Collaborative_Natural_Resource_Efforts_so_Hard_and_What_can_We_Do_to_Improve_it

Von Hippel, Eric (1986): Lead Users. A Source of novel product concepts. In: Management Science, 32(7): 791-805, https://evhippel.files.wordpress.com/2013/08/lead-users-paper-1986.pdf

Voss A., Asgari-Targhi M., Procter R. & Fergusson D. (2010): Adoption of e-Infrastructure services: configurations of practice. In: Philosophical Transactions A, 368(1926), September 2010, http://rsta.royalsocietypublishing.org/content/368/1926/4161

Voss A., Asgari-Targhi M., Procter R. *et al*. (2009): Adoption of e-Infrastructure Services: inhibitors, enablers and opportunities. In: Proc. 5th International Conference on e-Social Science, Cologne, Germany, 24-26 June 2009 http://www2.warwick.ac.uk/fac/sci/dcs/people/rob_procter/publications/adoption_final.pdf

W3C - DCAT - Data Catalog Vocabulary, http://www.w3.org/TR/vocab-dcat/

W3C - RDF - Resource Description Framework, http://www.w3.org/standards/techs/rdf#w3c_all

W3C - SKOS - Simple Knowledge Organization System, http://www.w3.org/2004/02/skos/

W3C - SPARQL Protocol and RDF Query Language, http://www.w3.org/standards/techs/sparql#w3c_all

W3C POWDER - Protocol for Web Description Resources, http://www.w3.org/TR/powder-dr/ (see also: http://www.w3.org/2007/powder/)

Waaijers L. & van der Graaf M. (2011): Quality of Research Data, an Operational Approach. In: D-Lib Magazine, 17(1/2), http://www.dlib.org/dlib/january11/waaijers/01waaijers.html

Wacha M. & Wisner M. (2011): Measuring Value in Open Access Repositories. In: The Serials Librarian, 61(3-4): 377-388, http://academiccommons.columbia.edu/download/fedora_content/download/ac:154085/CONTENT/0361526X.2011.580423.pdf

Waldrop, Mitchell (2008): Big data: Wikiomics. In: Nature 455: 22-25, http://www.nature.com/news/2008/080903/full/455022a.html

Wallis J.C. *et al*. (2007): Know Thy Sensor: Trust, Data Quality, and Data Integrity in Scientific Digital Libraries. 11th European Conference on Research and Advanced Technology for Digital Libraries, Budapest, 16-21 September 2007, http://escholarship.org/uc/item/4xx221vv

Wallis J.C. *et al*. (2008a): Moving archival practices upstream: An exploration of the life cycle of ecological sensing data in collaborative field research. In: International Journal of Digital Curation, 3(1), http://www.escholarship.org/uc/item/1cv6s3fn

Wallis J.C. *et al*. (2008b): An Exploration of the Life Cycle of eScience Collaboratory Data. iConference, UCLA, Los Angeles, 28.2.-1.3.2008, http://www.ischools.org/conference08/pc/PA10-4_iconf08.pdf

Wallis J.C., Rolando E. & Borgman C.L. (2013): If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. In: PLoS ONE, 8(7): e67332, http://dx.doi.org/10.1371/journal.pone.0067332

Wallrodt, John (2013): Beyond Tablet Computers as a Tool for Data Collection: Three Seasons of Processing and Curating Digital Data in a Paperless World [PARP:PS project]. In: Paperlessarchaeology.com, weblog, 25 March 2013 [draft of CAA-2013 paper with K. Dicus, L. Liebermann & G. Tucker], http://paperlessarchaeology.com/2013/03/25/beyond-tablet-computers-as-a-tool-for-data-collection-three-seasons-of-processing-and-curating-digital-data-in-a-paperless-world/

Wallrodt, John (2015): Videos of the workshop "Mobilizing the Past for a Digital Future: the Potential of Digital Archaeology", Wentworth Institute of Technology, Boston, 27-28 February 2015, http://paperlessarchaeology.com/2015/03/03/mobilizing-the-past-workshop-videos/

Waltman L. & Costas R. (2014): F1000 recommendations as a potential new data source for research evaluation: a comparison with citations. In: Journal of the Association for Information Science and Technology, 65(3): 433-445, http://onlinelibrary.wiley.com/doi/10.1002/asi.23040/full

Wang X., Liu C., Mao W. & Fang Z. (2015): The open access advantage considering citation, article usage and social media attention. In: Scientometrics, 103(2): 555-564; manuscript, http://arxiv.org/ftp/arxiv/papers/1503/1503.05702.pdf

Ward C., Freiman L., Jones S., Molloy L. & Snow K. (2011): Making Sense: Talking Data Management with Researchers. In: International Journal of Digital Curation, 6(2): 265-273, http://www.ijdc.net/index.php/ijdc/article/view/197

Ward, Kevin (2010): Research monographs: an overview, pp. 300-302, in: World Social Science Report. Knowledge Divides. UNESCO and International Social Science Council, http://unesdoc.unesco.org/images/0018/001883/188333e.pdf

Warinner C. & d'Alpoim Guedes J. (2015): Digitizing the Archaeobotanical Record, in: Marston J.M., d'Alpoim Guedes J. & Warinner C. (eds.): Method and Theory in Paleoethnobotany. University Press of Colorado

Warinner C., d'Alpoim Guedes J. & Goode D. (2011): Establishing Open-Access Online Reference Collections for Archaeobotanical Research. In: Vegetation History and Archaeobotany, 20(3): 41-244, (https://www.academia.edu/3497484/Paleobot.org_establishing_open-access_online_reference_collections_for_archaeobotanical_research)

Warwick C., Fisher C., Terras M. *et al.* (2009): iTrench: a study of user reactions to the use of information technology in field archaeology. In: Literary and Linguistic Computin,g 24(2): 211-223 (manuscript), http://discovery.ucl.ac.uk/13916/1/13916.pdf

Warwick C., Terras M., Galina I. *et al*. (2008): Library and information resources and users of digital resources in the humanities. In: Program: Electronic library and information systems, 42(1): 5-27; manuscript, http://eprints.ucl.ac.uk/13807/1/13807.pdf

Warwick C., Terras M., Huntington P., Pappa N. & Galina I. (2006): The LAIRAH Project: Log analysis of digital resources in the arts and humanities. Final report to the Arts and Humanities Research Council, http://www.ucl.ac.uk/slais/claire-warwick/publications/LAIRAHreport.pdf

Waters C.N., Zalasiewicz J.A., Williams M., Ellis M.A. & Snelling A.M. (eds., 2014): A Stratigraphical Basis for the Anthropocene. GSL Special Publications 395. Bath, UK: The Geological Society Publishing House, http://sp.lyellcollection.org/content/395/1.toc

Watts, Susan (2014): Society needs more than wonder to respect science. In: Nature 508 (7495): 151, http://www.nature.com/news/society-needs-more-than-wonder-to-respect-science-1.15012

Web of Science info (2011): Journal Expansion: New regional content, archived at: http://download.bioon.com.cn/upload/201101/18/171637lpbk3yn3knniplkv.attach.pdf

Weber N. & Thomer A. (2013): Lead, Lag or get out of the Index: Macroeconomic indicators of data use. Proceedings of the 2013 iConference. Ft. Worth, Texas, https://www.ideals.illinois.edu/bitstream/handle/2142/42076/487.pdf?sequence=2

Weber N.M., Baker K.S., Thomer A.K. *et al*. (2012a): Value and context in data use. Domain analysis revisited. In: Proceedings of ASIS&T 2012, Baltimore, 12-30 October 2012, https://www.asis.org/asist2012/proceedings/Submissions/168.pdf

Weber N.M., Palmer C.L. & Chao T.C. (2012b): Current Trends and Future Directions in Data Curation Research and Education. In: Journal of Web Librarianship, 6(4): 305-20, https://www.ideals.illinois.edu/bitstream/handle/2142/48759/JWL_Final_WeberEtAl.docx?sequence=2

Weber N.M., Thomer A.K., Mayernik M.S., Dattore B., Ji Z. & Worley S. (2013): The Product and System Specificities of Measuring Curation Impact. In: International Journal of Digital Curation, 8(2): 223-234, http://www.ijdc.net/index.php/ijdc/article/viewFile/8.2.223/330

Weber, Cornelia (2012): University Collections. In: European History Online (EGO), published by the Leibniz Institute of European History (IEG), Mainz, http://www.ieg-ego.eu/weberc-2012-en

Webometrics: Ranking web of repositories, http://repositories.webometrics.info

Weekes D., Sri Krishna S., Bakolitsa C. *et al*. (2010): TOPSAN: a collaborative annotation environment for structural genomics. In: BMC Bioinformatics 2010, 11:426, http://www.biomedcentral.com/1471-2105/11/426

Wellcome Trust (2013): The suitability of the CC-BY licence for research publications in the humanities and social sciences (HSS). Note of a workshop held on 24 April 2013, http://www.researchinfonet.org/wp-content/uploads/2013/02/Note-of-Wellcomeworkshop-on-CC-BY-in-hums-and-soc-sci-final.pdf

Weller, Toni (ed., 2013): History in the Digital Age. New York: Routeledge

Wells, E. Christian (2011): How can archaeologists improve the prospects for a sustainable world? American Anthropological Association, Archaeology Division, AAAnet weblog, 3 March 2011, http://blog.aaanet.org/2011/03/03/how-can-archaeologists-improve-the-prospects-for-a-sustainable-world/

Wessels B., Finn R.L., Linde P. *et al*. (2014): Issues in the development of open access to research data. In: Prometheus: Critical Studies in Innovation, 32(1): 49-66, https://pure.knaw.nl/portal/files/894399/08109028_2E2014_2E956505.pdf

Wesson C.B. & Cottier J.W. (2014): Big Sites, Big Questions, Big Data, Big Problems: Scales of Investigation and Changing Perceptions of Archaeological Practice in the Southeastern United States. In: Bulletin of the History of Archaeology, 24:16, http://www.archaeologybulletin.org/article/view/bha.2416

Wheatley, Paul (2012): Digital Preservation Cost Modelling: Where did it all go wrong? In: Open Preservation weblog, 29 June 2012, http://openpreservation.org/blog/2012/06/29/digital-preservation-cost-modelling-where-did-it-all-go-wrong/

Wheatley, Paul (2013): Digital Preservation and Data Curation Costing and Cost Modelling, http://wiki.opf-labs.org/display/CDP/Home

Whitcher-Kansa S. & Kansa E. (2009): Yes, it is all about you: User needs, archaeology and digital data. In: CSA Newsletter, Vol. XXII, No. 1, http://csanet.org/newsletter/spring09/nls0902.html

Whitcher-Kansa S. & Kansa E. (2011): Enhancing Humanities Research Productivity in a Collaborative Data Sharing Environment. White Paper to the NEH Division of Preservation and Access Advancing Knowledge. Alexandria Archive Institute, 27 June 2011, http://ux.opencontext.org/blog/wp-content/uploads/2011/06/white_paper_PK_50072.pdf

Whitcher-Kansa S. & Kansa E. (2015): Reflections on a Road Less Traveled: Alt-Ac Archaeology. In: Journal of Eastern Mediterranean Archaeology & Heritage Studies, 3(3): 293-298, http://www.jstor.org/stable/10.5325/jeasmedarcherstu.3.3.0293#full_text_tab_contents

White E.P., Baldridge E., Brym Z.T. *et al*. (2013): Nine simple ways to make it easier to (re)use your data. In: Ideas in Ecology and Evolution, 6(2): 1–10, https://peerj.com/preprints/7/

Whitlock M.C., McPeek M.A, Rausher M.D, Rieseberg L. & Moore A.J. (2010): Data archiving. In: American Naturalist, 175:145-146, http://www.jstor.org/stable/10.1086/650340

Whitlock, Michael C. (2011): Data archiving in ecology and evolution: best practices. In: Trends in Ecology and Evolution, 26(2): 61-65, http://wiki.datadryad.org/images/c/c4/WhitlockBestPractices.pdf

Whyte A., Job D., Giles S. & Lawrie S. (2008): Meeting Curation Challenges in a Neuroimaging Group. In: International Journal of Digital Curation, 3(1), http://www.ijdc.net/index.php/ijdc/article/view/74/53

Whyte, Angus (2012): Bringing data into the open repository. In: Digital Curation Centre weblog, 17 July 2012, http://www.dcc.ac.uk/blog/bringing-data-open-repository

Wicherts J.M., Bakker M. & Molenaar D. (2011): Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. In: PLoS ONE, 6(1): 1-7, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0026828

Wickham, Jackie (2010): Repository management: an emerging profession in the information sector. Online information meeting, London, 1–2 December, http://eprints.nottingham.ac.uk/1511/3/Online_paper_Jackie_Wickham.pdf

Wiesmann U., Hirsch Hadorn G., Hoffmann-Riem H. *et al*. (2008): Enhancing Transdisciplinary Research: A Synthesis in Fifteen Propositions, pp. 433-441, in: Handbook of Transdisciplinary Research. Hirsch Hadorn G. *et al*.: Handbook of Transdisciplinary Research. Dordrecht: Springer, http://www.transdisciplinarity.ch/d/Transdisciplinarity/documents/HB_Propositions.pdf

WikiArc - Wiki Archaeological Information Resource, http://www.wikiarc.org

Wikiquote: William Gibson: https://en.wikiquote.org/wiki/William_Gibson

Wilbanks, John (2013): Licence restrictions: A fool's errand. In: Nature 495: 440–441, 27 March 2013, http://www.nature.com/nature/journal/v495/n7442/full/495440a.html

Wiley, Steven (2009): Why Don't We Share Data? There are so, so many reasons—and they make a lot of sense. In: The Scientist - Magazine of the Life Sciences, 23(4): 33, http://classic.the-scientist.com/article/display/55535/

Willems W.J.H. (2007): The times they are a-changin': Observations on archaeology in a European context, pp. 5-23, in: Cooney, Gabriel (ed.): Archaeology in Ireland. A vision for the future. Royal Irish Academy Forum. Dublin 2007, http://www.willemwillems.com/uploads/6/8/1/6/6816862/willems2007d.pdf

Willems W.J.H. (2014): The future of world heritage and the emergence of transnational heritage regimes. In: Heritage & Society 7(2): 105-120, http://www.networkedheritage.org/wp-content/uploads/2014/12/Willems-HS7-2014.pdf

Willems, W.J.H. (2008): Language and Archaeology. In: Archaeologies: Journal of the World Archaeological Congress, 4(1): 179-181, https://openaccess.leidenuniv.nl/bitstream/handle/1887/14256/Willems2008b.pdf?sequence=1

Willett, Perry (2015): Merritt Service Update: February 2015. California Digital Library, CDL-INFO News, 18 March 2015, http://www.cdlib.org/cdlinfo/2015/03/18/merritt-service-update-february-2015/

Williams A., Wallin J., Yu H. *et al*. (2014): A Computational Pipeline for Crowdsourced Transcriptions of Ancient Greek Papyrus Fragments. 2nd Workshop on Big Humanities Data (27 October 2014), held in conjunction with the IEEE International Conference on Big Data (IEEE BigData 2014), Washington DC, USA, https://bighumanities.files.wordpress.com/2014/10/williams.pdf

Williams P., Stevenson I., Nicholas D., Watkinson A. & Rowlands I. (2009): The role and future of the monograph in Arts & Humanities research. In: Aslib Proceedings: New Information Perspectives, 61(1): 67-82, https://www.researchgate.net/publication/235286907_The_role_and_future_of_the_monograph_in_arts_and_humanities_research

Williford C. & Henry C. / CLIR - Council on Library and Information Resources (2012): One Culture: Computationally Intensive Research in the Humanities and Social Sciences. A Report on the Experiences of First Respondents to the Digging Into Data Challenge. CLIR: Washington, D.C., June 2012, http://www.clir.org/pubs/reports/pub151

Willis C., Greenberg J. & White H. (2012): Analysis and synthesis of metadata goals for scientific data. In: Journal of the American Society for Information Science and Technology, 63(8): 1505-1520

Wilson A.T. & Edwards B. (eds., 2015): Open Source Archaeology. Ethics and Practice. De Gruyter Open, http://www.degruyter.com/viewbooktoc/product/460080

Wilson G., Aruliah D.A., Brown C.T. *et al*. (2014): Best Practices for Scientific Computing. In: PLoS Biology 12(1): e1001745, http://dx.doi.org/10.1371/journal.pbio.1001745

Wilson, Scott (2014): Preserving and Curating Software. OSS Watch website, guidance material, 5 November 2014, http://oss-watch.ac.uk/resources/preservation

Wira-Alam, Andias (2014): Dublin Core Metadata for Research Data – Lessons Learned in a Real-World Scenario with datorium, pp. 64-73, in: DC-2014 – International Conference on Dublin Core and Metadata Appliations. Papers, Project Reports and Posters for DC-2014 in Austin, Texas, 8-11 October 2014, http://dcevents.dublincore.org/IntConf/dc-2014/paper/view/271/231

Wissenschaftsrat / German Council of Science and Humanities (2011a): Recommendations on Research Infrastructures in Humanities and Social Sciences. Drs. 10465-11, Berlin, 28 January 2011, http://www.wissenschaftsrat.de/download/archiv/10465-11_engl.pdf

Wissenschaftsrat / German Council of Science and Humanities (2011b): Recommendations on Scientific Collections as Research Infrastructures. Drs. 10464-11. Berlin, 28 January 2011, http://www.wissenschaftsrat.de/download/archiv/10464-11-11_engl.pdf

WissKI - Scientific Communication Infrastructure (Germany, DFG-funded project, 2009-2012, 2014-), http://wiss-ki.eu

Witt M., Carlson J., Brandt D.S & Cragin M.H. (2009): Constructing Data Curation Profiles. In: International Journal of Digital Curation, 4(3), http://www.ijdc.net/index.php/ijdc/article/view/137

Worldwide Protein Data Bank (international repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies), http://www.wwpdb.org

Wouters P. & Beaulieu A. (2006): Imagining e-science beyond computation, 48-70, in: Christine, H. (ed.), New infrastructures for knowledge production: Understanding e-science. London: Information Science Publishing, http://virtualknowledgestudio.nl/documents/_annebeaulieu/new-infrastructures-for-knowledge-production.pdf

Wouters P. & Costas R. (2012): Users, narcissism and control – tracking the impact of scholarly publications in the 21st century. SURF Foundation, http://research-acumen.eu/wp-content/uploads/Users-narcissism-and-control.pdf

Wouters P., Beaulieu A. Scharnhorst A. & Wyatt S. (eds., 2012): Virtual Knowledge. Experimenting in the Humanities and Social Sciences. Cambridge: MIT Press

Wouters P., Vann K., Scharnhorst A. *et al*. (2008): Messy shapes of knowledge – STS explores informatization, new media, and academic work. The Virtual Knowledge Studio, 319-351, in: Handbook of science and technology studies. MIT Press

Wouters, Paul (2007): Conservative culture or diverging identities? Studying resistance to technology in academia. E-Social Science conference 2007, Ann Arbor, Michigan, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.8644&rep=rep1&type=pdf

Woutersen-Windhouwer S., Brandsma R., Verhaar P. *et al*. (2009): Enhanced publications: Linking publications and research data in digital repositories. DRIVER & Amsterdam University Press, http://dare.uva.nl/document/150723

Wright S.J., Kozlowski W.A., Dietrich D. *et al*. (2013): Using Data Curation Profiles to Design the Datastar Dataset Registry. In: D-Lib Magazine 19(7-8), http://www.dlib.org/dlib/july13/wright/07wright.html

Wuchty S., Jones B. & Uzzi B. (2007): The Increasing Dominance of Teams in Knowledge Production. In: Science, vol. 316: 1036-1039, http://www.sciencemag.org/content/316/5827/1036

Xia J., Gilchrist S.B., Smith N. *et al*. (2012): A Review of Open Access Self-Archiving Mandate Policies. In: Libraries and the Academy, 12(1): 85-102, http://www.press.jhu.edu/journals/portal_libraries_and_the_academy/portal_pre_print/current/articles/12.1xia.pdf

Xia, Jingfeng (2007): Assessment of Self-archiving in Institutional Repositories: Across Disciplines. In: Journal of Academic Librarianship, 33(6): 647–654, http://ejournals.library.ualberta.ca/index.php/EBLIP/article/downloadSuppFile/1486/365

Xu L., Liu J. & Fang Q. (2011): Analysis on Open Access citation advantage: an empirical study based on Oxford Open journals, pp. 426-432, in: iConference '11, Proceedings of the 2011 iConference, Seattle, 11 February 2011, http://dl.acm.org/citation.cfm?id=1940819

Yakel E., Faniel I.M., Kriesberg A. & Yoon A. (2013): Trust in Digital Repositories. In: International Journal of Digital Curation, 8(1): 143-156, http://www.ijdc.net/index.php/ijdc/article/view/8.1.143

Yozwiak N.L., Schaffner S.F. & Sabeti P.C. (2015): Data sharing: make outbreak research open access. In: Nature 518: 477-479, http://www.nature.com/news/data-sharing-make-outbreak-research-open-access-1.16966

Zagora Archaeological Project (Andros island, Greece), http://www.powerhousemuseum.com/zagora/

Zamir, Dani (2013): Where have all the crop phenotypes gone? In: PLoS Biology, 11(6): e1001595, http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001595

Zeller, Jean-Daniel (2010): Cost of digital archiving: Is there a universal model? 8th European Conference on Digital Archiving, Geneva, 28-30 April 2010, http://regarddejanus.files.wordpress.com/2010/05/costsdigitalarchiving-_jdz_eca2010.pdf

Zenodo (data repository at CERN, related to OpenAIRE), http://www.zenodo.org

Zilinski L.D., Gobens A. & Briney K. (2015): University Data Policies and Library Data Services: Who Owns Your Data? In: ASIS&T Bulletin, August/September 2015, online, https://www.asist.org/publications/bulletin/aug-2015/who-owns-your-data/

Zimmerman, Ann (2003): Data Sharing and Secondary Use of Scientific Data: Experiences of Ecologists. Ph.D. Dissertation, University of Michigan, Ann Arbor/MI, http://deepblue.lib.umich.edu/bitstream/2027.42/61844/1/ann_zimmerman_dissertation_2003.pdf

Zimmerman, Ann (2007): Not by Metadata Alone: The Use of Diverse Forms of Knowledge to Locate Data for Reuse. In: International Journal on Digital Libraries, 7(1-2): 5-16, http://scientificdatasharing.com/wp-content/uploads/2011/02/Zimmerman_Not-by-metadata-alone_2007.pdf

Zimmerman, Ann (2008): New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. In: Science, Technology & Human Values, 33(5): 631-652, http://scientificdatasharing.com/wp-content/uploads/2011/02/Zimmerman_New-knowledge-from-old-data_2008.pdf

Zooniverse (“platform for people-powered research”), https://www.zooniverse.org/projects/

Zuccala A. & van Leeuwen T. (2011): Book Reviews in Humanities Research Evaluations. In: Journal of the American Society for Information Science and Technology, 62(10): 1979-91; preprint, http://www.illc.uva.nl/evaluating-humanities/BookReviewsPreprint.pdf

Zuccala, Alesia (2013): Evaluating the Humanities: Vitalizing ‘the forgotten sciences’. In: Research Trends, Issue 32, http://www.researchtrends.com/issue-32-march-2013/evaluating-the-humanities-vitalizing-the-forgotten-sciences/