D15.2: Report on the ARIADNE Linked Data Cloud

Authors:

Franca Debole, CNR-ISTI
Carlo Meghini, CNR-ISTI
Guntram Geser , SRFG
Douglas Tudhope, USW

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

Version: 1.0 (*final*) **27th January 2017**

Authors: **Franca Debole and Carlo Meghini, CNR-ISTI**

Guntram Geser , SRFG

Douglas Tudhope, USW

Contributing Partners **Ceri Binding, USW**

Sara Di Giorgio, MIBAC-ICCU

Achille Felicetti , PIN

Dimitris Gavrilis , ATHENA RC

Philipp Gerth, DAI

Maria Theodoridou, FORTH

Quality review: **Holly Wright, ADS**

Paola Ronzino, PIN

Table of content

Acronyms of ARIADNE partners

AIAC	Associazione Internazionale di Archeologia Classica (Italy)
ARHEO	Arheovest Timisoara Association (Romania)
ARUP-CAS	Archeologicky ustav AV CR, Praha, v.v.i. / Institute of Archaeology of the Academy of Sciences (Czech Republic)
Athena-DCU	Athena Research and Innovation Center in Information Communication and Knowledge Technologies / Digital Curation Unit (Greece)
CNR	Consiglio Nazionale delle Ricerche institutes, CNR-ISTI and CNR-ITABC (Italy)
CSIC-Incipit	Consejo Superior de Investigaciones Cientificas / Spanish National Research Council, Institute of Heritage Sciences (Spain)
CYI-STARC	The Cyprus Institute, Science and Technology in Archaeology Research Center
DAI	Deutsches Archäologisches Institut (Germany)
Discovery	The Discovery Programme LBG (Ireland)
FORTH-ICS	Foundation for Research and Technology Hellas, Institute of Computer Science (Greece)
INRAP	Institut National des Recherches Archéologiques Préventives (France)
KNAW-DANS	Netherlands Academy of Arts and Sciences, Data Archiving and Networked Services (Netherlands)
LeidenU	Leiden University, Faculty of Archaeology (Netherlands)
MiBAC-ICCU	Italian Ministry of Cultural Assets and Activities - Central Institute for the Union Catalogue (Italy)
MNM-NOK	Magyar Nemzeti Múzeum, Nemzeti Örökségvédelmi Központ / Hungarian National Museum, National Heritage Protection Centre (Hungary)
NIAM-BAS	National Institute of Archaeology with Museum of the Bulgarian Academy of Sciences (Bulgaria)
ÖAW-OREA	Österreichische Akademie der Wissenschaften, Institut für Orientalische und Europäische Archäologie (Austria)
PIN	PIN - Servizi Didattici e Scientifici per l'Università di Firenze s.c.r.l. (Italy)
SND	Swedish National Data Service (Sweden)
SRFG	Salzburg Research Forschungsgesellschaft m.b.H. (Austria)
USW	University of South Wales (United Kingdom)
ADS-UoY	Archaeology Data Service, University of York (United Kingdom)
ZRC-SAZU	Scientific Research Centre of the Slovenian Academy of Sciences and Arts, Institute of Archaeology (Slovenia)

Executive Summary

This report has been produced within the ARIADNE project as part of Work Package 15, “Linking Archaeological Data”. This document is a deliverable (D15.3) of the ARIADNE project (“Advanced Research Infrastructure for Archaeological Dataset Networking in Europe”), which is funded under the European Community's Seventh Framework Programme. It presents the results of the work carried out in Task 15.3 “ARIADNE Linked Data Cloud”. The overall objective of ARIADNE is to help making archaeological data better discoverable, accessible and re-useable. The project addresses the fragmentation of archaeological data in Europe and promotes a culture of open sharing and (re-)use of data across institutional, national and disciplinary boundaries of archaeological research. More specifically, ARIADNE implements an e-infrastructure for data interoperability, sharing and integrated access via a data portal. Linked Open Data can greatly contribute to these goals.

Lessons learned, recommendations and brief conclusions are included at the end of every section.

1 Introduction

Towards a web of archaeological Linked Open Data – a vision

The ARIADNE Linked Open Data "cloud" is envisioned as a web of semantically interlinked resources of and for archaeological research. Archaeology is a multi-disciplinary field of research, hence the web of Linked Data initiated by different projects, including ARIADNE, spans data resources of various domains and specialties, for example history and geography of the ancient world, classics, medieval studies, cultural anthropology and various data from the application of natural science methods to archaeological research questions (e.g. physical, chemical and biological sciences).

One of the main objectives of the ARIADNE project has been to provide the archaeological sector with a data infrastructure and portal for discovering and accessing datasets which are being shared by research institutions and digital archives located in different European countries. The infrastructure and portal are not stand-alone implementations but serve as a node in the ecosystem of e-infrastructure services for archaeology and various related disciplines, including other humanities as well as social, natural, environmental and life sciences. To become such a node, interoperability with external services is required and can be implemented based on the Linked Data approach.

Linked Data support in ARIADNE

WP15 supports the development of Linked Open Data within and beyond the project. The activities of this strand of work concerned:

- the metadata of the datasets registered in the ARIADNE data catalogue,
- vocabularies for the metadata describing registered datasets (e.g. mapping of existing vocabularies, support for the generation of vocabularies in SKOS),
- mapping of datasets to the core CIDOC CRM and extensions of the CRM created in ARIADNE,
- demonstrators generating and using Linked Data (e.g. metadata extracted from unstructured data such as grey literature, exploration of CIDOC CRM based data), and
- providing access to ARIADNE Linked Data for external application developers.

Thus the work centred on Linked Data related to data registration, enabling data integration via vocabularies and the CIDOC CRM ontology, demonstration of enhanced or new capabilities, and making the ARIADNE data catalogue and other results of these activities accessible through a graph database or "cloud" of Linked Data.

Current level of LOD adoption in archaeology

The last 10 years have seen substantial progress in LOD expertise, i.e. what is required to produce, publish and interlink LOD from cultural heritage collections (e.g. museum artefact collections). This expertise has been acquired mostly through experimental projects, and only a few cultural heritage datasets are effectively interlinked as yet. With regard to archaeological data specifically, few Linked Data datasets have been produced and hardly any show up on the well-known LOD Cloud diagram. In coming years a much wider uptake of the LOD approach in the domain is necessary, so that a rich web of data can emerge.

Requirements for a wider uptake

WP15 activities took into account factors that currently impede the development of a web of semantically interlinked archaeological data. Therefore the present report particularly addresses requirements for a wider uptake of a Linked Data approach in archaeology. The study of these requirements will be valuable for many who have taken an interest in Linked Open Data (LOD), would like an overview of the current situation in cultural heritage and archaeology, and recommendations on how to advance the availability and interlinking of LOD in this field.

Specific actions are recommended to:

- raise awareness of Linked Data,
- clarify the benefits and costs of Linked Data,
- enable non-IT experts use Linked Data tools,
- promote Knowledge Organization Systems as Linked Open Data,
- foster reliable Linked Data for interlinking,
- promote Linked Open Data for research.

Among the various requirements, the importance of fostering a community of LOD curators who take care for proper generation, publication and interlinking of archaeological datasets and vocabularies were highlighted.

Lessons learned in the development of LOD within ARIADNE

One finding is the critical importance of the subject vocabularies, e.g. the Getty Art and Architecture Thesaurus (AAT), combined with the CIDOC CRM ontology entities, which act as linking hubs for the web of data. This is the most obvious route to connection with external LOD. More work is needed on the identification of further linking hubs, for example the Period0 set of cultural periods. The mapping of datasets to such hubs requires domain knowledge, easy to use tools, and guidance for users who are carrying out such work for the first time. While recommended tools are helpful, fully automated mapping appears unlikely to achive quality results at the current time. There is much scope to explore the utility of LOD in practice, taking account of the objectives and requirements of different user communities. There is still a way to go before advanced uses of LOD will become applicable and beneficial in online research environments; more effort must be invested to make this happen. In order to motivate user organisations to work with Linked Data, exemplar working applications are needed that address a real user (scientific/research) need. Such exemplars might be end user applications or programmatic interfaces to the underlying LOD.

Building the ARIADNE LOD Cloud – lessons learned

While the Linked Open Data standards are essential for integrating data, the technology supporting such integration is still in its infancy. The ARIADNE LOD, comprised of LOD derived from the ARIADNE catalogue, is represented by three demonstrators and various vocabularies, and has resulted in the creation of about 32 million RDF triples. While any relational database can easily handle millions of records, the corresponding volume of RDF in a current triple store can cause serious efficiency problems as experienced in the experimentation with the ARIADNE Linked Data Cloud, and that this is the price to be paid for interoperability. More robust and efficient graph databases are required if we want to proceed towards Big Data as Linked Data. This is the first major lesson learned while implementing the ARIADNE Linked Data Cloud.

The second lesson comes from the graph data model. This model is intrinsically binary, which makes it difficult to express higher rank relations, and to easily implement data connection patterns. In the latter case, the patterns may involve data chains that span several arcs, and their definition and implementation is not trivial. Conversely, correlations between data items can be epitomized by such paths, which need to be detected, and this is a computationally very intensive task if the length of the paths go beyond 2-3 arcs. This fact has always been known from a theoretical point of view, but working with real data we could experience it in practice.

2 Vision, study summaries, and recommendations

This chapter summarises the research and development results presented in this report. It highlights a vision of a web of archaeological Linked Open Data (LOD), addresses the LOD principles and web of Linked Data (the "LOD Cloud"), the adoption of the LOD approach so far in archaeology, and requirements for a wider uptake in the sector. Moreover the chapter summarises the LOD development in ARIADNE and how the generated data is being made available beyond the project. The sections also provide recommendations on how to increase the adoption of the LOD approach in archaeology and lessons learned in the work on LOD in the ARIADNE project.

2.1 Archaeological Linked Open Data – a vision

This report envisions the emergence of a web of semantically interlinked resources of and for archaeological research based on the Linked Data approach. Over the next 5-10 years a web of Linked Open Data could be built that spans vocabularies and data of archaeological, cultural heritage and related fields of research.

About 10 years ago there were considerable doubts about the uptake of Semantic Web standards and technologies. Reasons for this doubt were centred on the still on-going standardisation work, little experience of implementation under real world conditions, and expected high costs of conversion of legacy metadata and knowledge organization systems (e.g. thesauri) to Semantic Web standards.

In recent years the Linked Data approach has seen substantial progress with regard to mature standards, available expertise and tools, and examples of data publication and linking. Recognition and uptake of the approach has grown far beyond the initially small pioneering groups of Linked Data developers. The Open Data movement has been an important driver for this development, particularly through the involvement of governmental and public sector agencies, who have promoted standards and implemented data catalogues and portals.

The Linked Data approach has been embraced by several research communities, for example, geo-spatial, environmental and some natural sciences (e.g. bio-sciences). Also the cultural heritage sector, particularly the library and museum domains, have been among the early adopters. Thus there is already potential for interlinking and enriching archaeological research data with specific information, as well as within a wider context.

Archaeology is a multi-disciplinary field of research, hence the web of Linked Open Data could include resources of various domains and specialties, for example history and geography of the ancient world, classics, medieval studies, cultural anthropology and various data from the application of natural sciences methods to archaeological research questions (e.g. physical, chemical and biological sciences). Also data of geo-spatial, environmental and earth sciences are relevant to several fields of archaeological research.

But wide and deep interlinking will require rich integration of conceptual knowledge (ontologies) and terminologies from different domains. Integration could be progressed based on use cases with a clear added value for archaeological and other research communities. Such use cases would support interdisciplinary research involving researchers in archaeology and other domains, natural history and environmental change, for instance.

As a multi-disciplinary area of research, archaeology could benefit greatly from a comprehensive web of Linked Open Data, involving data and vocabularies of all related disciplines. However, first there is still a lot of homework to do by research institutions, projects and archives so that an archaeological web of Linked Open Data will emerge and become interlinked with resources of other disciplines as well as relevant public sector information.

2.2 Study summaries and recommendations

2.2.1 Linked Open Data: Background and principles

Brief summary

The term Linked Data refers to principles, standards and tools for the generation, publication and and linking of structured data based on the W3C Resource Description Framework (RDF) family of specifications.

The basic concept of Linked Data was defined by Tim Berners-Lee in an article published in 2006. This concept helped to re-orientate and channel the initial grand vision of the Semantic Web into a productive new avenue. Previously the research and development community presented the Semantic Web vision as a complex stack of standards and technologies. This stack seemed always "under construction" and together with the difficult to comprehend Semantic Web terminology, created the impression of an academic activity with little real world impact.

In 2010 Berners-Lee's request for Linked *Open* Data aligned Linked Data with the Open Data movement. Since then, the quest for Linked Open Data (LOD) has become particularly strong in the governmental / public sector as well as initiatives for cultural and scientific LOD.

Linked Data principles include that a data publisher should make the data resources accessible on the Web via HTTP URIs (Uniform Resource Identifiers), which uniquely identify the resources, and use RDF to specify properties of resources and of relations between resources. In order to be Linked Data proper, the publishers should also link to URI-identified resources of other providers, hence add to the "web of data" and enable users to discover related information. And to be Linked Open Data the publisher must provide the data under an open license (e.g. Creative Commons Attribution [CC-BY] or release it into the Public Domain).

The Linked Data approach allows opening up "data silos" to the Web, interlinking of otherwise isolated data resources, and enables re-use of the interoperable data for various purposes. The landscape of archaeological data is highly fragmented. Therefore Linked Data are seen as a way to interlink dispersed and heterogeneous archaeological data and, based on the interlinking, enable discovery, access to and re-use of the data.

Building semantic e-infrastructure and services for a specific domain such as archaeology requires cooperation between domain data producers/curators, aggregators and service providers. Cooperation is necessary not only for sharing datasets through a domain portal (i.e. the ARIADNE data portal), but also to use common or aligned vocabularies (e.g. ontologies, thesauri) for describing the data so that it becomes interoperable.

In addition to the basic Linked Data principles there are also specific recommendations for vocabularies. Particularly important is re-using or extending wherever possible established vocabularies before creating a new one. The rationale for re-use is that different resources on the web of Linked Data which are described with the same or mapped vocabulary terms become interlinked. This makes it easier for applications to identify, process and integrate Linked Data. Moreover, re-use and extension of existing vocabularies can lower vocabulary development costs.

It is also recommended to provide metadata for Linked Data of datasets as well as vocabularies. The Vocabulary of Interlinked Datasets (VoiD) is often being used to provide such metadata. It is also good practice to register sets of Linked Data in a domain data catalogue and/or general registries such as the DataHub. Furthermore the publisher should announce the dataset via relevant mailing lists, newsletters etc. and invite others to consider linking to the dataset.

Linked Data should not be published "just in case". Rather publishers should consider the re-use potential and intended or possible users of their data. As Linked Data consumers they need to address the question of which data of others they could link to. These questions make clear the importance of joint initiatives for providing and interlinking datasets of certain domains such as archaeology.

Recommendations

- *Use the Linked Data approach to generate semantically enhanced and linked archaeological data resources.*
- *Participate in joint initiatives for providing and interlinking archaeological datasets as Linked Open Data.*
- *Choose datasets which allow generating value if made openly available as Linked Data and connected with other data, including linking of the datasets by others.*
- *Re-use existing Linked Data vocabularies wherever possible in order to enable interoperability.*
- *Describe the Linked Data with metadata, including provenance, licensing, technical and other descriptive information.*
- *Register the dataset in a domain data catalogue and/or general registries such as the DataHub. Also announce the dataset via relevant mailing lists, newsletters etc. and invite others to consider linking to the dataset.*

2.2.2 The Linked Open Data Cloud

Brief summary

The Linked Open Data Cloud is formed by datasets that are openly available on the Web in Linked Data formats and contain links pointing at other such datasets. One task of the ARIADNE project is to promote the emergence of a web of interlinked archaeological datasets which comply with the Linked Open Data (LOD) principles. It is anticipated that this web of archaeological LOD will become part of the wider LOD Cloud and interlinked with related other data resources.

The latest LOD Cloud diagram (2014) includes only few sets of cultural heritage LOD and they do not form a closely linked web of Linked Data. None of the datasets concerns archaeology specifically. Additional sets of cultural heritage Linked Data exist, a few of which are archaeological, but in 2014 they did not conform to the criteria for being included in the LOD Cloud diagram (e.g. the requirement of being connected via RDF links with at least one other compliant dataset).

Maybe the next version of the LOD Cloud diagram will contain some of the earlier and more recent sets of archaeological Linked Open Data. Hopefully this will include some relevant vocabularies which recently have been transformed to Linked Data in SKOS format. In 2014 the only cultural heritage vocabulary on the diagram was the Art & Architecture Thesaurus (AAT), which has the potential to become one of the core linking hubs for cultural heritage information in the LOD Cloud.

The LOD Cloud is not a single entity but represents datasets of different providers that are made available in different ways (e.g. LD server, SPARQL endpoint, RDF dump) and the resources may be unreliable, e.g. some SPARQL endpoints are off-line. There is no central management and quality control of the LOD Cloud. Webs of reliable and richly interlinked datasets are only present where there is a community of Linked Data producers and curators (e.g. in the areas of bio-medical & life sciences or libraries).

Cultural heritage is not yet an area of densly interlinked and reliable LOD resources; so far a community of cooperating LOD producers and curators has not solidified. Targeted activities to foster and support further publication and interlinking of datasets are required so that a web of archaeological, cultural heritage and other relevant data will become more established within the overall Linked Open Data Cloud.

Recommendations

- *Encourage more archaeological institutions and repositories to publish the metadata of their datasets (collections, databases) as Linked Open Data; also promote publication of domain and proprietary vocabularies of institutions as LOD.*
- *Foster the formation of a community of archaeological LOD producers and curators who generate, publish and interlink LOD, including linking/mapping between vocabularies.*

2.2.3 Adoption of the Linked Data approach in archaeology

Brief summary

In the areas addressed by this study, cultural heritage institutions are among the leading adopters of the Linked Data approach. The Ancient World and Classics research community is a front-runner of uptake on the research side, while there have been only few projects around Linked Data using archaeological research data.

This situation is due to considerable differences between cultural heritage institutions and research projects, and between projects in different domains of research. For cultural heritage institutions such as a libraries, archives and museums adoption of Linked Data is in line with their mission to make information about heritage readily available and relevant to different user groups, including researchers. Adoption has also been promoted by initiatives such as LOD-LAM, the International LOD in Libraries, Archives, and Museums Summit (since 2011). In the field of archaeological research there were no such initiatives or only at small scale, for example sessions at CAA conferences or national thematic workshops. But promotional activities, particularly at the national level, are important to reach archaeological institutes and research groups and make them aware of the Linked Data approach.

Adoption in the Ancient World and Classics research community is being driven by specialities such as numismatics and epigraphy, where there are initiatives to establish common descriptive standards based on Linked Data principles. The goal is to enable annotation and interlinking of information of special collections or corpora for research purposes. This community has led the way by focussing on certain types of artefacts (inscriptions, coins, ceramics and others), which provide clear advantages with regard to the ease of using the Linked Data approach.

A good deal of the recognition of the Ancient World and Classics research community being a front-runner in Linked Data stems from the Pelagios initiative. Pelagios provides a common platform and tools for annotating and connecting various textual resources (both the classical text and scholarly references) based on place references. Pelagios clearly demonstrates benefits of contributing and associating data derived from different contributors based on a light-weight Linked Data approach.

The data generated by the myriad forms of Archaeological fieldwork present a more difficult situation, in that a basic unit of research can be a site or an entire landscape, where archaeologists may document a variety of structures, cultural remains, artefacts and biological material, using a variety of methods. The heterogeneity of the archaeological data and the "site" as a focus of analysis presents a situation where the benefits of Linked Data, which would require semantic annotation of the variety of different data with common vocabularies, are less apparent. Therefore adoption of the Linked Data approach can be hardly found at the level of individual archaeological excavations and other fieldwork, but, in a few cases, community-level data repositories and databases of research institutes. Repositories and databases, not individual projects, should also in next years be the prime target when promoting the Linked Data approach.

All proponents of the Linked Data approach, including the ARIADNE Linked Data SIG as well as the directors of the Pelagios initiative, agree that much more needs to be done to raise awareness of the approach, promote uptake, and provide practical guidance and easy to use tools for the generation, publication and interlinking of Linked Data.

Recommendations

- *More needs to be done to raise awareness and promote uptake of the Linked Data approach for archaeological research data. In addition to sessions at international conferences, promote the approach to stakeholders such as archaeological institutes at the national level.*
- *The prime target when promoting the approach should be persistent data repositories and databases of research institutes (not individual projects).*
- *To drive uptake provision of practical guidance and easy to use tools for the generation, publication and interlinking of Linked Data is necessary.*
- *Promote the use of established and emerging semantic description and annotation standards for artefacts such as coins, inscriptions, ceramics and others; for biological remains of plants, animals and humans suggest using available relevant biological vocabularies (e.g. authoritative species taxons, life science ontologies, and others).*
- *Contribute to the Pelagios platform (where appropriate) or aim to establish similar high-visibility data linking projects for archaeological research data.*

2.2.4 Requirements for wider uptake of the Linked Data approach

Raise awareness of Linked Data

Brief summary

Linked Data enables interoperability of dispersed and heterogeneous information resources, allowing the resources to become more discoverable, accessible and re-useable. In the fragmented data landscape of archaeology this is substantial task. In the ARIADNE online survey, in addition to the expectations of the archaeological research community around the creation of a data portal, were cross-searching of data archives with innovative, more powerful search mechanisms. But such expectations were not necessarily associated with capabilities offered by Linked Data. Therefore the gap between advantages expected from advanced services and "buy in" and support of the research community for Linked Data must be closed by targeted actions.

A small survey of the AthenaPlus project (2013) indicated that cultural heritage organisations are already aware of Linked Data, but few had first-hand experience with such data. Among the expectations from connecting their own and external Linked Data resources, was increasing the visibility of collections and creating relations with various other information resources. Some respondents also considered possible disadvantages, e.g. loss of control over their own data or a decrease in data quality due to links to non-authoritative sources.

In the ARIADNE online survey (2013) "Improvements in linked data", i.e. interlinking of information based on Linked Data methods to enable better information services, was considered more helpful by repository managers than researchers. Researchers perceived interlinking of information as important, but may not see this as an area for their own research. Indeed, individual researchers and research groups should may not be thought of as a primary focus of Linked Data initiatives. Managers of digital archives for the research community and institutional repositories are much more relevant target groups. Furthermore data managers of large and long-term archaeological projects should be addressed as they will also consider required standards for data management and interlinking more thoroughly.

Recommendations

- *Address the highly fragmented landscape of archaeological data and highlight that Linked Data can allow dispersed and heterogeneous data resources become better integrated and accessible.*
- *Consider as primary target group of Linked Data initiatives not individual researchers but managers of digital archives and institutional repositories.*
- *Include also data managers and IT staff of large and long-term archaeological projects as they will also consider required standards for data management and interlinking more thoroughly.*

Clarify the benefits and costs of Linked Data

Brief summary

There is a widespread notion of an unfavourable ratio of costs compared to benefits of employing Semantic Web / Linked Data standards for information management, publication and integration. This notion should be removed as it is a strong barrier to a wider adoption of the Linked Data approach.

The basic assumption of Linked Data is that the usefulness and value of data increases the more readily it can combined with relevant other data. Convincing tangible benefits of Linked Data materialise if information providers can draw on own and external data for enriching services. There are examples for such benefits, e.g. in the museum context, but not yet for archaeological research data. Importantly, in the realm of research benefits of Linked Data are less about enhanced search services but research dividends, e.g. discovery of interesting relations or contradictions between data.

Linked Data projects typically mention some benefits (e.g. integration of heterogeneous collections, enriched information services), but very little is known about the costs of different projects. There is a clear need to document a number of reference examples, for example, what does it cost to connect datasets via shared vocabularies or integrate databases through mapping them to CIDOC CRM, and how does that compare to perceived benefits? Although vocabularies play a key role in Linked Data astonishing little is also known about the costs of employing various KOSs.

Some methods and tools appear to have reduced the cost of Linked Data generation considerably, OpenRefine or methods to output data in RDF from relational databases, for instance. As there is a proliferation of tools potential Linked Data providers need expert advice on what to use (and how to use it) for their purposes and specific datasets, taking account also of existing legacy systems and standards in use.

Recommendations

- *Proponents of the Linked Data approach should address the widespread notion of an unfavourable ratio of costs compared to benefits of employing Semantic Web / Linked Data standards.*
- *Major benefits of Linked Data can be gained from integration of heterogeneous collections/ databases and enhanced services through combining own and external data. But examples that clearly demonstrate such benefits for archaeological data are needed.*
- *In order to evaluate the costs, information about the cost factors and drivers should be collected and analysed. A good understanding of the costs of different Linked Data projects will help reduce the costs, for example by providing dedicated tools, guidance and support for certain tasks.*
- *More information would be welcome on how specific methods and tools have allowed institutions reducing the costs of Linked Data in projects of different types and sizes.*
- *General requirements for progress are more domain-specific guidance and reference examples of good practice.*

Enable non-IT experts use Linked Data tools

Brief summary

Showcase examples of Linked Data applications in the field of cultural heritage (e.g. museum collections) so far depended heavily on the support of experts who are familiar with the Linked Data methods and required tools (often their own tools). But such know-how and support is not necessarily available for the many cultural heritage and archaeology institutions and projects across Europe. A much wider uptake of Linked Data will require approaches that allow non-IT experts (e.g. subject experts, curators of collections, project data managers) do most of the work with easy to use tools and little training effort.

A number of projects have reported advances in this direction based on the provision of useful data mapping recipes and templates, proven tools, and guidance material. For example, the STELLAR Linked Data toolkit has been employed in several projects and appears to be useable also by non-experts with little training and additional advice.

Good tutorials and documentation of projects are helpful, but the need for expert guidance in various matters of Linked Open Data is unlikely to go away. For example, there are a lot of immature, not tried and tested software tools around. Therefore advice of experts is necessary on which tools are really proven and effective for certain tasks, and providers of such tools should offer practical tutorials and hands-on training, if required. Experienced practitioners can also help projects navigate past dead ends and steer project teams toward best practices.

Also more needs to be done with regard to integrating Linked Data vocabularies in tools for data recording in the field and laboratory. Like other researchers archaeologists typically show little enthusiasm to adopt unfamiliar standards and terminology, which is perceived as difficult, time-consuming, and may not offer immediate practical benefits.

Proposed tools therefore need to fit into normal practices and hide the semantic apparatus in the background, while supporting interoperability when the data is being published. Noteworthy examples are the FAIMS mobile data recording tools and the RightField tool for semantic annotation of laboratory spreadsheet data.

Recommendations

- *Focus on approaches that allow non-IT experts do most of the work of Linked Data generation, publication and interlinking with little training effort and expert support.*
- *Provide useful data mapping recipes and templates, proven tools and guidance material to enable reducing some of the training effort and expert support which is still necessary in Linked Data projects.*
- *Steer projects towards Linked Data best practices and provide advice on which methods and tools are really proven and effective for certain data and tasks.*
- *Current practices are very much focused on the generation of Linked Data of content collections. More could be done with regard to integrating Linked Data vocabularies in tools for data recording in the field and laboratory.*

Promote Knowledge Organization Systems as Linked Open Data

Brief summary

Knowledge Organization Systems (KOSs) such as ontologies, classification systems, thesauri and others are among the most valuable resources of any domain of knowledge. In the web of Linked Data KOSs provide the conceptual and terminological basis for consistent interlinking of data within and across fields of knowledge, enabling interoperability between dispersed and heterogeneous data resources.

The RDF family of specifications provides "languages" for Linked Data KOSs. The relatively lightweight language Simple Knowledge Organization System (SKOS) can be used to transform a thesaurus, taxonomy or classification system to Linked Data. KOSs that are complex conceptual reference models (or ontologies) of a domain of knowledge are typically expressed in RDF Schema (RDFS) or the Web Ontology Language (OWL). Linked Data KOSs are machine-readable which allows various advantages. For example a SKOSified thesaurus employed in a search environment can enhance search & browse functionality (e.g. facetted search with query expansion), while Linked Data ontologies can allow automated reasoning over semantically linked data.

Some years ago many KOSs were still made available as copyrighted manuals or online lookup pages. Recently open licensing of KOSs has become the norm and ever more existing KOSs are being prepared and published as Linked Open Data for others to re-use. Following the path-breaking library community, the initiative for KOSs as LOD is under way also in the field of cultural heritage and archaeology. Some international and national KOSs are already available as LOD, Iconclass, Getty thesauri (e.g. Arts & Architecture Thesaurus), several UK cultural heritage vocabularies, the PACTOLS thesaurus (France, but multi-lingual), and others.

But more still needs to be done for motivating and enabling owners of cultural heritage and archaeology KOSs to produce LOD versions and align them with relevant others, for example mapping proprietary vocabulary to major KOSs of the domain. Also more LOD KOSs for research specialities, such as the Nomisma ontology for numismatics, are necessary.

The sector of cultural heritage and archaeology could also benefit from a dedicated international registry for KOSs already available as LOD or in preparation. An authoritative registry could serve as an instrument of quality assurance and foster a community of KOSs developers who actively curate vocabularies. Such a registry could also allow announcing LOD KOSs projects so that duplication of work may be prevented and collaborative efforts promoted (e.g vocabulary alignments).

Recommendations

- *Foster the availability of existing Knowledge Organization Systems (KOSs) for open and effective usage, i.e. openly licensed instead of copyright protected, machine-readable in addition to manuals and online lookup pages.*
- *Provide practical guidance and suggest effective methods and tools for the generation, publication and linking of KOSs as Linked Open Data (LOD).*
- *Encourage institutional owners/curators of major domain KOSs (e.g. at the national level) to make them available as LOD.*
- *Promote alignment of major domain KOSs and mapping of proprietary vocabulary, e.g. simple term lists or taxonomies as used by many organizations, to such KOSs.*
- *Promote a registry for domain KOSs that supports quality assurance and collaboration between vocabulary developers/curators.*

Foster reliable Linked Data for interlinking

Brief summary

The core Linked Data principle arguably is that publishers should link their data to other datasets, because without such linking there is no "web of data". In practice this principle is often not followed, particularly also not in the field of cultural heritage and archaeology. This means that already produced Linked Data remains isolated, a web of data has not emerged yet. There are several reasons for this shortcoming. Obviously one factor is that only few projects so far have produced and exposed archaeological Linked Data. Developers of such data will also not consider popular Linked Data resources like DBpedia/Wikipedia as relevant candidates. Moreover there is the issue of reliability, that data one links to will remain accessible, which often they are not. Surveys found that many datasets present problems, for example SPARQL endpoints are often off-line or present errors.

With the increasing number of Linked Data resources their quality has become a core topic of the developer community. Detailed quality schemes and metrics are being elaborated and used to scrutinize resources and suggest improvements. The quality criteria essentially are about how users (humans and machines) can discover, understand and access Linked Data resources that are well-structured, accurate, up-to-date and reliable over time. Furthermore the resources should be well-documented, e.g. with regard to data provenance and policy/licensing. Ideally the result of the quality initiative will be easy to use tools that allow Linked Data curators monitor resources, detect and fix problems so that high-quality webs of data are being developed and maintained.

The lack of trustworthy resources in many quarters of the "web of data" makes clear that a community of curators is necessary who take care for reliable availability and interlinking of high-quality archaeological LOD datasets and vocabularies. A few domains already have such a community, the Libraries and Life Sciences domains, for instance. Also the Ancient World LOD community around the Pelagios initiative or the Nomisma community can be mentioned as examples of good practice. It appears that the domain of archaeology needs a LOD task force and a number of projects which demonstrate and make clear what is required for reliable interlinking of LOD.

Recommendations

- *Foster a community of LOD curators who take care for proper generation, publication and interlinking of archaeological datasets and vocabularies.*
- *Form a task force with the goal to ensure reliable availability and interlinking of LOD resources; LOD quality assurance and monitoring should be established.*
- *Sponsor a number of projects which demonstrate the interlinking and exploitation of some exemplary archaeological datasets as Linked Open Data.*

Promote Linked Open Data for research

Brief summary

Linked Open Data based applications that demonstrate considerable advances in research processes and outcomes could be a strong driver for a wider uptake of the LOD approach in the research community. Current examples of Linked Data use for research purposes rarely go beyond semantic search and retrieval of information. This has not gone unnoticed by researchers who expect relevance of Linked Open Data also for generating and validating or scrutinizing knowledge claims. To allow for such uses a tighter integration of discipline-specific vocabularies and effective Linked Data tools and services for researchers are required.

Expectations of reseach-focused applications of LOD in the field of cultural heritage and archaeology often relate to the CIDOC CRM as an integrating framework. The CIDOC CRM is recognised as a common and extendable ontology that allows semantic integration of distributed datasets and addressing research questions beyond the original, local context of data generation. Notably, in the ARIADNE project several extensions of the CIDOC CRM have been created or enhanced, e.g. CRMarchaeo, an extension for archaeological excavations, and extensions for scientific observations and argumentation (CRMsci and CRMinf).

To meet expectations such as automatic reasoning over a large web of archaeological data many more (consistent) conceptual mappings of databases to the CIDOC CRM would be necessary. Linked Data applications then might demonstrate research dividends such as detecting inconsistencies, contradictions, etc. in scientific statements (knowledge claims) or suggesting new, maybe interdisciplinary lines of research based on surprising relationships between data.

Recommendations

- *LOD based applications that enable advances in archaeological research processes and outcomes may foster uptake of the LOD approach by the research community.*
- *LOD based applications for research will have to demonstrate advantages over or other benefits than already established forms of data integration and exploitation.*
- *Develop LOD based services that go beyond semantic search and retrieval of information and also support other research purposes.*
- *Build on the CIDOC CRM and available extensions to exploit conceptually integrated LOD.*

2.2.5 Linked Data development in ARIADNE

Brief summary

The developmental ARIADNE Linked Data work described in this chapter has focused on the production of (and support for) SKOS subject vocabularies, mappings between those vocabularies and the Art & Architecture Thesaurus, in order to provide a multilingual capability, and the mappings of datasets to the CIDOC-CRM. Furthermore three advanced case studies with demonstrators are presented that generate and use Linked Data based on the CIDOC CRM and key subject vocabulary hubs: coins, wooden material and sculptures.

The first two case studies involve information extraction from text reports in addition to mapping datasets, while the third explores external linking beyond the immediate ARIADNE datasets. Exploratory work on mining of Linked Data and NLP techniques are described but both are research areas with potential for much further work. The transformation of the metadata of the datasets registered in the ARIADNE data catalogue to Linked Data is described in the next chapter, as are the details of the ARIADNE Linked Data service.

The demonstrators are still being finalised at the time of this deliverable but will be available for general use via the ARIADNE Portal. For the reasons discussed in the early chapters, the case studies are experimental investigations of the future use cases that are afforded by Linked Data technology; they result in (working) research demonstrators rather than actual operational systems. They illustrate the kinds of possibilities for cross search and the semantic integration of diverse kinds of datasets and text reports that Linked Data and the related semantic technologies make possible.

One obvious finding from the experience to date is the critical importance of the subject vocabularies (e.g. the AAT) combined with the CIDOC CRM ontology entities, which act as linking hubs in the web of data. More work is needed on the identification of further linking hubs and consequent semantic enrichment of the Linked Data to relevant external datasets. One example of a potential linking hub is the Period0 set of cultural periods which can be used by providers of various archaeological and other cultural heritage datasets.

Necessary for the widespread uptake of the Linked Data approach is the availability of a variety of mapping and alignment software for different contexts, together with evaluative studies and guidelines as to their use. Beyond that, to motivate user organisations to devote scarce resources to working with Linked Data, some exemplar working applications are needed that address a real user (scientific/research) need. Such applications should offer a user interface that is easy and attractive to work with, one that does not require programming skills or detailed knowledge of the underlying data schema or ontology structure.

It should not necessarily be assumed that the end-application directly operates over a (Linked Data) triple store. There are advantages in doing so for data updates and external connections and it is an obvious route. However, periodic harvesting of Linked Data is a possibility for applications that have reasons to employ a wider range of programming platforms. Another possibility is for Linked Data providers to consider exposing programmatic web services for application developers (in addition to a SPARQL endpoint), assuming that an appropriate set of of use cases for the services can be identified.

Lessons learned

- *Mapping of datasets to established domain KOSs (in our case CIDOC CRM, AAT and others) allows their integration within and beyond the catalogue of a data portal.*

- *State-of-the-art linking hubs will play an increasingly important role in the web of LOD, comprehensive domain thesauri as the AAT as well as specialised vocabularies like the Nomisma thesaurus.*
- *The mapping of datasets to such hubs requires domain knowledge, easy to use tools, and guidance of users who carry out such work for the first time. While recommender tools are helpful, fully automated mapping appears unlikely to achive quality results at the current time.*
- *The ARIADNE portal and pilot demonstrators show that this work is worth the effort. But there is still a way to go before advanced uses of LOD will become applicable and beneficial in online research environments; more effort must be invested to make this happen.*
- *There is much scope to explore the utility of LOD in practice, taking account of the objectives and requirements of different user communities. The best ways to provide and employ LOD will largely depend on their specific contexts (museum collections, data archives or research platforms, for instance), together with the anticipated use cases. In order to motivate user organisations to work with Linked Data, exemplar working applications that address a real user (scientific/research) need would be very helpful.*

2.2.6 ARIADNE LOD Cloud

Brief summary

The ARIADNE registry holds metadata of data resources from the content providers. These metadata are being collected and enriched with an aggregator (MORe) and included in the ARIADNE data catalogue. ARIADNE makes the catalogue and other data generated in demonstrators available as Linked Open Data (LOD); thereby the ARIADNE LOD can become part of a web of Linked Data of archaeological and related other information resources.

This work within ARIADNE involved the use of a suitable RDF store and graph database for the Linked Data generation and linking efforts. The project has experimented with two such technologies, Virtuoso and Blazegraph, to perform archaeologically relevant SPARQL queries on the generated Linked Data, and to allow updates of datasets using the SPARQL 1.1 Graph Store HTTP Protocol. Based on this preliminary work, a scalable implementation that can efficiently support the publication and use of the ARIADNE LOD has been designed and realized to offer three different services: the Linked Open Data Server, the Demonstrators, and the Mapping and Ontology Server.

The Linked Open Data Server provides access to a large RDF dataset, which comprises of several graphs of archaeological datasets and can be queried via a SPARQL endpoint. The Demonstrators have been developed to exemplify the capability of Linked Data based item-level data integration to support answering archaeological research questions. They represent three different subject areas of archaeology: coins, sculptures and wooden material. For each a number of datasets have been integrated based on mappings to the CIDOC CRM (and recent extensions) and use of other domain vocabularies. The Mapping and Ontology Server provides information about the mappings and the vocabularies (ontologies, thesauri) involved in the ARIADNE LOD Cloud.

The current ARIADNE LOD Cloud is just the initial stage of an information space that is expected to grow in terms of data, vocabularies, services and users. Experiments to exploit the ARIADNE LOD have just started, with promising results as shown by the Demonstrators. Planned future work will aim to proceed with linking the available Linked Data to relevant other datasets. To promote interlinking, the ARIADNE LOD will be announced via relevant mailing lists, newsletters etc. of the

Linked Data community in the field of archaeology and cultural heritage. A number of Linked Data developers will also be contacted directly to suggest and discuss interlinking with their or other available datasets in the web of LOD.

Lessons learned

While the Linked Open Data standards are essential for integrating data, the technology supporting such integration is still in its infancy. The ARIADNE LOD, comprising of LOD of the ARIADNE catalogue, three demonstrators and various vocabularies sum up to about 32 million RDF triples. While any relational database can easily handle millions of records, the corresponding amount of RDF in a current triple store can cause serious efficiency problems as experienced in the experimentation with the ARIADNE Linked Data Cloud. It is becoming apparent that this is the price to be paid to have interoperability. More robust and efficient graph databases are required if we want to proceed towards Big Data as Linked Data. This is the first lesson that we have learned while implementing the ARIADNE Linked Data Cloud.

The second lesson comes from the graph data model. This model is intrinsically binary, hence makes it difficult to express higher rank relations, and to easily implement data connection patterns. In the latter case, the patterns may involve data chains that span several arcs, and their definition and implementation is not trivial. Conversely, correlations between data items can be epitomized by such paths, which need to be detected, and this is a computationally very intensive task if the length of the paths go beyond 2-3 arcs. This fact has always been known from a theoretical point of view, but working with real data we could experience it in practice.

3 Linked Open Data: Background and principles

This chapter introduces the Linked Open Data approach, describing the development of the approach, the Linked Data principles, standards and good practices for datasets and vocabularies. The chapter also suggests what adopters of the Linked Data approach should consider first, and describes the main steps in the Linked Data lifecycle.

3.1 LOD – A brief introduction

Linked Data are Web-based data that are machine-readable and semantically interlinked based on World Wide Web Consortium (W3C) recommended standards, *in primis* the Resource Description Framework (RDF) family of specifications but also others. Linked *Open* Data are such data resources that are freely available under an open license (e.g. Creative Commons Attribution - CC-BY) or in the Public Domain.

The Linked Data standards allow the creation, publication and linking of metadata and knowledge organization systems (KOSs) in ways that make the semantics (meaning) of data elements and terms clear to humans and machines. Linked Data are linked semantically based on explicit, typed relations between the data resources.

The semantic web of Linked Data essentially is about relationships between information resources such as collections of digital content. The metadata of digital collections (or other sets of data items), describe different facets of the resources, e.g. what, where, when, who, etc. For such facets knowledge organization systems (KOSs) such as thesauri provide concepts and terms.

The W3C recommended Linked Data standards provide the basis of a semantic web infrastructure that facilitates domain-independent interoperability of data. Building on the standards, domain-based metadata and knowledge models are needed to enable interoperability and rich interlinking between data of specific domains such as cultural heritage and archaeological research.

The requirements for semantic interoperability are considerable. In the case of data sets of archaeological projects, stored in different digital archives, the metadata of the data packages must be converted to Resource Description Framework (RDF) and include terms of shared vocabulary, which also must be available as Linked Data (e.g. in the Simple Knowledge Organization System – SKOS format). Data curators thus need to become familiar with new standards and tools to generate, publish and connect Linked Data. But it does no mean that they must abandon established databases, because tools are available to output RDF data from existing databases (RDB2RDF tools).

Building semantic e-infrastructure and services for a specific domain requires cooperation between domain data producers/curators, aggregators and service providers. Cooperation is necessary not only for sharing datasets through a domain portal (i.e. the ARIADNE data portal), but also to use common or aligned vocabularies (e.g. ontologies, thesauri) for describing the data so that it becomes interoperable. For example, in ARIADNE the data providers agreed to map vocabulary which they use for their dataset metadata to the comprehensive and multi-lingual Art & Architecture Thesaurus (AAT), which is available as Linked Open Data.

ARIADNE also recommends the CIDOC Conceptual Reference Model (CRM) as a common ontology for data integration based on Linked Data. The CIDOC CRM has been developed specifically for describing cultural heritage knowledge and data. Archaeology partly overlaps with this domain as well as needs modelling of additional conceptual knowledge, for example, to describe observations of an excavation (e.g. stratigraphy). The ARIADNE Reference Model comprises the core CIDOC CRM and a set of enhanced and new extensions, including for the archaeological excavation process (CRMarchaeo) and built structures such as historic buildings (CRMba)[1].

3.2 Historical and current background

The basic concept of Linked Data has been defined by Tim Berners-Lee, the inventor of the World Wide Web, in an article published in 2006 (Berners-Lee 2006). The concept helped to re-orientate and channel the initial grand vision of the Semantic Web into a productive new avenue. In an update 2010 of the initial article on Linked Open Data Berners-Lee aligned it with the Open Data movement (Berners-Lee 2010).

In a historical perspective it is worth noting that Berners-Lee since 1998 had addressed various "Design Issues" of the Semantic Web on the website of the World Wide Web Consortium – W3C (Berners-Lee 1998-). In 2001 the vision of a Semantic Web reached a wider audience with a highly influential article in the *Scientific American* (Berners-Lee, Hendler & Lassila 2001). The widely quoted "Semantic Web Statement" of the dedicated W3C Activity (started in 2001) included: *"The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications".* [2]

Previous to Berners-Lee's Linked Data article (2006) the research and development community presented the Semantic Web vision as a complex stack of standards and technologies. This stack seemed always "under construction" and together with the difficult to comprehend Semantic Web terminology created the impression of an academic activity with little real world impact.

The re-branding of the Semantic Web as Linked Data and the moderate definition of such data was a brilliant communicative coup. It signalled a re-orientation which was welcomed by many observers, including business-oriented information technology consultants (e.g. PricewaterhouseCoopers 2009; Hyland 2010). In 2009, a paper co-authored by Berners-Lee on *"Linked Data – the story so far"* summarised: *"The term Linked Data refers to a set of best practices for publishing and connecting structured data on the Web. These best practices have been adopted by an increasing number of data providers over the last three years, leading to the creation of a global data space containing billions of assertions - the Web of Data"* (Bizer, Heath & Berners-Lee 2009). However the authors also noted some issues in Linked Data, in particular, the quality and open licensing of Linked Data required to allow for data integration.

In 2010 Berners-Lee's request for Linked Open Data aligned the Linked Data with the Open Data movement (Berners-Lee 2010), which has become particularly strong in the governmental / public sector. In this sector Open Data are seen as a means to ensure trust through transparency and make publicly funded information available (Huijboom & Van den Broek 2011; Geiger & Lucke 2012)[3]. In this context Linked Open Data are recognized as just the right approach to expose and connect

---

[1] Description of the ARIADNE Reference Model and individual extensions (including reference document, presentation, RDFS encoding) is available at http://www.ariadne-infrastructure.eu/Resources/Ariadne-Reference-Model

[2] Since December 2013, the W3C Semantic Web Activity is subsumed under the W3C Data Activity which *"has a larger scope; new or current Working and Interest Groups related to 'traditional' Semantic Web technologies are now part of that Activity"* (http://www.w3.org/2001/sw/). In the course of this shift, the quoted "vision" statement has been removed (replaced by some other, rather vague lines).

[3] The international development of open governmental data is tracked and measured by the Open Data Barometer project, http://opendatabarometer.org

existing legacy data silos as well as enable re-use of data for new services. The same rationale applies to the cultural heritage sector with its heavily publicly-funded institutions.

The Open Data movement has also renewed and strengthened the interest of governmental and public sector institutions to improve and integrate their knowledge organization systems (KOSs). One major goal here is enabling access to governmental, cultural and scientific information resources across different organizational departments, institutions and domains (Hodge 2014).

3.3 Linked Data principles and standards

3.3.1 Linked Data basics

In 2006, Berners-Lee published the basic article on Linked Data in which he summarised in four principles how to "grow" the Semantic Web (Berners-Lee 2006). In these principles Uniform Resource Identifiers (URIs) and the W3C Resource Description Framework (RDF), which requires the use of URIs, are key standards to follow, which we describe in a commentary to Berners-Lee's Linked Data principles below. The basic principles are:

1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

This sounds simple, but what are these URIs, RDF and SPARQL?

***URIs**:* Linked Data use Uniform Resource Identifiers[4] as globally unique identifiers for any kind of linkable "resources" such as abstract concepts or information about real-world objects. More precisely, Linked Data should use dereferencable HTTP URIs, which allow a web client look up an URI using the HTTP protocol and retrieve the information resource (content, metadata, description of term, etc.). URIs are the key element of Linked Data statements which are formed according to the RDF model (see below). It is important to design and serve URIs properly, following best practices.[5] The persistence of URIs is a crucial part of the whole setup of the "web of data", especially concerning the required trust in the reliability of Linked Data sources.

***RDF**:* Linked Data is based on the W3C Resource Description Framework (RDF) model.[6] The RDF model uses subject-predicate-object statements (the so called "triples") which employ dereferencable URIs for describing data items. The predicate of an RDF statement defines the property of the relation that holds between two items. This allows for setting typed links between the items which make explicit the semantics of the relations. A searchable web of Linked Data can be created if data providers publish the items of their datasets as HTTP URIs and related items are connected

[4] Uniform Resource Identifier (URI): Generic Syntax, RFC 3986 / STD 66 (2005) specification, http://tools.ietf.org/html/std66; W3C (2004) Recommendation: Architecture of the World Wide Web (Volume 1), 15 December 2004, http://www.w3.org/TR/webarch/#identification

[5] W3C (2008): Cool URIs for the Semantic Web, http://www.w3.org/TR/cooluris/; the "10 rules for persistent URIs" suggested in ISA (2012); and Arwe (2011) on how to cope with un-cool URIs.

[6] W3C (2014) Recommendation: RDF 1.1 Concepts and Abstract Syntax, 25 February 2014, https://www.w3.org/TR/rdf11-concepts/

through links of RDF statements. For example, one dataset may contain information about archaeological sites in a region, another dataset about data deposits of excavations, another about archaeologists so that one can search at which sites excavations have been conducted, where what kind of the data is available, who from institutions was involved, etc.

SPARQL*:* The SPARQL Protocol and RDF Query Language (SPARQL)[7] allows for querying and manipulating RDF graph content in an RDF store or on the Web, including federated queries across different RDF datasets.

3.3.2 Linked Open Data

In 2010, Berners-Lee added a section on *"Is your Linked Open Data 5 Star?*" to the Linked Data article of 2006 (Berners-Lee 2006). This section addressed the missing principle of openness of the data.

Berners-Lee's 5 star scheme of Linked Open Data[8]:

*	Available on the web (whatever format) *but with an open licence, to be Open Data*
**	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
***	as (2) plus non-proprietary format (e.g. CSV instead of excel)
****	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
*****	All the above, plus: Link your data to other people's data to provide context

Some comments may be appropriate to relate this scheme to the 2006 definition of Linked Data and explain some points which may be misunderstood:

Available on the web (whatever format)*:* The phrase "on the web" as used in the Semantic Web community does not necessarily mean a webpage, but any information resource that has an URI (Uniform Resource Identifier) and can be linked and accessed and, possibly, acted upon. However the standard example is a simple HTML page that presents information and includes links to other content (e.g. stored on a local server). *(whatever format):* Means that at the first, 1-star level or step towards Linked Open Data it is not seen as important that the content may be difficult to re-use (e.g. a PDF of a text document or a JPEG image of a diagram).

Open licensing*:* Concerning the important issue of explicit open licensing Berners-Lee notes: *"You can have 5-star Linked Data without it being open. However, if it claims to be Linked Open Data then it does have to be open, to get any star at all."* He does not suggest any particular "open license" like Creative Commons (CC0, CC-BY and others)[9] or Open Data Commons (PDDL, ODC-By, ODbL)[10].

[7] W3C (2013) Recommendation: SPARQL 1.1 Overview, 21 March 2013, http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/

[8] See also the "5 ★ Open Data" website which provides more detail and examples, http://5stardata.info

[9] Creative Commons, https://creativecommons.org/licenses/

[10] Open Data Commons, http://opendatacommons.org/licenses/

Machine-readable structured data*:* In contrast to the first statement "(whatever format)", here Berners-Lee emphasises that the data should not be "canned" (i.e. not an image scan/PDF of a table) but open for re-use by others (i.e. the actual table in Excel or CSV data).

Non-proprietary format*:* This criterion is about preventing dependence on proprietary data formats and software to read the data. However it is somewhat at odds with the widespread use of proprietary formats such as Excel spreadsheets. For example, many potential users will be capable of re-using such spreadsheets, and it is unlikely that data providers would convert their data to CSV (Comma Separated Values) just to comply with the criterion. Therefore the primary criterion is that the data should not be "canned" and, secondary, provided in an easy to re-use format.

Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff*:* While the criteria above address the openness of data/content in terms of format and license, here we enter the realm of Linked Data, e.g. URIs "*to identify things, so that people can point at your stuff"* when they form RDF statements (as described in the section above).

Link your data to other people's data to provide context*:* The highest level of Linked Open Data demands interlinking through RDF own data with other Linked Data resources to create an enriched web of information. The RDF links connect data from different sources into a graph that enables applications (e.g. a Linked Data browser) to navigate between them and use their information for providing services.

In summary:

- The criteria for earning the first three stars relate to "open data" in terms of data format and licensing; notably the first three stars can be earned without employing W3C standards and techniques.
- The next level, 4-star data clearly points to these standards and techniques (RDF, SPARQL and others), while 5-star data requires interlinking own data with resources of others so that a rich web of data can emerge.
- Surprisingly, Berners-Lee did not address metadata and knowledge organization systems, although they can be subsumed under "structured data". However, in response to some criticism he added: *"Yes, there should be metadata about your dataset. That may be the subject of a new note in this series."*
- To emphasise again the importance of open licensing, Berners-Lee states: *"Linked Data does not of course in general have to be open (…). You can have 5-star Linked Data without it being open. However, if it claims to be Linked Open Data then it does have to be open, to get any star at all."*

3.3.3 Metadata and vocabulary as Linked Data

Above we noted that Berners-Lee's Linked Open Data principles do not mention metadata and knowledge organization systems (KOSs), arguably to avoid addressing such more formalized structures of Linked Data. They come in two variants of "vocabularies": 1) metadata schema for content collections, and 2) knowledge organization systems (KOSs) that provide concepts for metadata records of collection items.

Metadata schemas define a set of elements (and properties) for describing the items. For example, the 15 elements of the Dublin Core Metadata Element Set (e.g. creator, title, subject, publisher, etc.)[11] are often used for metadata records of cultural products. KOSs (e.g. thesauri) are being used

[11] Dublin Core Metadata Element Set, Version 1.1, 2012-06-14, http://dublincore.org/documents/dces/

to select values for the element fields in metadata records (e.g. the subject/s of a paper). The structure and content of both metadata schemas and KOSs can be represented as Linked Data.

Among the KOSs, thesauri and classifications systems (or taxonomies) are mostly represented in the W3C Simple Knowledge Organization System (SKOS) format[12]. A thesaurus in this format can be used to state that one concept has a broader or narrower meaning than another, or that it is a related concept, or that various terms are labels for a given concept.

KOSs that are complex conceptual reference models (or ontologies) of a domain of knowledge are typically expressed in RDF Schema (RDFS)[13] or the Web Ontology Language (OWL)[14], which allow for some automated reasoning over the semantically interlinked resources.

Besides the mentioned KOSs, there are gazetteers of geographical locations (e.g. GeoNames[15]) and so called authority files of major institutions, for example, for names of persons (e.g. VIAF)[16]. At the lowest level of complexity are flat lists of terms and glossaries (term lists including description of the terms).

3.3.4 Good practices for Linked Data vocabularies

Because of the core role of knowledge organization systems (KOSs) for Linked Data, developers recommend additional good practices for such vocabularies (e.g. Heath & Bizer 2011 [section 5.5]; W3C 2014 [vocabulary checklist]). Vocabularies should of course follow the basic Linked Data principles, e.g. use dereferenceable HTTP URIs so that clients can retrieve descriptions of the concepts/terms[17]. The first specific rule for vocabularies is to re-use or extend wherever possible established vocabulary before creating a new one. The rationale for re-use is that different resources on the web of Linked Data which are described with the same vocabulary terms become interlinked. This makes it easier for applications to identify, process and integrate Linked Data.

Moreover, re-use and extension of existing vocabularies can lower vocabulary development costs. Extension here means that vocabulary developers re-use terms from one or more widely employed vocabularies (which usually represent common types of entities) and define proprietary terms (in their own "namespace") for representing aspects that are not covered by these vocabularies.

It is generally recommended that publishers of Linked Data sets (e.g. metadata of content collections), should also make their often proprietary vocabulary (e.g. thesaurus, term list) available in Linked Data format. As Janowicz *et al*. (2014) note, *"querying Linked Data that do not refer to a vocabulary is difficult and understanding whether the results reflect the intended query is almost impossible"*. The authors suggest a 5-star rating for vocabularies:

- One star is assigned if a Web-accessible human-readable description of the vocabulary is available (e.g. a webpage or PDF documenting the vocabulary),

[12] W3C (2009) Recommendation: SKOS Simple Knowledge Organization System, 18 August 2009, https://www.w3.org/2004/02/skos/

[13] W3C (2014) Recommendation: RDF Schema 1.1, 25 February 2014, http://www.w3.org/TR/rdf-schema/

[14] W3C (2012) Recommendation: OWL 2 Web Ontology Language Document Overview (Second Edition), 11 December 2012, https://www.w3.org/TR/2012/REC-owl2-overview-20121211/

[15] GeoNames, http://www.geonames.org

[16] VIAF - Virtual International Authority File (combines multiple name authority files into a single name authority service), https://viaf.org

[17] W3C (2008) Working Group Note: Best Practice Recipes for Publishing RDF Vocabularies, 28 August 2008, https://www.w3.org/TR/swbp-vocab-pub/

- Two stars can be earned if the vocabulary is available in an appropriate machine-readable format, for instance a thesaurus in SKOS format or an ontology in RDFS or OWL,
- Three stars will receive a vocabulary that also has links to other vocabularies (for example, a mapping between proprietary terms to corresponding terms of widely employed thesauri),
- Four stars are due if also machine-readable metadata about the vocabulary is available (e.g. author/s, vocabulary language, version, license),
- Finally, 5 stars are reserved if the vocabulary is also linked to by other vocabularies, which demonstrates external usage and perceived usefulness.

The criteria for the third and fifth star concern linking of vocabularies. Such linking requires that vocabulary owners/publishers produce a mapping between their vocabulary concepts/terms, ontology classes or properties and other vocabularies, which should be done by subject experts. In the case of thesauri in SKOS format such mappings for example are skos:exactMatch (two concepts have equivalent meaning), skos:closeMatch (similar meaning), skos:broadMatch and skos:narrowMatch (broader or narrower meaning). For ontologies RDF Schema (RDFS) and the Web Ontology Language (OWL) define link types which represent correspondences between entity classes and properties (e.g. rdfs:subClassOf, rdfs:subPropertyOf).

3.3.5 Metadata for sets of Linked Data

Linked Data resources are assets which, like any other valuable information resource, should be described with machine-processible metadata. Linked Data resources include data, metadata and vocabularies, and links established between them (link-sets). For example, a mapping between two vocabularies is a valuable link-set which should be documented with metadata and provided to an appropriate registry. The metadata should provide descriptive, technical, provenance and licensing information such as:

- What kind of resource is available in terms of content, format, etc. (e.g. a thesaurus, in SKOS format, serialized in JSON[18]),
- Who created / provides it (author/s, publisher) and other provenance information (e.g. version, last update etc.),
- Licensing: explicit license or waiver statements should be given; for LOD “open licenses” such as Creative Commons (CC0, CC-BY) or Open Data Commons (PDDL, ODC-By) can be considered as adequate,
- Where and how can the resource be accessed (e.g. an HTML webpage, RDF dump, SPARQL endpoint for querying the data).

One widely used vocabulary for describing RDF datasets and links between them (link-sets) is the Vocabulary of Interlinked Datasets - VoiD (Alexander *et al*. 2009)[19]. Schmachtenberg *et al*. (2014a) in their survey of the Linked Open Data Cloud in 2014 found that of 1014 identified datasets 140 (13.46%) were described with VoiD. Most users of VoID were providers of Linked Data in the categories Government, Geographic, and Life Sciences. In the humanities for example the Pelagios initiative for linking of Ancient World resources based on the places they refer requests data

[18] JSON - JavaScript Object Notation (is a lightweight data-interchange format), https://en.wikipedia.org/wiki/JSON

[19] W3C (2011) Interest Group Note: Describing Linked Datasets with the VoID Vocabulary, 3 March 2011, http://www.w3.org/TR/void/

providers to make available a VoID file; the file describes the dataset (mappings of place references to one or more gazetteers), publisher, license etc., and contains the link from which Pelagios can get the dateset[20].

The Networked Knowledge Organization Systems (NKOS) Task Group of the Dublin Core Metadata Initiative (DMCI) has been working on a Dublin Core based metadata schema for vocabularies/KOSs. One important function of this schema is description of KOSs in vocabulary registries or repositories (Golub *et al*. 2014). The suggested Dublin Core Application Profile - NKOS AP has been released for discussion in 2015 (Zeng & Žumer 2015). For providing metadata of ontologies the Vocabulary of a Friend (VOAF)[21] is often being used. For example, the Linked Open Vocabularies (LOV) registry uses VOAF (and dcterms) for describing registered ontologies, i.e. vocabularies in RDFS or OWL (Vandenbussche *et al*. 2015).

3.4 What adopters should consider first

Adopters of the Linked Data approach should first think about what they wish to achieve by publishing one or more datasets as Linked Data. If the goal is primarily making data available as Open Data there are simpler solutions, for example providing the data as a downloadable CSV file[22]. For Linked Data the goal generally is enrichment of data and services by interlinking own data with data of other providers. Adopters therefore should consider which own data will generate most value if available as and interlinked with other Linked Data.

Linked Data should not be published "just in case". Rather publishers should consider the re-use potential and intended or possible users of their data. As Linked Data consumers they need to address the question of which data of others they could link to.

These questions make clear the importance of joint initiatives for providing and interlinking datasets of certain domains. Particularly small institutions should look for and connect to a relevant initiative. A framework for collaboration on Linked Data can ensure value generation, for example, by using common vocabularies. Linked Data developers should also ensure institutional commitment and support, i.e. an official project with a clear mandate, allocated staff and resources (cf. Smith-Yoshimura 2014f).

Linked Data adopters of all sizes will best start with a small targeted project that does not require a lot of resources. The project should allow gaining first-hand experience in Linked Data and provide potential for taking next steps. Obviously creating HTTP URIs for the selected data is an essential step towards interlinking it based on RDF. Exposing local data identifiers as HTTP URIs allows opening up a database so that others can link to and reference/cite the data.

Large institutions such as governmental agencies may benefit from streamlining with the Linked Data approach internal processes for sharing and integration of data of different departments and closely related organisations. Such institutions are also often those which publish major controlled vocabularies which others can use to connect data (Archer *et al.* 2014: 55-56).

[20] Pelagios: Joining Pelagios, https://github.com/pelagios/pelagios-cookbook/wiki/Joining-Pelagios

[21] VOAF - Vocabulary of a Friend, http://lov.okfn.org/vocommons/voaf/v2.3/

[22] See Heath (2010) for a comparison between providing a CSV file vs. Linked Data.

3.5 Mastering the Linked Data lifecycle

The previous sections present the principles, standards and good practices of Linked Data, but do not describe how such data are actually generated, published and interlinked. This study does not intend providing a guidebook for mastering the so called "lifecycle" of Linked Data, the different steps that are necessary to get to and benefit from such data. In brief, the main steps are:

- *Select a relevant dataset:* Chose a dataset which allows generating value if made available as RDF data and linked to other LOD, including linking of the dataset by others. The publisher should of course be able to provide the data under an open license or place it in the public domain.
- *Clean and prepare the source data:* Bring the source data in a shape that it is easy to manipulate and convert to RDF, addressing issues of data quality such as missing values, invalid values, duplicate records, etc. The OpenRefine[23] tool is recommended for this task.
- *Design the URIs of the data items:* Follow suggested good practice for designing the structure of the URIs (e.g. W3C 2008; ISA 2012).
- *Define the target data model:* Re-use an existing model that is being used in the domain (e.g. CIDOC CRM for cultural heritage data) or create one re-using concepts from widely employed vocabularies; re-use will aid data interoperability and decrease development effort/costs.
- *Transform the data to RDF:* In the transformation the source data (e.g. data tables) are converted to a set of RDF statements (graph-based representation) according to the defined target model. Many tools are available that allow transformation of almost any data format and database (e.g. CSV, Excel, relational databases) to RDF.[24]
- *Store and publish the RDF data:* The generated RDF data is typically stored in an RDF database (triple store) where it can be accessed via a web server or queried at an SPARQL endpoint; the data is also often published as a so called "RDF dump" (a RDF dataset made available for download).
- *Link to other RDF data on the Web:* According to the Linked Data principles publishers should link to other datasets to create an enriched web of Linked Data. Therefore relevant linking targets need to be identified which can add value (i.e. where relationships exist between data) and are well maintained. Publishers may be aware of such datasets in their domain or search existing registries (e.g. DataHub) to identify relevant datasets. If there is a relevant dataset, the publisher must decide which properties from established domain or general Linked Data vocabularies to use for the linking.
- *Describe, register and promote the dataset:* The publisher of a set of Linked Data should describe the dataset with metadata (including provenance, licensing, technical and other descriptive information) which can be attached to the dataset. It is also good practice to register the dataset in a domain data catalogue and general registries such as the DataHub. Furthermore the publisher should announce the dataset via relevant mailing lists, newsletters etc. and invite others to consider linking to the dataset.

There are many introductory and advanced level guides available that describe how to generate, publish, link and use Linked Data: As introductory level guides Bauer & Kaltenböck (2012), Hyland & Villazón-Terrazas (2011) and W3C (2014) can be suggested. Advanced "cookbooks" are the EUCLID

[23] OpenRefine, http://openrefine.org

[24] W3C wiki: Converter to RDF, http://www.w3.org/wiki/ConverterToRdf

curriculum[25], Heath & Bizer (2011), Morgan *et al*. (2014); Ngonga Ngomo *et al*. (2014), van Hooland & Verborgh (2014) and Wood *et al*. (2014).

Concerning useful tools such as RDF converters, Linked Data editors, RDF databases, etc. the W3C wiki provides an extensive tool directory[26]. Some projects describe selected tools they recommend for different tasks of the Linked Data lifecycle, for example, the projects LATC (various tools)[27] and LOD2 (mainly tools of the project partners)[28]. But adopters of the Linked Data approach should seek additional expert advice on which tools are proven and effective for their data and certain tasks.

3.6 Brief summary and recommendations

Brief summary

The term Linked Data refers to principles, standards and tools for the generation, publication and and linking of structured data based on the W3C Resource Description Framework (RDF) family of specifications.

The basic concept of Linked Data has been defined by Tim Berners-Lee in an article published in 2006. This concept helped to re-orientate and channel the initial grand vision of the Semantic Web into a productive new avenue. Previously the research and development community presented the Semantic Web vision as a complex stack of standards and technologies. This stack seemed always "under construction" and together with the difficult to comprehend Semantic Web terminology created the impression of an academic activity with little real world impact.

In 2010 Berners-Lee's request for Linked *Open* Data aligned the Linked Data with the Open Data movement. Since then the quest for Linked Open Data (LOD) has become particularly strong in the governmental / public sector as well as initiatives for cultural and scientific LOD.

The Linked Data principles include that a data publisher should make the data resources accessible on the Web via HTTP URIs (Uniform Resource Identifiers), which uniquely identify the resources, and use RDF to specify properties of resources and of relations between resources. In order to be Linked Data proper, the publishers should also link to URI-identified resources of other providers, hence add to the "web of data" and enable users to discover related information. And to be Linked Open Data the publisher must provide the data under an open license (e.g. Creative Commons Attribution [CC-BY] or release it into the Public Domain).

The Linked Data approach allows opening up "data silos" to the Web, interlink otherwise isolated data resources, and enable re-use of the interoperable data for various purposes. The landscape of archaeological data is highly fragmented. Therefore Linked Data are seen as a way to interlink dispersed and heterogeneous archaeological data and, based on the interlinking, enable discovery, access to and re-use of the data.

Building semantic e-infrastructure and services for a specific domain such as archaeology requires cooperation between domain data producers/curators, aggregators and service providers. Cooperation is necessary not only for sharing datasets through a domain portal (i.e. the ARIADNE data portal), but also to use common or aligned vocabularies (e.g. ontologies, thesauri) for describing the data so that it becomes interoperable.

---

[25] EUCLID - Educational Curriculum for the Usage of Linked Data, http://euclid-project.eu

[26] W3C wiki: Tools, http://www.w3.org/2001/sw/wiki/Tools

[27] LATC - LOD Around The Clock (EU, FP7-ICT, 9/2010-8/2012), http://latc-project.eu

[28] LOD2 - Creating Knowledge out of Interlinked Data (EU, FP7-ICT, 9/2010-8/2014), http://lod2.eu

In addition to the basic Linked Data principles there are also specific recommendations for vocabularies. Particularly important is re-using or extending wherever possible established vocabularies before creating a new one. The rationale for re-use is that different resources on the web of Linked Data which are described with the same or mapped vocabulary terms become interlinked. This makes it easier for applications to identify, process and integrate Linked Data. Moreover, re-use and extension of existing vocabularies can lower vocabulary development costs.

It is also recommended to provide metadata for Linked Data of datasets as well as vocabularies. The Vocabulary of Interlinked Datasets (VoiD) is often being used for providing such metadata. It is also good practice to register sets of Linked Data in a domain data catalogue and/or general registries such as the DataHub. Furthermore the publisher should announce the dataset via relevant mailing lists, newsletters etc. and invite others to consider linking to the dataset.

Linked Data should not be published "just in case". Rather publishers should consider the re-use potential and intended or possible users of their data. As Linked Data consumers they need to address the question of which data of others they could link to. These questions make clear the importance of joint initiatives for providing and interlinking datasets of certain domains such as archaeology.

Recommendations

- *Use the Linked Data approach to generate semantically enhanced and linked archaeological data resources.*
- *Participate in joint initiatives for providing and interlinking archaeological datasets as Linked Open Data.*
- *Choose datasets which allow generating value if made openly available as Linked Data and connected with other data, including linking of the datasets by others.*
- *Re-use existing Linked Data vocabularies wherever possible in order to enable interoperability.*
- *Describe the Linked Data with metadata, including provenance, licensing, technical and other descriptive information.*
- *Register the dataset in a domain data catalogue and/or general registries such as the DataHub. Also announce the dataset via relevant mailing lists, newsletters etc. and invite others to consider linking to the dataset.*

4 The Linked Open Data Cloud

This chapter describes what has been termed the LOD Cloud and is generally illustrated with the LOD Cloud diagram of interlinked datasets. Some available figures for the state of the LOD Cloud are presented and also some issues highlighted. Furthermore an overview of cultural heritage LOD present on the LOD Cloud diagram and other known cultural heritage LOD, including archaeological LOD, is being given.

4.1 LOD Cloud figures

The Linked Open Data (LOD) Cloud is formed by datasets that are openly available on the Web in Linked Data formats and contain links pointing at other such datasets. The latest LOD Cloud figures and visualization have been published online in August 2014 (Schmachtenberg *et al*. 2014a [statistics online], 2014b [paper]). They are based on information collected through a crawl of the Linked Data web in April 2014. The crawl found 1014 datasets of which 569 (56%) linked to at least one other dataset; the 569 datasets were connected by in total 2909 link-sets. The remaining datasets were only targets of RDF links, and therefore at the periphery of the "cloud", or they were isolated. Of the 569 core LOD Cloud datasets 374 were registered in the DataHub.[29] The latest comparable figures to the ones reported by Schmachtenberg *et al*. (2014a/b) are based on the DataHub metadata of datasets from September 2011 (Jentzsch *et al*. 2011)[30].

Below we summarize some results of Schmachtenberg *et al*. (2014a and 2014b, of which the latter compares the figures of 2011 and 2014) which give an impression of the adoption of the Linked Data principles:

- *Increase in datasets:* There has been a substantial increase in identified datasets: 2011: 294 LD datasets registered in the DataHub; 2014: 1014 datasets identified through a crawl of the web of Linked Data. With 530 datasets the largest group in 2014 was the newly introduced category of social web/networking. These datasets describe people profiles and social relations amongst people. Among the established categories three showed a large growth in number of dataset, Government (2011: 49; 2014: 183), Life Sciences (2011: 41; 2014: 83) and User-generated content (2011: 20; 2014: 48).
- *Linking of datasets:* 445 (43.89%) of the 1014 datasets did not set any out-gowing RDF links, 176 (17.36%) linked to one other dataset, 106 (10.45%) to two datasets, 127 (12.52%) to 3-5 datasets, 81 (7.99%) to 6-10 datasets, and 79 (7.79%) even to more than 10 datasets.
- *A less centralized LOD Cloud:* In 2014 the web of linked data appeared to be less centralized. In 2011 the cross-domain Linked Data resource DBpedia.org clearly occupied the centre of the LOD Cloud. In 2014 also GeoNames was used widely and there were some category-specific linking hubs (e.g. data.gov.uk in the category Goverment). Most interconnected were resources of the category Publications (e.g. RKB Explorer datasets) and of the category Life Sciences (e.g. Bio2RDF datasets).
- *Use of vocabularies:* The 2014 survey discovered in total 649 vocabularies. 271 vocabularies (41.76%) were "non-proprietary", defined as used by at least two datasets. Among these

[29] DataHub (Open Knowledge Foundation), http://datahub.io

[30] State of the LOD Cloud, 19/09/2011, http://lod-cloud.net/state/

vocabularies, RDF and RDFS aside, the most used were FOAF[31] (701 datasets used it) and Dublin Core[32] (568 datasets used it). A special analysis showed that among the 378 "proprietary" vocabularies (defined as used by only one dataset) only 19.25% were fully and 8% partially dereferencable; 72.75% had term URIs which were not dereferencable at all. One or more proprietary vocabularies were used by 241 datasets (23.17% of the total).

- *Metadata for sets of Linked Data:* For 35.77% of all sets of Linked Data in 2014 machine-readable provenance and other metadata were provided (most often in Dublin Core, DCTerms or MetaVocab), about the same percentage than in 2011 (36.63%). Only about 8% provided machine-readable licensing information, mostly dc:license/dc:rights and cc:license. Hence lack of metadata for sets of Linked Data remains an issue.

4.2 (Mis-)reading the LOD diagram

In the years 2007-2011 a diagram of the LOD Cloud has been produced based on datasets registered in the DataHub. The latest version of the diagram has been published in August 2014[33] and in addition to the DataHub information uses the results of a crawl of the Linked Data Web in April 2014 (Schmachtenberg *et al*. 2014a/b, as summarized above). The LOD Cloud diagram has grown enormously, too large to present it here.

The criteria for including a dataset in the LOD Cloud diagram are[34]:

- There must be resolvable http:// (or https://) URIs.
- They must resolve, with or without content negotiation, to RDF data in one of the popular RDF formats (RDFa, RDF/XML, Turtle, N-Triples).
- The dataset must contain at least 1000 triples.
- The dataset must be connected via RDF links to at least one other dataset in the diagram, by using URIs from that dataset or vice versa; at least 50 links are required.
- Access of the entire dataset must be possible via RDF crawling, an RDF dump or a SPARQL endpoint.

The LOD Cloud diagrams that since 2007 have been produced based on these criteria showed some linking hubs, but in 2014 there still were many rather isolated datasets (e.g. linked to only one other Linked Data resource). Yet the LOD Cloud diagrams have often been misleadingly referenced as presenting a compact "web of data" or "a huge web-scale RDF graph" (cf. the critique by Hogan & Gutierrez 2014). Also the researchers who published the latest figures on the LOD Cloud state: *"By setting RDF links, data providers connect their datasets into a single global data graph which can be navigated by applications and enables the discovery of additional data by following RDF links"* (Schmachtenberg *et al*. 2014a).

What must be added is that the "single global data graph" is patchy (as described above) and that relevant applications for end-users are hardly available. There are Linked Data browsers[35] which,

[31] FOAF - Friend-of-a-Friend (defines terms for describing persons, their activities and their relations to other people and object), http://xmlns.com/foaf/spec/

[32] Dublin Core Metadata Initiative (DCMI) Metadata Terms, http://dublincore.org/documents/dcmi-terms/

[33] The Linking Open Data cloud diagram 2014, by M. Schmachtenberg, C. Bizer, A. Jentzsch and R. Cyganiak, available at: http://lod-cloud.net

[34] cf. The Linking Open Data cloud diagram, http://lod-cloud.net

however, seem not to be in wider use, arguably because of a lack of interlinked data that are relevant for user communities. Research oriented developers have created search engines based on crawled and semantic Web Data (e.g. Sindice [service ended in 2014], Swoogle, Watson). These engines are of little use for non-experts. They serve as research tool to better understand the Linked Data landscape. Research based on crawled Web data has become a specialty and is conducted around resources such as the Common Crawl[36].

The LOD Cloud is not a single entity but represents datasets of different providers that are made available in different ways (e.g. LD server, SPARQL endpoint, RDF dump) and often with low reliability. For example, Buil-Aranda *et al*. (2013) found that of 427 public SPARQL endpoints registered in the DataHub the providers of only one-third gave descriptive metadata. Half of the endpoints were off-line and only one third was available more than 99% of the time during a monitoring of 27 months; the support of SPARQL features and performance for generic queries was varied.

Public SPARQL endpoints could form a distributed infrastructure for federated queries[37] of relevant data of different sources (Rakhmawati *et al*. 2013). Thereby views across the different datasets could be provided, allowing researchers to explore the data. But this depends on reliable maintenance of the datasets and SPARQL endpoints by the service providers. Instead of querying the "single global graph" or just a number of LD datasets, the typical approach is to pull the data into one data repository and run queries over this database. This approach is impractical for any but a small number of datasets (or datasets of a small size), especially if only some interlinking between the datasets is of interest.

For intelligent searching, question answering and reasoning over Linked Data much more is necessary than providing SPARL endpoints or pulling a number of datasets into one graph database. One approach is "reason-able views" of Linked Data which has been developed by researchers of Ontotext and demonstrated with the FactForge service[38] (Kiryakov *et al*. 2009; Damova 2010; Simov & Kiryakov 2015). A reason-able view is constructed by assembling different datasets and vocabularies into a compound set of Linked Data, produce mappings between instance data of the datasets, and create a single ontology for querying the compound dataset using SPARQL. The ontology is created based on mappings between the vocabularies and/or an upper-level ontology, in the case of FactForge: PROTON[39]. Damova & Dannells (2011) illustrate the approach with a "museum reason-able view" including mappings between CIDOC CRM and PROTON, CIDOC CRM and Swedish Open Cultural Heritage (K-samsök)[40], and information of the Gothenburg City Museum transformed to RDF. Also existing mappings of DBPedia and GeoNames to PROTON were included. A reason-able view provides a controlled environment of integrated datasets to exploit existing and newly created sets of Linked Data, reduce development costs and risks of unreliable datasets.

There is no central management of LOD Cloud, the assumed "huge web-scale RDF graph", but (some) areas for which a community of developers produces and interlinks relevant resources and creates applications for the purposes of the intended end-users. In such cases network effects in the web of Linked Data are being achieved. Such effects do not result automatically from merely putting more

[35] LOD Browser Switch (offers a set of browsers), http://browse.semanticweb.org

[36] Common Crawl, http://commoncrawl.org

[37] W3C (2013) Recommendation: SPARQL 1.1 Federated Query, 21 March 2013, http://www.w3.org/TR/sparql11-federated-query/

[38] Ontotext: FactForge, http://ontotext.com/factforge-links/

[39] Ontotext: PROTON, http://ontotext.com/products/proton/

[40] Swedish Open Cultural Heritage (K-samsök): http://www.ksamsok.se/in-english/

datasets into the LOD cloud, actual interlinking is required to generate a web of Linked Data. One example of effective linking is the Linked Data community of the bio-medical and life sciences. In this area the Bio2RDF[41] project has created 35 Linked Data sets of existing databases and interlinked some of them. Another well-curated area is Linked Data of the library community. Cultural heritage or archaeology is not yet an area of densly interlinked information. So far a community of cooperating LOD producers, curators and integrators has not emerged.

4.3 Cultural heritage in the LOD Cloud

The latest LOD Cloud diagram (August 2014) provides an indicator for the state of cultural heritage Linked Data. So far only few cultural heritage LD datasets show up on the diagram, and they do not form a closely linked web of LD. None of the datasets concerns archaeology specifically. Some more cultural heritage LD sets exist, also a few archaeological datasets. But they did not conform to the criteria for being included in the LOD Cloud diagram, e.g. the requirement of being connected via RDF links with at least one other compliant dataset (see section above).

Below we first list the cultural heritage datasets which conform to the criteria, not including datsets of the library sector (e.g. Bibliothèque nationale de France [data.bnf.fr] or Deutsche Nationalbibliothek [DNB]):

- *Europeana LOD:* mentioned in the first place because it is the largest cultural heritage LD dataset (20 million records) and comprises of records of museums, archives and libraries across Europe[42].
- *Swedish Open Cultural Heritage (K-samsök):* a web service that harvests metadata from the databases of cultural heritage organisations in Sweden and allows creating LD based information services[43].
- *Archives Hub Linked Data:* the Archives Hub[44] aggregates and allows searching across descriptions of archival collections held at over 250 institutions in the UK (a search of the portal for "archaeology" produces over 1000 hits). Linked Data of a sub-set of the aggregated descriptions has been produced by the LOCAH project (2010-2011)[45].
- *British Museum - Semantic Web Collection Online*: provides Linked Data access to the same collection records as the Museum's web presented Collection Online; the data has also been organised using the CIDOC CRM[46].
- *Amsterdam Museum:* has been the first museum in the Netherlands to convert its complete museum collection database (over 70,000 records) to RDF; the data includes links to two Getty

[41] Bio2RDF: Linked Data for the Life Sciences, http://bio2rdf.org

[42] Europeana Linked Data, http://labs.europeana.eu/api/linked-open-data/introduction/; a search on the Europeana website for "archaeology" shows that the providers of most related content are the Swedish National Heritage Board (812,971 items) and the UK Portable Antiquities Scheme (236,627). ARIADNE partners are also present: German Archaeological Institute / ARACHNE (183,683 items), Archaeology Data Service, UK (34,197) and Data Archiving and Networked Services, Netherlands (6456).

[43] Swedish Open Cultural Heritage (K-samsök): http://www.ksamsok.se/in-english/; see also: DataHub, http://datahub.io/dataset/swedish-open-cultural-heritage

[44] Archives Hub, http://archiveshub.ac.uk

[45] Archives Hub – LOCAH, http://data.archiveshub.ac.uk

[46] British Museum - Semantic Web Collection Online, http://collection.britishmuseum.org

thesauri (AATNed [Dutch version] and ULAN), GeoNames, and DBPedia pages (De Boer *et al*. 2012 and 2013)[47].

- *Art & Architecture Thesaurus (AAT)* of the Getty Research Institute: The only cultural heritage KOS on the 2014 LOD diagram; meanwhile two other Getty KOSs have become available: Thesaurus of Geographic Names (TGN) and Union List of Artist Names (ULAN); the Cultural Objects Name Authority (CONA) was expected to follow in Fall 2015 but seems to require more effort than expected[48].

The second list below presents further cultural heritage and archaeological datasets in Linked Data formats that are registered in the DataHub or of which we know from searching various other sources. The list is certainly not comprehensive, because there have been quite some cultural heritage projects that trialled the Linked Data approach, however the whereabouts of the created Linked Data are often unclear. The Linked Data resources listed below are roughly ordered according to their relevance in the context of our study:

- *Archaeology Data Service (ADS):* ADS Linked Open Data initially has been produced in the STELLAR project by converting databases and CSV files to RDF, using the CRM-EH ontology; this RDF data is available from a SPARQL endpoint[49]. According to their annual report 2014/2015 ADS now also have LOD of deposited project archives, including the projects Roman Amphora[50] and Colonisation of Britain (see Cripps 2014 for background); the number of LOD triples in 2015 was 2,531,302, up from 680,500 in the previous reporting period (ADS 2015: 26). Notably, ADS also consume LOD from external sources to populate own metadata (e.g. Ordnance Survey geographic data[51]).
- *Data Archiving and Networked Services (DANS):* DANSlabs has produced LOD of metadata records of more than 25,000 data sets stored in the DANS-EASY digital archive, which includes the E-Depot for Dutch Archaeology; this was done 2013 in a demonstration project, but the LOD (with little cross-linking) is accessible via their SPARQL endpoint under an Open Data Commons license[52].
- *CLAROS - The World of Art on the Semantic Web:* the data of this international collaboration comes from major Classics collections, including from ARIADNE partner DAI; the data has been prepared for a search portal based on CIDOC CRM modelling; the data service is maintained by the University of Oxford's e-Research Centre and offers a SPARQL endpoint[53].
- *Cultura Italia:* provides metadata of a number of Italian heritage institutions; offers a SPARQL endpoint for the metadata; also the PICO thesaurus is available for download[54].

[47] Amsterdam Museum in Europeana Data Model RDF, http://semanticweb.cs.vu.nl/lod/am; see also: DataHub, http://datahub.io/dataset/amsterdam-museum-as-edm-lod

[48] Getty Vocabularies LOD, http://vocab.getty.edu

[49] ADS Linked Open Data, http://data.archaeologydataservice.ac.uk; STELLAR project, http://archaeologydataservice.ac.uk/research/stellar/

[50] Roman Amphorae: a digital resource (University of Southampton, 2005; updated 2014), http://archaeologydataservice.ac.uk/archives/view/amphora_ahrb_2005/

[51] Ordnance Survey (UK), http://data.ordnancesurvey.co.uk

[52] DANSlabs: EASY Metadata as Linked Open Data Demo, http://dans-labs.github.io/easy-lod/

[53] CLAROS: Data, http://data.clarosnet.org

[54] Cultura Italia: Dati, http://dati.culturaitalia.it

- *English Heritage Places:* contains metadata for about 400,000 nationally important places as recorded by English Heritage[55]; also seven English Heritage and other UK thesauri are registered in the DataHub, but for those we refer to the LD versions produced in the SENESCHAL project[56].
- *Pleiades:* a gazetteer for ancient world studies operated by the Institute for the Study of the Ancient World (USA)[57]; Pleiades URIs are used in the digital classics network Pelagios to interconnect scholarly ancient world resources through the places they refer to; the Pelagios project provides services and tools to allow scholars annotate, aggregate, access and display the place references[58].
- *Nomisma:* provides as LOD an ontology for describing coins and several numismatics datasets of the American Numismatic Society and institutions in Europe; a SPARQL endpoint is available[59].
- *Portable Antiquities Scheme:* PAS data of finds in the UK has been linked to LD resources of the Ordnance Survey (national mapping service), Pleiades (gazetteer), British Museum, Nomisma and DBpedia[60] (cf. Pett 2014a/b).
- *LinkedARC.net*[61]: Frank Lynam (Trinity College Dublin), produced Linked Data of data of excavations at Priniatikos Pyrgos (Crete), modelled primarily using CIDOC CRM and its type values link to terms of the FISH Archaeological Objects Thesaurus, British Museum and Getty vocabularies. The project is particularly interesting as it demonstrated the integration of excavation data of American and Irish groups of archaeologists, applying the Locus-Pail method of excavation and MoLAS single-context method respectively.
- *MONDIS:* a dataset about monument damages developed in the Czech research project MONDIS; includes their diagnostic Monument Damage Ontology (Cacciotti & Valach J. 2015)[62].
- *MisMuseos.net:* a "semantic catalog" of museums in Spain and their information about art works and artists[63]; the solution builds on the GNOSS social and semantic platform (Maturana *et al.* 2013).
- *Musei Italiani:* a list of geo-referenced museums in Italy; that for museum categories the dataset links to DBpedia and for places to GeoNames[64].
- *ReLoad - Repository for Linked Open Archival Data:* a project of the Archivio Centrale dello Stato, Istituto per i Beni culturali dell'Emilia-Romagna and regesta.exe (2010-2013), the project developed ontologies for archival data sources and produced a LOD dataset of several archival inventories; ReLoad provides a SPARQL endpoint[65].

[55] English Heritage Places, DataHub information: http://datahub.io/dataset/englishheritage_places

[56] Heritage Data: Vocabularies, http://www.heritagedata.org/blog/vocabularies-provided/

[57] Pleiades, http://pleiades.stoa.org

[58] Pelagios, http://commons.pelagios.org

[59] Nomisma, http://nomisma.org/datasets

[60] Portable Antiquities Scheme, http://finds.org.uk

[61] Linkedarc.net, http://linkedarc.net; datasets, https://datahub.io/dataset/linkedarc

[62] MONDIS project, http://www.mondis.cz; DataHub information: http://datahub.io/dataset?q=mondis

[63] MisMuseos.net, DataHub information: http://datahub.io/dataset/mismuseos-gnoss

[64] Musei Italiani, http://www.linkedopendata.it/datasets/musei

[65] ReLoad, http://labs.regesta.com/progettoReload/, see also their project description for the LODLAM 2013 Summit challenge, http://summit2013.lodlam.net/2012/12/01/challenge-entry-reload-repository-for-linked-open-archival-data/

Some of the datasets listed above may show up on the next version of the LOD Cloud diagram, most likely those which are maintained and employed by a dedicated group of developers and users like the Nomisma ontology and datasets and the Pleiades gazetteer, for instance.

The Art & Architecture Thesaurus (AAT) as a linking hub

Already on the 2014 LOD Cloud diagram was the Art & Architecture Thesaurus (AAT) which the Getty Research Institute in February 2014 released as LOD. The multilingual AAT contains over 40,000 concepts and over 350,000 terms for describing objects of visual art, architecture, other material heritage, archaeology, conservation, archival materials, etc. The AAT has the potential to become one of the core linking hubs for cultural heritage information in the Linked Open Data Cloud. In a survey on Linked Data of the AthenaPlus project half of the 24 project partners said they intend to link to the AAT and other Getty thesauri when they are available as LOD (AthenaPlus 2013b: 10). When the AAT was released as LOD, among the initiatives that started using it was Europeana. Europeana partners who already use AAT terms were invited to re-submit their metadata so that their old AAT term labels (provided as a simple text string) could be automatically replaced by the new AAT URIs (Charles & Devarenne 2014). This enables linking to information of others on the web who use these URIs. This is also possible if data providers map their local vocabulary to the AAT. In ARIADNE the data providers mapped terms of vocabularies (e.g. national thesauri or own term lists) which they use for their dataset metadata to appropriate terms of the AAT, using SKOS mappings (e.g. skos:exactMatch, skos:closeMatch and others).

4.4 Brief summary and recommendations

Brief summary

The Linked Open Data Cloud is formed by datasets that are openly available on the Web in Linked Data formats and contain links pointing at other such datasets. One task of the ARIADNE project is to promote the emergence of a web of interlinked archaeological datasets which comply with the Linked Open Data (LOD) principles. It is anticipated that this web of archaeological LOD will become part of the wider LOD Cloud and interlinked with related other data resources.

The latest LOD Cloud diagram (2014) includes only few sets of cultural heritage LOD and they do not form a closely linked web of Linked Data. None of the datasets concerns archaeology specifically. Some more sets of cultural heritage Linked Data sets exist, also a few archaeological, but in 2014 they did not conform to the criteria for being included in the LOD Cloud diagram (e.g. the requirement of being connected via RDF links with at least one other compliant dataset).

Maybe the next version of the LOD Cloud diagram will contain some of the earlier and more recent sets of archaeological Linked Open Data. Hopefully this will include some relevant vocabularies which recently have been transformed to Linked Data in SKOS format. In 2014 the only cultural heritage vocabulary on the diagram was the Art & Architecture Thesaurus (AAT), which has the potential to become one of the core linking hubs for cultural heritage information in the LOD Cloud.

The LOD Cloud is not a single entity but represents datasets of different providers that are made available in different ways (e.g. LD server, SPARQL endpoint, RDF dump) and the resources are often unreliable, e.g. many SPARQL endpoints are off-line. There is no central management and quality control of the LOD Cloud. Webs of reliable and richly interlinked datasets are only present where there is a community of Linked Data producers and curators (e.g. in the areas of bio-medical & life sciences or libraries).

Cultural heritage or archaeology is not yet an area of densly interlinked and reliable LOD resources; so far a community of cooperating LOD producers and curators has not emerged. Targeted activities to foster and support further publication and interlinking of datasets are required so that a web of archaeological, cultural heritage and other relevant data will emerge within the overall Linked Open Data Cloud.

Recommendations

- *Encourage archaeological institutions and repositories to publish the metadata of their datasets (collections, databases) as Linked Open Data; also promote publication of domain and proprietary vocabularies of institutions as LOD.*
- *Foster the formation of a community of archaeological LOD producers and curators who generate, publish and interlink LOD, including linking/mapping between vocabularies.*

5 Adoption of the Linked Data approach in archaeology

Since about 10 years the Semantic Web / Linked Data standards, methods and tools have become more mature and applicable. Cultural heritage institutions have been among the leading adopters of the Linked Data approach, mainly to better interlink domain resources and, in some cases, to enrich their online information with information of popular resources such as DBpedia/Wikipedia content. With regard to Linked Data of archaeological project archives and databases there have been only few projects, with arguably limited recognition by the wider archaeological research community. At the same time, there has been a boom in Linked Data projects in the Ancient World and Classics research community. This chapter describes and aims to explain this situation in greater detail.

5.1 Adoption by cultural heritage institutions

Institutions of the cultural heritage sector, particularly libraries and museums, are among the leading adopters of the Linked Data approach. In an international survey for institutional implementers of Linked Data services by OCLC Research in 2015, seventy-one institutions from 16 countries (45% USA) reported in total 168 Linked Data projects (Smith-Yoshimura 2016). The survey had a focus on libraries, but also some other organisations participated (e.g. American Numismatic Society, The British Museum, Europeana Foundation). Two-thirds of the projects were completed (i.e. a service implemented).

In the area of museums one pioneering project was Finnish Museums on the Semantic Web (Hyvönen *et al*. 2002)[66], followed by many others, in recent years for example the Amsterdam Museum (De Boer *et al*. 2012 and 2013)[67], British Museum[68], Peter the Great Museum of Anthropology and Ethnography in St. Petersburg (Ivanov 2011), Russian Museum in St. Petersburg (Mouromtsev *et al*. 2015) and Smithsonian American Art Museum (Szekely *et al*. 2013).[69]

Archives appear to be less advanced in the application of Linked Data. Their initial steps focus on bringing legacy finding aids online while providing access to the archival records and material still often requires much digitisation work. In recent years there has been some progress in standardisation that will help in moving towards Linked Data. For example, efforts by the Experts Group on Archival Description (EGAD, since 2012) to make the Encoded Archival Description (EAD, 2002) standard more data-centric in EAD3 (2015) and better connect it with Encoded Archival Context – Corporate Bodies, Persons and Families (EAC-CPF, 2010) and other standards[70] (Gueguen *et al*. 2013; Pitti *et al*. 2014).

Currently the archive community seeks to establish guidelines for structuring archival Linked Data resources with the new standards, build support for editing and publication into archival tools (e.g. ease adding identifiers of authorities), and derive good practice from the experience of first projects in the field (Gracy & Lambert 2014; Gracy 2015). Examples of pioneer projects are LOCAH - Linked

[66] The Semantic Computing Research Group (SeCo) at Aalto University (Finland), who led the project, continues to be a leader in Linked Data applications for cultural heritage resources, http://seco.cs.aalto.fi

[67] Amsterdam Museum as Linked Open Data in the Europeana Data Model Amsterdam Museum, http://semanticweb.cs.vu.nl/lod/am

[68] British Museum - Semantic Web Collection Online, http://collection.britishmuseum.org

[69] Some other examples are listed on the Museums and the Machine-processable Web wiki, http://museum-api.pbworks.com/w/page/21933420/Museum%C2%A0APIs

[70] Encoded Archival Description (official site), http://www.loc.gov/ead/

Archives and Linking Lives (2010-2012)[71] (Stevenson 2012) and ReLoad - Repository for Linked Open Archival Data (2010-2013)[72] (Mazzini & Ricci 2011). The LiAM - Linked Archival Metadata project (2012-2013)[73] provides a guidebook that helps applying Linked Data approaches to archival description (Morgan *et al*. 2014).

While there exists no comprehensive overview of cultural heritage Linked Data projects, studies which describe several examples (e.g. Edelstein *et al*. 2013a/b) typically do not include archaeological projects. But there is a significant difference between cultural heritage institutions and research organisations and projects. Cultural heritage institutions such as libraries, archives and museums are motivated by a service ethos, the mission to make information about heritage readily available. Researchers are primarily interested to publish research results, while still little academic reward can be gained from sharing the data underlying the results. Therefore Linked Data of legacy datasets may be easier to promote than data of current research, where first the objective of "open data" in general needs to be addressed (ARIADNE 2015e: chapter 4; Carver & Lang 2013).

5.2 Low uptake for archaeological research data

In the cultural heritage sector there have been initiatives promoting the Linked Data approach, for example, LOD-LAM, the International LOD in Libraries, Archives, and Museums Summit (since 2011)[74], or the Linked Heritage project[75] which disseminated guidance for Linked Data to museums in Europe.[76] In the field of archaeological research there were no such initiatives or only at small scale, for example, sessions at CAA conferences or national thematic workshops. But promotional activities, particularly at the national level, are important to reach archaeological institutes and research groups and make them aware of the Linked Data approach. For example, in France the Consortium MASA[77] aims to provide archaeologists with vocabularies and tools to improve the interoperability of their data via Linked Data standards. MASA is one of the ten consortium of the HUMA-NUM research infrastructure which focus on particular resources and fields of (digital) humanities research[78].

In ARIADNE a Linked Data Special Interest Group (SIG)[79] has been formed that acts as an interface with the wider Linked Data community, communicating developments between the community and ARIADNE (and *vice versa*), looking for synergy, and relevant common use cases. Participants of the first meeting of the ARIADNE Linked Data SIG (2013) noted a still low uptake or even awareness of

---

[71] LOCAH - Linked Archives and Linking Lives (UK, 2010-2012, Archives Hub), http://locah.archiveshub.ac.uk

[72] ReLoad - Repository for Linked Open Archival Data (Italy, 2010-2013, Archivio Centrale dello Stato, Istituto per i Beni culturali dell'Emilia-Romagna and regesta.exe), http://labs.regesta.com/progettoReload/; see also their project description for the LODLAM 2013 summit (ReLoad 2013).

[73] LiAM - Linked Archival Metadata project (USA, 2012-2013, led by Tufts University, Digital Collections and Archives), http://sites.tufts.edu/liam/

[74] LOD-LAM, http://lodlam.net

[75] Linked Heritage (EU, ICT-PSP, 2011-2013), http://www.linkedheritage.eu

[76] A strong impact have also had the cultural heritage aggregation projects such as Cultura Italia (http://dati.culturaitalia.it); Swedish Open Cultural Heritage (K-samsök, http://www.ksamsok.se/in-english/), and of course Europeana, which has published one of the largest Linked Data sets comprising records of museums, archives and libraries across Europe (http://labs.europeana.eu/api/linked-open-data/introduction/).

[77] MASA - Mémoire des Archéologues et des Sites Archéologiques, http://masa.hypotheses.org

[78] HUMA-NUM: Consortiums, http://www.huma-num.fr/consortiums

[79] ARIADNE Linked Data SIG, http://www.ariadne-infrastructure.eu/Community/Special-Interest-Groups/Linked-Data

the Linked Data approach by archaeological research and other organisations. The participants saw a clear need of raising awareness of advantages offered by Linked Data and promoting further adoption in the sector. Furthermore, to leverage the creation and interlinking of Linked Data resources, practical guidance and easy to use tools are necessary.

In the second meeting of the ARIADNE Linked Data SIG (2014), Leif Isaksen, the chair of the CAA Semantic SIG[80], characterized the current phase of archaeological Linked Data as *"a period of experimentation"*. Group members expected that from this experimentation some projects will pave the way to a broader adoption and increasing utility of Linked Data in archaeology.

The requirements for a wider uptake recognised by the ARIADNE Linked Data SIG are also emphasised by the community that aims to interlink information about the ancient world. In 2012 the 3-day Linked Ancient World Data Institute meeting (LAWDI 2012) brought together projects and interested new users in this field. The meeting report notes: *"Essentially all LAWDI participants were eager to show resources that provide stable URIs or to ask for advice on what is currently available. But both the participants in and organizers of LAWDI recognize the need to take active steps to grow the number of high-quality digital resources. That will require ongoing outreach as well as clear examples of how Linked Open Data benefits both creators and users"* (Elliott, Heath & Muccigrosso 2012: 45).

From the Linked Ancient World Data Institute (LAWDI) meetings in 2012 and 2013 a collection of 30 articles originated which illustrates the adoption of the Linked Data approach in the Ancient World research community and what it takes to move from concept to actual implementation and operation (Elliott, Heath & Muccigrosso 2014). The papers cover a wide range of cultural objects, topics and information resources including, among others, cuneiform tablets, epigraphy, numismatics, prosopography (information about people), ancient and classical literature, publication of bibliographies and reviews, location/mapping services, historical periodization, integration of historical-geographic information, and more.

5.3 The Ancient World research community as a front-runner

At the "Linked Pasts" colloquium, which was organised by the Pelagios project at King's College London (20-21 July 2015), one topic was the importance to demonstrate benefits of using Linked Open Data. LOD developers in research fields of ancient history and classics were recognised being closer to this goal than early adopters in archaeology. As summarized in an article on the ARIADNE website: *"Of most interest to ARIADNE were the reasons Classics has been more successful than other cultural heritage domains (i.e. archaeology generally) at successfully implementing LOD. This was stated as primarily down to a lack of resources, heterogeneity of data, and (therefore) difficulty demonstrating clear benefits"* (ARIADNE 2015d). When we ask why some fields of Ancient World and Classics research are more advanced than Archaeology with regard to Linked Data, the heterogeneity of data in archaeological project archives and databases indeed is a major factor.

Advantage of specialties

While archaeologists unearth and document a large variety of built structures, cultural artefacts and biological remains, related Ancient World and Classics research specialties typically focus on one type of artefacts such as inscriptions (epigraphy), coins (numismatics), ceramics, and others. Consequently in these (smaller) research communities it is easier to establish and promote the use of common

[80] CAA Semantic SIG, https://groups.google.com/forum/#!forum/caa-semantic-sig

description standards. These standards are applied to databases of artefact collections, which have often been created (at least in part) from finds of archaeological excavations. The difference generally is that in archaeology the basic unit of research and analysis is the archaeological site, while research in specialities of Ancient World and Classics builds on collections or, in the case of texts, a corpus.

One leading example among the specialties is the international Nomisma[81] collaboration (since 2010) that develops description standards for coins (e.g. the Nomisma Ontology which provides stable URIs for numismatic concepts and entities), produces Linked Data sets of major collections, and shares them under open licenses. One reference implementation is Online Coins of the Roman Empire (OCRE)[82] of the American Numismatic Society (Gruber *et al*. 2013; Meadows & Gruber 2014).

The ontology and Linked Open Data methodologies established by Nomisma are employed by several other numismatics resources, for example, Antike Fundmünzen Europa[83], a web-based coins database developed by the Romano-Germanic Commission of the German Archaeological Institute (Tolle & Wigg-Wolf 2016). The Commission also coordinates the European Coin Find Network - ECFN and several joint meetings of ECFN and Nomisma have been organised[84].

Concerning pottery datasets the Kerameikos[85] initiative follows lessons learned in the development of Nomisma and aims to develop a thesaurus that defines domain concepts with URIs and RDF for representing and sharing pottery data across disparate systems. The initiative has been introduced with a paper at the CAA 2014 conference in Paris that demonstrates the potential (Gruber & Smith 2015), followed by a roundtable on LOD applied to pottery databases at the CAA 2015 conference in Siena (Gruber *et al*. 2015). Initially Kerameikos focuses on concepts within Greek black- and red-figure pottery, to be extended to other fields of pottery studies. See also the case study presented by Thiery (2014) on a LOD approach to simian ware, linking potters, pots and places.

Another broad field of research is inscriptions (epigraphy), where the Europeana Network of Ancient Greek and Latin Epigraphy (EAGLE)[86] project has achieved a substantial advance (Casarosa *et al*. 2014; Liuzzo 2014 and 2016). This includes a conceptual and a metadata model based on CIDOC CRM and TEI/EpiDoc, respectively (EAGLE 2015), and a set of vocabularies for classical epigraphy in SKOS format[87].

Coins, pottery and inscriptions are but three examples chosen because they concern material artefacts familiar to archaeologists. Other examples of LOD oriented initiatives concern the domain of ancient and classical texts. For example, the Standards for Networking Ancient Prosopographies (SNAP)[88] project defines annotation conventions and builds a single virtual authority list for referencing ancient people, brought together from different authoritative lists of persons and names.

[81] Nomisma, http://nomisma.org

[82] Online Coins of the Roman Empire (OCRE), http://numismatics.org/ocre/

[83] Antike Fundmünzen in Europa (AFE), http://afe.fundmuenzen.eu

[84] European Coin Find Network (ECFN), http://www.ecfn.fundmuenzen.eu

[85] Kerameikos, http://kerameikos.org

[86] Europeana Network of Ancient Greek and Latin Epigraphy - EAGLE (EU, ICT-PSP, 4/2013-3/2016), http://www.eagle-network.eu

[87] EAGLE vocabularies (Material, Type of inscription, Execution technique, Object type, Decoration, Dating criteria, State of preservation), http://www.eagle-network.eu/resources/vocabularies/

[88] Standards for Networking Ancient Prosopographies – SNAP (UK AHRC funded project, 2014-2015), http://snapdrgn.net

A focus on common description standards for certain types of Ancient World artefacts and texts does of course not mean ignoring their relations with other subject areas and common issues. As the "Linked Ancient World Data: Relating the Past" panel at the Digital Humanities 2016 conference explains, these projects *"are also concerned with issues far beyond their primary subject area: the interoperability of bibliographical references, citations of ancient sources, encoding of date and time, events and actors, material objects and their curatorial history all contribute to the study and understanding of the ancient world (and mutatis mutandis of any other). All also recognise that there is no firm demarcation between the cultures of the Mediterranean in the classical period, nor between the worlds and cultures bordering them in time and space"* (Linked Ancient World Data 2016).

Important to note is that all Linked Data efforts mentioned are about artefacts and texts, while a large segment of archaeological research concerns biological remains of humans, animals and plants. However, biological vocabularies are not developed by archaeologists, but by taxonomists (with regard to species names)[89], Biodiversity Information Standards (TWDG)[90], who develop Life Science Identifiers (LSID) and vocabularies for biodiversity information, and expert groups that produce relevant biological ontologies which are shared via the BioPortal[91]. While authoritative species names are widely used by archaeobotanists and zooarchaeologists, other standards such as biological ontologies seem to be employed seldom. Indeed, we found only example where such an ontology, the Uber Anatomy Ontology (UBERON)[92] has been used in a zooarchaeological Linked Data project (Kansa *et al*. 2014; Whitcher-Kansa 2015).

Pelagios as a common platform

The strongest impression of the Ancient World research community being a front-runner in humanities LOD comes from Pelagios[93], which since 2011 supports connecting various scholarly resources through the places and other geographic entities they refer to. Pelagios is a loose confederation of many organisations and projects that have agreed to use for such references the Open Annotation[94] RDF vocabulary and URIs of gazetteers of the ancient world geography, *in primis* Pleiades[95] but also others (e.g. iDAI.gazetteer[96], Digital Atlas of the Roman Empire[97], Vici.org[98] and others). Among the currently 21 dataset contributors of Pelagios are the ARIADNE partners German Archaeological Institute (iDAI.objects database with 87,735 references concerning 5363 places) and Fasti Online (with 686 references concerning 256 places)[99].

Pelagios aggregates the annotations, which are hosted by the data providers (often in the form of an RDF dump), and makes them available through a map-based search interface and an API so that

[89] A major integrator in this field is the Catalogue of Life, http://www.catalogueoflife.org

[90] TDWG - Biodiversity Information Standards, http://www.tdwg.org

[91] BioPortal (US National Center for Biomedical Ontology), https://bioportal.bioontology.org

[92] UBERON - Uber Anatomy Ontology, http://uberon.org

[93] Pelagios, http://commons.pelagios.org

[94] Open Annotation Collaboration, http://www.openannotation.org

[95] Pleiades, http://pleiades.stoa.org

[96] iDAI.gazetteer (German Archaeological Institute), http://gazetteer.dainst.org

[97] Digital Atlas of the Roman Empire (Department of Archaeology and Ancient History, Lund University, Sweden), http://dare.ht.lu.se

[98] Vici.org - Archaeological Atlas of Antiquity (community-based gazetteer), http://vici.org

[99] Pelagios: Datasets, http://pelagios.org/peripleo/pages/datasets

developers can build on the data. The annotation platform Recogito aids the process of identifying places referred to in individual digital texts and maps and linking them to a gazetteer, supported by an automated suggestion system (Simon *et al*. 2015). Currently in development is Peripleo, a tool to explore the growing pool of data as a whole and to progressively filter and drill down to individual records (Simon *et al*. 2016).

Isaksen *et al*. (2014) address several factors which determined the success of the Pelagios initiative. Among the most important arguably are the lightweight Linked Data approach, focus on geographical references as the most common feature of the various data resources, quick demonstration of benefits from associating contributors' data, and the sustained funding by the Andrew W. Mellon Foundation (since 2013, currently by a grant until 2018[100]). But they also note, *"we are at the tip of the iceberg even in this case as the overwhelming majority of classicists and classical archaeologists have never heard of Linked Open Data"* (Isaksen *et al*. 2014).

In summary, major factors that contribute to an advanced position of the Ancient World research community in the application of the Linked Data approach are: a) there are groups who develop and promote description standards in certain specialities, and b) there is a common platform (Pelagios) that allows linking of information based on a light-weight approach. Archaeological projects can benefit from this development, for example, use the Nomisma description standards for coin finds.

[100] Initial funding in 2011-2012 by JISC (UK) and grants for special projects in 2014-2015 by AHRC (UK) and Open Knowledge Foundation.

5.4 Brief summary and recommendations

Brief summary

In the areas addressed by this study, cultural heritage institutions are among the leading adopters of the Linked Data approach. The Ancient World and Classics research community is a front-runner of uptake on the research side, while there have been only few projects for Linked Data of archaeological research data.

This situation is due to considerable differences between cultural heritage institutions and research projects, and between projects in different domains of research. For cultural heritage institutions such as a libraries, archives and museums adoption of Linked Data is in line with their mission to make information about heritage readily available and relevant to different user groups, including researchers. Adoption has also been promoted by initiatives such as LOD-LAM, the International LOD in Libraries, Archives, and Museums Summit (since 2011). In the field of archaeological research there were no such initiatives or only at small scale, for example sessions at CAA conferences or national thematic workshops. But promotional activities, particularly at the national level, are important to reach archaeological institutes and research groups and make them aware of the Linked Data approach.

Adoption in the Ancient World and Classics research community is being driven by specialities such as numismatics and epigraphy, where there are initiatives to establish common description standards based on Linked Data principles. The goal here is to enable annotation and interlinking of information of special collections or corpora for research purposes. The focus on certain types of artefacts (inscriptions, coins, ceramics and others) provide clear advantages with regard to the promotion of the Linked Data approach within and among the relatively small research communities of the specialities.

A good deal of the recognition of the Ancient World and Classics research community being a front-runner in Linked Data also stems from the Pelagios initiative. Pelagios provides a common platform and tools for annotating and connecting various scholarly resources based on place references. Pelagios clearly demonstrates benefits of contributing and associating data of the different contributors based on a light-weight Linked Data approach.

Archaeology presents a more difficult situation, in that the basic unit of research is the site, where archaeologists unearth and document a large variety of built structures, cultural artefacts and biological material. The heterogeneity of the archaeological data and the site as focus of analysis present a situation where the benefits of Linked Data, which would require semantic annotation of the variety of different data with common vocabularies, are not apparent. Therefore adoption of the Linked Data approach can be hardly found at the level of individual archaeological excavations and other fieldwork, but, in a few cases, community-level data repositories and databases of research institutes. Repositories and databases, not individual projects, should also in next years be the prime target when promoting the Linked Data approach.

All proponents of the Linked Data approach, including the ARIADNE Linked Data SIG as well as the directors of the Pelagios initiative, agree that much more needs to be done to raise awareness of the approach, promote uptake, and provide practical guidance and easy to use tools for the generation, publication and interlinking of Linked Data.

Recommendations

- *More needs to be done to raise awareness and promote uptake of the Linked Data approach for archaeological research data. In addition to sessions at international conferences, promote the approach to stakeholders such as archaeological institutes at the national level.*
- *The prime target when promoting the approach should be community-level data repositories and databases of research institutes (not individual projects).*
- *To drive uptake provision of practical guidance and easy to use tools for the generation, publication and interlinking of Linked Data is necessary.*
- *Promote the use of established and emerging semantic description and annotation standards for artefacts such as coins, inscriptions, ceramics and others; for biological remains of plants, animals and humans suggest using available relevant biological vocabularies (e.g. authoritative species taxons, life science ontologies, and others).*
- *Contribute to the Pelagios platform (where appropriate) or aim to establish similar high-visibility data linking projects for archaeological research data.*

6 Requirements for wider uptake of the Linked Data approach

Linked Open Data (LOD) allow for semantic interoperability of dispersed and heterogeneous data resources. Despite this potential LOD is not produced and applied yet by many research institutions and projects in the archaeological sector. The sections of this chapter address different requirements and approaches for fostering a wider uptake of the Linked Data approach for archaeological research data. The aim is to present the current state with regard to impediments, potential drivers and exemplary projects, and for each area of identified requirements provide practical recommendations for Linked Data developers and other stakeholders.

6.1 Raise awareness of Linked Data

Linked Data enable interoperability of dispersed and heterogeneous information resources, allowing the resources to become better discoverable, accessible and re-useable. In a fragmented data landscape as present in the sector of archaeology this is substantial value proposition. Indeed, in an ARIADNE online survey on top of the expectations of about 500 researchers, research directors and other respondents from a data portal were cross-searching of data archives with innovative, more powerful search mechanisms (ARIADNE 2014a: 114, about 500 respondents).

But such expectations are not necessarily associated with capabilities offered by Linked Data. Therefore the gap between advantages expected from advanced data services and “buy in” and support of the research community for Linked Data must be closed by targeted actions. This section addresses the situation of a highly fragmented landscape of archaeological data, presents some available results on the awareness of Linked Data by cultural heritage organisations and archaeologists, and suggests whom to consider as priority target groups for Linked Data initiatives.

6.1.1 Fragmentation of archaeological data

The ARIADNE “First Report on Users’ Needs” (ARIADNE 2014a) identified major general factors that impede the uptake of the Linked Data approach in the domain of archaeological research. The results of the literature review, pilot interviews and online survey made clear that the archaeological data landscape is characterized by high fragmentation due to several factors. These factors include, but are not limited to

- diverse organisational settings (research institutes, heritage management agencies, museums and others) in which data are collected and managed,
- data management practices that are predominantly focused on individual projects, rather than an institutional or domain oriented perspective (e.g. “project archives”, one per excavation site, stored on a file servers, etc.),
- a low level of open sharing of research data, due to lack of recognition and rewards for making the data available, the additional work effort for documenting data sets for proper archiving, and lack of community archives in many countries.

The situation does not present favourable conditions for the integration and linking of archaeological data sets through data e-infrastructures such as ARIADNE. Therefore ARIADNE encourages initiatives to establish state-of-the-art community-level data archives in countries where they are missing at present. This suggestion is in line with the development that research funders increasingly demand data management & access plans with the goal to make the generated research data openly accessible through digital archives (open data mandates).

Research projects will have to think about data management from the start, including where to deposit their data, required metadata, and licensing agreements. Also some scientific journals now require a data availability statement, i.e. that the data which underpins published research is available in an accessible archive. However with regard to promoting archaeological Linked Data the primary focus must not necessarily be individual researchers, research groups and projects. Because data produced by projects will increasingly be deposited in accessible data archives, according to sector standards with regard to metadata and vocabularies.

6.1.2 Current awareness of Linked Data

Results for cultural heritage organisations

It is worthwhile having an indication of the current state of awareness and knowledge of Linked Open Data (LOD) at cultural heritage organisations, some of which may curate archaeological artefacts among other objects and content. The AthenaPlus project[101] conducted a survey among partners and other organisations about their awareness of LOD and existing initiatives, how they get information about LOD, and if they already use LOD (AthenaPlus 2013b). 28 questionnaires were returned by respondents of organisations located in 16 EU countries. The respondents worked at museums, libraries, archives, data aggregators and other organisations, including ministries, governmental agencies, university research centres and IT service organisations. Thus a rather small number of responses from diverse organisations were received. The survey results were as follows:

Questions	***Yes***	***No***
Are you or your organisation familiar with the concept of Linked Open Data (LOD)?	25	3
Do you or your organisation know of any LOD projects or initiatives in your country in the field of cultural heritage?	19	9
Have you or your organisation had experience of using LOD in connection with your collections?	6	22
Have you or your organisation had experience of publishing LOD in connection with your collections?	4	24
Does your organisation plan to publish LOD in the near future?	21	7
Does your organisation plan to connect with new LOD sources in the near future? (1 did not answer this question)	14	13

In summary, most respondents to the AthenaPlus survey said that they (or their organisation) are familiar with Linked Open Data and knew of related projects and initiatives in their country. But only few had first-hand experience with LOD. At the same time, most had plans to publish and/or consume LOD in the near future.

Sixteen respondents answered an open question on their expectations from connecting own data with LOD resources. According to the survey authors the most common expectations related to *"enlarging accessibility of data in a broader context, increasing the visibility of collections, extend the*

[101] AthenaPlus (EU, CIP Best Practice Network, 3/2013-8/2015), http://www.athenaplus.eu

semantic relations between various collections, development of cross-domain interdisciplinary networks of knowledge, possibility of re-contextualizing the resources for improved research infrastructure. Recognized as an added value for the own collections was the possibility to enrich own data via (inter)national connections. One reply mentioned the prospect of easy access to valuable information for scientific research and the purpose to create educational apps."

Some respondents also considered possible disadvantages, which included loss of control over own published data, a decrease in data quality due to links to non-qualified sources, or an overload of links which might cause a loss of visibility and/or accessibility.

ARIADNE results for archaeology

One observer of the Semantic Web community notes: *"In contrast to the cultural heritage sector aka museums, the Semantic Web has seen less uptake in archaeology. This could be because archaeologists tend to focus on analysis and recording of the data rather than dissemination. Experiences are mostly limited to spreadsheets, relational databases and/or spatial data management. Many academic archaeologists remain protective of their data especially when it has not been published in traditional media. The complexity of combining siloed resources may be overwhelming"* (Solanki 2009).

However, researchers are not necessarily the primary target group of Linked Data awareness raising actions. The online survey reported in ARIADNE's "First Report on Users' Needs" (ARIADNE 2014a [April 2014]) had one question about how helpful researchers and data managers perceive different services ARIADNE might provide. Among nine options there was "Improvements in linked data", defined as *"interlinking of information based on Linked Data methods (i.e. methods of publishing structured data so that it can be interlinked)"*.

Not surprisingly, this option was at the bottom of the researchers' list of perceived helpfulness, only the service option "Content recommendations based on collaborative filtering, rating and similar mechanisms" fared worse. But of the over 470 researchers who answered the question still 37% thought "Improvements in linked data" could be "very helpful" and 43% "rather helpful" (ARIADNE 2014a: 114). The good results for "Improvements in linked data" indicate that interlinking of research results is generally relevant to researchers and, arguably, that quite some researchers had already heard about Linked Data as a novel way of interlinking information.

An additional survey addressed repository managers that are a considerably smaller target group than researchers. The survey received 52 sufficiently filled questionnaires, hence a good response but certainly not representative. The managers were asked if their repository and clients could benefit from services ARIDANE might provide, presenting the same list of service options as the survey of researchers. Among the managers who answered the question (32), the option "Improvements in linked data" fared better: it came in on position five of the nine options with 39% "very helpful" and 39% "rather helpful". The favourite was "Services for Geo-integrated data", 52% "very helpful", 32% "rather helpful" (ARIADNE 2014a: 141).

The repository managers in general were more sceptical about potential improvements, but they appreciated "Improvements in linked data" considerably more than the researchers. As noted, the results for the data managers are far from representative. But we think that they are indicative and add to our view that data managers are a more relevant target group for the Linked Data approach than researchers. Data managers are active in different contexts, digital archives of the research community, repositories of individual institutions (e.g. university, research center), and large archaeological projects in need of systematic and long-term data management. Within ARIADNE, consultancy and training for Linked Data has been mainly given to managers of institutional data resources with regard to vocabularies that are being used for the metadata of the resources, e.g. related to the mapping of the vocabularies to the Art & Architecture Thesaurus.

In the ARIADNE portals survey for the "Second Report on Users' Needs" (ARIADNE 2015a) 23 experts of project partners (18 of which archaeologists) studied existing information portals, defined as websites that provide access to content of more than one institution or project. The aim was to identify good practices and give further ideas for the development of the ARIADNE data portal. Some participants considered Linked Data for integrating information within the portal and linking to external resources. The statements addressed the potential of the Linked Data approach as well as the current lack of awareness of the benefits of such data; also the need of high-quality Linked Data was mentioned (ARIADNE 2015a: 103-104).

The suggestions of the survey participants concerning Linked Data were summarised in three recommendations for the ARIADNE data portal and evaluated by project partners (28 experts) with regard to their relevance and time-horizon (ARIADNE 2015e: 282-287). Among the top-ranked of all 34 recommendations of the portals survey was *"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)"*. 79% of the evaluators considered this as relevant and 86% thought that it might be achieved within the formal duration of the project (until January 2017). The evaluators were less confident with regard to encouraging a wider uptake of LOD principles among archaeological institutions and projects, but about 60% expected that the project will promote this.

6.1.3 Brief summary and recommendations

Brief summary

Linked Data enable interoperability of dispersed and heterogeneous information resources, allowing the resources to become better discoverable, accessible and re-useable. In the fragmented data landscape of archaeology this is substantial value proposition. In the ARIADNE online survey on top of the expectations of the archaeological research community from a data portal were cross-searching of data archives with innovative, more powerful search mechanisms. But such expectations are not necessarily associated with capabilities offered by Linked Data. Therefore the gap between advantages expected from advanced services and "buy in" and support of the research community for Linked Data must be closed by targeted actions.

A small survey of the AthenaPlus project (2013) indicated that cultural heritage organisations are already aware of Linked Data, but few had first-hand experience with such data. Among the expectations from connecting own and external Linked Data resources were increasing the visibility of collections and creating relations with various other information resources. Some respondents also considered possible disadvantages, e.g. loss of control over own data or a decrease in data quality due to links to non-qualified sources.

In the ARIADNE online survey (2013) "Improvements in linked data", i.e. interlinking of information based on Linked Data methods to enable better information services, was considered more helpful by repository managers than researchers. Researchers of course perceive interlinking of information as important, but may not see this as an area for own activity. Indeed, we think individual researchers and research groups should not be a primary focus of Linked Data initiatives. Managers of digital archives of the research community and institutional repositories are much more relevant target groups. Furthermore data managers of large and long-term archaeological projects should be addressed as they will also consider required standards for data management and interlinking more thoroughly.

Recommendations

- *Address the highly fragmented landscape of archaeological data and highlight that Linked Data can allow dispersed and heterogeneous data resources become better integrated and accessible.*
- *Consider as primary target group of Linked Data initiatives not individual researchers but managers of digital archives and institutional repositories.*
- *Include also data managers and IT staff of large and long-term archaeological projects as they will also consider required standards for data management and interlinking more thoroughly.*

6.2 Clarify the benefits and costs of Linked Data

One targeted action to help close the current Linked Data adoption gap in the archaeological sector could be removing the widespread notion of an unfavourable ratio of costs compared to benefits of employing Semantic Web / Linked Data standards for information management, publication and integration. While the standards have matured and become much better applicable this notion is still prevalent and a barrier to wider adoption of the Linked Data approach.

6.2.1 The notion of an unfavourable cost/benefit ratio

In a paper titled *"Is Participation in the Semantic Web Too Difficult?"*, published in 2002, the authors emphasised the need of lowering the entry barrier for cultural heritage organisations, especially small ones, by offering significant added value and advantages over established ways of content management and publication (Haustein & Pleumann 2002). The authors note that initial steps towards the Semantic Web will require some extra effort and, therefore, *"the system needs to ensure that this cost is outweighed by the gain for the content provider. This gain should not count too much on the network effect of the Semantic Web, because this effect might take some time to really pay off. Instead, the gain has to be immediately visible to the content provider."*

In the DigiCULT Forum thematic issue *"Towards a Semantic Web for Heritage Resources"* (2003) the position paper stressed that it is difficult to legitimate investment of institutions in the Semantic Web, because over the next five years it would bring little benefit (Ross 2003). A DigiCULT Forum assessment in 2004 of the readiness of heritage institutions for several e-culture technologies argued that Semantic Web technologies would be adopted primarily by large institutions in a longer-term perspective of 6 or more years (Geser 2004).

With regard to an archaeological semantic Web Julian Richards in 2006 noted an increase in online available documents and archives so that "*there should be no shortage of content with which to build such a web"*; however *"archaeology could get left behind if the rewards for creating the mark-up necessary to make the Semantic Web a reality are only evident in the commercial sector. The sector is currently more likely to participate in Berners-Lee's vision through the creation of semantic mark-up for information about monument access arrangements, opening hours and facilities for the tourism industry than for academic research*" (Richards 2006: 977).

Reasons for the doubts of a quick adoption of Semantic Web standards and technologies included still on-going standardization work, need for specialist knowledge, little experience of implementation under real world conditions and, in particular, expected high costs of conversion of legacy metadata and knowledge organization systems such as thesauri to Semantic Web standards.

6.2.2 Lack of cost/benefit evaluation

Unfortunately, little effort has been invested so far to make clear cost / benefit ratios of different levels and ways in which Linked Data can be produced and employed. Among the exceptions is a model that considers "pay-off points" of five escalating levels at which information can be formalized (Isaksen *et al*. 2010a/b). The purpose of the model is to encourage a step-wise adoption of Linked Data principles, including for small-scale data sources (i.e. "small tail" data sets). The authors consider that *"(at least) five escalating levels of semantic formalization can be identified, each with differing requirements and benefits for the implementer: i. Literal Standardization, ii. Instance URI generation, iii. Canonical URI mapping, iv. RDF generation, and iv. Database-schema-to-Ontology mapping"* (Isaksen *et al*. 2010a).

In this scheme (i) means the creation and use of a locally defined restricted vocabulary (e.g. list of terms or thesaurus), (ii) the creation of web-accessible unique identifiers for the proprietary vocabulary terms, and (iii) mapping of the terms to established concepts/terms of an acknowledged authority. The suggested approach seems at odds with the Linked Data principle that projects should wherever possible re-use established vocabulary, however "normalization" of terms will often be necessary when attempting to integrate different legacy datasets. This was the case in the Roman Ports in the Western Mediterranean Project (Isaksen *et al*. 2009) to which the authors refer in the discussion of the suggested scheme of semantic formalization.

The authors emphasise *"that Linked Data – hitherto seen as the simplest semantic approach – is relatively advanced in this scheme. We argue that data providers should be encouraged to migrate towards full semantic formalization only as their requirements dictate, rather than all at once. Such an approach acts as both a short and long-term investment in semantic approaches, in turn encouraging increased community engagement. We also propose that for such processes to be accessible to data-curators with low technical literacy, assistive software must be created to facilitate these steps"* (Isaksen *et al*. 2010a).

The authors also address benefits and costs (or, rather, requirements) of the different levels of semantic formalization, although only generically. For example, that RDF generation allows machines to exploit the URI linkage for data aggregation and discovery, but requires a basic grasp of ontological modelling, selection and/or creation of predicate URIs, tools or scripting for the RDF generation, and maybe new/unfamiliar RDF data storage mechanisms.

The suggested approach of a stepwise migration towards Linked Data seems reasonable. But without a method for evaluating the "pay-offs" in terms of the cost/benefit ratio, and a number of reference examples, it will remain theoretical and of little help in driving "buy in" of potential Linked Data providers.

The key point of the approach is to look for different levels at which Linked Data can be employed. In this regard Eric Kansa of the archaeological data publication platform Open Context provides a helpful discussion of what can be considered as medium and high-level routes to Linked Data (above the low-level semantic formalizations mentioned by Isaksen *et al*.).

Kansa (2014a) sees the medium-level route in annotation and cross-referencing of data using shared controlled vocabularies, while the high-level is represented by employing the CIDOC CRM to align datasets based on shared conceptual modelling (level iv. "Database-schema-to-Ontology mapping" in the model suggested by Isaksen *et al*. 2010a). Referring to experiences from Open Context projects Kansa is convinced *"that vocabulary alignment can help researchers more, at least in the near-term, than aligning datasets to elaborate semantic models (via CIDOC-CRM)".* At least it allows reaching *"some lower-hanging, easier to reach fruit in our efforts to make distributed data work better together"* and *"meet more immediate research needs"*.

One example of such a project employed annotations to common vocabularies to enable the integration and comparison of zooarchaeological datasets from 17 sites (in total over 294,000 records of bone specimens). Each dataset had its own organization (schema) and used somewhat different proprietary vocabulary/terminology. The project annotated dataset-specific taxonomic categories with Web URIs for animal taxa curated by the Encyclopedia of Life[102], annotated classifications of bone elements with concepts of the Uber Anatomy Ontology (UBERON)[103], and employed a vocabulary developed by Open Context for bone fusion, sex determinations and standard measurements. The vocabulary alignments provided the basis for data integration and comparison across the different datasets (Arbuckle *et al*. 2014; Kansa *et al*. 2014; Whitcher-Kansa 2015).

Concerning the CIDOC CRM, the high-level route of aligning datasets based on shared conceptual modelling, despite its increasing adoption little is known about the cost / benefit ratio. While considerable benefits have been reported in some cases, the cost side is usually not addressed.

For example, Jordal *et al*. (2012) report benefits and new opportunities opened up by the CRM-based integration of ethnographic collections held by the Museum of Cultural History in Oslo. Connecting the collections via a CRM-based model allows the curators integrated access to the legacy catalogues and databases, and the model also guides the registration of new items. The integration of the collections also *"gives a better basis for telling a story for each artefact"*, and *"provides a possibility to do research on the objects with as complete, accurate and rich data as possible"*.

Other institutions have achieved a lot by applying the CIDOC CRM to integrate large and heterogeneous datasets, enable advanced search on their website, and participate in cultural heritage web portals. One outstanding example in this regard is Arachne, the central object database of the German Archaeological Institute (DAI) and the Archaeological Institute of the University of Cologne[104]. The CIDOC CRM based internal integration of data allows advanced exploration of a mass of heterogeneous information resources. Arachne also participates in CLAROS - Classical Art Research Online Services (launched in May 2011)[105] which provides a portal for searching several sources for Classical studies based on the Linked Data approach and CIDOC CRM.

Oldman & Rahtz (2014) highlight that the CLAROS project *"established the credentials of the CIDOC CRM standard as a semantic framework that can harmonise data from many different institutions while providing a richer environment (when compared to its digital sources) in which to explore and research cultural heritage data"*. But the CLAROS Linked Data based search environment offers rather limited research functionality. The ResearchSpace project[106], in which Dominic Oldman serves as principal investigator, aims to enable advanced exploration and research of CIDOC CRM mediated cultural heritage data.

---

[102] Encyclopedia of Life, http://eol.org

[103] UBERON - Uber Anatomy Ontology, http://uberon.org

[104] Arachne, http://arachne.uni-koeln.de

[105] CLAROS, http://www.clarosnet.org; http://data.clarosnet.org

[106] ResearchSpace, http://www.researchspace.org

6.2.3 Collecting examples of benefits and costs

Benefits of Linked Data

The basic assumption of Linked Data is that the usefulness and value of data increases the more readily it can combined with relevant other data. The Linked Data approach of using stable URIs, typed RDF links and common vocabulary greatly supports benefits from bringing together related information. Berners-Lee described benefits of Linked Data with phrases such as *"to provide context"* or that users *"can discover more things"* (Berners-Lee 2006 and addition on 5-star data in 2010).

Indeed, convincing tangible benefits of Linked Data materialise if information providers can draw on own and external data for enriching services. A prominent early example is that the BBC used DBpedia (Wikipedia Linked Data)[107] und MusicBrainz Linked Data[108] to enrich the information of their music pages (Kobilarov *et al*. 2009; Raimond *et al*. 2013 report on BBC's use of Linked Data for other services). An example from the museum world is the Smithsonian American Art Museum (SAAM) that enriches their artist pages with identifiers of the Getty Union List of Artist Names (ULAN) and information from DBpedia and *New York Times* Linked Data (Szekely *et al*. 2013; Zaino 2013).

Szekely *et al*. (2013) summarize the benefits for the SAAM as follows: *"the linked data provides access to information that was not previously available. The Museum currently has 1,123 artist biographies that it makes available on its website; through the linked data, we identified 2,807 links to people records in DBpedia, which SAAM personnel verified. The Smithsonian can now link to the corresponding Wikipedia biographies, increasing the biographies they offer by 60%. Via the links to DBpedia, they now have links to the New York Times, which includes obituaries, exhibition and publication reviews, auction results, and more. They can embed this additional rich information into their records, including 1,759 Getty ULAN identifiers, to benefit their scholarly and public constituents."*

This suggests that the benefit of Linked Data may somehow be calculated based on the increase in richness of information services per dataset added, also considering different beneficiaries such as (in this example) art historians, journalists and people generally interested to learn about artists and art works.

Similar examples should be collected or developed as Linked Data use cases for datasets of archaeological research projects and archives/collections. It seems clear that popular Linked Data resources like Wikipedia may not be appropriate for purposes of archaeological research. But there are other resources, for example, among the extensive Linked Data of the bio-sciences which might be exploited for relevant research use cases concerning human, animal or plant remains (e.g. the example of zooarchaeological Linked Data reported in Kansa *et al*. 2014).

But some differences between benefits of enriching via Linked Data museum or archive information and integrating research data should be noted. Cultural heritage institutions can benefit from making their collections more meaningful and relevant to end-users by adding external contextual information (links to related content). In a web of richly interlinked information the in-coming links can also leverage usage of own content. This is fully in line with the institutions' mission to communicate contextualised cultural heritage to an as wide as possible audience.

In the realm of research the benefits of Linked Data should be reflected in terms of research dividends that can be gained by interlinking data. Such dividends for example are discovery of

[107] DBpedia, http://wiki.dbpedia.org

[108] LinkedBrainz - MusicBrainz in RDF and SPARQL http://linkedbrainz.org

relations between research data worth exploring further, combination of data from different projects in ways that enable interesting new lines of research, different views on data from various disciplinary perspectives suggesting interdisciplinary approaches, etc. (see the discussion of search vs. research in *Section 6.6*).

Costs of Linked Data

In order to evaluate the costs of Linked Data providers, information about the different cost factors and drivers should be collected. A good understanding of the costs of different Linked Data projects may help to possibly reduce the costs, for example, by providing dedicated tools, guidance and support for certain task.

The costs in general concern the acquisition of the expertise and the work effort and tools required for the actual generation, publication and interlinking of the data. Basic steps in the process are to select relevant data, clean it, design the URIs, convert the data to RDF, store and make it accessible, map proprietary terms to established domain vocabulary, and find and create links to related data on the Web[109] (see *Section 3.5*).

For the process steps information about the costs should be collected and analysed, taking account of projects of different types and sizes. As an example of required information: In the MultimediaN E-Culture project several legacy datasets from different institutions have been converted to Linked Data and integrated (Omelayenko 2008): It was found that nearly every dataset required some dataset-specific code to be written. But by identifying and separating conversion rules that could be re-used the overall effort was reduced considerably. Nevertheless, it has been estimated that a skillful professional who uses a state-of-the-art conversion support tool (in this case, AnnoCultor) needed around four weeks to transform a major museum database, creating for this purpose a dedicated converter of 50-100 conversion rules plus some custom code.

Some new methods and tools have reduced considerably the costs of data conversion, publication, annotation and linking. For example, Van Hooland *et al*. (2012a) of the Free Your Metadata initiative[110] argue that the interactive data cleaning and transformation tool OpenRefine[111] *"has made data cleaning and reconciliation available for the masses"*. Clearly data cleaning, trans-formation and reconciliation (matching entities with other Linked Data) are essential steps in Linked Data generation. The authors illustrate the case with metadata of the Cooper-Hewitt National Design Museum, New York and the Powerhouse Museum, Sydney (Van Hooland *et al.* 2012a and 2012b).

Numerous other tools are available ranging from tools for specific tasks to comprehensive Linked Data generation, management and publication platforms. The proliferation of tools means that potential Linked Data providers need expert advice on what to use (and how to use it) for their purposes and specific datasets, taking account also of existing legacy systems, standards in use, etc.

Particularly relevant in this context are approaches that allow exploiting legacy databases and avoid keeping and managing RDF data separately in a dedicated database (triple store). Various solutions are available to output data in RDF from existing databases (Sahoo *et al*. 2009; Michel *et al*. 2013)[112]. This requires a mapping of the database to RDF, which may be created automatically (for simple databases) but more often needs an expert mapping to a domain ontology in RDF Schema or OWL.

[109] W3C (2014) Working Group Note: Best Practices for Publishing Linked Data, 9 January 2014, https://www.w3.org/TR/ld-bp/

[110] Free Your Metadata, http://freeyourmetadata.org

[111] OpenRefine, http://openrefine.org

[112] One example is D2RQ - Accessing Relational Databases as Virtual RDF Graphs, http://d2rq.org

As an example of an archeological database, the Laboratoire Archéologie et Territoires, Université de Tours - CNRS, France aims to open up their ArSol - Archives du Sol (Soil Archives) system[113] based on a mapping of concepts of the relational database to the CIDOC CRM. This mapping is being used to query the database employing SPARQL-to-SQL rewrites (Le Goff E. *et al* 2015; Marlet *et al*. 2016). The approach avoids the extract-transform-load (ETL) process for exporting data in an RDF store and for updating it when data changes. The researchers employ the Ontop[114] platform developed by the Knowledge Representation meets Databases (KRDB) research group at the University of Bozen-Bolzano (Bagosi *et al*. 2014). The same approach and platform is being used by the EPNet project[115] (Calvanese *et al*. 2015; Calvanese *et al*. 2016).

Effective and easy-to-use tools are of utmost importance for reducing the costs of core tasks of Linked Data generation, publication and linking. But advice on how to best approach other tasks such as URI design or vocabulary selection is critical as well.

Here is not the place to address all steps in the so called lifecycle of Linked Data from data selection to RDF publication and use, particularly because cost figures are hard to come by. As an example, a study by PricewaterhouseCoopers for the Interoperability Solutions for European Public Administrations programme looked into business models for linked open government data services (Archer *et al*. 2013). One of their research questions therefore concerned the costs of the Linked Data services, including development, maintenance and promotion.

The study investigated 14 cases but did not bring out the cost structure of the Linked Data activities because most respondents did not separately account for this. Only the German National Library gave figures for specific development tasks and on-going work for Linked Data provision[116]: Initial development including mappings between internal database format and RDF vocabularies, implementation of data conversions, and standards related work consumed 221 person days; the estimated effort for maintenance was 1 FTE (full-time equivalent) but for the bibliographic services which included the supply of Linked Data; the cost specifically for the latter remained unclear (Archer *et al.* 2014: 3, 30 and 58).

A final important point, the discussion on costs of Linked Data in general (including above) centres on the data and vocabulary providers. But in the Linked Data ecology also the costs of potential users need to be considered. As one respondent to a discussion on why data providers should carry the costs of publishing Linked Data emphasised, *"in the current state of the world, it comes with added costs for the consumers as well. Most developers don't know much about RDF and surrounding tools and standards, so they have to learn about it in order to consume your dataset. These costs can easily outweigh potential benefits. Of course, the mission of the linked data community is to change that fact by popularizing RDF technologies and standards, so that might not be true anymore 5 years from now"* (Samwald 2010). Another respondent seconded this by adding, *"I don't mean to say Linked Data is not the way forward, I just don't think it's yet a representation that large numbers of people would feel comfortable or capable of working with, given what they currently know, what they currently do, and they culturally currently do it…"* (Hirst 2010).

---

[113] ArSol - Archives du Sol (Soil Archives), http://arsol.univ-tours.fr

[114] Ontop, http://ontop.inf.unibz.it

[115] EPNet - Production and Distribution of Food during the Roman Empire: Economic and Political Dynamics (ERC Advanced Grant project, 3/2014-2/2019), http://www.roman-ep.net

[116] Linked Data Service of the German National Library, http://dnb.de/EN/lds

Costs of knowledge organization systems

Knowledge organization systems (KOSs), including forms such as thesauri (terminology), taxonomies (classification systems) and ontologies (conceptual reference models) play a key role in Linked Data. Indeed without the semantics of KOSs a web of meaningful Linked Data cannot be built. Therefore it is astonishing that little is known about the costs of employing KOSs.

As an example, in a special issue of the Bulletin of the Association for Information Science and Technology published 2014 (ASIS&T 2014) on the economics of KOSs none of the five articles gives an example of the actual or estimated costs of a KOS. However, Denise Bedford in this bulletin elaborates in detail the assets and liabilities different types of "taxonomies" (her term for KOSs) generate, for example a flat list of terms vs. a thesaurus. Bedford also gives an overview of general categories of costs involved, but states: *"The actual costs of any taxonomy project are tied to its organizational context and the scope and scale of the effort. It is not possible or advisable to say that a typical thesaurus project can be completed for $100,000 or for $500,000 because there is no 'typical thesaurus' "* (Bedford 2014: 20).

Lack of solid knowledge about the costs of employing KOSs has a long "tradition" in the Semantic Web (Linked Data) community. For example, Tim Berners-Lee, Wendy Hall and Nigel Shadbolt, key figures of the community, in their paper "The Semantic Web Revisited" (Shadbolt *et al*. 2006) address the issue of costs but can only give "naïve but reasonable assumptions". They consider that in some application *"the costs – no matter how large – will be easy to recoup. For example, an ontology will be a powerful and essential tool in well-structured areas such as scientific applications. In certain commercial applications, the potential profit and productivity gain from using well-structured and coordinated vocabulary specifications will outweigh the sunk costs of developing an ontology and the marginal costs of maintenance. In fact, given the Web's fractal nature, those costs might decrease as an ontology's user base increases. If we assume that ontology building costs are spread across user communities, the number of ontology engineers required increases as the log of the user community's size. The amount of building time increases as the square of the number of engineers. These are naïve but reasonable assumptions for a basic model. The consequence is that the effort involved per user in building ontologies for large communities gets very small very quickly".* They go on discussing the difference between deep and shallow ontologies, requiring *"considerable effort"* (for the ontological conceptualization) and (unspecified) *"effort but over much simpler sets of terms and relations"* in the case of shallow ontologies (Shadbolt *et al*. 2006: 99).

Hepp (2007) addresses economic and other issues that constrain the development, adoption and maintenance of useful ontologies and other KOSs. He notes that KOSs are regarded as central building blocks of the Semantic Web, and much has been written about the benefits of using them, but that there are substantial disincentives for building and adopting relevant KOSs. He discusses interesting general assumptions, but also does not give a single cost figure.

Hepp assumes that KOSs exhibit positive network effects, hence their perceived utility will increase with the number of users. But convincing people to invest effort into building or using them is difficult in the initial phase in which there is no or only a small user base. The utility for early adopters is low, whereas adoption may require a higher effort than in a later phase of diffusion when practical use cases and expertise are available. At that point a KOS may also be more elaborated and cover better the intended domain of knowledge. Particularly interesting are Hepp's empirically confirmed assumptions concerning the relation between the expressiveness of a vocabulary (ontology) and the size of the community that will adopt it.

Basically, the more expressive the ontology, the smaller the user community will be, because of the effort necessary to comprehend and apply it (arguably the CIDOC CRM is such a case as discussed in *Section 6.3.3*). In practice this comes down to the fact that *"useful ontologies must be small enough to have reasonable familiarization and commitment costs and big enough to provide substantial added value for using them"* (Hepp 2007: 94), where big enough means both sufficient coverage of the intended domain and the existing user base. Arguably this is why small vocabularies such as FOAF and Dublin Core (dcterms) are most widely used in sets of Linked Data (Schmachtenberg 2014a; see also Coyle 2013 on the use of Dublin Core in LOD).

Excellent work on the costs of creating KOSs has been done by the ONTOCOM project[117]. But their highly elaborated model of cost factors and drivers does not include the cost of actually employing a KOS for purposes such as data transformation and linking (cf. Simperl *et al*. 2012).

6.2.4 Brief summary and recommendations

Brief summary

There is a widespread notion of an unfavourable ratio of costs compared to benefits of employing Semantic Web / Linked Data standards for information management, publication and integration. This notion should be removed as it is a strong barrier to a wider adoption of the Linked Data approach.

The basic assumption of Linked Data is that the usefulness and value of data increases the more readily it can combined with relevant other data. Convincing tangible benefits of Linked Data materialise if information providers can draw on own and external data for enriching services. There are examples for such benefits, e.g. in the museum context, but not yet for archaeological research data. Importantly, in the realm of research benefits of Linked Data are less about enhanced search services but research dividends, e.g. discovery of interesting relations or contradictions between data.

Linked Data projects typically mention some benefits (e.g. integration of heterogeneous collections, enriched information services), but very little is known about the costs of different projects. There is a clear need to document a number of reference examples, for example, what does it cost to connect datasets via shared vocabularies or integrate databases through mapping them to CIDOC CRM, and how does that compare to perceived benefits? Although vocabularies play a key role in Linked Data astonishing little is also known about the costs of employing various KOSs.

Some methods and tools appear to have reduced the cost of Linked Data generation considerably, OpenRefine or methods to output data in RDF from relational databases, for instance. As there is a proliferation of tools potential Linked Data providers need expert advice on what to use (and how to use it) for their purposes and specific datasets, taking account also of existing legacy systems and standards in use.

Recommendations

- *Proponents of the Linked Data approach should address the widespread notion of an unfavourable ratio of costs compared to benefits of employing Semantic Web / Linked Data standards.*

[117] Ontology Cost Estimation with ONTOCOM, http://ontocom.sti-innsbruck.at

- *Major benefits of Linked Data can be gained from integration of heterogeneous collections/ databases and enhanced services through combining own and external data. But examples that clearly demonstrate such benefits for archaeological data are needed.*
- *In order to evaluate the costs, information about the cost factors and drivers should be collected and analysed. A good understanding of the costs of different Linked Data projects will help reduce the costs, for example by providing dedicated tools, guidance and support for certain tasks.*
- *More information would be welcome on how specific methods and tools have allowed institutions reducing the costs of Linked Data in projects of different types and sizes.*
- *General requirements for progress are more domain-specific guidance and reference examples of good practice.*

6.3 Enable non-IT experts use Linked Data tools

There are already several showcase examples of Linked Data application in the field of cultural heritage (e.g. museum collections) which, however, depended heavily on the support of experts who are familiar with the Linked Data methods and required tools. A much wider uptake of Linked Data will require approaches that allow non-IT experts do most of the work with easy to use tools and little training effort. A number of projects have reported advances in this direction based on data mapping recipes, supportive tools and guidance material. Further progress may be achieved by integrating Linked Data vocabularies in tools for data recording in the field and laboratory.

6.3.1 Linked Data tools: there are many and most are not useable

Linked Data tools is a field of software development that is largely dominated by academic research groups and individual developers (e.g. in the context of a PhD thesis). While produced under the open source banner, their work rarely leads to mature, maintained and serviced tools or services. There is a lot of obviously immature and abandoned software of such developers on open source software platforms (e.g. GitHub, SourceForge and others) or project websites. Often the aim seems not to be a working solution but a number of publications around the tool or service 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 higher academic recognition of publications impacts negatively on the curation and long-term availability of software that is produced in this context (Todorov 2012).

Some academic open source projects are successful because they find a community of dedicated developers or are developed further by a commercial spin-off, but relevant others would need institutional support and curation to ensure sustainability (Katz *et al*. 2014; Wilson 2014). In some respects the development of semantic tools presents a quasi-Darwinian pattern of survival of the fittest. The field of semantic Wikis may serve as a representative case: A section of Semanticweb.org lists 37 semantic Wiki projects[118] of which 30 (80%) appear to be defunct or are inactive since long. Such lists are very helpful because seldom software project websites indicate that work on a tool has been discontinued or maybe superseded by another project, on a new website and renamed tool. In most cases of still available software it remains unclear if the tool has been completed and is usable, or is an unstable prototype with limited functionality, bugs, etc.

[118] Semanticweb.org: Semantic Wiki projects, http://semanticweb.org/wiki/Semantic_wiki_projects

The LOD Around the Clock (LATC) project warns that a lot of open source Linked Data software tools are not completed, well-tested and stable. The developers often lose interest in a project *"leaving users stranded without improvements or support"* (LATC 2012: 10-11, includes a list of questions to consider in the evaluation of relevant tools). LATC, LOD2[119] and other projects present selected tools for different phases of the Linked Data life cycle, but the selection is often informed by what project participants have on stock. Moreover tools suggested by projects completed two or three years ago may already be superceded by new ones with features that are improved in some respects.

In short, new entries in the realm of Linked Data should look which tools are being used by similar other projects and consult with experts in the field which ones will fit best for their data and goals.

6.3.2 Need of expert support

Arguably all Linked Data showcases in the field of cultural heritage so far depended heavily on the support of experts who are familiar with the required methods and tools, often their own. Many projects have been by experts together with museums, starting with the path-breaking Finnish Museums on the Semantic Web project (Hyvönen *et al*. 2002) up to more recent projects at the Amsterdam Museum (de Boer *et al*. 2012 and 2013), Gothenburg City Museum (Damova & Dannells 2011), Peter the Great Museum of Anthropology and Ethnography in St Petersburg (Ivanov 2011), Russian Museum in St. Petersburg (Mouromtsev *et al*. 2015), Smithsonian American Art Museum (Szekely *et al*. 2013), natural history museums in the Natural Europe project (Skevakis *et al*. 2013), and others.[120] One reason for the strong presence of museums is that they wish to make their collections more accessible to the public, and may more easily do this by drawing on popular resources such as Wikipedia via DBpedia Linked Data.

A much wider generation and use of cultural heritage and archaeology Linked Data, especially also for research purposes, requires appraochs that allow non-experts to do the work with easy to use tools and little training effort. But this may remain an illusory goal. As Eric Morgan, the lead researcher of the Linked Archival Metadata (LiAM) notes: *"Linked data might be a 'good thing', but people are going to need to learn how to work more directly with it"* (Morgan 2014). He suggests practical tutorials, hands-on training on how Linked Data can be put into practice, and hackathons involving practitioners and Linked Data specialists.

In short, turning substantial legacy collections or research datasets into Linked Data resources will hardly be possible without support of specialists, at least for some steps in the process. As a summary of a discussion on skills required for Linked Data puts it, *"Realistically, for many people, expertise needs to be brought in. Most organisations do not have resources to call upon. Often this is going to be cheaper than up-skilling – a steep learning curve can take weeks or months to negotiate whereas someone expert in this domain could do the work in just a few days"* (Stevenson 2011).

6.3.3 The case of CIDOC CRM: from difficult to doable

A special case of a difficult adoption process is the CIDOC Conceptual Reference Model, which is a core for cultural heritage information exchange and integration. The CIDOC CRM is an ontology represented in RDF Schema (RDFS) and considered as a key integrator of heterogeneous datasets in

[119] LOD2 - Creating Knowledge out of Interlinked Data (EU, FP7-ICT, 2010-2014), http://lod2.eu

[120] Some other examples are listed on the Museums and the Machine-processable Web wiki, e.g. Auckland Museum (New Zealand); British Museum (UK), Harvard Art Museums (USA); National Maritime Museum (UK) and others, http://museum-api.pbworks.com/w/page/21933420/Museum%C2%A0APIs

the emerging web of cultural heritage Linked Data. The ontology became an official ISO standard in 2006 (ISO 21127:2006, updated in 2014), which is but one factor that contributed to its wider adoption in the cultural heritage sector, including archaeology.

The increasing use of the CIDOC CRM in recent cultural heritage Linked Data projects is noteworthy. In its early days the CIDOC CRM was perceived as difficult to apply by researchers and practitioners who were not involved in its development and related demonstration projects. For example, in the SCULPTEUR project (2002-2005) museum databases were mapped to the CRM to implement concepts-based cross-collections search & retrieval. The implementers reported that *"mapping is complex and time consuming. The CRM has a steep learning curve, and performing the mapping requires a good understanding of both ontological modelling as well as the source metadata system. Eventually the assistance of a CRM expert was required to complete and validate the mappings"* (Sinclair *et al*. 2005).

Indeed, the CIDOC CRM is a complex ontology that requires a good understanding of its event-centric modelling approach as well as how to apply, extend or specialise the ontology for a particular use case, if required. Researchers of the BRICKS project (2004-2007) noted the abstractness of the CRM concepts and lack of technical specification as factors that could impede the goal of enabling interoperability across heterogeneous databases (Nußbaumer & Haslhofer 2007; see also Nußbaumer *et al*. 2010).

Similar statements can be found elsewhere, for example, one respondent to Leif Isaksen's survey on cultural heritage and archaeology Semantic Web projects wrote: *"CIDOC CRM is bloody hard to understand and use with zero tool support available at the time. Museum bods are understandably not knowledge engineers, so require lots of support"* (in Isaksen 2011: 203). On the other hand, Dominic Oldman (2012) notes that some of the issues pertain to *"a lack of domain knowledge by those creating cultural heritage web applications. The CRM exposes a real issue in the production and publication of cultural heritage information about the extent to which domain experts are involved in digital publication and, as a result, its quality (…) The CRM requires real cross disciplinary collaboration to implement properly – and this type of collaboration is difficult."*

Meanwhile a number of exemplary CIDOC CRM use cases, available documentation and sharing of know-how among practitioners have enabled more projects large and small applying the ontology. However newcomers will still often need expert guidance, as has been given to ARIADNE partners by FORTH-ICS' Centre for Cultural Informatics on modeling scientific archaeological data[121].

6.3.4 Progress through data mapping tools and templates

Projects on databases of heritage collections reported considerable difficulties in getting to Linked Data and archaeological research datasets arguably pose even greater challenges. For example, the datasets that were mapped in the Roman Ports in the Western Mediterranean Project are described as follows: *"While the datasets all pertain to the same domain, they frequently employ mixed taxonomies and are heterogeneously structured. Normalization is rare, uncertainty frequent and variant spellings common. Different recording methodologies have also given rise to alternative quantification and dating strategies. In other words, it is a typical real-world mixed-context situation"* (Isaksen *et al*. 2009).

[121] Cf. ARIADNE (2014b), website: Modeling scientific data: workshop report, 12 September 2014, http://www.ariadne-infrastructure.eu/News/Modeling-scientific-data

But a number of projects have reported advances toward the goal of enabling non-experts apply semantic standards and tools. The data mapping tools that were developed and employed in the Roman Ports project *“have proven remarkably successful against a broad range of sample datasets from four different countries (UK, Spain, France, Italy). The most important achievement has been to enable domain experts to provide data derived in different contexts as ontology-compliant Linked Data extremely quickly and sustainably. Previous attempts to produce homogeneous RDF have generally required a lengthy and expensive mapping process against one or two large resources. We feel that making it possible for ‘the long tail’ of archaeological data is a vital task in the Linked Data revolution”* (Isaksen *et al*. 2009).

Similarly, the Linked Data toolkit developed in the STELLAR[122] project has been reported to allow non-expert users mapping and extracting archaeological datasets to XML/RDF conforming to CIDOC CRM, CRM-EH (English Heritage) or CLAROS CRM Objects concepts and relations. The toolkit comprises of an open source software tool (Stellar Console) and a set of customizable templates. The approach taken was to identify a set of commonly occurring patterns in domain datasets and the CIDOC CRM, and express them in a set of mapping templates.

Tudhope *et al*. (2013) note that with the CIDOC CRM the same semantics underlying cultural heritage datasets can be mapped in different ways, which raises barriers for semantic interoperability the CRM aims to enable. CRM adopters needed mapping guidelines and templates for general use cases in their domain (e.g. archaeology). Therefore the STELLAR project made available a facility for user-defined templates as well as helpful tutorials with worked examples[123] (Binding *et al*. 2015 present in detail the template use for archaeological datasets and a case study with non expert users).

The STELLAR templates have been adapted and used by other projects. For example, the ArcheoInf project[124] aimed to develop a database that combines and integrates, through mappings to CIDOC CRM, data of archaeological surveys and excavations conducted by German university institutes of classical archaeology. Adapted STELLAR templates allowed exporting datasets tagged with CIDOC CRM mappings in XML/RDF (Carver 2013; Carver & Lang 2013). Other projects that employed the STELLAR toolkit for Linked Data generation were Colonisation of Britain (digitisation and semantic enhancement of a major research archive)[125] and the SKOSification of the thesaurus used with ZENON, the online public access catalog of the German Archaeological Institute (Romanello 2012).

6.3.5 Need to integrate shared vocabularies into data recording tools

We will also need to see more progress with regard to integrating Linked Data vocabularies in data recording tools. It is widely held that archaeologists exhibit an aversion to use unfamiliar semantics and prefer to develop their own vocabulary. The argument typically is that this is necessary because of their specific research questions. Frederick W. Limp even thinks 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”* (Limp 2011: 278).

[122] STELLAR - Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources project (UK, AHRC-funded project, 2010-2011), http://hypermedia.research.southwales.ac.uk/kos/stellar/

[123] Hypermedia Research Unit, University of South Wales: STELLAR Applications, http://hypermedia.research.southwales.ac.uk/resources/STELLAR-applications/

[124] ArcheoInf project, http://www.ub.tu-dortmund.de/archeoinf/

[125] Archaeogeomancy.net (2014): Colonisation of Britain, 30 May 2014, http://www.archaeogeomancy.net/2014/05/colonisation-of-britain/

The reticence to use vocabularies that are based on semantic standards is augmented by a perception that this can be difficult, time consuming and have no immediate practical benefit. The team of Open Context in the development their archaeological data publication platform collected views and practical experiences of many archaeologists, cultural resource management professionals, museum curators and others. The results across all participants suggested *"little motivation or interest in having researchers 'markup' their own data to align these data with more general Web or semantic standards"*. Rather project participants *"generally saw this as a somewhat abstract goal, disconnected from their immediate needs, and usually felt such semantic and standards alignment stood too far outside of their area of expertise"* (Kansa & Whitcher-Kansa 2011: 5-6).

The Federated Archaeological Information Management Systems project (FAIMS, Australia) in workshops with potential users found that archaeologists would appreciate tools that allow high flexibility and customization to accommodate their established research practices. Little enthusiasm was perceived for adopting common data standards and terminology, e.g. to record an agreed set of attributes about excavation contexts or artefacts (Ross *et al*. 2013: 111-114).

The results made the FAIMS team rethink their approach to semantic interoperability, which was initially planned to build around a stable (if extensible) core of data standards, data schemata and user interfaces. To accommodate both flexibility and interoperability, FAIMS mobile data recording software now provides sophisticated tools to map data to shared vocabularies as it is created. As they describe the tools, *"Using an approach borrowed from IT localization, interface text, including the names of entities (e.g., 'stratigraphic unit'), attributes (e.g., 'soil color'), and controlled-vocabulary values ('Munsell 5YR'), can be saved and exported using widely-shared terminology (including uniquely identified terms in an ontology) but displayed using the preferred language of an individual project (e.g., 'stratigraphic unit' can display as 'context'). Second, open-linked data URIs can be embedded in all entities, attributes, and controlled-vocabulary values (linking, e.g., species to the Encyclopedia of Life, or places to Pleiades). Finally, data can be systematically transformed or amplified during export, a final opportunity for mapping to shared ontologies or linking to URIs. These approaches balance the flexibility required by archaeologists with the ability to produce interoperable data"* (Ross 2015).

Similar tools are necessary for describing data recorded in laboratory work. One such tool is RightField[126]. The open source tool (implemented in Java) has been developed at the School of Computer Science, University of Manchester (UK) together with other bioinformatics research groups (Wolstencroft *et al*. 2011; Wolstencroft 2012). RightField allows scientists easy semantic annotation of spreadsheet data with common vocabulary of their area of research using simple drop-down lists. For each annotation field, a range of allowed terms from a chosen vocabulary can be specified. Vocabularies can either be imported from a local system or a registry/repository of vocabularies in SKOS, RDFS or OWL (e.g. the BioPortal for biological vocabularies). The generated semantic information (and its provenance) is all held within the spreadsheet. Data sharing initiatives can use RightField to generate and distribute a spreadsheet template to laboratory scientists and collect and integrate the data and semantic annotations.

[126] RightField, http://www.rightfield.org.uk

6.3.6 Brief summary and recommendations

Brief summary

Showcase examples of Linked Data applications in the field of cultural heritage (e.g. museum collections) so far depended heavily on the support of experts who are familiar with the Linked Data methods and required tools (often their own tools). But such know-how and support is not necessarily available for the many cultural heritage and archaeology institutions and projects across Europe. A much wider uptake of Linked Data will require approaches that allow non-IT experts (e.g. subject experts, curators of collections, project data managers) do most of the work with easy to use tools and little training effort.

A number of projects have reported advances in this direction based on the provision of useful data mapping recipes and templates, proven tools, and guidance material. For example, the STELLAR Linked Data toolkit has been employed in several projects and appears to be useable also by non-experts with little training and additional advice.

Good tutorials and documentation of projects are helpful, but the need for expert guidance in various matters of Linked Open Data is unlikely to go away. For example, there are a lot of immature, not tried and tested software tools around. Therefore advice of experts is necessary on which tools are really proven and effective for certain tasks, and providers of such tools should offer practical tutorials and hands-on training, if required. Experienced practitioners can also help projects navigate past dead ends and steer project teams toward best practices.

Also more needs to be done with regard to integrating Linked Data vocabularies in tools for data recording in the field and laboratory. Like other researchers archaeologists typically show little enthusiasm to adopt unfamiliar standards and terminology, which is perceived as difficult, time-consuming, and may not offer immediate practical benefits.

Proposed tools therefore need to fit into normal practices and hide the semantic apparatus in the background, while supporting interoperability when the data is being published. Noteworthy examples are the FAIMS mobile data recording tools and the RightField tool for semantic annotation of laboratory spreadsheet data.

Recommendations

- *Focus on approaches that allow non-IT experts do most of the work of Linked Data generation, publication and interlinking with little training effort and expert support.*
- *Provide useful data mapping recipes and templates, proven tools and guidance material to enable reducing some of the training effort and expert support which is still necessary in Linked Data projects.*
- *Steer projects towards Linked Data best practices and provide advice on which methods and tools are really proven and effective for certain data and tasks.*
- *Current practices are very much focused on the generation of Linked Data of content collections. More could be done with regard to integrating Linked Data vocabularies in tools for data recording in the field and laboratory.*

6.4 Promote Knowledge Organization Systems as Linked Open Data

Knowledge Organization Systems (KOSs) such as ontologies, classification systems, thesauri and others are among the most valuable resources of any domain of knowledge. Because of the large variety of cultural artefacts and contexts the cultural heritage sector is particularly rich in KOSs. In the web of Linked Data KOSs are infrastructural components which provide the conceptual and terminological basis for consistent interlinking of data within and across fields of knowledge. They can serve as bridges which enable interoperability between dispersed and heterogeneous data resources. Therefore KOSs should be openly available and of course in appropriate Linked Data formats.

Most Linked Open Data KOSs are being developed from existing systems. The development requires collaboration of domain and technical experts, or domain experts with the required mix of knowledge and skills. As John Unsworth once put it for KOSs in general, *"In some form, the semantic web is our future, and it will require formal representations of the human record. Those representations – ontologies, schemas, knowledge representations, call them what you will – should be produced by people trained in the humanities. Producing them is a discipline that requires training in the humanities, but also in elements of mathematics, logic, engineering, and computer science. Up to now, most of the people who have this mix of skills have been self-made, but as we become serious about making the known world computable, we will need to train such people deliberately. There is a great deal of work for such people to do – not all of it technical, by any means. Much of this map-making will be social work, consensus-building, compromise. But even that will need to be done by people who know how consensus can be enabled and embodied in a computational medium. Consensus-based ontologies (in history, music, archaeology, architecture, literature, etc.) will be necessary, in a computational medium, if we hope to be able to travel across the borders of particular collections, institutions, languages, nations, in order to exchange ideas"* (Unsworth 2002).

6.4.1 Knowledge Organization Systems (KOSs)

Knowledge organization systems (KOSs) can take different forms, e.g. glossary, thesaurus, classification scheme, ontology (Souza *et al*. 2012; Bratková & Kučerová 2014). A KOS may be used by institutions in many countries, mainly in one country or as a "home-grown" vocabulary only by one institution. Most KOSs are being used as controlled vocabularies to select preferred terms, names or other "values" for certain fields of metadata records. For example, a subjects thesaurus provides terms for the subjects of documents or a gazetteer provides names and geo-coordinates for places. An ontology provides a conceptual model of a domain of knowledge (e.g. the CIDOC Conceptual Reference Model).

Some years ago many KOSs were still made available as copyrighted manuals in PDF format or as simple online lookup pages. Recently open licensing of KOSs has become the norm and ever more existing KOSs are being prepared and published as Linked Open Data for others to re-use.

The RDF family of specifications provides "languages" for KOSs such as Simple Knowledge Organization System (SKOS), RDF Schema (RDFS) and Web Ontology Language (OWL). The relatively lightweight language SKOS[127] can be used to transform a thesaurus, taxonomy or classification system to Linked Data; it can of course also be used to build a new KOS, if necessary. Released as a W3C recommendation in 2009, the language has been adopted by many KOS owners/developers to

[127] W3C (2009) Recommendation: SKOS Simple Knowledge Organization System, 18 August 2009, https://www.w3.org/2004/02/skos/

transform ("SKOSify") controlled vocabularies for use in the web of Linked Data. KOSs that are complex conceptual reference models (or ontologies) of a domain of knowledge are typically expressed in RDF Schema (RDFS)[128] or the Web Ontology Language (OWL)[129].

KOSs in the mentioned languages are machine-readable which allows various advantages. For example a SKOSified thesaurus employed in a search environment can enhance search & browse functionality (e.g. facetted search with query expansion), while Linked Data ontologies can allow automated reasoning over semantically linked data.

6.4.2 Cultural heritage vocabularies in use

Before looking into the development of cultural heritage and archaeological KOSs as Linked Data it will be good to have a view on the current used of KOSs in these fields. For cultural heritage a study of the AthenaPlus project gives an impression, and for archaeology the varity of vocabulary usage by ARIADNE data partners may be indicative for the situation.

AthenaPlus study of vocabularies in use

AthenaPlus (2013a) collected and analysed information on 52 cultural heritage vocabularies that are in use at 33 organisations in Europe. The main results of the study can be summarised as follows:

- Most of the vocabularies are thesauri or classification systems with a more or less complex hierarchical structure. Some are flat lists of terms which may combine terms from different terminologies.
- Most of the organisations use an own vocabulary developed in-house, often with no reference to standards (e.g. ISO thesauri standards)[130]; this group includes national-level organisations.
- Multi-lingual vocabularies are rare, only a few vocabularies have concepts in more than one language.
- The vocabularies are mainly used for indexing and as a query feature of an online database.
- Most vocabularies have unique identifiers for the concepts, and only few management systems do not allow to export them from the local dabase (e.g. in a CSV-file).
- The situation concerning copyrights (licensing) is varied, some vocabularies are free of rights, some organisations apply a Creative Commons license, others have not sought to clarify copyrights yet.

Some of the vocabularies may be used by archives and museums that hold archaeological artifacts among other cultural heritage objects, but few seem to be relevant for archaeological research data sets due to lack of specific terms for this domain.

Vocabulary use by ARIADNE partners

The pattern of vocabulary use by ARIADNE data partners is roughly similar to the results of the AriadnePlus study (cf. ARIADNE 2013):

---

[128] W3C (2014) Recommendation: RDF Schema 1.1, 25 February 2014, http://www.w3.org/TR/rdf-schema/

[129] W3C (2012) Recommendation: OWL 2 Web Ontology Language Document Overview (Second Edition), 11 December 2012, https://www.w3.org/TR/2012/REC-owl2-overview-20121211/

[130] ISO thesauri standards: ISO 2788:1974/1986 (monolingual), ISO 5964:1985 (multilingual), or ISO 25964-1/2:2011 (thesauri and interoperability with other vocabularies).

- Three partners use international and/or multi-lingual vocabularies (more than two languages):
  - European Language Social Science Thesaurus (ELSST)[131],
  - General Multilingual Environmental Thesaurus (GEMET)[132] and part of the Tree of Life taxonomy for wood species[133],
  - PACTOLS thesaurus (multi-lingual)[134].
- Four partners use national standard vocabularies
  - Geological Survey of Ireland (classifications for geology, petrology and soils)[135], Placenames Database of Ireland[136], Irish National Monuments Service monument class list[137], Artefact classification[138],
  - Swedish Monument type vocabulary[139],
  - Archeologisch Basisregister (ABR, Netherlands)[140],
  - PICO thesaurus[141] and SITAR vocabularies (Italy)[142].
- Seven partners use proprietary controlled vocabularies (thesauri, term lists),
- Three partners currently do not use controlled vocabularies.

Some of the vocabularies mentioned are already available in SKOS (e.g. GEMET since many years) or such a version is in preparation (see below).

6.4.3 Development of KOSs as Linked Open Data

The first generation of cultural heritage Semantic Web projects (started about 15 years ago) often used major vocabularies such as the Getty thesauri, Iconclass (Netherlands Institute for Art History) and others for "research purposes", i.e. without allowance to share publicly vocabulary Linked Data

---

[131] ELSST is a broad-based, multilingual thesaurus for the social sciences. It is currently available in 12 languages: Czech, English, Danish, Finnish, French, German, Greek, Lithuanian, Norwegian, Romanian, Spanish and Swedish, http://elsst.ukdataservice.ac.uk

[132] GEMET (EIONET/European Environment Agency), http://www.eionet.europa.eu/gemet/

[133] Tree of Life (TOL) project, http://tolweb.org/tree/

[134] PACTOLS - Peuples, Anthroponymes, Chronologie, Toponymes, Oeuvres, Lieux et Sujets (Fédération et ressources sur l'Antiquité (FRANTIQ, France), http://pactols.frantiq.fr

[135] Geological Survey of Ireland, http://www.gsi.ie

[136] Placenames Database of Ireland, http://www.logainm.ie/en/

[137] Irish National Monuments Service monument class list, http://webgis.archaeology.ie/NationalMonuments/WebServiceQuery/Lookup.aspx

[138] National Museum of Ireland: Artefacts, http://www.museum.ie/en/list/artefacts.aspx

[139] See http://www.fmis.raa.se (lämningstyp) and Swedish National Heritage Board (2014), extended by the Swedish National Data Service (SND) with keywords researchers use when depositing data with SND.

[140] Archeologisch Basisregister (Cultural Heritage Agency of the Netherlands), http://cultureelerfgoed.nl/dossiers/archis-30/archeologisch-basisregister-plus

[141] PICO thesaurus (Central Institute for the Union Catalogue - ICCU, Italy; terms in Italian and English, but not archaeology-specific), http://purl.org/pico/thesaurus_4.2.0.skos.xml

[142] SITAR Project Data Model & DataSet (Soprintendenza Speciale per i Beni Archeologici di Roma), https://www.academia.edu/5029017/MiBACT-SSBAR_SITAR_Project_Data_Model_presentation_at_the_ARIADNE_Workshop_in_Pisa_7-8.11.2013_

they produced from parts of such resources. The move to Open and Linked Data vocabularies was initiated by the library community, for example the US Library of Congress (since 2009)[143], OCLC (worldwide library cooperative)[144] and others. In recent years the owners of major vocabularies for the humanities and cultural heritage followed.

In 2012 Iconclass, the widely used classification system for visual content of cultural works (e.g. iconography), was made available as Linked Open Data[145]. In 2014/2015 the Getty Research Institute released three of their vocabularies as Linked Open Data: Art & Architecture Thesaurus (AAT), Thesaurus of Geographic Names (TGN) and Union List of Artist Names (ULAN); the Cultural Objects Name Authority (CONA) was intended to follow in Fall 2015 but seems to require more effort than expected.[146]

In the UK the SENESCHAL project (2013-2014)[147] transformed several cultural heritage vocabularies of English Heritage, Royal Commission on the Ancient and Historical Monuments of Scotland (RCAHMS) and Royal Commission on the Ancient and Historical Monuments of Wales (RCAHMW) to SKOS and made them available online[148] (Binding & Tudhope 2016). SENESCHAL built on the experience and tools developed in the STAR and STELLAR projects (2007-2011)[149]. The goal of the project was to make it easier for vocabulary providers to publish their vocabularies as Linked Data and for users to index their data with uniquely identified terms of the SKOSified vocabularies. The project developed RESTful web services that facilitate concept searching, browsing, suggestion and validation. Furthermore browser-based widgets (predefined user interface controls) are available that allow for embedding the vocabularies in web pages and web forms to better index data and improve search applications.

Many others have also already transformed their vocabularies to SKOS or developed new ones based on the standard. Some examples relevant for archaeological data are: The PACTOLS thesaurus[150] of the Fédération et ressources sur l'Antiquité (FRANTIQ), France, is a multi-lingual thesaurus that focuses on antiquity and archaeology from prehistory to the industrial age (terms in French, English, German, Italian, Spanish, Dutch, and some Arabic).

In the Netherlands the Rijksdienst Cultureel Erfgoed (Cultural Heritage Agency) have produced SKOS versions of their Archeologisch Basisregister (ABRr+) and other thesauri[151]. Some of them have been used in ARIADNE to explore the extraction of (meta-)data from Dutch fieldwork reports based on

---

143 Library of Congress: Linked Data Service, http://id.loc.gov; Library of Congress Subject Headings (LCSH), MARC Code Lists, Thesaurus of Graphic Materials, AFS Ethnographic Thesaurus and others.

144 OCLC (worldwide library cooperative): Linked Data, http://oclc.org/developer/develop/linked-data.en.html; available: Dewey Decimal Classification (DDC), Virtual International Authorities File (VIAF), Faceted Application of Subject Terminology (FAST) and WorldCat.

145 Iconclass as Linked Open Data, http://www.iconclass.org/help/lod

146 Getty Vocabularies as Linked Open Data, http://www.getty.edu/research/tools/vocabularies/lod/index.html

147 SENESCHAL - Semantic Enrichment Enabling Sustainability of Archaeological Links (UK, AHRC-funded project, 2013-2014), http://hypermedia.research.southwales.ac.uk/kos/seneschal/

148 HeritageData, http://www.heritagedata.org

149 STAR - Semantic Technologies for Archaeological Resources (UK, AHRC-funded project, 2007-2010), http://hypermedia.research.southwales.ac.uk/kos/star/; STELLAR - Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources (UK, AHRC-funded project, 2010-2011), http://hypermedia.research.southwales.ac.uk/kos/stellar/

150 PACTOLS (Peuples, Anthroponymes, Chronologie, Toponymes, Œuvres, Lieux et Sujets), http://pactols.frantiq.fr

151 Rijksdienst Cultureel Erfgoed: Erfgoedthesaurus, http://www.erfgoedthesaurus.nl

named entity recognition (ARIADNE 2015c). In Sweden the Riksantikvarieämbetet (National Heritage Board) aims to translate their vocabularies (e.g. the Swedish monuments types thesaurus) to SKOS and release them as Linked Open Data. This work is under way in their Digital Archaeological Workflow programme, 2013-2018 (Smith 2015: 219).

Examples of Linked Data vocabularies for research specialities are the Nomisma ontology for numismatics[152], the set of vocabularies for epigraphy developed by the EAGLE project[153], and the multi-lingual vocabulary for dendrochronological data based on the Tree Ring Data Standard (TRiDaS) standard[154]. The vocabuarly has been developed by Data Archiving and Networked Services (DANS, Netherlands), with support by ARIADNE. The vocabulary is being employed for the Digital Collaboratory for Cultural Dendrochronology[155] (Jansma 2013) and available also to other users.

As the case of dendrochronology reminds us, Linked Data vocabularies for archaeological data are of course not limited to cultural artefacts. Such vocabularies are also needed for describing biological remains of humans, animals and plants. There are many relevant biological vocabularies available in Linked Data formats shared on the BioPortal[156], and may increasingly be used by archaeological institutions and projects to integrate datasets. One example is a project that employed concepts of the Uber Anatomy Ontology (UBERON)[157] for zooarchaeological data (Kansa *et al*. 2014; Whitcher-Kansa 2015).

An interesting case where a vocabulary of an established system is being transformed to SKOS is TAXREF, the French national taxonomic reference for fauna, flora and fungus (Callou *et al*. 2015). TAXREF is being used for the National Inventory of Natural Heritage (INPN)[158], and the Archaeozoological and Archaeobotanical Inventories of France (I2AF) database[159] (Callou *et al*. 2009 and 2011). TAXREF and the databases are maintained by the French National Museum of Natural History (MNHN), the I2AF in collaboration with a multi-institute network of bioarchaeologists[160].

In addition to publishing TAXREF in SKOS it is intended to set up a Web service allowing to query the taxonomy and retrieve results in different formats such as XML/RDF and JSON. Furthermore there

[152] Nomisma ontology, http://nomisma.org/ontology

[153] EAGLE vocabularies (Material, Type of inscription, Execution technique, Object type, Decoration, Dating criteria, State of preservation), http://www.eagle-network.eu/resources/vocabularies/

[154] Tree Ring Data Standard (TRiDaS), vocabularies: http://www.tridas.org/vocabularies/

[155] Digital Collaboratory for Cultural Dendrochronology - DCCD, http://dendro.dans.knaw.nl, see also: https://vkc.uu.nl/vkc/dendrochronology/

[156] BioPortal (US National Center for Biomedical Ontology), https://bioportal.bioontology.org

[157] UBERON - Uber Anatomy Ontology (http://uberon.org) is a cross-species anatomy ontology that represents body parts, organs and tissues in a variety of animal species, with a focus on vertebrates; it includes relationships to taxon-specific anatomical ontologies, allowing integration of functional, phenotype and expression data; see Mungall *et al*. (2012).

[158] Inventaire National du Patrimoine Naturel / National Inventory of Natural Heritage (Muséum national d'Histoire naturelle), http://inpn.mnhn.fr

[159] Inventaires archéozoologiques et archéobotaniques de France (I2AF), https://inpn.mnhn.fr/espece/inventaire/I100

[160] GDR 3644 BioArchéoDat, Sociétés, biodiversité et environnement: données et résultats de l'archéozoologie et de l'archéobotanique sur le territoire de la France, http://archeozoo-archeobota.mnhn.fr/spip.php?article236&lang=fr

are plans to create mappings to other KOSs such as the NCBI Organismal Classification[161], the GeoSpecies ontology[162], the ENVO environment ontology[163], GeoNames and others.

The I2AF database is being populated with data on flora and fauna from archaeological investigations carried out in French territories. When data from archaeological reports is imported into I2AF, it is aligned to TAXREF and a thesaurus of cultural periods (the oldest records date back to the Middle Palaeolithic). In 2015 I2AF contained 180,000 data items concerning 2700 animal and 1100 plant species. The data was based on more than 3200 references, 85% “grey literature” such as excavations reports, specialist studies and other material, referring to 4700 archaeological sites and 46,600 contexts (pits, well, stratigraphic units etc.).

6.4.4 KOSs registries

With the growth of the World Wide Web since the 1990s ever more KOSs have been published on the Web. Initially they were provided as text documents or simple HTTP pages for looking up vocabulary terms. More recently vocabularies were implemented as databases in XML, and with RDF they can not only be published on the Web but become part of the web of Linked Data. Indeed, major vocabularies are important hubs in this web, for example, the AGROVOC thesaurus for the agriculture and food sector (which is aligned with 16 other vocabularies)[164]. The W3C Library Linked Data Incubator Group envisage that major vocabularies can play an important role in the Web of Data as *value vocabularies,* provided that they are expressed with the unique identifiers (URIs) required for their use in Linked Data (Isaac *et al*. 2011).

The proliferation of KOSs (in various formats) has led to the creation of registries that provide information about vocabularies, relevant for one or all sectors, collected by the registry and/or submitted by vocabulary owners/developers (Golub & Tudhope 2009; Golub *et al*. 2014). As an example of a domain registry, Agricultural Information Management Standards (AIMS) maintain a catalogue of vocabularies for the agriculture and food sector (about 120 vocabularies)[165]. The largest multi-domain registry is the BARTOC - Basel Register of Thesauri, Ontologies & Classifications[166] of the Basel University Library (Switzerland). The registry was launched in 2013 and documents over 1800 KOSs (Ledl & Voß 2016); it also briefly describes and links to 70 other, more specialized vocabulary registries. On BARTOC vocabularies can be searched and filtered based on several categories, including type, topic, language, location, access (e.g. free or licensed), and format (e.g. CSV, XML, JSON, RDF, SKOS). For 139 vocabularies a SKOS version seems to be available (7.5% of 1846 entries as of 19/7/2016).

If we look for registries of KOSs in Linked Data formats specifically, there is the Linked Open Vocabularies (LOV) registry which currently documents 560 ontologies (Vandenbussche *et al*. 2015)[167]. LOV does not register thesauri or other terminology resources, but general and domain ontologies in RDFS or OWL, which others may wish to re-use as a whole or only certain classes and properties. An example of a comprehensive domain registry of ontologies is the BioPortal[168], which

[161] NCBI Organismal Classification, https://bioportal.bioontology.org/ontologies/NCBITAXON

[162] GeoSpecies ontology, https://bioportal.bioontology.org/ontologies/GEOSPECIES

[163] Environment Ontology, https://bioportal.bioontology.org/ontologies/ENVO

[164] AGROVOC Linked Open Data, http://aims.fao.org/standards/agrovoc/linked-open-data

[165] Vocabularies, Metadata Sets and Tools (VEST) registry: KOS, http://aims.fao.org/vest-registry/vocabularies

[166] BARTOC, http://www.bartoc.org

[167] LOV - Linked Open Vocabularies (LOV), http://lov.okfn.org

[168] BioPortal, http://bioportal.bioontology.org

documents over 300 biological/bio-medical vocabularies that can be browsed and downloaded; the portal also shows mappings between classes in different ontologies.

For cultural heritage and archaeology Linked Data vocabularies a comprehensive international registry does not exist as yet. At the national level the Forum on Information Standards in Heritage (FISH) provides a list of British vocabularies that can be consulted online and/or downloaded as CSV or PDF; for nine vocabularies available in SKOS format FISH links to the Heritage Data server implemented by the SENESCHAL project[169]. In Finland the Finnish Ontology Library Service (ONKI)[170] includes KOSs of the cultural sector (Hyvönen, Viljanen *et al*. 2008; Suominen *et al*. 2014). In the Netherlands the CATCH vocabulary and alignment repository[171] once aimed to cover vocabularies of the cultural heritage domain (van der Meij *et al*. 2010).

At present it is difficult to identify vocabularies such as thesauri or ontologies for cultural heritage and archaeology that are already available in Linked Data formats (SKOS, RDFS, OWL) or are work in progress. A KOS registry could help finding potentially relevant vocabulary resources for re-use as a whole or for selecting relevant concepts/terms. As Lang *et al*. note, *"Tackling this lack of a common repository for storing archaeological vocabularies with a persistent identifier for each concept will be one of the main issues of the SKOS-community in the future"* (Lang *et al*. 2013). This issue has not been solved as yet. It may also be questioned if it makes sense to implement a registry or repository specifically for cultural heritage and archaeology Linked Data vocabularies. Maybe an available registry of all kinds of Linked Data resources like the DataHub is a sufficient or even better solution?

At this stage, arguably a solution should be preferred that supports community building of developers and users of Linked Data vocabularies. Registration is but one important function (for which the DataHub may do), but as or even more important is fostering a community that values high-quality and actively curated vocabularies. Because many published vocabularies do not conform to the Linked Data principles, e.g. lack dereferencable HTTP URIs for retrieving descriptions of KOS concepts/terms. Schmachtenberg *et al*. (2014b) found that of 375 proprietary vocabularies (defined as being used by only one dataset) only 19% were fully and 8% partially dereferencable, 73% had term URIs not dereferencable at all. Only 21% set links to one or more other vocabularies.

One reason for the weakness of proprietary vocabularies is that the rapid uptake of the Linked Data approach by many data providers has not been accompanied by training and support for proper vocabulary modelling. Corcho *et al*. (2015) note a general preference of light-weight vocabularies (e.g. FOAF) and combinations thereof. Such vocabularies may be designed badly or, even, be *"Frankenstein ontologies"*, i.e. concepts cobbled together inconsistently from different vocabularies. Providing support for proper Linked Data vocabulary creation therefore is seen as *"one of the main challenges that the ontology engineering field will have to address"* (Corcho *et al*. 2015: 16).

In this challenge, a KOS registry could serve as an instrument of quality control, improvement and confirmation. Zimmermann (2010) suggested a quality assessment process for Linked Data vocabularies in which some criteria can be checked automatically (e.g. dereferencable URIs) while others require judgement by domain experts, e.g. clear labels and description of each term, adequacy of the complexity and granularity of the KOS to intended uses.

[169] Forum on Information Standards in Heritage (FISH): http://heritage-standards.org.uk/fish-vocabularies/; see also Heritage Data: Vocabularies provided, http://www.heritagedata.org/blog/vocabularies-provided/

[170] ONKI - Finnish Ontology Library Service (currently 87 KOSs of which 13 are relevant for the domain of culture and cultural heritage), http://onki.fi; see also: http://finto.fi/en/

[171] CATCH Vocabulary and alignment repository demonstrator, http://www.cs.vu.nl/STITCH/repository/

A useful feature of a KOS registry would also be that Linked Data vocabulary projects can be announced so that duplication of work may be prevented and collaborative efforts fostered. A registry may also promote joint activities such as vocabulary alignments, vocabulary-level links which increase the interoperability of datasets based on terms that are common across them.

6.4.5 Brief summary and recommendations

Brief summary

Knowledge Organization Systems (KOSs) such as ontologies, classification systems, thesauri and others are among the most valuable resources of any domain of knowledge. In the web of Linked Data KOSs provide the conceptual and terminological basis for consistent interlinking of data within and across fields of knowledge, enabling interoperability between dispersed and heterogeneous data resources.

The RDF family of specifications provides "languages" for Linked Data KOSs. The relatively lightweight language Simple Knowledge Organization System (SKOS) can be used to transform a thesaurus, taxonomy or classification system to Linked Data. KOSs that are complex conceptual reference models (or ontologies) of a domain of knowledge are typically expressed in RDF Schema (RDFS) or the Web Ontology Language (OWL). Linked Data KOSs are machine-readable which allows various advantages. For example a SKOSified thesaurus employed in a search environment can enhance search & browse functionality (e.g. facetted search with query expansion), while Linked Data ontologies can allow automated reasoning over semantically linked data.

Some years ago many KOSs were still made available as copyrighted manuals or online lookup pages. Recently open licensing of KOSs has become the norm and ever more existing KOSs are being prepared and published as Linked Open Data for others to re-use. Following the path-breaking library community, the initiative for KOSs as LOD is under way also in the field of cultural heritage and archaeology. Some international and national KOSs are already available as LOD, Iconclass, Getty thesauri (e.g. Arts & Architecture Thesaurus), several UK cultural heritage vocabularies, the PACTOLS thesaurus (France, but multi-lingual), and others.

But more still needs to be done for motivating and enabling owners of cultural heritage and archaeology KOSs to produce LOD versions and align them with relevant others, for example mapping proprietary vocabulary to major KOSs of the domain. Also more LOD KOSs for research specialities, such as the Nomisma ontology for numismatics, are necessary.

The sector of cultural heritage and archaeology could also benefit from a dedicated international registry for KOSs already available as LOD or in preparation. An authoritative registry could serve as an instrument of quality assurance and foster a community of KOSs developers who actively curate vocabularies. Such a registry could also allow announcing LOD KOSs projects so that duplication of work may be prevented and collaborative efforts promoted (e.g vocabulary alignments).

Recommendations

- *Foster the availability of existing Knowledge Organization Systems (KOSs) for open and effective usage, i.e. openly licensed instead of copyright protected, machine-readable in addition to manuals and online lookup pages.*
- *Provide practical guidance and suggest effective methods and tools for the generation, publication and linking of KOSs as Linked Open Data (LOD).*
- *Encourage institutional owners/curators of major domain KOSs (e.g. at the national level) to make them available as LOD.*

- *Promote alignment of major domain KOSs and mapping of proprietary vocabulary, e.g. simple term lists or taxonomies as used by many organizations, to such KOSs.*
- *Promote a registry for domain KOSs that supports quality assurance and collaboration between vocabulary developers/curators.*

6.5 Foster reliable Linked Data for interlinking

The principles for Linked Data include that publishers should link their data to other datasets. In practice this principle is often not followed, particularly also not in the field of cultural heritage and archaeology. There are several reasons for this shortcoming, in the first place arguably a lack of relevant, high-quality and reliable other datasets. Without such resources a web of archaeological Linked Open Data will not emerge. For building this web a community of curators is necessary who take care for proper generation, publication and interlinking of LOD datasets and vocabularies.

6.5.1 Current lack of interlinking

The Linked Data principles are meant to enable and drive the linking of information in an open "web of data". The core principle in this regard is that publishers should link their data to other people's data to provide users with more context and allow them to discover related information (Berners-Lee's principle 4). This principle is often not followed: In the 2014 LOD Cloud survey of the 1014 identified datasets 445 (43.89%) did not set any out-gowing RDF links; they were either only the target of RDF links from other datasets or were isolated. 176 datasets (17.36%) linked to one other dataset, 106 (10.45%) to two and 287 (28.30%) to three or more datasets, 79 (7.79%) even to more than 10 (Schmachtenberg *et al*. 2014a).

Also in the area of cultural heritage and archaeology few projects so far obey to Berners-Lee's principle 4, which means that already produced Linked Data is highly fragmented, a web of data has not emerged yet.

Andrea d'Andrea (2012) argues that in this area interlinking with other available resources has not been considered sufficiently. He looked into six projects, three of which had an archaeological or classical studies focus, but found that they did not provide links to additional external Linked Data or attempted to integrate data of different domains. As one obstacle d'Andrea sees the lack of a standardised approach or at least authoritative recommendations on how to implement the fourth Linked Data principle in the cultural heritage sector. For example, the CIDOC-CRM LOD Recommendation for Museums mainly addresses URIs (Crofts, Doerr & Nyman 2011; ICOM 2011; CIDOC 2012).

The lack of interlinking is confirmed by Leif Isaksen (2011) who for his dissertation surveyed 40 projects which employed semantic technologies. The sample comprises of projects in the fields of cultural heritage, archaeology and classical studies. Among the 36 data-focused projects (i.e. not only providing an ontology), the majority used URIs to express data (Linked Data principle 1), while just half also had dereferencable HTTP URIs (principle 2). 16 projects expressed their data as RDF (principle 3), but just five linked to external URIs as well (principle 4). (Isaksen 2011: 64)

In a case study Isaksen also explored approaches for enhancing with Linked Data methods projects which created data interoperability in a centralised and often closed system (Isaksen 2011, chapter 7). He concludes that enhancement will often be impractible because such projects typically have been small-to-medium scale in terms of number of participants and datasets. In such projects the effort required of project partners to convert and work with data in the unfamiliar Semantic Web formats would not compare well with the achievable "analytical return" on investment. A pay-off would only materialize in a decentralized landscape of Open Linked Data where network effects can drive addition and interlinking of more datasets.

6.5.2 Why is there a lack of interlinking?

There are several reasons for the neglect of the fourth Linked Data principle in the field of archaeology. Obviously one major reason is that only few projects so far have produced and exposed archaeological Linked Data. Therefore the issue for archaeology is not a "needle in a haystack" problem. Some Linked Data researchers assume that there is a difficulty to identify in the Linked Data Cloud resources which are worth to link with (e.g. Nikolov & d'Aquin 2011; Nikolov *et al*. 2012), but such a problem does not exist for archaeology and most other scientific domains.

Developers of archaeological Linked Data projects will also not consider popular Linked Data resources like DBpedia / Wikipedia as relevant candidates. But showcase examples of linking to other, scientific resources are missing or not well known. For example, the Open Context data publication platform reports linking zooarchaeological data with Encyclopedia of Life animal taxa and Uber Anatomy Ontology (UBERON) concepts (Kansa *et al*. 2014; Whitcher-Kansa 2015).

Andreas Blumauer (2013) thinks that the low level of external linking in most domains is due to two reasons: 1) there is *not much domain-specific knowledge and data* in the LOD Cloud, except for the biological domain (created by the Bio2RDF initiative, among others) and some high-quality "micro LOD clouds" which have been developed by dedicated domain projects; 2) many datasets of the LOD cloud *are not maintained* in a professional manner and hence not trustworthy for sustainable interlinking. Furthermore Blumauer notes that there is often a lack of clear open data licensing.

Smith-Yoshimura (2014c and 2016) notes a number of barriers or challenges institutional implementers of Linked Data services mentioned in the OCLC Research surveys 2014 and 2015. Among the most cited issues when trying to consume or link to other Linked Data sets were:

- What is published as Linked Data is not always reusable or lacks URIs,
- Understanding how others data is structured,
- Easy aligning not possible (e.g. important authority terms are missing),
- Vocabulary mapping proves to be difficult (e.g requires a lot of manual work, issues with level of specificity of terms),
- Lack of useful "off the shelf" tools (e.g. with regard to visualisation),
- Datasets not being updated,
- Size of RDF dumps and volatility of data format of dumps,
- Service reliability, e.g. unstable SPARQL endpoints.

Other barriers included: lack of Linked Data sets of local interest, licenses more restrictive than CC-By or ODC-BY, insufficient internal resources to incorporate available Linked Data into routine workflows.

6.5.3 Need of reliable Linked Data resources

The web of Linked Data will emerge from the publication and interlinking of ever more resources of different providers. This means a shift from a model of single, authoritative and mostly static metadata records to a distributed approach in which statements about items of interest (e.g. research objects) can come from different resources. Therefore the quality and continued availability of the resources is paramount for the overall working of the web of Linked Data.

The benefits of Linked Data will not materialize if computer applications cannot reliably use it for specific purposes. But many studies have shown that basic Linked Data principles and additional best practices suggested by leading developers are often not followed (e.g. Duan *et al*. 2011; Hogan *et al*. 2010; Hogan *et al*. 2012; Schmachtenberg *et al*. 2014a/b).

Interlinking with Linked Data of other providers requires that one can trust that their data and services are reliable with regard to criteria of quality. However the Linked Open Data Cloud is a mix of resources, some of which may not fulfil requirements with regard to content (e.g. incomplete), others are not reliable with regard to maintenance. Buil-Aranda *et al*. (2013) found that of 427 public SPARQL endpoints registered in the DataHub half were off-line and only one third were almost always available during a monitoring of 27 months.

Recent figures available from LODStats[172] show that most Linked Data resources simply are not reliable. LODStats processes RDF datasets from the DataHub, data.gov and publicdata.eu data catalogs to produce statistical overviews of the state the data web (Auer *et al*. 2012b; Ermilov *et al*. 2016). In May 2016 LODStats identified 9960 datasets of which 7112 (71.5%) presented problems; 6712 of in total 9416 RDF dumps having errors (71.28%) and 400 of in total 544 SPARQL endpoints with errors (73.53%).

The issue of reliability of resources for linking is emphasised by many data providers, including from the cultural heritage sector where authoritative information and well maintained services are essential. For example authors of the library domain stress: *"The main problem for the linked data web is dealing with reliability: Is the data correct and do processes exist that guarantee a high data quality? Who is responsible for it? Of the same importance is reliability in time: Is a resource stable enough to be citable, or will it be gone at some point? These questions are of special importance in the context of research, where citability is essential, and for higher-level services that are based on this kind of data"* (Hannemann & Kett 2010).

With the increasing number of Linked Data resources their quality has become a core topic of semantic web conference sessions and dedicated workshops. Ever more detailed schemes and metrics for Linked Data quality are being elaborated and used to scrutinize resources and suggest improvements, if required (e.g. Assaf & Senart 2012; Auer *et al*. 2013 [chapter 7]; Behkamal 2014; Fürber & Hepp 2010a/b and 2011a[173]; PlanetData 2012; Zaveri *et al*. 2013). As a novelty, Hoxha *et al*. (2011) base their framework on principles of "green engineering", e.g. that it is better to prevent waste than to treat or clean up after it is formed. The approach works particularly well with regard to re-use of resources and alignment with actual user demand.

The Linked Data quality schemes tend to centre on adherence to good practices with regard to data and technical standards. But also general criteria are being addressed, for example, that LD resources should be easy to find and assess with regard to relevance and trustworthiness, e.g. well-documented in a general or domain registry, including data description, transparent data policy, data provenance information, and others.

[172] LODStats (Agile Knowledge Engineering and Semantic Web Group at University of Leipzig, Germany), http://stats.lod2.eu

[173] See also the related website http://semwebquality.org and the Data Quality Management Vocabulary (Fürber & Hepp 2011b) and Data Quality Constraints Library (Fürber *et al*. 2011)

While different approaches are being used, the quality criteria essentially are about how users (humans and machines) can discover, understand and access Linked Data resources that are well-structured, accurate, up-to-date and reliable over time. Ideally the result of the current efforts will be easy to use tools that allow Linked Data curators monitor resources, detect and fix problems so that high-quality webs of data are being developed and maintained.

6.5.4 Foster a community of archaeological LOD curators

The lack of trustworthy resources in many quarters of the "web of data" makes clear a core requirement for high-quality Linked Open Data: a community of curators who ensure reliable availability and interlinking of LOD datasets and vocabularies.

One domain of good Linked Data curation practices which could be followed are the Life Sciences. Ten years ago the Life Sciences Semantic Web was described as full of *"semantic creep – timid, piecemeal and ad hoc adoption of parts of standards by groups that should be stridently taking a leadership role for the community"* (Good & Wilkinson 2006). Meanwhile the domain has advanced substantially towards a more integrated area of the web of LOD. One outstanding example is the Bio2RDF[174] community which created and/or interlinked 35 datasets. The Bio2RDF datasets are one of the densest clusters present on the LOD diagram[175].

The importance of LOD curation becomes clear when considering that also a lot of life and bio-sciences related Linked Data produced as yet remains isolated and difficult to integrate. Hasnain *et al*. (2015) catalogued 137 public SPARQL endpoints of relevant Linked Data providers and tried to link concepts and properties of the resources. They found that most resources could not be easily mapped because there was very little vocabulary and URI re-use, i.e. vocabularies which might bridge between the resources were not present. Also shortcomings of URIs are noted as a lot could not be deferenced and many datasets included orphan URIs (i.e. "type"-less URI instances).

If the domain of archaeological research aspires to grow a rich and robust web of LOD within the overall LOD Cloud, it will have to foster and support a community of curators who take care for proper generation, publication and interlinking of LOD datasets and vocabularies. This community could benefit from good practices demonstrated by the Ancient World LOD community mobilised and integrated by Pelagios and research object centred initiatives such as Nomisma (see *Section 5.3*).

6.5.5 Brief summary and recommendations

Brief summary

The core Linked Data principle arguably is that publishers should link their data to other datasets, because without such linking there is no "web of data". In practice this principle is often not followed, particularly also not in the field of cultural heritage and archaeology. This means that already produced Linked Data remains isolated, a web of data has not emerged yet. There are several reasons for this shortcoming. Obviously one factor is that only few projects so far have produced and exposed archaeological Linked Data. Developers of such data will also not consider popular Linked Data resources like DBpedia/Wikipedia as relevant candidates. Moreover there is the issue of reliability, that data one links to will remain accessible, which often they are not. Surveys found that many datasets present problems, for example SPARQL endpoints are often off-line or present errors.

[174] Bio2RDF: Linked Data for the Life Sciences, http://bio2rdf.org

[175] Cf. the Linking Open Data cloud diagram, http://lod-cloud.net

With the increasing number of Linked Data resources their quality has become a core topic of the developer community. Detailed quality schemes and metrics are being elaborated and used to scrutinize resources and suggest improvements. The quality criteria essentially are about how users (humans and machines) can discover, understand and access Linked Data resources that are well-structured, accurate, up-to-date and reliable over time. Furthermore the resources should be well-documented, e.g. with regard to data provenance and policy/licensing. Ideally the result of the quality initiative will be easy to use tools that allow Linked Data curators monitor resources, detect and fix problems so that high-quality webs of data are being developed and maintained.

The lack of trustworthy resources in many quarters of the "web of data" makes clear that a community of curators is necessary who take care for reliable availability and interlinking of high-quality archaeological LOD datasets and vocabularies. A few domains already have such a community, the Libraries and Life Sciences domains, for instance. Also the Ancient World LOD community around the Pelagios initiative or the Nomisma community can be mentioned as examples of good practice. It appears that the domain of archaeology needs a LOD task force and a number of projects which demonstrate and make clear what is required for reliable interlinking of LOD.

Recommendations

- *Foster a community of LOD curators who take care for proper generation, publication and interlinking of archaeological datasets and vocabularies.*
- *Form a task force with the goal to ensure reliable availability and interlinking of LOD resources; LOD quality assurance and monitoring should be established.*
- *Sponsor a number of projects which demonstrate the interlinking and exploitation of some exemplary archaeological datasets as Linked Open Data.*

6.6 Promote Linked Open Data for research

Archaeological data and knowledge present a great challenge for Linked Data. This challenge stems from the multi-disciplinarity of the research on archaeological sites and objects (Vavliakis *et al*. 2012). A web of Linked Data based on cross-domain and domain-specific ontologies and terminologies can allow addressing better archaeological research questions, which require integration of knowledge and data of different domains.

Today benefits of Linked Open Data are mainly framed, and sometimes demonstrated, in terms of advanced search services based on the semantic linking between related datasets. This may appeal to cultural heritage institutions as it allows making their collections better discoverable and more relevant by adding external contextual information.

While such search services are also important to researchers, a focus on data search arguably does not strongly promote the generation of Linked Open Data of research datasets. Research groups and institutions will be much more attracted by demonstrated research dividends of semantically interlinked and integrated data. Such dividends could for example result from combining data from several projects in ways that enable interesting new lines of research, or views on data from different disciplinary perspectives suggesting interdisciplinary approaches. Researchers also need effective tools, usable by non-IT experts, to benefit from Linked Data in the research process, e.g. explore and exloit semantic relations between datasets or between publications and related data.

Established ways of data integration for research follow other paradigms than Linked Data. For example data shared by researchers in a database with research tools implemented on top, e.g. the

Paleobiology Database for which Fossilworks provides data query and analysis tools[176]. Or a stand-alone database with sophisticated modelling and interactive web interfaces such as ORBIS - The Stanford Geospatial Network Model of the Roman World[177]. ORBIS allows calculating the effort (time, financial expense) associated with different types of travel in antiquity (Meeks & Grossner 2012; Scheidel 2015). Applications of Linked Open Data for research will have to demonstrate advantages over or other benefits than already established forms of data integration and exploitation.

6.6.1 A Linked Open Data vision (2010)

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 increasing adoption of the Linked Data approach for sharing library, government and scientific data, and 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 highlighted discipline-specific vocabularies (e.g. thesauri, ontologies), which need to be integrated so that a searchable web of scientific data can emerge. Furthermore he noted that integration of Linked Data tools in scientific work environments was missing. So far Bizer's vision is not realised, but has four further years to materialize until 2020.

6.6.2 LOD for research: The current state of play

Efforts for cultural heritage LOD so far have been invested mainly on publishing various museum collections, often linked to DBpedia/Wikipedia. Concerning special collections an outstanding example is the numismatics databases that participate in the Nomisma initiative[178]. Also a few

[176] Fossilworks, http://fossilworks.org

[177] ORBIS - The Stanford Geospatial Network Model of the Roman World, http://orbis.stanford.edu

[178] 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.

archaeological datasets have been published as Linked Data, for example, in the STELLAR project Linked Data of project archives deposited with the Archaeology Data Service (ADS)[179]. Special mention deserves that the Getty Research Institute has published their major cultural heritage thesauri as LOD[180], and also other widely employed international and national vocabularies have become available as LOD, e.g. Iconclass[181], UK thesauri made available by the SENESCHAL project[182], the PACTOLS thesaurus[183], and others.

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*. 2011b; Elliott *et al*. 2014; May *et al*. 2015). In total, however, not many domain LOD datasets have been produced and effectively interlinked as yet.

If there is a substantial further increase in published and interlinked LOD datasets, semantic search and browse applications will allow discovery and retrieval of related content/data. But such an advance will mainly concern data aggregation, search and access, use of LOD for other research purposes is not implied. By use for research purposes we mean capability to address research questions and validate or scrutinize knowledge claims. The lack of such capability has not gone unnoticed by researchers and data managers who expect relevance of the LOD approach also in this direction.

For example a researcher who tried using museum Linked Data sets for an art historical study suggests cultural heritage institutions *"to seek out research uses of their data, and not limit their thinking to mere aggregation and dissemination (…). Creating LOD is hard enough for these institutions, so with some more utilities for individual researchers to take advantage of the complex data expressions and queries offered by LOD, hopefully it will be easier for GLAMs to design their data offerings to better support the kind of detailed research that these data projects keep promising to enable"* (Lincoln 2016 [note: GLAMS is an acronym for Galleries, Libraries, Archives and Museums]).

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. For example, using the LD created by the Portable Antiquity Scheme, the British Museum and ADS, and look at what we can actually learn by combining these datasets. Are they even compatible? What makes datasets compatible for interoperability? How compatible must they be in order to generate new and useful information? Does interoperability actually confound the results, as we don't understand how best to filter it? It's one thing to keep putting LOD out there, but we need to partner in a focussed way with domain experts to start answering these questions, begin building best practice on how to actually use LD"* (J. Charno, H. Wright and J. Richards, ADS, statement in the consultation on the ARIADNE innovation agenda).

---

[179] Archaeology Data Service: The STELLAR project, http://archaeologydataservice.ac.uk/research/stellar/; ADS Linked Open Data, http://data.archaeologydataservice.ac.uk

[180] 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.

[181] ICONCLASS as Linked Open Data, http://www.iconclass.org/help/lod

[182] Heritage Data - Linked Data Vocabularies for Cultural Heritage, http://www.heritagedata.org

[183] PACTOLS - Peuples, Anthroponymes, Chronologie, Toponymes, Oeuvres, Lieux et Sujets, http://frantiq.mom.fr/thesaurus-pactols

Also researchers of the data publication platform Open Context emphasise, *"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). Open Context has been working on projects with researchers and institutions that involve Linked Data. For example, one project focused on zooarchaeological datasets documenting early agricultural communities in Anatolia. The datasets have been made comparable by linking and annotating them according to animal taxa published by the Encyclopedia of Life[184] and to morphological concepts of the Uber Anatomy Ontology[185] (Kansa *et al*. 2014; Whitcher-Kansa 2015). This is a rare example where archaeological data has been interlinked with a scientific KOS, although not supporting research tasks beyond searching objects.

The need to progress from LOD based content/data search to research-focused applications is also stressed by the e-science and linked science communities that want to see LOD support the process of research, including scientific workflows, computing and analysis (Bechhofer *et al*. 2011; Kauppinen *et al*. 2013). 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.

6.6.3 Search vs. research

Some examples will be useful to illustrate the difference between searching archaeological information based on LOD and research-focused LOD applications. The Getty Research Institute has made available their major cultural heritage thesauri as LOD. Patricia Harpring, Managing Editor of the Getty Vocabulary Program, describes a scenario where these vocabularies would aid discovery of related information:

"Let's imagine that a researcher finds an interesting article online about the historical use of incense burners in Mexico. To explore the topic further today would require many hours or days of research; however, LOD will enable a new generation of search engines to follow the links between data sources to deliver more complete answers in much less time. In this use case, the AAT [Art & Architecture Thesaurus] could provide variant spellings, synonyms in other languages for 'incense burners,' and the narrower concept 'censers' with its variant terms, enabling the researcher to instantaneously discover numerous museum sites and articles on this topic. The AAT hierarchy could also focus the search on censers attributed to Pre-Columbian cultures. The user could explore geographic regions where these censers were created through TGN [Thesaurus of Geographic Names] place names, hierarchies, and linked maps. The names and biographies in ULAN [Union List of Artist Names] could lead the user to pertinent information about artists and patrons associated with the creation of the censers. CONA [Cultural Objects Name Authority], which ideally will have subject indexing, could provide links to photographs, paintings, or even YouTube videos portraying usage of censers (see an entertaining video of a 'monster censer' at Santiago de Compostela, Spain)" (Harpring 2014).

Achieving this scenario for a lot of cultural heritage information would be a great advance in the discovery of related information. As Harpring notes, it would allow finding more complete answers to search questions in much less time. However, this is about search, not research.

Beck (2010) addresses future research-focused archaeological applications of LOD. One example is sequences of pottery styles which are being used to establish a framework for dating archaeological

[184] Encyclopedia of Life, http://eol.org

[185] UBERON - Uber Anatomy Ontology, http://uberon.org

contexts, e.g. stratigraphic layers of an excavation. Beck envisions that interlinked LOD of pottery classifications and documentation of excavations would allow identifying inconsistencies in the published archaeological record.

"In addition to many other things pottery provides essential dating evidence for archaeological contexts. However, pottery sequences are developed on a local basis by individuals with imperfect knowledge of the global situation. This means there is overlap, duplication and conflict between different pottery sequences which are periodically reconciled (…). This is the perennial process of lumping and splitting inherent in any classification system. Updated classifications and probable dates allow us to re-examine our existing classifications. One can reason over the data to find out which contexts, relationships and groups are impacted by a change in the dating sequences either by proxy or by logical inference (a change in the date of a context produces a logical inconsistency with a stratigraphically related group). (…) Publicly deposited RDF data should be linked data: this means that all the primary data archives are linked to their supporting knowledge frameworks (such as a pottery sequence). When a knowledge framework changes the implications are propagated through to the related data dynamically".

This scenario is very demanding as it includes machine-based reasoning over LOD pottery classifications interlinked with information in many datasets of excavations which contain dating of stratigraphic layers of excavations based on pottery finds. The pottery classification system (or, more likely, different systems) would have to be available as Linked Data (based on SKOS or OWL), and the pottery based datings in the excavation datasets described consistently in a common format, and the datasets of course also published as Linked Data.

While unrealistic, the scenario touches upon crucial issues of stablility and change of knowledge frameworks. If they are "living" frameworks that support the on-going research and knowledge creation process, there is always some addition and modification going on. One extreme example is species taxonomies where revisions are conducted regularly and produce more or less intensive "revision shocks" which impact on the documentation of species and even critical measures such as species protection and conservation (Vences *et al*. 2013). Hepp (2007) addresses conceptual dynamics in domains of knowledge and the issue of long update cycles of formalized knowledge organization systems. Thus new and arguably most interesting concepts in current research will not be present for long in domain thesauri or ontologies. Furthermore there is the issue of different classifications of the same research objects which, ideally, would co-exist in a knowledge system or interlinked systems (cf. Madsen 2004: 41, in the context of archaeological reference collections).

Visions of research-focused archaeological applications of LOD, like Beck's example, expect such applications to allow automatic reasoning over a web of many interlinked data resources. In this quasi artificial intelligence scenario Linked Data applications would identify inconsistencies, contradictions, etc. in scientific statements (knowledge claims) or, as a positive example, present surprising relationships between data worth exploring further. Thus Linked Data applications would carry out some tasks that can be subsumed under research rather than search, e.g. detect relevant relationships between data or scientific statements that are contradictory.

6.6.4 Examples of research-oriented Linked Data projects

There are already some Linked Data projects which aim to go beyond simple search functionality. But not many and not necessarily in archaeology. We describe two examples, one in the field of social history and another concerning Classical Studies.

Dutch Ships and Sailors[186]: As an example of LOD in the field of social history, the Dutch Ships and Sailors project has brought together four datasets on Dutch maritime history as five-star Linked Data. End of March 2014 the Linked Data comprised of 25 million RDF triples, divided over 33 named graphs. Around 1.5 million links connected the datasets as well as linked to external sources; for example 180,000 links to external historical newspaper articles were established and 2500 geographical entities matched to GeoNames entities (De Boer *et al*. 2014 and 2015). The project presented a number of examples of how the data can be used for historical research on the socio-economic realities of the 18th Century, for example lists of persons who embarked on different types of ships, analysis of the birth provinces of sailors on Dutch East India Company ships over multiple years, etc. In a follow-up project further datasets have been added to the initial Dutch Ships and Sailors Cloud (de Boer & Leinenga 2014; Entjes 2015).

EPNet Project[187]: Aims to provide historians with data resources and tools for investigating the Roman trade system based on Latin and Greek inscriptions on amphoras for food transportation. In collaboration with experts of the history of the Roman economy the project has specified an ontology of domain knowledge which represents the way the data are being understood by scholars, how they are connected, and how they relate to the literature and current research practices. The main section of the ontology is a specialisation of the CIDOC CRM while other sections build on the metadata model of the EAGLE project (EAGLE 2015), EpiDoc[188] for the encoding of editions of ancient texts/documents (inscriptions, papyri, manuscripts), FaBiO[189] for bibliographic references, and others. The EPNet ontology is meant to be "functional to research", e.g. support researchers in the exploration of hypotheses and question established narratives (Calvanese *et al*. 2015; Calvanese *et al*. 2016). Initial data resources are the rich database of Roman amphorae and their associated epigraphy (i.e. stamps and tituli) of the Centre for the Study of Provincial Interdependence in Classical Antiquity, University of Barcelona[190], the Epigraphic Database Heidelberg[191], and the Pleiades gazetteer and graph of ancient places[192].

6.6.5 CIDOC CRM as a basis for research applications

Expectations of reseach-focused applications of LOD in the field of archaeology and other cultural heritage research often relate to the CIDOC CRM as an 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[193] (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.

[186] Dutch Ships and Sailors (Clarin IV project, 4/2013-3/2014), http://dutchshipsandsailors.nl

[187] EPNet - Production and Distribution of Food during the Roman Empire: Economic and Political Dynamics (ERC Advanced Grant project, 3/2014-2/2019), http://www.roman-ep.net

[188] EpiDoc: Epigraphic Documents in TEI XML, http://epidoc.sf.net

[189] FaBiO - FRBR-aligned Bibliographic Ontology, http://vocab.ox.ac.uk/fabio

[190] CEIPAC database, http://ceipac.ub.edu

[191] Epigraphic Database Heidelberg, http://edh-www.adw.uni-heidelberg.de

[192] Pleiades, http://pleiades.stoa.org

[193] ResearchSpace, http://www.researchspace.org

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 scientific documentation of finds, sites and monuments, at the level of detail and precision required by researchers of the heritage sciences[194]. Recent extensions of the CIDOC CRM cover scientific observation and argumentation (CRMsci and CRMinf). Thus CIDOC CRM based modelling of scientific processes and documentation of observations can enable integration of scientific information and argumentation (knowledge claims).

The CIDOC CRM developer community invites data sharing and integration projects to use the ontology to describe the meaning and context of their information objects so that research e-infrastructure and services can provide homogeneous 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 applications. What they see as inadequate is the traditional information aggregation and integration approach based on fixed "core" metadata fields which are artificial generalizations that do not mediate the contextual knowledge of the data providers such as research institutes and museums (Doerr & Oldman 2013; Oldman *et al*. 2014).

The vision of the CIDOC CRM developer community goes well beyond enabling cultural heritage institutions to provide structured access to collection objects. Archaeological and other heritage 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.

This objective has been addressed by the development of the ARIADNE Reference Model which is based on the CIDOC CRM and enhanced or new extensions (e.g. CRMarchaeo for archaeological excavations)[195]. The aim of semantic integration of research data requires that the participants produce 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 RDF format which can be shared as Linked Data on the Web.

The challenge of enabling effective mappings has been addressed by an innovative solution, the SYNERGY Reference Model (Doerr *et al*. 2014b). 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-useable mapping cases. Several ARIADNE have already used the Mapping Memory Manager[196] module of SYNERGY to define complex correspondences between entities of their and other databases and the conceptual classes provided by the extended CIDOC CRM (ARIADNE 2016a; Doerr *et al*. 2016; Gerth *et al*. 2016).

At large scale this approach will allow reaping the expected benefits only in the medium to long term, when many databases are mapped to the extended CIDOC CRM. However, mapping of a few related databases may demonstrate significant advantages of CIDOC CRM based integration in the short-term, possibly promoting further mappings.

[194] 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

[195] See the overview and description of the CIDOC-CRM extensions at: http://www.ics.forth.gr/isl/CRMext/

[196] Mapping Memory Manager - 3M (FORTH-ICS), http://www.ics.forth.gr/isl/3M

6.6.6 Brief summary and recommendations

Brief summary

Linked Open Data based applications that demonstrate considerable advances in research processes and outcomes could be a strong driver for a wider uptake of the LOD approach in the research community. Current examples of Linked Data use for research purposes rarely go beyond semantic search and retrieval of information. This has not gone unnoticed by researchers who expect relevance of Linked Open Data also for generating and validating or scrutinizing knowledge claims. To allow for such uses a tighter integration of discipline-specific vocabularies and effective Linked Data tools and services for researchers are required.

Expectations of reseach-focused applications of LOD in the field of cultural heritage and archaeology often relate to the CIDOC CRM as an integrating framework. The CIDOC CRM is recognised as a common and extendable ontology that allows semantic integration of distributed datasets and addressing research questions beyond the original, local context of data generation. Notably, in the ARIADNE project several extensions of the CIDOC CRM have been created or enhanced, e.g. CRMarchaeo, an extension for archaeological excavations, and extensions for scientific observations and argumentation (CRMsci and CRMinf).

To meet expectations such as automatic reasoning over a large web of archaeological data many more (consistent) conceptual mappings of databases to the CIDOC CRM would be necessary. Linked Data applications then might demonstrate research dividends such as detecting inconsistencies, contradictions, etc. in scientific statements (knowledge claims) or suggesting new, maybe interdisciplinary lines of research based on surprising relationships between data.

Recommendations

- *LOD based applications that enable advances in archaeological research processes and outcomes may foster uptake of the LOD approach by the research community.*
- *LOD based applications for research will have to demonstrate advantages over or other benefits than already established forms of data integration and exploitation.*
- *Develop LOD based services that go beyond semantic search and retrieval of information and also support other research purposes.*
- *Build on the CIDOC CRM and available extensions to exploit conceptually integrated LOD.*

7 Linked Data development in ARIADNE

The ARIADNE project promotes a culture of open sharing and (re-)use of archaeological data across institutional, national and disciplinary boundaries of archaeological research. Linked Open Data can greatly contribute to this goal. Therefore ARIADNE recognises Linked Data as a key approach for data sharing and interoperability. One strand of the project work supports the development of such data. The activities in this strand of work concerned

- the metadata of the datasets registered in the ARIADNE data catalogue,
- vocabularies for the metadata describing registered datasets (e.g. mapping of existing vocabularies, support for the generation of vocabularies in SKOS),
- mapping of datasets to the core CIDOC CRM and extensions of the CRM created in ARIADNE,
- demonstrators generating and using Linked Data (e.g. metadata extracted from unstructured data such as grey literature, CIDOC CRM based datasets), and
- providing access to ARIADNE Linked Data for external application developers.

Thus the work mainly centred on Linked Data related to data registration, enabling data integration via vocabularies and the CIDOC CRM ontology, demonstration of enhanced or new capabilities (e.g. enhanced cross-searching of data resources), and preparing the ground for linking of resources also beyond the ARIADNE pool of resources. The ARIADNE data catalogue and other results of the activities listed above are included in the ARIADNE graph database and accessible through a SPARQL endpoint (see *Chapter 8*). The sections below describe the activities in greater detail, including the Linked Data methods and tools that have been applied, enhanced or newly developed by ARIADNE researchers and developers.

7.1 The ARIADNE catalogue as Linked Open Data

The key component of the ARIADNE e-infrastructure is the dataset registry/catalogue. In the registry data providers describe their resources (data sets, collections, etc.) based on a common model, the ARIADNE Catalogue Data Model (ACDM)[197]. The ACDM builds on the W3C's Data Catalog Vocabulary (DCAT)[198] which has been designed to facilitate interoperability between data catalogs published on the Web. The ACDM extends DCAT taking account of requirements of describing archaeological data resources. The ARIADNE registry/catalogue holds metadata of data resources, the project does not collect, store and curate primary research data – which are tasks of the data providers (e.g. community data archives or institutional repositories). The metadata is being collected and enriched with the MoRe (Metadata & Object Repository) aggregator[199] and included in the ARIADNE data catalogue. ARIADNE makes the catalogue and other data generated in the project available as Linked Open Data. This means that other service/application developers can query the data as well as interlink it with other LOD. Thereby the ARIADNE LOD can become part of a Linked Data "cloud" of archaeological and related other information resources.

[197] ARIADNE Catalogue Data Model (ACDM), http://support.ariadne-infrastructure.eu

[198] W3C (2014) Recommendation: DCAT - Data Catalog Vocabulary, 16 January 2014, http://www.w3.org/TR/vocab-dcat/

[199] MoRe (Metadata & Object Repository), http://more.dcu.gr; also registration of single datasets with the metadata entered manually is possible.

7.2 Work on vocabularies as Linked Data

Project partners conducted various work concerning vocabularies as Linked Data. This includes

- Generation of SKOS versions of existing or newly developed vocabularies,
- Development of a toolset for vocabulary mapping and mapping of subject vocabularies which partners use for data indexing to a major common vocabulary, the Art & Architecture Thesaurus,
- Use of vocabularies to support Natural Language Processing (e.g. metadata extraction from archaeological "grey literature",
- Mapping of datasets to the core CIDOC CRM and extensions of the CRM created in ARIADNE,
- Demonstrators using Linked Data (e.g. CIDOC CRM based datasets) and demonstrating enhanced or new capabilities (e.g. enhanced cross-searching of data resources).

This work and results achieved are described in the sections that follow.

7.2.1 Vocabularies in SKOS

Vocabularies such as taxonomies and thesauri are essential knowledge structures and terminology of domains of knowledge. ARIADNE is a project and therefore not in a position to publish and maintain vocabularies. This must be done by the institutions who own the vocabularies. However some partners and associated organisations own and/or manage national or other major vocabularies, which are being used in ARIADNE. Below we briefly describe vocabularies that have been transformed to SKOS previously, in parallel to or within the ARIADNE project, including the number of mappings to the Art & Architecture Thesaurus (which is described in the next section):

- *Italian Ministry of Cultural Assets and Activities / Central Institute for the Union Catalogue (ICCU) – PICO thesaurus*[200]*:* A large thesaurus related to culture and cultural heritage (Italian and English) which is being used for the data of CulturaItalia[201]; a small number of about 200 terms concern archaeology of which most have been mapped to the AAT.
- *German Archaeological Institute (DAI) vocabularies:* The Institute has vocabularies for different entities (e.g. books, collections, inscriptions, buildings and structures, multi-part monuments, topographic objects) from which about 400 concepts, already in SKOS and previously mapped to the AAT, are being used in ARIADNE. Work is ongoing to harmonize the different DAI thesauri to one common standard, the iDAI.vocab[202].
- *Major UK thesauri*[203]: In the SENESCHAL project (UK, AHRC-funded project, 2013-2014), running in parallel to ARIADNE, the project partner University of South Wales (Hypermedia Research Group) helped UK heritage institutions – Historic England and the Royal Commissions on Ancient & Historical Monuments of Scotland (RCAHMS) and Wales (RCAHMW) make their vocabularies

[200] PICO thesaurus (MiBAC-ICCU, Italy), http://purl.org/pico/thesaurus_4.2.0.skos.xml

[201] Cultura Italia: Dati, http://dati.culturaitalia.it

[202] iDAI.vocab: This is a group of 14 thesauri of monolingual archaeological terminology aimed to collect and organise the terminology used in information services of the German Archaeological Institute. The thesauri are in different languages (Arabic, Chinese, English, Farsi, French, German, Greek, Hungarian, Italian, Portuguese, Russian, Spanish, Turkish, Ukrainian) and of varied size (ranging from below 100 to several thousand terms). The German thesaurus, which is already mapped to the AAT, serves as the central hub to and through which the other thesauri are linked. iDAI.vocab, http://archwort.dainst.org

[203] Heritage Data - Linked Data Vocabularies for Cultural Heritage, http://www.heritagedata.org

available in SKOS format as Linked Open Data. In ARIADNE the Archaeology Data Service employs five Historic England thesauri of which about 850 concepts have been mapped to the AAT.

- *Fédération et ressources sur l'Antiquité (FRANTIQ, France) – PACTOLS thesaurus*[204]: A large multi-lingual thesaurus which focuses on antiquity and archaeology from prehistory to the industrial age; terms in French, English, German, Italian, Spanish, Dutch, and (some) Arabic). ARIADNE has a cooperation agreement with FRANTIQ on the deployment of PACTOLS in the project. Over 1600 PACTOLS concepts which the ARIADNE partner Institut National des Recherches Archéologiques Préventives (Inrap, France) uses in their catalogue of archaeological reports (DOLIA) have been mapped to the AAT.
- In the Netherlands, Data Archiving and Networked Services (DANS) provide a list of monument types (Archeologische complextypen) for describing Dutch archaeological excavations. The types are managed by the Rijksdienst voor het Cultureel Erfgoed (RCE)[205]. These have recently been expressed as SKOS. About 450 concepts have been mapped to the AAT.
- The most detailed classification system available for Irish Monument types is the class list developed by the National Monuments Service (NMS). This is a hierarchical list which was used in the classification of sites and monuments that formed part of the Archaeological Survey of Ireland. It has been expressed in SKOS as part of the LoCloud project[206]. Over 480 concepts have been mapped to the AAT.
- AIAC's FASTI Online uses a flat list of monument types in the "advanced" search interface. The set of FASTI concepts are published online with URIs[207]. About 130 concepts have been mapped to the AAT.

Within the ARIADNE project data providers, with support by the University of South Wales (Hypermedia Research Group), created or transformed/enhanced existing vocabularies in/to SKOS format:

- *Data Archiving and Networked Services (DANS, Netherlands) – Dendrochronology multi-lingual vocabulary*: With help from ARIADNE, DANS and collaborators have restructured and enhanced the Tree Ring Data Standard (TRiDaS). TRiDaS[208] is used to describe the data resulting from all kinds of dendrochronological analysis. The multilingual vocabulary, which has recently been expressed in SKOS, is being employed for the Digital Collaboratory for Cultural Dendro-chronology[209] (Jansma 2013) and available also to other users. Some 336 concepts have been mapped to the AAT.

- *Italian Ministry of Cultural Assets and Activities / Central Institute for the Union Catalogue (ICCU) – Reperti Archeologici (RA) Thesaurus*[210]*:* A pictorial thesaurus describing archaeological finds. This has been expressed as SKOS during ARIADNE using the STELLAR toolkit. About 1100 concepts of this vocabulary have been mapped to the AAT.

---

[204] PACTOLS (Peuples, Anthroponymes, Chronologie, Toponymes, Œuvres, Lieux et Sujets), http://pactols.frantiq.fr

[205] See: http://cultureelerfgoed.nl/dossiers/archis-30/archeologisch-basisregister-plus

[206] Irish Monuments http://vocabulary.locloud.eu/Irish_Monuments/

[207] FASTI Online, see http://www.fastionline.org/data_view.php, and for an example of a concept with URI see http://www.fastionline.org/concept/attributetype/monument

[208] TRiDaS - The Tree Ring Data Standard, http://www.tridas.org

[209] Digital Collaboratory for Cultural Dendrochronology - DCCD, http://dendro.dans.knaw.nl; project website: http://vkc.library.uu.nl/vkc/dendrochronology/

[210] Reperti Archeologici (RA) Thesaurus, http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/74; http://vast-lab.org/thesaurus/ra/vocab/index.php

7.2.2 Mapping of subject vocabularies

The main goal of the mapping between vocabularies in the ARIADNE project has been to enable searching of relevant data resources which are being held by archives in different countries. Bringing together the original resource metadata does not allow for effective searching of relevant resources, because the providers use terms from subject vocabularies in different languages and, if in the same language, often use different terms for the same subject.

To enable cross-searching of data resources mapping of terms was necessary. But the ARIADNE project has 15 data providers and many others expressed interest to make data resources searchable through the ARIADNE portal. There is no scalable approach for direct, many-to-many mapping between terms in several vocabularies. Therefore it was decided to use an appropriate common vocabulary as intermediary "hub" onto which data providers map their subject terms (the so called switching language approach). The content-rich and multi-lingual Art & Architecture Thesaurus (AAT) of the Getty Research Institute has been selected as the central hub of the mapping. The AAT is available as Linked Open Data in SKOS, published unter the Open Data Commons Attribution License (ODC-By) 1.0[211].

The AAT contains over 40,000 concepts and over 350,000 terms, organised in seven facets (and 33 hierarchies as subdivisions): Associated concepts, Physical attributes, Styles and periods, Agents, Activities, Materials, Objects and optional facets for time and place (Harpring 2016). The AAT's scope is broader than archaeology, encompassing visual art, architecture, other material heritage, archaeology, conservation, archival materials, etc., but contains many useful high level archaeological concepts, particularly in the Built Environment, Materials and Objects hierarchies.

Vocabulary mapping tools

For the mapping the project partner University of South Wales (Hypermedia Research Group) developed an interactive tool which enables subject experts to produce SKOS mapping relationships (e.g. broadMatch or closeMatch) between their vocabulary terms and the AAT terms (Binding & Tudhope 2016). The tool is a lightweight browser based application that presents concepts from chosen source and target vocabularies side by side, exposing additional contextual evidence to allow the user to make a more informed choice when deciding on potential mappings. The tool is for vocabularies already expressed in RDF/SKOS and can work directly with the data – querying external SPARQL endpoints rather than storing any local copies of complete vocabularies. The set of mappings developed can be saved locally, reloaded and exported to a number of different output formats (JSON for use in ARIADNE). The tool is provided open source and the software code is available on GitHub[212]. A second mapping approach has been developed for source vocabularies that are smaller term lists and not yet expressed in RDF. Such term lists are often available or can be easily represented in a spreadsheet. A standard template with example mappings was designed to support domain experts in the mapping of terms to the target vocabulary. A CSV transformation produces the representation of the mappings in RDF/JSON format[213].

---

[211] Getty Vocabularies as Linked Open Data, http://www.getty.edu/research/tools/vocabularies/lod/index.html

[212] Vocabulary Matching Tool, http://heritagedata.org/vocabularyMatchingTool/; source code for local download and installation, https://github.com/cbinding/VocabularyMatchingTool

[213] ARIADNE subject mappings: Spreadsheet template and conversion, https://github.com/cbinding/ARIADNE-subject-mappings

Mappings conducted

The application of the tools and the "hub" approach have first been tested and evaluated in an exploratory pilot (Binding & Tudhope 2016). Terms of five subject vocabularies employed by ARIADNE data providers were mapped to the AAT and the semantic linkage used for retrieval experiments. The vocabularies are: a flat list of monument types employed in Fasti Online (in English), terminology for types of archaeological sites of the Central Institute for the Union Catalogue, Italy (in Italian), Archeologische complextypen of the Rijksdienst Cultureel Erfgoed (in Dutch, employed by Data Archiving and Networked Services, Netherlands), relevant terms of the archaeological dictionary of the German Archaeological Institute (in German), and Historic England's Thesaurus of Monument Types (in English, employed by the Archaeology Data Service, UK). The study demonstrated advantages of the approach by performing mediated cross-search over archaeological datasets from different countries with semantic expansion across the multilingual vocabularies.

By June 2016, concepts from 25 vocabularies employed by 11 project partners were already mapped to the AAT; six partners each employed concepts from 1 vocabulary, two partners each from 2 vocabularies, and the other three partners from 4, 5 and 6 vocabularies. In terms of structure and size the vocabularies varied from a small term list for a particular dataset to standard national vocabularies with a large number of concepts. 15 of the vocabulary mappings were conducted with the spreadsheet template (or a similar partner spreadsheet), 2 using the online interactive mapping tool (i.e. when the source vocabulary was available in RDF/SKOS) and 8 using the partner's own (intellectual/manual) resources.

In total 5823 mappings were conducted, with mappings of individual partners ranging from a few up to over 1600 terms. To give some examples: The Institute of Archaeology of the Scientific Research Centre of the Slovenian Academy of Sciences and Arts (Slovenia) mapped 93 terms for archaeological site records in their ARKAS - Arheološki kataster Slovenije system to the AAT; the Data Archiving and Networked Services (Netherlands) and collaborators mapped 336 concepts of the vocabulary of the Digital Collaboratory for Cultural Dendrochronology, the Discovery Programme (Ireland) 486 concepts of the Irish Monument Types thesaurus, the Institut National des Recherches Archéologiques Préventives (France) 1634 concepts of the PACTOLS thesaurus which are being used by their catalogue of archaeological reports (DOLIA).

Very few terms could not be mapped to the AAT. 50% of the mapping relations were skos: exactMatch, 18% skos:closeMatch, 27% skos:broadMatch and 5% skos:narrowMatch (one partner also did a few skos:relatedMatch mappings). As expected there was only a small number of skos:narrowMatch mappings, i.e. where the ATT was more specialised than the partners' vocabularies. An ARIADNE project deliverable is available which describes the mappings in greater detail (ARIADNE 2016b).

The ARIADNE data catalogue employs the MoRe (Metadata & Object Repository) aggregator[214] to harvest the metadata provided by the project partners utilising the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). A bespoke AAT subject enrichment service has been developed that applies the partner vocabulary mappings (in JSON format) to the partner subject metadata and derives an AAT concept (both preferred label and URI) to augment the subject metadata in the data catalogue. For example, 773,600 of the Archaeology Data Service or 6131 records of Fasti Online have been enriched in this way. The catalogue metadata is supplied to the ARIADNE portal, where the search functionality can use the AAT based terminology "hub" to retrieve metadata of different

[214] MoRe (Metadata & Object Repository) aggregator, http://more.dcu.gr

data providers who mapped related subject terms to the AAT. A search on a term originating from any one vocabulary can utilize the mediating structure to route through to terms from other vocabularies (which may be expressed in different languages) and retrieve the identified data records.

7.2.3 Metadata for vocabularies and mappings in SKOS

Concerning the vocabularies and mappings between them in Linked Data format it would be beneficial having metadata for these products. In the SENESCHAL project University of South Wales (Hypermedia Research Unit) produced VoID (Vocabulary of Interlinked Datasets)[215] metadata of each of the UK thesauri which have been transformed to Linked Data in RDF/SKOS. This metadata and links to example resources have been published in the DataHub[216]. Also datasets of mappings between vocabularies are valuable semantic assets for which metadata about versions, authorship, licensing, etc. would be necessary for users and machines, for example to distinguish between different mappings produced for large vocabularies. ARIADNE partners who own vocabularies in SKOS and have produced mappings to the AAT have been recommended to follow the good practice exemplified by University of South Wales (Hypermedia Research Group).

7.3 What – Where – When as Linked Data

On the ARIADNE data portal the core services for cross-searching the different resources for relevant information are based on the "What - When - Where" approach. The approach has been successfully demonstrated in the ARENA portal for searching archaeological sites and monuments of six European countries[217]. In a nutshell, "What" concerns the subjects, "Where" the geographical locations, and "When" the periods (named cultural periods and date ranges) for which users wish to find relevant data. This information is provided by the data providers in the metadata of the resources they register in the ARIADNE catalogue.

The ARIADNE data portal allows searching across the various data resources based on subjects, location and date ranges (chronology). In the portal this has been implemented as subject-based search, map-based search and a timeline feature. The implementation of the search & browse services is not based on Linked Data, but such data for subjects, location and chronology is being prepared, particularly for future linking to external Linked Data resources as well as external developers who wish to query the ARIADNE Linked Data and/or link it with other data.

7.3.1 What (subjects)

Linked Data for the subjects contained in the metadata partners have provided to the ARIADNE data catalogue has been produced through the mapping of concepts to the Art & Architecture Thesaurus (as described in the sections above.

[215] W3C (2011) Interest Group Note: Describing Linked Datasets with the VoID Vocabulary, 3 March 2011, http://www.w3.org/TR/void/

[216] HeritageData on DataHub, http://datahub.io/dataset?q=heritagedata

[217] ARENA - Archaeological Records of Europe - Networked Access project (2001-2004, and 2009-2010 in the context of DARIAH), http://ads.ahds.ac.uk/arena/search/

7.3.2 Where (places)

"Where" concerns geographic information which can mean just names of places, areas, regions, etc., or names together with geo-referencing (lat./long coordinates). In the ARIADNE survey on expectations for data portal services map-based search was a clear "must have" (cf. ARIADNE 2015e: 278-289). Therefore the dataset metadata in the ARIADNE catalogue in addition to place names should include standard lat./long. coordinates to allow for map-based search of relevant resources on the data portal. As the common standard ARIADNE adopted WGS84 (World Geodetic System 1984)[218]. Most data providers already had WGS84 based coordinates. In cases where the original metadata contained only place names the data providers employed the GeoNames gazetteer to derive coordinates for the names.

The database of the GeoNames[219] gazetteer is integrating geographical data such as names of places in various languages, elevation, population and others from various sources. All lat./long. coordinates are in WGS84 (World Geodetic System 1984). The GeoNames data is available through a number of web services and a daily database export. The data is provided free of charge under a Creative Commons Attribution license (CC-BY). It contains over 10 million geographical names and consists of over 9 million unique features whereof 2.8 million populated places and 5.5 million alternate names.

GeoNames is available as Linked Open Data and one of the core linking hubs of the Linked Data Cloud. Therefore ARIADNE sees GeoNames as the core gazetteer for Linked Data based linking with external data resources based on place names and other geographical information. GeoNames covers modern places and other geographical information, which is also generally used by archaeologists in the documentation of fieldwork, reports and publications. However archaeological material also often includes ancient/historical place names and other geographical references. For such references ARIADNE itends to collaborate with the Pelagios initiative which employs the Pleidades and other Ancient World gazetteers. The ARIADNE partners German Archaeological Institute and Fasti Online already participate in the Pelagios project (see *Section 5.3*).

7.3.3 When (chronology)

In archaeology the "when" of sites and objects is typically given as a cultural periods and date-ranges. In the ARIADNE survey on expectations for the data portal services the archaeological researchers considered searching data resources based on cultural periods and date-ranges as particularly important (cf. ARIADNE 2015e: 278-289).

To enable such searching, data partners have to give in their metadata the period terms which they use and the absolute date ranges (start/end dates) which apply to each term for their country/regions. The period terms and date ranges are often defined in standard national periodizations but also proprietary controlled period lists derived from authoritative sources are possible. For example, the Archeologisch Basisregister (ABR) of the Cultural Heritage Agency of the Netherlands or MIDAS Heritage for the UK provide standard national periodizations.

A cultural period as elaborated in archaeological and historical research has temporal and geographical boundaries, defined by some characteristics which set it apart from the previous and later period in a chronology. Named period search on the ARIADNE data portal, for example "Roman" returns results for period AD43 to AD410 from UK datasets and results for period 10BC to AD450

[218] World Geodetic System 1984 (WGS 84), http://earth-info.nga.mil/GandG/wgs84/

[219] GeoNames, http://www.geonames.org

from Dutch datasets; however date-range/timeline-based search, e.g. 10BC to AD40 return Roman results from Dutch datasets and Iron Age results from UK datasets.

On Linked Data for cultural periods ARIADNE collaborates with the PeriodO project[220]. PeriodO is building a system for collecting, organising and referencing definitions of periods based on URIs. The periods are provided through an online application as well as a downloadable set of Linked Data. The PeriodO approach is to gather individual period assertions made by authoritative scholarly sources about the temporal and spatial boundaries of periods in particular research contexts, retaining the provenance of the assertions, e.g. scholarly book or paper (Rabinowitz 2014; Golden & Shaw 2015 and 2016).

But the PeriodO system also includes established national periodizations. ARIADNE has produced from available periodizations a set of cultural periods and their time ranges from the Paleolithic to Modern times for 24 European countries (in total 659 periods)[221]. The periods set has been incorporated in the PeriodO system which allows stable linking of data based on the persistent URIs assigned by PeriodO. To use the PeriodO URIs in ARIADNE an enrichment service is being developed and included in the MoRe aggregator which will attach the URIs when processing the metadata harvested from data providers.

Through the PeriodO system also other projects can use periods provided by ARIADNE and others. ARIADNE promotes the use of PeriodO URIs to allow for wider interlinking of data based on periods/chronologies. The PeriodO project is funded until 2018 by a grant of the US Institute of Museum and Library Services.

7.4 Use of vocabularies in NLP and data mining

Vocabularies are also important in natural language processing and data mining tasks. The sections below describe such uses in research and development carried out in ARIADNE.

7.4.1 Natural Language Processing

In ARIADNE also research and development on Natural Language Processing (NLP) of archaeological content has been explored with the aim of making text-based resources more discoverable and useful (ARIADNE 2015c). This work of researchers of the Archaeology Data Service, University of South Wales (Hypermedia Research Group) and Leiden University (Faculty of Archaeology) focused specifically on the “grey literature” of archaeological investigations.

The partners have explored machine learning and rule-based approaches. Here we focus on the work on ruled-based methods in which vocabularies in Linked Data format have been used. In this work the OPTIMA semantic annotation system of the Hypermedia Research Group has been used. OPTIMA performs the NLP tasks of Named Entity Recognition, Relation Extraction, Negation Detection and Word-Sense Disambiguation using hand-crafted rules and terminological resources (Vlachidis 2012; Vlachidis *et al*. 2013; Vlachidis & Tudhope 2015a). The system uses the GATE (General Architecture for Text Engineering) framework, Ontology Based Information Extraction (OBIE) and several other techniques.

OPTIMA contributed to the Semantic Technologies for Archaeological Research (STAR) project, a pioneer in the use of NLP for extraction of metadata and linking of archaeological grey literature and

---

[220] PeriodO - Periods, Organized, http://perio.do; see also https://wiki.digitalclassicist.org/PeriodO

[221] ARIADNE set of cultural periods in the PeriodO system, http://n2t.net/ark:/99152/p0qhb66

digital archive databases based on English Heritage terminology vocabularies and the CIDOC CRM (Tudhope *et al.* 2011b; Vlachidis *et al*. 2012).

The NLP work in ARIADNE builds upon the experiences of STAR but targets "grey literature" also in other languages. This faces challenges of different vocabularies (e.g. with regard to structure) as well as differences in language characteristics. The address these challenges grey literature in Dutch has been chosen using thesauri of the Rijksdienst Cultureel Erfgoed. The original SKOSified thesauri were not suitable for supporting Ontology Based Information Extraction (OBIE) approaches, due to the incapacity of the GATE ontology tool to parse (understand) broader/narrower term relationships. Therefore transformation of the thesauri to OWL-Lite (ontology) was necessary.

With regard to language characteristics particularly compound noun forms present a challenge for the usual "whole word" matching mechanisms. Compound noun forms examples might include "beslagplaat" where both "beslag" and "plaat" are known to the vocabulary and also "aardewerk-magering" where aardewerk (pottery) is known but "magering" is not.

But the current pilot system has achieved some promising semantic enrichment of Dutch grey literature reports, concerning artefacts (such as "aardewerk") and other concepts including time periods. In order to overcome the "whole word" restrictions mechanisms operating on part matching are being explored. Negation detection is another aspect that has been explored during ARIADNE (Vlachidis *et al*. 2015b); it is important to distinguish whether the text indicates that evidence of some archaeological issue has or has not been found during an excavation. Expansion of NLP for extraction, indexing and linking of data/metadata from other European language grey literature is intended. Critical for good results in general is the availability of rich and well-structured vocabularies, but even in such cases some modification may be required to conduct NLP with optimal results.

7.4.2 Mining of Linked Data

ARIADNE partner Leiden University, in collaboration with the associated partner Free University Amsterdam, examined the feasibility of mining archaeological Linked Data, for example, to detect relevant patterns in the graph-structure of such data.

In the first years of the project, started in February 2013, no archaeological Linked Data was produced in the project. But an examination of a few datasets available elsewhere showed that they largely consisted of flat data structures with descriptive metadata values (ARIADNE 2015b). Mining of such data is unlikely to yield archaeologically interesting patterns. Indeed, interviews with domain experts indicated a strong interest in archaeological contexts, which means rich information generated in fieldwork. Particularly interesting would be spatio-temporal patterns between archaeological contexts.

Therefore the research group decided to work on information in the Dutch archaeological protocol SIKB 0102, called digital "pakbon" (package slip), developed and maintained by the Stichting Infrastructuur Kwaliteitsborging Bodembeheer (SIKB) / Foundation Infrastructure for Quality Assurance of Soil Management[222]. The SIKB 0102 has been introduced a few years ago (first version in 2010). It specifies which mandatory information about excavations and finds has to be provided as an XML document when depositing data in the E-Depot for Dutch Archaeology (managed by

[222] Stichting Infrastructuur Kwaliteitsborging Bodembeheer: Protocol 0102 Archeologie, http://sikb.nl/datastandaarden/richtlijnen/protocol-0102

ARIADNE partner Data Archiving and Networked Services - DANS)[223]. With regard to terminology the thesauri in the Archeologisch Basisregister (ABR+) of the Rijksdienst Cultureel Erfgoed (Cultural Heritage Agency)[224] have to be used.

While the amount of "pakbonnen" is growing each one still is an isolated entity and the XML documents as such cannot be used for semantic integration and mining of the information. Therefore the research group developed a Linked Data version of the SIKB 0102 (pakbon-ld), which incorporates its set of archaeological concepts and properties, but restructured and expanded to exploit the graph structure[225]. This version has been modelled in CIDOC CRM including the English Heritage extension (CRM-EH) which contains archaeology-specific concepts and relations. Moreover ABR+ thesauri in SKOS have been prepared for use in the transformation of SIKB 0102 XML documents to Pakbon Linked Data. Once these foundations were completed, a tool for automatic conversion has been developed[226]. With this tool 73 SIKB 0102 XML documents from the E-Depot for Dutch Archaeology have been translated and stored in the graph database together with the CIDOC CRM, CRM-EH and ABR+ vocabularies.

So far the results of mining this resource with SPARQL queries have been encouraging from a technical point of view, but far from useful from an archaeological perspective (e.g. trivial or conflicting results). It appears that the detection of archaeologically meaningful patterns requires an iterative interaction of researchers with query results from a database of still richer data than the "pakbonnen" provide. But the project now has a model and tool for converting documentation of fieldwork in the Netherlands to Linked Data and include it in the web of archaeological Linked Data.

[223] E-depot for Dutch Archaeology, http://www.edna.nl

[224] Rijksdienst Cultureel Erfgoed: Archeologisch Basisregister, http://abr.erfgoedthesaurus.nl

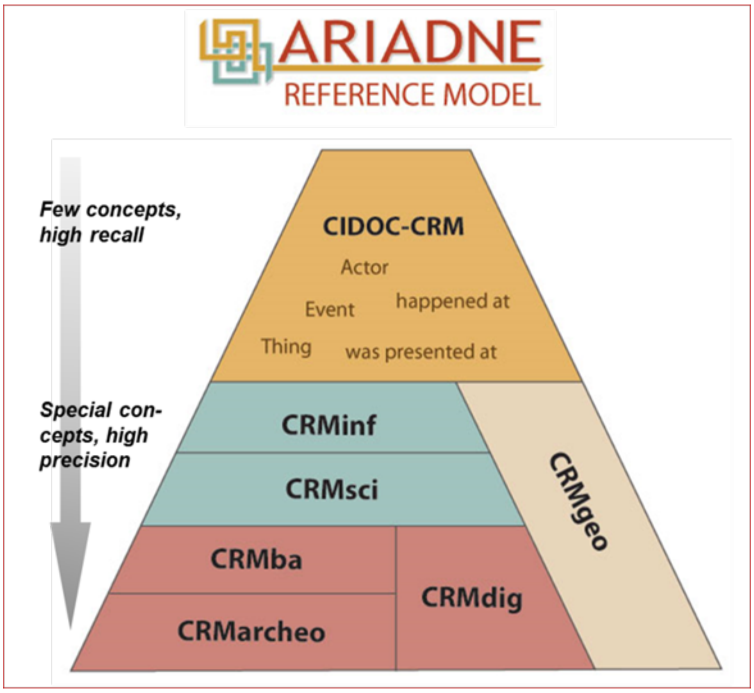
[225] Wilke Xander (VU Amsterdam, SPINlab): Pakbon Linked Data, http://pakbon-ld.spider.d2s.labs.vu.nl/home

[226] Wilke Xander (VU Amsterdam, SPINlab): Linked Data translation of the SIKB archaeological protocol 0102 (aka Pakbon), https://github.com/wxwilcke/pakbon-ld

7.5 CIDOC CRM extensions and mappings

ARIADNE recommends the CIDOC Conceptual Reference Model (CRM)[227] as a common ontology for data integration, discovery and access based on Linked Data, including the more ambitious goal to support research-oriented applications (see *Section 6.6.5*).

The CIDOC CRM has been developed specifically for describing and facilitating the exchange and integration of cultural heritage knowledge and data. Archaeology partly overlaps with this domain as well as needs modelling of additional conceptual knowledge, for example, to describe observations of an excavation (e.g. stratigraphy). The ARIADNE Reference Model comprises the core CIDOC CRM and a set of enhanced and new extensions, including the archaeological excavation process (CRMarchaeo) and built structures such as historic buildings (CRMba).



The table below gives an overview of the extensions to the CIDOC CRM which have been created or enhanced in the ARIADNE[228]:

○ CRMgeo: spatio-temporal model that articulates relations between the standards of the geospatial and the cultural heritage communities (integrates CRM with OGC standards; applications such as GeoSPARQL)	New extension, v1.0, April 2013
○ CRMdig: model of digitisation processes, to encode metadata about the steps and methods of production ("provenance") of digital representations such as 2D, 3D or animated models (validated in several projects)	Enhanced extension, v3.2, August 2014

[227] CIDOC - Conceptual Reference Model (CIDOC-CRM), http://www.cidoc-crm.org

[228] Description of the ARIADNE Reference Model and individual extensions (including reference document, presentation, RDFS encoding) is available at http://www.ariadne-infrastructure.eu/Resources/Ariadne-Reference-Model; see also http://www.ics.forth.gr/isl/CRMext/

○ CRMsci: model for integrating metadata about scientific observation, measurements and processed data (validated in archaeology, biodiversity and geology cases)	Enhanced extension, v1.2.2, August 2014
○ CRMinf: model for integrating data with scholarly argumentation and inference making in descriptive and empirical sciences (being validated with scholarly annotations); harmonized with CRMsci	New extension, v0.7, February 2015
○ CRMarchaeo: model for integrating metadata about the archaeological excavation process (introduces concepts of stratigraphy and excavation); being validated by archaeological records	New extension, v1.4, April 2016
○ CRMba: model for investigating historic and prehistoric buildings, the relations between building components, functional spaces, topological relations and construction phases through time and space; harmonized with CRMarchaeo	New extension, v1.4, April 2016
○ ARIADNE Reference Model: CIDOC CRM + set of new or enhanced extensions	ARIADNE Reference Model, v1.0, April 2016

The ARIADNE Reference Model is intended to allow the accurate documentation of complex entities and relations of archaeological/scientific observations and analysis, data integration and search, involving reasoning over the distributed data and knowledge. This however depends on the interest of data providers to map their databases to relevant parts of the conceptual reference model, which some ARIADNE partners have already done and others are considering (ARIADNE 2016a).

CRM mapping tool

A new tool, the Mapping Memory Manager (3M)[229] has been developed by ARIADNE partner Foundation for Research and Technology Hellas, Institute of Computer Science (FORTH-ICS, Greece) to facilitate the mapping of databases to the extended CIDOC CRM and the validation of the mapping; mappings can be exported in CRM compliant RDF. The mapping process is supported by the X3ML Mapping Framework that ensures the integrity and preservation of the "meaning" of the initial data (Minadakis *et al*. 2016).

Mapping of databases

Several partner databases (DB schemas) have been mapped with the 3M tool to relevant parts of the extended CIDOC CRM. Some of the mappings have been used in pilot applications which demonstrate advantages of the extended CRM (see below). The following three examples illustrate representative mappings:

dFMRÖ - Digitale Fundmünzen der Römischen Zeit in Österreich (Digital Coin-finds of the Roman Period in Austria)[230]: The dFMRÖ is a relational database of pre-Roman and Roman Imperial period coins found in Austria and Romania (75,565 records of coin finds), developed by the Numismatics Research Group at the Austrian Academy of Sciences. The database schema of the dFMRÖ was mapped to CIDOC CRM, using also the CRMdig extension and a specialized extension for coins covering the need to map categorical information (Doerr *et al*. 2016). The database provided a good

[229] Mapping Memory Manager - 3M (FORTH-ICS), http://www.ics.forth.gr/isl/3M

[230] dFMRÖ - Digitale Fundmünzen der Römischen Zeit in Österreich (ÖAW Numismatic Research Group), http://www.oeaw.ac.at/antike/index.php?id=358

example for mapping of a large class of well-defined traditional databases where there is a need to address and separate both categorical and factual information. Results have been employed together with other datasets in the coins demonstrator.

Athenia Agora excavation database: This database (over 280,000 data items) presented a case of highly contextualized research data. The most relevant parts of the database schema were mapped by a researcher of the German Archaeological Institute to CIDOC CRM, using the extensions CRMarchaeo and CRMsci. The mapping results have been used together with other datasets in the sculptures demonstrator.

SITAR - Archaeological Territorial Informative System of Rome[231]: The SITAR system manages different types of data sets including information about monuments, archaeological finds, survey and conservation work, archival documents, bibliographic references and others. A mapping between the SITAR database schema and the concepts of CIDOC CRM and CRMarchaeo has been carried out by the ARIADNE partner Italian Ministry of Cultural Assets and Activities (Central Institute for the Union Catalogue) in cooperation with domain experts of the Soprintendenza Speciale per il Colosseo, il Museo Nazionale Romano e l'Area Archeologica di Roma, and the Department of Computer Science of the University of Verona.

Also the ACDM model of the ARIADNE data registry/catalogue has been mapped to the CIDOC CRM and a set of integrated queries implemented in order to validate the adequacy of the models. This mapping is being used to support data integration both at the catalogue and at the item level. The enhanced capability provided by the ARIADNE Reference Model is being demonstrated in item-level pilot applications.

7.6 Demonstrators using CRM-based Linked Data

Three pilot applications are being developed to demonstrate the capability of the extended CRM to support Linked Data use cases of item-level data integration, discovery and access. The demonstrators concern different objects (coins, sculptures, wooden material) and are implemented by different partners. It is planned to integrate the pilot demonstrators in the ARIADNE data portal, including a menu of exemplar queries for portal users.

The coins demonstrator

The pilot application has been led by FORTH-ICS and demonstrated the item-level integration process of information about coins from five datasets based on the extended CIDOC CRM, Nomisma ontology (numismatics vocabularies)[232] and Art & Architecture Thesaurus (Felicetti, Gerth *et al*. 2016). The demonstrator employed the core CIDOC CRM, the extension CRMdig and a small coin-specific extension modelling categorical information.

The following datasets have been used in the demonstrator:

- dFMRÖ - Digitale Fundmünzen der Römischen Zeit in Österreich (Digital Coin-finds of the Roman Period in Austria), online MySQL database (source: Numismatics Research Group at the Austrian Academy of Sciences);

[231] SITAR - Sistema Informativo Territoriale Archeologico di Roma, http://www.archeositarproject.it

[232] Nomisma ontology, http://nomisma.org/ontology

- MuseiD-Italia documentation of several coins collections of Italian museums integrated in CulturaItalia (source: Italian Ministry of Cultural Assets and Activities - Central Institute for the Union Catalogue);
- A subset of numismatics records (1670) from the Fitzwilliam Museum (Cambridge) database prepared in the COINS project (COINS - Combat On-line Illegal Numismatic Sales, 2007-2009, see Jarrett *et al*. 2011; COINS was led by PIN-VastLab, the Coordinator of the ARIADNE project);
- Coins data records (630) from the Soprintendenza Archeologica di Roma (SAR) database – prepared in the COINS project;
- Documentation of coin finds (517) in the iDAI.field research database of the Pergamon project, with detailed information about the archaeological context (source: German Archaeological Institute).
- Natural Language Processing techniques were employed by University of South Wales (Hypermedia Research Group) to extract numismatic information from a sample set of six reports from the ADS Grey Literature library to demonstrate the potential of NLP for data integration. The resulting data was expressed in the same CIDOC CRM, AAT and Nomisma form used for the datasets. It was successfully integrated into the FORTH-ICS demonstrator and it was found that the NLP techniques had identified items from the report text not explicitly mentioned in the site record metadata.

The demonstrator aimed at item-level integration of the diverse coin datasets in an environment where users can effectively query and receive combined results coming from the different datasets. To enable such a search environment four of the datasets were mapped with FORTH-ICS' Mapping Memory Manager (3M) to the ARIADNE Reference Model and transformed to RDF format; the MuseiD-Italia data was already in CIDOC-CRM RDF form, compatible with the ARIADNE Reference Model. In addition mapping of terms in dataset records to the Art & Architecture Thesaurus (AAT) and Nomisma ontology (both available as Linked Data) was necessary to enable integrated searching of the coins documentation.

The pilot application employs the Blazegraph RDF graph database[233] and the user interface is based on the Metaphacts platform[234]. The platform implements the Fundamental Categories and Relationships for intuitive querying CIDOC CRM based repositories, described in Tzompanaki & Doerr (2012). Users can formulate queries by selecting from six basic categories and the relations between them without the need to be familiar with the underlying schema. The results of the queries are coming from the different datasets, and it is possible to refine the search with a facet view.

The coin demonstrator has shown that datasets of different origin, language, property, and of heterogeneous information can be successfully integrated by relying on the CIDOC CRM. The relative homogeneity of the coin class of objects has made the mapping and conversion work relatively easy. But validity of the methodological approach can be assumed for any type of archaeological object.

The sculptures demonstrator

This demonstrator has been developed by researchers of the German Archaeological Institute (Gerth *et al*. 2016a/b). The researchers produced and explored a dataset of semantic data from five different databases based on the CIDOC CRM, including the extensions CRMsci and CRMarchaeo for describing scientific data acquisition and archaeological excavation processes. Furthermore the demonstrator used the object-oriented version of Functional Requirements for Bibliographic Records

[233] Blazegraph, https://www.blazegraph.com

[234] Metaphacts, http://www.metaphacts.com

(FRBRoo)[235] for describing bibliographical records and the Basic Geo vocabulary[236] for simple geometry description. The researchers developed a prototypical implementation of the different standards for archaeological research regarding time, space, actors, literature and other entities covered by domain-specific vocabulary.

The following datasets have been used in the demonstrator:

- German Archaeological Institute: Arachne[237] and data from the iDAI.field instance of the Chimtou project[238],
- British Museum: Semantic Web Collection Online[239],
- Oxford Roman Economy Project: Stone Quarries Database[240],
- American School of Classical Studies in Athens: Athenian Agora Excavation data[241].

The pilot application presents a case of integration of various datasets with different origins (museum catalogue, object database, excavation database, research results). The data resources are provided with different services and interfaces and therefore required a novel strategy for integration, based on CIDOC CRM. The data of the British Museum could be accessed directly via its SPARQL endpoints and integrated by using a SPARQL federated query; the British Museum has the data already organised based on CIDOC CRM. Arachne's data could be exported via an OAI-PMH interface, which provides RDF/XML using CIDOC CRM. The other data exports were transformed to XML and imported into FORTH-ICS' Mapping Memory Manager. The 3M editor was used to describe the datasets with CIDOC CRM and transform the data into RDF format.

To enable a unified search environment for all datasets it was also necessary to harmonize differing CIDOC CRM mappings as well as map terms to a common reference vocabulary, e.g. archaeological terminology to the AAT and places to the iDAI.gazetteer.

The Linked Data has been stored in a Blazegraph graph database (triple store) to perform archaeologically relevant SPARQL queries on the data to showcase the possibilities of the approach. The search interface has been implemented with Metaphacts on top of the Blazegraph triple store and allows accessing the data in a wiki system.

An object-centric and a sites-based view into the cloud of archaeological linked data have been explored. The research questions in the object-centric view concerned comparable objects by applying the same parameters. For example one object-centric query was about a fragmentary head of a Satyr that was found in Chimtou. The sites-based view concerned quarries, for example quarries where white marble was produced. Here search questions were about all possible sculptures from a specific quarry (Pentelli), and literature that describes objects which are made out of the marble of that quarry. The approach demonstrated the advantages of the extended CIDOC CRM for research as queries to answer archaeological questions could be run successfully over to integrated datasets.

[235] FRBRoo model, v2.1, February 2015, http://www.cidoc-crm.org/frbr_drafts.html

[236] Basic Geo (WGS84 lat/long) Vocabulary, https://www.w3.org/2003/01/geo/

[237] Arachne, the central object database of the German Archaeological Institute and the Archaeological Institute of the University of Cologne, http://arachne.uni-koeln.de

[238] Deutsches Archäologisches Institut, Simitthus / Chimtou (Tunesien) Projekt, http://www.dainst.org/projekt/-/project-display/33904

[239] British Museum - Semantic Web Collection Online, http://collection.britishmuseum.org

[240] Oxford Roman Economy Project (Oxford University): Stone Quarries Database http://oxrep.classics.ox.ac.uk/databases/stone_quarries_database/

[241] Agora Excavations (American School of Classical Studies in Athens), http://agora.ascsa.net

The wooden material demonstrator

The wooden material demonstrator is being developed by University of South Wales (Hypermedia Research Group) in collaboration with ADS, DANS and SND. It aims to investigate the potential for Natural Language Processing information extraction techniques to achieve a degree of semantic interoperability between archaeological datasets and the textual content of grey literature reports. Thus the aim is to extract more specific information from the reports than is available in the metadata alone. Similar NLP methods will be employed to those used in the Coins demonstrator described above. The work builds on the techniques developed for the UK STAR Project (Tudhope *et al*. 2011b; Vlachidis et al. 2015). Output will be expressed as RDF using the same CIDOC CRM model as used for the Coins Demonstrator with mappings made to the AAT.

The case study has a broad theme relating to wooden material including shipwrecks, with a focus on indications of types of wooden material, samples taken, wooden objects with dating from dendrochronological analysis, etc. The work is ongoing and will be reported in the forthcoming ARIADNE deliverable D15.3 (ARIADNE 2017b). The intention is to draw on both English and Dutch language datasets and grey literature reports, together with Swedish archaeological reports. The end result will be a SPARQL pilot demonstrator of the technical possibilities, operating over a Linked Data expression of the output, which will offer cross search over both the datasets and text reports. It is intended that the demonstrator will explore possibilities for a more (archaeology) user-centred application interface (using the 'widget' techniques developed in the SENESCHAL project) than a plain SPARQL endpoint.

7.7 Brief summary and lessons learned

Brief summary

The developmental ARIADNE Linked Data work described in this chapter has focused on the production of (and support for) SKOS subject vocabularies, mappings between those vocabularies and the Art & Architecture Thesaurus, in order to provide a multilingual capability, and the mappings of datasets to the CIDOC-CRM. Furthermore three advanced case studies with demonstrators are presented that generate and use Linked Data based on the CIDOC CRM and key subject vocabulary hubs: coins, wooden material and sculptures.

The first two case studies involve information extraction from text reports in addition to mapping datasets, while the third explores external linking beyond the immediate ARIADNE datasets. Exploratory work on mining of Linked Data and NLP techniques are described but both are research areas with potential for much further work. The transformation of the metadata of the datasets registered in the ARIADNE data catalogue to Linked Data is described in the next chapter, as are the details of the ARIADNE Linked Data service.

The demonstrators are still being finalised at the time of this deliverable but will be available for general use via the ARIADNE Portal. For the reasons discussed in the early chapters, the case studies are experimental investigations of the future use cases that are afforded by Linked Data technology; they result in (working) research demonstrators rather than actual operational systems. They illustrate the kinds of possibilities for cross search and the semantic integration of diverse kinds of datasets and text reports that Linked Data and the related semantic technologies make possible.

One obvious finding from the experience to date is the critical importance of the subject vocabularies (e.g. the AAT) combined with the CIDOC CRM ontology entities, which act as linking hubs in the web of data. More work is needed on the identification of further linking hubs and consequent semantic enrichment of the Linked Data to relevant external datasets. One example of a potential linking hub is the Period0 set of cultural periods which can be used by providers of various archaeological and other cultural heritage datasets.

Necessary for the widespread uptake of the Linked Data approach is the availability of a variety of mapping and alignment software for different contexts, together with evaluative studies and guidelines as to their use. Beyond that, to motivate user organisations to devote scarce resources to working with Linked Data, some exemplar working applications are needed that address a real user (scientific/research) need. Such applications should offer a user interface that is easy and attractive to work with, one that does not require programming skills or detailed knowledge of the underlying data schema or ontology structure.

It should not necessarily be assumed that the end-application directly operates over a (Linked Data) triple store. There are advantages in doing so for data updates and external connections and it is an obvious route. However, periodic harvesting of Linked Data is a possibility for applications that have reasons to employ a wider range of programming platforms. Another possibility is for Linked Data providers to consider exposing programmatic web services for application developers (in addition to a SPARQL endpoint), assuming that an appropriate set of of use cases for the services can be identified.

Lessons learned

- *Mapping of datasets to established domain KOSs (in our case CIDOC CRM, AAT and others) allows their integration within and beyond the catalogue of a data portal.*
- *State-of-the-art linking hubs will play an increasingly important role in the web of LOD, comprehensive domain thesauri as the AAT as well as specialised vocabularies like the Nomisma thesaurus.*
- *The mapping of datasets to such hubs requires domain knowledge, easy to use tools, and guidance of users who carry out such work for the first time. While recommender tools are helpful, fully automated mapping appears unlikely to achive quality results at the current time.*
- *The ARIADNE portal and pilot demonstrators show that this work is worth the effort. But there is still a way to go before advanced uses of LOD will become applicable and beneficial in online research environments; more effort must be invested to make this happen.*
- *There is much scope to explore the utility of LOD in practice, taking account of the objectives and requirements of different user communities. The best ways to provide and employ LOD will largely depend on their specific contexts (museum collections, data archives or research platforms, for instance), together with the anticipated use cases. In order to motivate user organisations to work with Linked Data, exemplar working applications that address a real user (scientific/research) need would be very helpful.*

8 ARIADNE LOD Cloud

8.1 The ARIADNE LOD Cloud – in brief

The ARIADNE Linked Open Data Cloud (ALDC) is a web of data that encompasses relevant vocabulary parts of the wider LOD cloud, such as the CIDOC CRM, Art & Architecture Thesaurus (AAT), national and other vocabularies as well as instance data of archaeological and other cultural heritage datasets. The core linking "hubs" are the CIDOC CRM and AAT as they are the main vehicles for linking to/from the ARIADNE catalogue metadata.

The ARIADNE metadata repository is an integrated semantic network, an aggregation of the data produced through the process of mapping and transformation of each data provider's source database to the common target ARIADNE Catalogue Data Model (ACDM). Furthermore the ACDM has been mapped to the CIDOC CRM to enable applications that employ catalogue information and item level information of various datasets, for example sets of Linked Data with CIDOC CRM mapping of the pilot demonstrators. The various Linked Data generated in the project, including links to external resources, is brought together in a Linked Data graph database which forms the basis of the ARIADNE LOD Cloud (ALDC). The database content is accessible via a SPARQL endpoint to internal and external application developers.

There are several reasons for bringing together all the available data in the ALDC:

- *Shareability:* By using *de facto* standards such as those promoted by the W3C under the umbrella of the Semantic Web, the data in the ARIADNE information space are made universally accessible from a unique point.
- *Interoperability:* By using CIDOC CRM the data in the ARIADNE information space are made as interoperable as possible. Coupled with the technical interoperability supported by the Semantic Web languages (RDF, RDFS, SKOS), this semantic interoperability provides maximum re-usability.
- *Scientific discovery:* Besides the two reasons above, the ALDC represents an attempt of bringing together several kinds of archaeological data, related by subject, temporal and geo-spatial overlapping. These data *potentially* enable scientists to address research questions that could not be addressed based on the individual resources. As will be discussed in due course, this potential is being explored to see whether it can *actually* provide new scientific knowledge.

It must be stressed that the current ALDC is the initial stage of an information space that is expected to grow in terms of data, vocabularies, services and users. The role of the ARIADNE project has been to set up this information space and to endow it with a first portfolio of valuable data, vocabularies and services. But, if really successful, the ALDC will never be completed. Rather, it will continue to grow and evolve, reflecting the growth and the evolution of Linked Data generation and usage by the archaeological research and data management community.

The next sections are organised as follows: First the ALDC architecture is introduced, highlighting the logical components that make up the overall system. Each component is then described in the subsequent sections, emphasizing the content of the component in terms of data, vocabularies and mappings. Furthermore the strategy followed to make the ALDC discoverable on the web is presented. The final section summarises and provides some lessons learned in the work on the ARIADNE LOD Cloud.

8.2 Architecture

Figure 1 presents the architecture of ARIADNE LOD Cloud (ALDC) in a simplified, diagrammatic form:

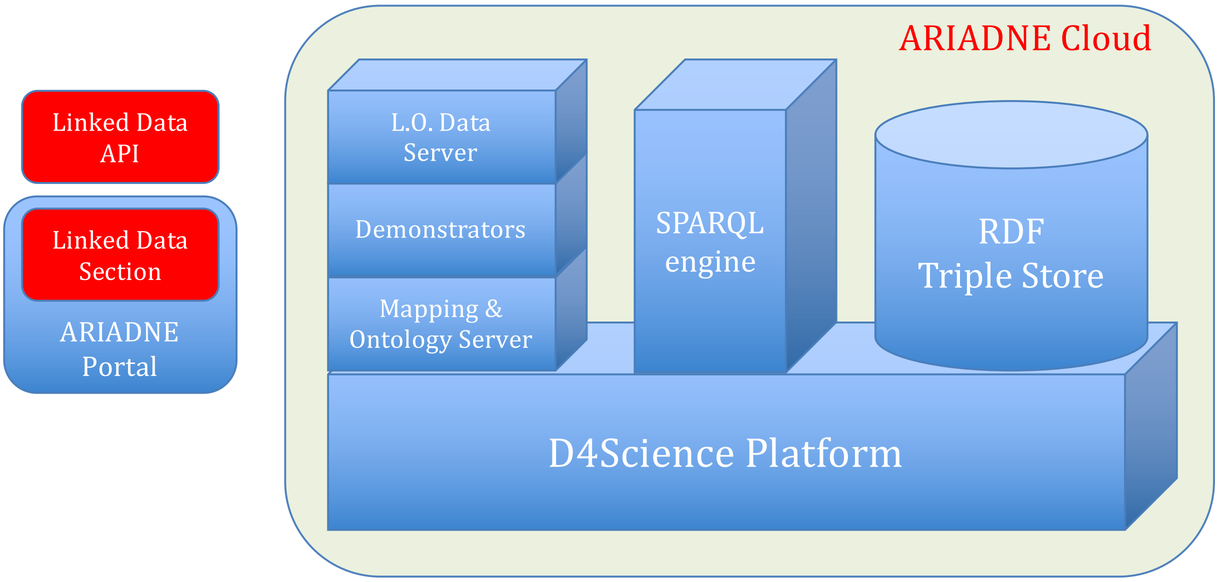


Figure 1: Architecture of the ARIADNE LOD Cloud system

The architecture is shown within the largest box labelled “ARIADNE Cloud”. It comprises of hardware and software components that together realize the ALDC. The services of the ALDC can be accessed in two different ways, indicated in the Figure by the boxes outside the “ARIADNE Cloud”:

- Humans can use the *Linked Data Section* of the ARIADNE Portal, which enables them to obtain vocabularies and mappings, use the CIDOC CRM based Linked Data demonstrators, and access data via a SPARQL interface;
- Software agents can use the *Linked Data API* to issue SPARQL queries against the underlying triple store, thereby obtaining the requested data in one of the formats supported.

The architecture of the ALDC consists of the following components:

- *D4Science Platform:* The D4Science Platform is a hybrid data infrastructure offering services to support the activity of researchers. At present it connects 2500+ researchers in 44 countries, integrating over 50 heterogeneous data providers. With 99.7% service availability it provides access to over a billion records in repositories worldwide and executes over 13,000 models & algorithms per month. In the context of ARIADNE, the platform is being used for running the semantic technologies that support the ALDC (triple store and SPARQL Engine). It also relieves the ALDC developers from the burden of implementing low-level services such as authentication, memory management, security and the like. In addition, the platform allows easy installation, configuration, management and operation of the Demonstrators. Finally, it offers a distributed and scalable file system, accessible through a user-friendly interface, for hosting and accessing data that are not ingested in the triple stores, such as mappings.
- *SPARQL engine and RDF triple store:* The semantic technologies employed by the ALDC are a SPARQL engine and an RDF triple store operated by the SPARQL engine. These are deployed on a virtual machine installed on and operated by the D4Science platform. The triple store hosts the datasets included in the ALDC, along with the ontologies defining the classes and properties used

in these datasets. The technology employed for these two components is the Virtuoso Universal Server, in its open-source edition[242] and the Blazegraph graph database[243].

- The services for the users of the ALDC, whether humans or software agents, are offered by the following components:
  - *Linked Open Data Server:* Provides access to the ARIADNE Linked Data which comprises of ARIADNE catalogue data (based on the ACDM, which is also mapped to the CIDOC CRM) and data of the Demonstrators (see below). The server is technically implemented as a SPARQL endpoint, endowed with a programmatic and an end-user interface. Both interfaces receive SPARQL queries, execute those queries against the underlying SPARQL Engine, and return the results to the user in the appropriate format, depending on the selected access channel.
  - *Demonstrators:* Exemplify the capability of Linked Data based item-level data integration to support answering archaeological research questions. They represent three different subject areas of archaeology: coins, sculptures and wooden material. For each a number of datasets have been integrated based on mappings to the CIDOC CRM (and recent extensions) and use of other domain vocabularies.
  - *Mapping and Ontology Server:* Is a file system-like interface for browsing and downloading the mappings and the ontologies involved in the ALDC. This interface is exclusively for human users and accessible from a Virtual Research Environment implemented on top of the D4Science platform. The interface is being provided for the sole purpose of browsing and accessing mappings and ontologies, while the service for discovering such resources is offered by the Linked Open Data Server.

A detailed description of the contents of each component is given below.

From a technical point of view, the ALDC architecture includes many other components, required for the proper operations of those listed above. The D4Science platform itself includes dozens of open source components, which are integrated into the platform. But these components are not shown as they implement internal services not directly perceived by the users and as such outside of the scope of this presentation.

8.3 The Linked Open Data Server

The ARIADNE Linked Open Data Server runs a large RDF dataset, consisting of several RDF graphs, each corresponding to an archaeological dataset. All graphs are expressed in the vocabulary of the CIDOC CRM, including recent extensions of the ontology. The main datasets (graphs) are the dataset of the ARIADNE Catalogue records and the datasets of the Demonstrators.

ARIADNE Catalogue dataset

- This dataset contains the data of all catalogue records, expressed in RDF and based on two different vocabularies: the ARIADNE Catalogue Data Model (ACDM) and the CIDOC CRM. The ACDM-based records describe the data resources that are being made accessible by the ARIADNE data providers through the ARIADNE Portal. These descriptions have been directly imported from the MORe data aggregation infrastructure supporting the ARIADNE Catalogue service. The CRM-based versions of the descriptions have been generated by first creating the ACDM to CRM

[242] https://virtuoso.openlinksw.com

[243] https://www.blazegraph.com/product/

mappings and then applying those mappings to the ACDM-based descriptions. The CRM-based descriptions have been produced to enable a higher data interoperability, as is demonstrated by one of the demonstrators in the ALDC (see the Coins demonstrator below).

- In addition to the ACDM/CRM-based descriptions of the catalogue records there are descriptions of datasets resulting from the item-level integration of datasets generated and used by the Demonstrators; these descriptions are also expressed in ACDM-CRM.

ARIADNE Demonstrators datasets

In addition to the catalogue-level data, the Linked Open Data Server includes the datasets of the Demonstrators. Here we feature only the datsets of the three main Demonstrators (Coins, Sculptures, Wooden Material), which are briefly described in the next section. Descriptions of other demonstrators, and the datasets used by them, are given in the D14.2 Pilot Deployment Experiments.

- *Coins demonstrator:* This dataset results from the item-level integration of information about coins from five datasets based on the CRM, Nomisma ontology, and Art & Architecture Thesaurus. The demonstrator employs the core CRM, the extension CRMdig and a small coin-specific extension modelling categorical information. The integrated datasets are:
    - dFMRÖ - Digitale Fundmünzen der Römischen Zeit in Österreich (Digital Coin-finds of the Roman Period in Austria), is a relational database of pre-Roman and Roman Imperial period coins found in Austria and Romania (75,565 records of coin finds), developed by the Numismatics Research Group at the Austrian Academy of Sciences;
    - MuseiD-Italia documentation of several coins collections of Italian museums integrated in CulturaItalia;
    - A subset of numismatics records (1670) from the Fitzwilliam Museum (Cambridge) database from the COINS project (2007-2009, led by PIN);
    - Coins data records (630) from the Soprintendenza Archeologica di Roma (SAR) database, also from the COINS project;
    - Documentation of coin finds (517) in the iDAI.field research database of the Pergamon project, with detailed information about the archaeological context;
    - The result of knowledge extraction using Natural Language Processing methods from a collection of textual documents about coins.
- *Sculptures demonstrator:* A set of data from five different databases based on the CRM, CRMsci and CRMarchaeo, using the Basic Geo vocabulary and the object-oriented version of Functional Requirements for Bibliographic Records (FRBRoo) for describing bibliographical records. The dataset comprises of sculptures data from:
    - British Museum: Semantic Web Collection Online (is mapped to the core CRM and includes links to BM vocabularies), was accessed directly via its SPARQL endpoints and integrated by using a SPARQL federated query;
    - Arachne, data exported via an OAI-PMH interface, which provides RDF/XML using CIDOC-CRM;
    - iDAI.field database of the Chimtou project, transformed to XML and imported into FORTH's 3M tool, described with CIDOC-CRM and transformed to RDF;

- Oxford Roman Economy Project: Stone Quarries Database, RDF generation as above;
- Athenia Agora excavation DB (over 280,000 data items), mapped using the extensions CRMarchaeo and CRMsci; the most relevant parts of the database schema have been mapped to CRM, also using CRMarchaeo and CRMsci.

o *Wooden Material demonstrator:* A dataset with a broad theme relating to wooden material including shipwrecks, with a focus on indications of types of wooden material, samples taken, wooden objects with dating from dendrochronological analysis, etc. The data has been extracted from archaeological datasets and grey literature reports in different languages and expressed using the CIDOC CRM and mappings made to the AAT. The integrated datasets are:
  - Digital Collaboratory for Cultural Dendrochronology (DCCD) dataset, an extract of the international DCCD database facilitated by DANS;
  - Dendrochronology Database of the Vernacular Architecture Group (UK), 2016. Archaeology Data Service (doi: 10.5284/1039454);
  - Cruck Database of the Vernacular Architecture Group (UK), 2015. ADS (doi: 10.5284/1031497);
  - Newport Medieval Ship. N. Nayling (Univ. Wales Trinity St David) & T. Jones (Newport Museums and Heritage Service), 2014. ADS (doi: 10.5284/1020898);
  - Mystery Wreck Project (Flower of Ugie). Hampshire and Wight Trust for Maritime Archaeology, 2012. ADS (doi: 10.5284/1011899);
  - Data extracted via NLP from 25 archaeological grey literature reports in Dutch, English and Swedish (reports provided by ADS, DANS and SND).

The rationale for uniting all datasets, the datasets of the ARIADNE Catalogue, the main Demonstrators and others in the ARIADNE LOD Cloud is twofold: the accessibility of the LOD datasets from a single source is clearly an advantage for researchers, and there is the ambition of supporting research questions in archaeology that could not be addressed based on individual collections. The Demonstrators are first experiments on the discovery of knowledge across several different datasets; the experimentation is ongoing.

Connections

There exist several connections amongst the Linked Data graphs addressed above. All Catalogue-level data are expressed in the same vocabularies (ACDM, CIDOC CRM), and link to the same external Linked Data vocabularies. This includes the SKOS version of the Art & Architecture Thesaurus (AAT) which is employed as the backbone of the ARIADNE subjects terminology “hub”. Other thesauri in SKOS format are involved through the mapping of terms used in data provider records to the AAT, for example, the multi-lingual PACTOLS thesaurus and Historic England thesauri. Figure 2 presents an ACDM based Catalogue-level description of a coin dataset using AAT concepts.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description
rdf:about="http://schemas.cloud.dcu.gr/#acdm:ariadne/acdm:ariadneArchaeologicalResource/acdm:dataset">
    ...
   <rdf:Description rdf:about="http://.../acdm:dataset/acdm:ariadneSubject">
```

```
  <rdf:Description rdf:about="http://.../acdm:dataset/acdm:ariadneSubject/acdm:derivedSubject">
    <skos:prefLabel>coins (money)</skos:prefLabel>
    <dc:source>http://vocab.getty.edu/aat/300037222</dc:source>
  </rdf:Description>
</rdf:Description>
<rdf:Description rdf:about="http:// ... /acdm:dataset/acdm:ariadneSubject_2">
  <rdf:Description rdf:about="http://.../acdm:dataset/acdm:ariadneSubject_2/acdm:derivedSubject">
    <skos:prefLabel>archaeological sites</skos:prefLabel>
    <dc:source>http://vocab.getty.edu/aat/300000810</dc:source>
  </rdf:Description>
</rdf:Description>
```

Figure 2: Example of an ACDM-based description of a dataset

All item-level data of the demonstrators are expressed in the CIDOC CRM vocabulary, and link to external vocabularies employed by the demonstrators. For example, terms in coins datasets are linked to the Nomisma thesaurus or toponyms in sculptures datasets are linked to the iDAI.gazetteer. Demonstrators also use external datasets, for example the sculptures demonstrator links to data in the British Museum’s Semantic Web Collection Online.

Catalogue-level and item-level data are linked to each other by employing specific properties of the CIDOC CRM. For example, coin data are linked to ARIADNE catalogue records by adding to each coin a triple linking it to the dataset where the information about the coin belongs. This connection is established through the CRM property *P67i_is_referred_to_by*. The type of the triple that implements the linking between a coin record and an ACDM record is:

The coin (subject):	*E22_Man-Made_Object ->*
The CRM property (predicate)	*P67i_is_referred_to_by ->*
The ACDM record (object)*:*	*E73_Information_Object*

Moreover, NLP results are linked to the coins through terms of the Nomisma.org vocabulary and then to the ARIADNE catalogue records through the links between coins and records as described above.

In this way information in the catalogue dataset is integrated with other datasets (e.g. datasets of coins, wooden material, sculptures, etc.) allowing to query the Linked Data at different levels of information, catalogue information as well as item specific information.

To give some figures of the current ARIADNE LOD Cloud: The dataset of the ARIADNE catalogue has 20+ million RDF triples, the Coins demonstrator 1+ million triples, the Sculptures demonstrator 5+ million triples, and the Wooden Material demonstrator 1+ million triples. The ingested vocabularies amount to 4+ million triples of which the AAT is the largest part. Thus the ARIADNE LOD Cloud at present contains a total of about 32 million triples.

8.4 The Demonstrators

The Demonstrators represent three different subject areas of archaeology, coins, sculptures and wooden material. The datasets that are being employed by the Demonstrators are described above. The datasets have been harmonized, where necessary, using the CIDOC CRM (and recent extensions), transformed into RDF graphs and ingested into the ARIADNE LOD Cloud. The Demonstrators are described in greater detail in the deliverable D14.2 Pilot Deployment Experiments and the deliverable D15.3 Semantic Annotation and Linking.

The Demonstrators will become accessible to end-users through a dedicated Linked Data Section on the ARIADNE Portal. They have been developed to exemplify the capability of Linked Data based item-level data integration to support answering archaeological research questions. This capability builds on the mapping of datasets to the CIDOC CRM (including recent extensions) and other domain vocabularies (i.e. AAT, Nomisma and others). Here we give a brief account of some promising results that have been obtained in demonstrators.

The Coins Demonstrator can illustrate important points that are present also in other demonstrators. The Coins Demonstrator employs datasets of different providers (including results of NLP of archaeological grey literature), mappings to the CIDOC CRM (and CRMdig extension), and other domain vocabularies (AAT, Nomisma). Furthermore it presents a case that shows the potential of querying, in the ARIADNE LOD Cloud, this item-level data together with catalogue-level data.

Queries across the datasets of the Coins Demonstrator show useful results for researchers. Queries that are trivial to be answered by each dataset separately become relevant for a researcher when they are executed across several datasets, and the results combined by the researcher. For example searches such as *Find coins minted in the same place/area, Find coins minted by the same authority (e.g. Antonianus), Find coins produced in the same period (e.g. the same century), Find coins made from specific material (e.g. bronze), etc.* Moreover, item-level and catalogue-level data can be queried simultaneously, e.g. *Find the publishers of all collections that contain bronze antoninianus*.

The Sculptures Demonstrator has the same general characteristic but involves some different aspects. For example, the datasets include data from excavations and instead of grey literature reports the large Zenon bibliographic database of the German Archaeological Institute is involved. Consequently the Sculptures Demonstrator employs the CRM extensions CRMarchaeo and CRMsci and Functional Requirements for Bibliographic Records (FRBRoo), along with other vocabularies (e.g. the AAT and the iDAI.gazetteer). Also this demonstrator shows advanced capability to support answering archaeological research questions. For example, queries over the datasets concerned quarries where white marble was produced, all possible sculptures from a specific quarry, and literature that describes objects which are made out of the marble of that quarry.

The wooden material Demonstrator also shares the general characteristics with a particular focus on the integration of grey literature textual reports in different languages with datasets on a dendrochronological theme. The complexity of the underlying semantic framework based on the CIDOC CRM and Getty AAT is shielded from the user by the Web application user interface. The Demonstrator highlights the potential for archaeological research that can interrogate grey literature reports in conjunction with datasets. Queries concern wooden objects (e.g. *samples of beech wood keels)*, optionally from a given date range, with automatic expansion over hierarchies of wood types.

8.5 The Mapping and Ontology Server

The Mapping and Ontology Server provides information about the mappings and the vocabularies (ontologies, thesauri) involved in the ARIADNE LOD Cloud.

The following mappings of datasets to the CIDOC CRM (and extensions) are available:

- Schemas of the Italian Central Institute for Catalogue and Documentation for archaeological finds (RA) and monuments and complexes (MA/CA) mapped to the CRM, using, where required, more specialised classes and properties of CRM extensions (provided by ICCU);
- Database schema and concepts of SITAR, the Archaeological Territorial Informative System of Rome mapped to the CRM and CRMarchaeo (ICCU in cooperation with other institutions);
- dFMRÖ (coins database) mapped to CRM, CRMdig and a specialized extension for coins, used in the Coins demonstrator (ÖAW);
- iDAI.field database of the Pergamon project mapped to CRM, CRMarchaeo and CRMsci, used in the Coins demonstrator (DAI);
- iDAI.field database of the Chimtou project including stone objects and archaeological contexts, mapped as above and used in the Sculpture demonstrator (DAI);
- Athenia Agora excavation database (over 280,000 data items), mapped as above and used in the Sculptures demonstrator (DAI);
- Digital Collaboratory for Cultural Dendrochronology (DCCD) dataset, an extract facilitated by DANS, mapped to the CRM (USW);
- Dendrochronology Database of the Vernacular Architecture Group (UK), 2016 (doi: 10.5284/1039454), provided by ADS, mapped to the CRM (USW);
- Cruck Database of the Vernacular Architecture Group (UK), 2015 (doi: 10.5284/1031497), provided by ADS, mapped to the CRM (USW);
- Newport Medieval Ship. N. Nayling & T. Jones, 2014 (doi: 10.5284/1020898), dataset provided by ADS, mapped to the CRM (USW);
- Mystery Wreck Project (Flower of Ugie). Hampshire and Wight Trust for Maritime Archaeology, 2012 (doi: 10.5284/1011899), dataset provided by ADS, mapped to the CRM (USW);
- Animal Bone Evidence South England (doi:10.5284/1000102), dataset provided by ADS, mapped to the CRM and extensions and used in an Animal Remains demonstrator (DAI);
- Holozängeschichte der Tierwelt Europas (doi:10.13149/001.mcus7z-2), dataset provided by IANUS, mapped and used as above (DAI).

The following ontologies are available as references:

- CIDOC CRM core. Version 5.0.4, December 2011;
- CRMarchaeo. Model for integrating metadata about the archaeological excavation process; introduces concepts of stratigraphy and excavation. Version 1.4, April 2016;
- CRMsci. Model for integrating metadata about scientific observation, measurements and processed data. Version 1.2.3, April 2016;

- CRMdig. Model of digitisation processes, to encode metadata about the steps and methods of production (“provenance”) of digital representations such as 2D, 3D or animated models. Version 3.2.1, April 2016;
- CRMba. Model for investigating historic and prehistoric buildings, the relations between building components, functional spaces, topological relations and construction phases through time and space; harmonized with CRMarchaeo. Version 1.4, April 2016;
- CRMgeo. Spatio-temporal model that integrates CRM and OGC standards. Version 1.2, February 2015;
- CRMinf. Model for integrating data with scholarly argumentation and inference making in descriptive and empirical sciences; harmonized with CRMsci. Version v0.7, February 2015;
- Functional Requirements for Bibliographic Records, FRBRoo encoded in RDFS. Version 2.4, June 2016.

The following thesauri in SKOS are available as references:

- AAT - Art & Architecture Thesaurus (Getty);
- PACTOLS thesaurus (Peuples, Anthroponymes, Chronologie, Toponymes, Œuvres, Lieux et Sujets) of the Fédération et ressources sur l’Antiquité, France. A large multi-lingual thesaurus which focuses on antiquity and archaeology from prehistory to the industrial age; terms in French, English, German, Italian, Spanish, Dutch, and (some) Arabic). Over 1600 PACTOLS concepts, used by Inrap in their catalogue of archaeological reports (DOLIA), have been mapped to the AAT;
- Historic England thesauri (Forum on Information Standards in Heritage – FISH), thesauri in SKOS provided by HeritageData (SENESCHAL project). ADS, employs five of the thesauri (monuments, components, building-material, maritime-craft, fish objects) of which about 850 concepts have been mapped to the AAT;
- PICO thesaurus (ICCU): A large thesaurus of terms related to culture and cultural heritage (Italian and English) which is being used for the data of CulturaItalia; a number of terms concern archaeology which have been mapped to the AAT;
- Italian Archaeological Finds Vocabulary / Reperti Archeologici (RA) Thesaurus, a thesaurus describing archaeological finds (ICCU);
- RCE Archeologisch Basisregister - ABRr+ thesauri (Rijksdienst Cultureel Erfgoed, Netherlands), about 450 concepts of monument types (Archeologische complextypen) have been mapped by DANS to the AAT;
- Irish Monument Types thesaurus (National Monuments Service), a hierarchical list of concepts expressed in SKOS as part of the LoCloud project;
- iDAI.vocab: group of 14 thesauri of archaeological terminology in different languages and of varied size; the German thesaurus, mapped to the AAT, serves as the central hub to and through which the other thesauri are linked;
- iDAI.Gazetteer: provides over 1 million entries describing modern and ancient places that are of interest to the archaeologists and also acts as a hub by linking other gazetteers like Geonames and Pleiades;
- Dendrochronology multi-lingual vocabulary of the Digital Collaboratory for Cultural Dendrochronology, developed and recently expressed in SKOS by DANS;

- EAGLE epigraphy vocabularies (Material, Type of inscription, Execution technique, Object type, Decoration, Dating criteria, State of preservation);
- Nomisma ontology of numismatic concepts and entities (Nomisma.org).

8.6 Promotion of external use

One of the core principles of Linked Open Data is linking of published datasets to others which generates an expanding and increasingly rich web of Linked Data. Promotion of linking relevant datasets to the ARIADNE LOD by external developers is planned to include documentation of the data in relevant registries, targeted dissemination of information about the available data, and direct discussion with a number of interested developers.

Data registration: Documenting sets of LOD in relevant registries makes it easier for application developers to identify, evaluate and link to relevant datasets. The Vocabulary of Interlinked Data Sets (VoID) is most often being used to describe and register sets of LOD. In VoID a dataset is a collection of data, published and maintained by a single provider, available as RDF, and accessible, for example, through a SPARQL endpoint. Figure 3 illustrates a VoID description of the ARIADNE LOD:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix void: <http://rdfs.org/ns/void#> .
:ARIADNE-LOD  a void:Dataset;
    dcterms:title "ARIADNE registry";
    dcterms:publisher "ARIADNE Project";
    foaf:homepage <http://registry.ariadne-infrastructure.eu>;
    dcterms:description "A registry of data for archaeological research";
    dcterms:license <http://opendatacommons.org/licenses/by/>;
    void:sparqlEndpoint <http://ariadne2.isti.cnr.it/sparql>;
    …
```

Figure 3: VoID description of the ARIADNE registry

The final ARIADNE LOD will be registered in the Data Hub (datahub.io), where also some resources employed by ARIADNE can be found (e.g. the Getty AAT, English Heritage thesauri, and others); other registries and platforms (e.g. Github, Wikidata) are being considered.

Targeted dissemination: Announcements and other information about the available LOD will be disseminated via relevant mailing lists, newsletters etc. of the Linked Data community in the fields of archaeology, cultural heritage, classical studies, history and other humanities.

Direct consultation with developers: A number of Linked Data application developers of institutions and projects will be contacted directly to suggest and discuss interlinking with their or other available datasets in the web of LOD.

8.7 Brief summary and lessons learned

Brief summary

The ARIADNE registry holds metadata of data resources from the content providers. These metadata are being collected and enriched with an aggregator (MORe) and included in the ARIADNE data catalogue. ARIADNE makes the catalogue and other data generated in demonstrators available as Linked Open Data (LOD); thereby the ARIADNE LOD can become part of a web of Linked Data of archaeological and related other information resources.

This work within ARIADNE involved the use of a suitable RDF store and graph database for the Linked Data generation and linking efforts. The project has experimented with two such technologies, Virtuoso and Blazegraph, to perform archaeologically relevant SPARQL queries on the generated Linked Data, and to allow updates of datasets using the SPARQL 1.1 Graph Store HTTP Protocol. Based on this preliminary work, a scalable implementation that can efficiently support the publication and use of the ARIADNE LOD has been designed and realized to offer three different services: the Linked Open Data Server, the Demonstrators, and the Mapping and Ontology Server.

The Linked Open Data Server provides access to a large RDF dataset, which comprises of several graphs of archaeological datasets and can be queried via a SPARQL endpoint. The Demonstrators have been developed to exemplify the capability of Linked Data based item-level data integration to support answering archaeological research questions. They represent three different subject areas of archaeology: coins, sculptures and wooden material. For each a number of datasets have been integrated based on mappings to the CIDOC CRM (and recent extensions) and use of other domain vocabularies. The Mapping and Ontology Server provides information about the mappings and the vocabularies (ontologies, thesauri) involved in the ARIADNE LOD Cloud.

The current ARIADNE LOD Cloud is just the initial stage of an information space that is expected to grow in terms of data, vocabularies, services and users. Experiments to exploit the ARIADNE LOD have just started, with promising results as shown by the Demonstrators. Planned future work will aim to proceed with linking the available Linked Data to relevant other datasets. To promote interlinking, the ARIADNE LOD will be announced via relevant mailing lists, newsletters etc. of the Linked Data community in the field of archaeology and cultural heritage. A number of Linked Data developers will also be contacted directly to suggest and discuss interlinking with their or other available datasets in the web of LOD.

Lessons learned

While the Linked Open Data standards are essential for integrating data, the technology supporting such integration is still in its infancy. The ARIADNE LOD, comprising of LOD of the ARIADNE catalogue, three demonstrators and various vocabularies sum up to about 32 million RDF triples. While any relational database can easily handle millions of records, the corresponding amount of RDF in a current triple store can cause serious efficiency problems as experienced in the experimentation with the ARIADNE Linked Data Cloud. It is becoming apparent that this is the price to be paid to have interoperability. More robust and efficient graph databases are required if we want to proceed towards Big Data as Linked Data. This is the first lesson that we have learned while implementing the ARIADNE Linked Data Cloud.

The second lesson comes from the graph data model. This model is intrinsically binary, hence makes it difficult to express higher rank relations, and to easily implement data connection patterns. In the latter case, the patterns may involve data chains that span several arcs, and their definition and implementation is not trivial. Conversely, correlations between data items can be epitomized by such paths, which need to be detected, and this is a computationally very intensive task if the length of the paths go beyond 2-3 arcs. This fact has always been known from a theoretical point of view, but working with real data we could experience it in practice.

9 References and relevant other sources

5 ★ Open Data (details Berners-Lee's 5-star scheme of Linked Open Data with examples and explains benefits of and some issues in providing such data), http://5stardata.info

Acheson, Phoebe (2014): Linked Open Bibliographies in Ancient Studies. ISAW Paper 7.2, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Agosti M., Conlan O., Ferro N. *et al*. (2013): Interacting with Digital Cultural Heritage Collections via Annotations: The CULTURA Approach. DocEng'13, Florence, Italy, September 10–13, 2013, http://www.digitalmeetsculture.net/wp-content/uploads/2013/12/Interacting-with-Digital-Cultural-Heritage-Collections-via-Annotations.pdf

Agricultural Information Management Standards (AIMS): Vocabularies, Metadata Sets and Tools (VEST) registry: KOS, http://aims.fao.org/vest-registry/vocabularies

AGROVOC Linked Open Data, http://aims.fao.org/standards/agrovoc/linked-open-data

Alexander K., Cyganiak R., Hausenblas M. & Zhao J. (2009): Describing Linked Datasets. On the Design and Usage of voiD, the "Vocabulary of Interlinked Datasets". In: Proceedings of the Linked Data on the Web (LDOW'09) workshop, Madrid, Spain, 20 April 2009. http://ceur-ws.org/Vol-538/ldow2009_paper20.pdf

Allemang D. & Hendler J. (2011): Semantic Web for the Working Ontologist. Effective Modeling in RDFS and OWL Second Edition. Morgan Kaufmann

Almas B., Babeu A. & Krohn A. (2014): Linked Data in the Perseus Digital Library. ISAW Paper 7.3, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Almeida B., Roche C. & Rute C. (2016): Terminology and ontology development in the domain of Islamic archaeology, pp. 147-156, in: Erdman-Thomsen H., Pareja-Lora A. & Nistrup Madsen B. (2016): Term Bases and Linguistic Linked Open Data. TKE 2016 - 12th International conference on Terminology and Knowledge Engineering. Copenhagen Business School, http://openarchive.cbs.dk/handle/10398/9323

Aloia N., Papatheodorou C., Gavrilis D., Debole F. & Meghini C. (2014): Describing Research Data: A Case Study for Archaeology, pp. 768–775, in: Meersman R. *et al*. (eds.): On the Move to Meaningful Internet Systems: OTM 2014 Conferences. Springer (LNCS 8841); preprint, https://www.academia.edu/19889230/Describing_Research_Data_A_Case_Study_for_Archaeology

Amsterdam Museum in Europeana Data Model RDF, http://semanticweb.cs.vu.nl/lod/am

Ancient World Mapping Centre (AWMC / University of North Carolina): Antiquity À-la-carte and public map tiles, http://awmc.unc.edu/wordpress/alacarte/

Anichini F. & Gattiglia G. (2012): MappaOpenData. From web to society. Archaeological open data testing, pp. 54-56, in: Opening the Past: Archaeological Open Data, MapPapers 3-II, http://mappaproject.arch.unipi.it/wp-content/uploads/2011/08/Pre_atti_online3.pdf

Antike Fundmünzen in Europa (web-based coins database developed by the Romano-Germanic Commission of the German Archaeological Institute), http://afe.fundmuenzen.eu

Arbuckle S., Whitcher-Kansa S., Kansa E., Orton D. *et al*. (2014): Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals across Neolithic Turkey. In: PLoS ONE, 9(6): e99845, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0099845

Archaeogeomancy.net (2014): Colonisation of Britain, 30 May 2014, http://www.archaeogeomancy.net/2014/05/colonisation-of-britain/

Archaeology Data Service (2015): ADS / Internet Archaeology Annual Report, 1.8.2014–31.7.2015, http://archaeologydataservice.ac.uk/attach/annualReports/ADS%20Annual%20Report%202014-15.pdf

Archaeology Data Service: Linked Open Data, http://data.archaeologydataservice.ac.uk

Archaeology Data Service: The STELLAR project, http://archaeologydataservice.ac.uk/research/stellar/

Archaeotools - Data mining, facetted classification and E-archaeology (UK, e-Science Research Grant, 2007-2009), http://archaeologydataservice.ac.uk/research/archaeotools

ArcheoInf - Informationszentrum für die Archäologie (Germany, DFG-funded project, 2008-), http://archeoinf.tu-dortmund.de

Archeologisch Basisregister (Rijksdienst Cultureel Erfgoed / Cultural Heritage Agency of the Netherlands), http://abr.erfgoedthesaurus.nl

Archer P., Dekkers M., Goedertier S., Harzard N. & Loutas N. (2013): Study on business models for Linked Open Government Data (BM4LOGD). Study prepared for the ISA programme by PwC EU Services, 23 November 2013, https://joinup.ec.europa.eu/community/semic/document/study-business-models-linked-open-government-data-bm4logd

Archives Hub, http://archiveshub.ac.uk

Archives Hub: LOCAH - Linked Archives and Linking Lives projects (2010-2012), http://locah.archiveshub.ac.uk

ARENA - Archaeological Records of Europe - Networked Access project (2001-2004, and 2009-2010 in the context of DARIAH), http://ads.ahds.ac.uk/arena/search/

ARIADNE - Linked Data SIG (2013): First Meeting, EAA 2013 Conference, Pilsen, 4 September 2013, http://www.ariadne-infrastructure.eu/Community/Special-Interest-Groups/Linked-Data

ARIADNE - Linked Data SIG (2014): Second Meeting, CAA 2014 Conference, Paris, 23 April 2014, http://www.ariadne-infrastructure.eu/Community/Special-Interest-Groups/Linked-Data

ARIADNE (2013): D3.2 Report on Project Standards (November 2013), http://www.ariadne-infrastructure.eu/Resources/D3.2-Report-on-project-standards

ARIADNE (2014a): D2.1 First Report on Users' Needs (April 2014), http://www.ariadne-infrastructure.eu/Resources/D2.1-First-report-on-users-needs

ARIADNE (2014b): Modeling scientific data: workshop report, 12 September 2014, http://www.ariadne-infrastructure.eu/News/Modeling-scientific-data

ARIADNE (2014c): The Way Forward to Digital Archaeology in Europe. November 2014, http://www.ariadne-infrastructure.eu/Media/Files/Ariadne-Booklet-The-Way-Forward-to-Digital-Archaeology-in-Europe

ARIADNE (2015a): D2.2 Second Report on Users' Needs (February 2015), http://www.ariadne-infrastructure.eu/content/view/full/1188

ARIADNE (2015b): D16.1 First Report on Data Mining (March 2015), http://www.ariadne-infrastructure.eu/Resources/D16.1-First-Report-on-Data-Mining

ARIADNE (2015c): D16.2 First Report on Natural Language Processing (May 2015), http://www.ariadne-infrastructure.eu/Resources/D16.2-First-Report-on-Natural-Language-Processing

ARIADNE (2015d): ARIADNE at Linked Pasts: Checking in on the state of the art for Linked Open Data and Cultural Heritage. ARIADNE news, 7 August 2015, http://www.ariadne-infrastructure.eu/News/ARIADNE-at-Linked-Pasts

ARIADNE (2015e): D2.3 Preliminary Innovation Agenda and Action Plan (November 2015), http://www.ariadne-infrastructure.eu/Resources/D2.3-Preliminary-Innovation-Agenda-and-Action-Plan

ARIADNE (2016a): D14.1 Extended CRM (April 2016), http://www.ariadne-infrastructure.eu/Resources/D14.1-Extended-CRM

ARIADNE (2016b): D15.1 Report on Thesauri and Taxonomies (August 2016), http://www.ariadne-infrastructure.eu/Resources

ARIADNE (2017a): D14.2 Pilot Deployment Experiments (January 2017), will be available at http://www.ariadne-infrastructure.eu/Resources

ARIADNE (2017b): D15.3 Report on Semantic Annotation and Linking (January 2017), will be available at http://www.ariadne-infrastructure.eu/Resources

ARIADNE Catalogue Data Model (ACDM), http://support.ariadne-infrastructure.eu

ARIADNE Datasets Registry, http://registry.ariadne-infrastructure.eu

ARIADNE: Ariadne Reference Model (set of CIDOC CRM extensions, including reference document, presentation, RDFS encoding), http://www.ariadne-infrastructure.eu/Resources/Ariadne-Reference-Model

Aroyo L., Hyvönen E. & van Ossenbruggen J. (eds., 2007): Cultural Heritage on the Semantic Web. Proceedings of the workshop co-located with the 6th International Semantic Web Conference, Busan, Korea, http://www.cs.vu.nl/~laroyo/CH-SW/ISWC-wp9-proceedings.pdf

ArSol - Archives du Sol (Soil Archives) project, http://arsol.univ-tours.fr

Arwe, John (2011): Coping with Un-Cool URIs in the Web of Linked Data. Presented at the Linked Enterprise Data Patterns Workshop. Data-driven Applications on the Web, Cambridge, 6 December 2011, http://www.w3.org/2011/09/LinkedData/ledp2011_submission_5.pdf

ASIS&T (2014): Special section Economics of Knowledge Organization Systems. ASIS&T - Bulletin of the Association for Information Science and Technology, 40(4): 13-42, http://asis.org/Bulletin/Apr-14/Bulletin_AprMay14_Final.pdf

Aspöck E. & Geser G. (2014): What is an archaeological research infrastructure and why do we need it? - Aims and challenges of ARIADNE. In: Proceedings of the 18th International Conference on Cultural Heritage and New Technologies (CHNT 18), Vienna, November 2013, http://www.chnt.at/wp-content/uploads/Aspoeck_Geser_2014.pdf

Assaf A. & Senart A. (2012): Data Quality Principles in the Semantic Web. ICSC'12 Proceedings of the 2012 IEEE Sixth International Conference on Semantic Computing; preprint: arXiv:1305.4054 [cs.DL], http://arxiv.org/ftp/arxiv/papers/1305/1305.4054.pdf

AthenaPlus (2013a): First release GLAM sector reference terminologies. Project deliverable 4.1, September 2013, http://www.athenaplus.eu/getFile.php?id=187

AthenaPlus (2013b): Review on Linked Open Data Sources. Project deliverable 4.2, October 2013, http://www.athenaplus.eu/getFile.php?id=190

AthenaPlus (EU, CIP Best Practice Network, 3/2013-8/2015), http://www.athenaplus.eu

Auer S., Bühmann L., Dirschl C. *et al.* (2012a): Managing the Life-Cycle of Linked data with the LOD2 Stack. ISWC 2012 - 11th International Semantic Web Conference, Boston, USA, 11-15.11.2012, http://iswc2012.semanticweb.org/sites/default/files/76500001.pdf (also: http://svn.aksw.org/lod2/Paper/ISWC2012-InUse_LOD2-Stack/public.pdf)

Auer S., Demter J., Martin M. & Lehmann, J. (2012b): LODStats - An Extensible Framework for High-performance Dataset Analytics. Proceedings of the EKAW 2012 – 18th International Knowledge Engineering and Knowledge Management Conference, Galway City, Ireland, 8-12 October 2012, http://svn.aksw.org/papers/2011/RDFStats/public.pdf

Bagosi T., Calvanese D., Hardi J. *et al*. (2014): The Ontop Framework for Ontology Based Data Access (OBDA), pp. 67-77, in: CSWS 2014 -The Semantic Web and Web Science - 8th Chinese Conference, Wuhan, China, 8-12 August 2014, Springer; pre-print, http://www.inf.unibz.it/~calvanese/papers/bago-etal-CSWS-2014.pdf

Barbera N., Meschini F., Morbidoni C. & Tomasi F. (2012): Annotating digital libraries and electronic editions in a collaborative and semantic perspective, pp. 46-57, in: Agosti M. *et al*. (eds): Digital Libraries and Archives. 8th Italian Research Conference (IRCDL 2012), CCIS 354, Heidelberg. Springer, http://dspace.unitus.it/bitstream/2067/2331/1/paper_annotation_last.pdf

BARTOC - Basel Register of Thesauri, Ontologies & Classifications (Basel University Library, Switzerland), http://www.bartoc.org

Basharat A., Abro B., Arpinar I.B., & Rasheed K. (2016): Semantic Hadith: Leveraging Linked Data Opportunities for Islamic Knowledge. In: LDOW2016 - 9th Workshop on Linked Data on the Web, Montreal, Canada, 12 April 2016, http://events.linkeddata.org/ldow2016/papers/LDOW2016_paper_06.pdf

Battenfeld I., Beckmann I., Schultze J. & Türk H. (2009): Unifying Archaeological Databases using Triples [ArcheoInf], pp. 281-284, in: Proceedings of COINFO '09 - Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology, Beijing, China, IEEE, http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5361890

Bauer F. & Kaltenböck M. (2012): Linked Open Data: The Essentials. A Quick Start Guide for Decision Makers. REEP & Semantic Web Company. Vienna: edition mono, http://www.semantic-web.at/LOD-TheEssentials.pdf

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

Beck, Anthony (2010): Dig the new breed, Part III – wrapping it all up. In: Open Knowledge Blog, 11 June 2010, http://blog.okfn.org/2010/06/11/dig-the-new-breed-part-iii-wrapping-it-all-up/

Bedford, Denise (2014): Understanding and Managing Taxonomies as Economic Goods and Services. In: ASIS&T Bulletin, 40(4):15-22, https://www.asist.org/publications/bulletin/apr-14/

Behkamal, Behshid (2014): Metrics Driven Framework for LOD Quality Assessment. ESWC 2014 - The Semantic Web: Trends and Challenges. Lecture Notes in Computer Science 8465, pp. 806-816, http://2014.eswc-conferences.org/sites/default/files/phdpaper_17.pdf

Benefiel R. & Sprenkle S. (2014): Herculaneum Graffiti Project. ISAW Paper 7.4, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Bénel, Aurélien (2015): Semiotic Issues and Perspectives on Modeling Cultural Artifacts Revisiting 1970's French Criticisms on 'New archaeologies', pp. 57-64, in: SWASH 2016 - 1st Workshop on Semantic Web for Scientific Heritage, Portoroz, Slovenia, 1 June 2015, http://ceur-ws.org/Vol-1364/sw4sh-2015.pdf

Bergman, Michael K. (2014): A Decade in the Trenches of the Semantic Web. AI3 weblog, 16 July 2014, http://www.mkbergman.com/1771/a-decade-in-the-trenches-of-the-semantic-web/

Berman M.L., Mostern R. & Southall H. (eds., 2016): Placing Names: Enriching and Integrating Gazetteers. Bloomington: Indiana University Press (Series: The Spatial Humanities), http://www.iupress.indiana.edu/product_info.php?products_id=808056

Berners-Lee T., Hendler J. & Lassila O. (2001): The Semantic Web. In: Scientific American, May 2001, http://www.sciam.com/2001/0501issue/0501berners-lee.html

Berners-Lee, Tim (1998–): Design Issues, http://www.w3.org/DesignIssues/

Berners-Lee, Tim (2006): Linked Data, http://www.w3.org/DesignIssues/LinkedData.html

Bikakis N., Tsinaraki C., Gioldasis N., Stavrakantonakis I. & Christodoulakis S. (2013): The XML and Semantic Web Worlds: Technologies, Interoperability and Integration. A survey of the State of the Art. In: Semantic Hyper/Multimedia Adaptation. Studies in Computational Intelligence, Vol. 418, 319-360, http://www.dblab.ntua.gr/~bikakis/papers/XMLSemanticWebSurvey.pdf

Binding C. & Tudhope D. (2016): Improving Interoperability using Vocabulary Linked Data. In: International Journal on Digital Libraries, 17(1): 5-21; accepted manuscript, http://hypermedia.research.glam.ac.uk/media/files/documents/2015-09-14/IJDL2015-binding-tudhope-P.docx

Binding C., Charno M., Jeffrey S., May K. & Tudhope D. (2015): Template Based Semantic Integration: From Legacy Archaeological Datasets to Linked Data. In: International Journal on Semantic Web and Information Systems, 11(1), 1-29. IGI Global, www.igi-global.com. Posted by permission of the publisher. http://hypermedia.research.southwales.ac.uk/media/files/documents/2015-09-14/tudhope-paper_IJSWIS111.pdf

Binding C., Tudhope D., Vlachidis A. *et al*. (2016): ARIADNE: A Research Infrastructure for Archaeology. In: Journal on Computing and Cultural Heritage (forthcoming).

Binding, Ceri (2010): Implementing archaeological time periods using CIDOC CRM and SKOS, pp. 273-287, in: Aroyo L., Antoniou G., Hyvönen E. *et al*. (eds.): ESWC 2010 - The Semantic Web: Research and Applications. Springer (LNCS 6088); preprint, http://www.researchgate.net/profile/Ceri_Binding/publication/225153456_Implementing_Archaeological_Time_Periods_Using_CIDOC_CRM_and_SKOS/links/0deec536b3f5384be7000000.pdf

Binding, Ceri (2014): 5 star data – achieving the 5th star. NKOS 2014 - 13th European Networked Knowledge Organization Systems Workshop, London, 11 September 2014, https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2014/programme.html

Bio2RDF: Linked Data for the Life Sciences, http://bio2rdf.org

BioPortal (US National Center for Biomedical Ontology, provides access to over 300 biological/bio-medical vocabularies), https://bioportal.bioontology.org

Bizer C., Heath T. & Berners-Lee T. (2009): Linked Data - the story so far. In: International Journal on Semantic Web and Information Systems, 5(3): 1-22; preprint, http://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf

Bizer, Chris (2010): Data Linking, pp. 34-43, in: GRDI2020 - Global Research Data Infrastructures: Towards a 10-year vision for global research data infrastructures, http://www.grdi2020.eu/Repository/FileScaricati/9a85ca56-c548-47e4-8b0e-86c3534ad21d.pdf

Blackwell C. & Crane G. (2009): Cyberinfrastructure, the Scaife Digital Library and classics in a digital age. In: Digital Humanities Quarterly 3(1), http://digitalhumanities.org/dhq/vol/3/1/000035/000035.html

Blackwell C. & Smith D.N. (2014): The Homer Multitext and RDF-Based Integration. ISAW Paper 7.5, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Blumauer, Andreas (2013): The LOD cloud is dead, long live the trusted LOD cloud. In: Semantic-Web.at weblog, (7 June 2013, http://blog.semantic-web.at/2013/06/07/the-lod-cloud-is-dead-long-live-the-trusted-lod-cloud/

Booth, David (2010): Resource Identity and Semantic Extensions: Making Sense of Ambiguity. Semantic Technology Conference, San Francisco, 25-June-2010, http://dbooth.org/2010/ambiguity/paper.html

Bozic B. & Gordea S. (2014): Enhancing the Local Value of Thematic Cultural Tourism. PATCH workshop: The Future of Experiencing Cultural Heritage, part of the IUI 2014 - Int. Conf. on Intelligent User Interfaces, Haifa, Israel, 24-27 February 2014, http://patch2014.files.wordpress.com/2012/07/submission-13-version-of-dec-24-10_08.pdf

Bratková E. & Kučerová H. (2014): Knowledge Organization Systems and Their Typology. In: Revue of Librarianship, 25 (supplementum 2): 1-25, http://oldknihovna.nkp.cz/knihovna142_suppl/1402sup01.htm

Brewster C.A. & O'Hara K. (2004): Knowledge Representation with Ontologies: The Present and Future. IEEE Intelligent Systems, January/February 2004, https://www.inf.unibz.it/~franconi/papers/ieee-intelligent-systems-04.pdf

Brewster C.A. & O'Hara K. (2007): Knowledge representation with ontologies: Present challenges - Future possibilities. International Journal of Human-Computer Studies, 65(7): 563-568, https://www.semanticscholar.org/paper/Knowledge-representation-with-ontologies-Present-Brewster-O%27Hara/69b7951abd61c63bb04636f7f51df8f2675d7417/pdf

Brewster C.A., Iria J., Ciravegna F. & Wilks Y. (2005): The Ontology: Chimaera or Pegasus. Proceedings of the Dagstuhl Seminar on Machine Learning for the Semantic Web, February 2005, http://eprints.aston.ac.uk/83/1/dagstuhl05.pdf

British Museum - Semantic Web Collection Online, http://collection.britishmuseum.org

Buil-Aranda C., Hogan A., Umbrich J. & Vandenbussche P.Y. (2013): SPARQL Web-Querying Infrastructure: Ready for Action? In: The Semantic Web - ISWC 2013: 12th International Semantic Web Conference, Sydney, 21-25 October 2013, Proceedings, Part 2: 277-293, http://aidanhogan.com/docs/epmonitorISWC.pdf

Busch, Joseph A. (2005): Making the business case for taxonomy (September 27, 2005), http://www.taxonomystrategies.com/presentations/BusinessCase.ppt

Byrne G. & Goddard L. (2010): The strongest link: Libraries and Linked Data. In: DLib Magazine, 16(11/12), http://www.dlib.org/dlib/november10/byrne/11byrne.html

Byrne K. & Klein E. (2009): Automatic Extraction of Archaeological Events from Text. CAA 2009 - Computer Applications in Archaeology, Williamsburg, Virginia, USA, http://homepages.inf.ed.ac.uk/kbyrne3/docs/byrneKleinCAA2009.pdf

Byrne, Kate (2006): Tethering Cultural Data with RDF. In Proceedings of the 2006 Jena Users Conference, Bristol, http://homepages.inf.ed.ac.uk/kbyrne3/docs/juc2006.pdf

Byrne, Kate (2008a): Relational Database to RDF Translation in the Cultural Heritage Domain. School of Informatics, University of Edinburgh, May 2008, http://homepages.inf.ed.ac.uk/s0233752/docs/rdb2rdfForCH.pdf

Byrne, Kate (2008b): Having Triplets – Holding Cultural Data as RDF. IACH2008 - Workshop on Information Access to Cultural Heritage, ECDL 2008, Aarhus, Denmark, 18 September 2008, http://homepages.inf.ed.ac.uk/kbyrne3/docs/iach08kfb.pdf

Byrne, Kate (2009): Putting Hybrid Cultural Data on the Semantic Web. Journal of Digital Information (JoDI), 10(6), http://homepages.inf.ed.ac.uk/kbyrne3/docs/jodi09kfb.pdf

CAA Semantic SIG, https://groups.google.com/forum/#!forum/caa-semantic-sig

Cacciotti R. & Valach J. (2015): The MONDIS project Semantic Web and the protection of historic buildings, pp. 307-313, in: Proceedings of Digital Heritage 2015, Granada, Volume 2, http://dx.doi.org/10.1109/DigitalHeritage.2015.7419512

Cacciotti R., Blasko M. & Valach J. (2014): A diagnostic ontological model for damages to historical construction. In: Journal of Cultural Heritage, 16(1): 40-48; preprint, https://www.academia.edu/10541233/A_diagnostic_ontological_model_for_damages_to_historical_constructions

Callou C., Baly I., Gargominy O. & Rieb E. (2011): National Inventory of Natural Heritage website: recent, historical and archaeological data. In: The SAA Archaeological Record, 11(1): 37-40, http://alexandriaarchive.org/bonecommons/archive/files/kroeger_etal_icaz_saa_jan2011_f5bf7cdac2.pdf

Callou C., Baly I., Martin C. & Landais E. (2009): Base de données I2AF: Inventaires archéozoologiques et archéobotaniques de France. In: Archéopages, Issue 26, Juillet 2009, 64-73, http://amenageurs.inrap.fr/userdata/c_bloc_file/13/13661/8449_fichier_pratiques-26.pdf

Callou C., Michel F., Faron-Zucker C., Martin C. & Montagnat J. (2015): Towards a shared reference thesaurus for studies on history of zoology, archaeozoology and conservation biology, pp. 15-22, in: SWASH 2016 – 1st Workshop on Semantic Web for Scientific Heritage, Portoroz, Slovenia, 1 June 2015, http://ceur-ws.org/Vol-1364/sw4sh-2015.pdf

Calvanese D., Liuzzo P., Mosca A., Remesal J., Rezk M. & Rull G. (2016): Ontology-based data integration in EPNet: Production and distribution of food during the Roman Empire, pp. 212–229, in: Mining the Humanities: Technologies and Applications. Engineering Applications of Artificial Intelligence, Volume 51, May 2016; preprint, https://www.semanticscholar.org/paper/Ontology-based-data-integration-in-EPNet-Calvanese-Liuzzo/3fad69e4e6a68f59b769042340c582e3d59d1f0b/pdf

Calvanese D., Mosca A., Remesal J., Rezk M. & Rull G. (2015): A 'Historical Case' of Ontology-Based data Access, pp. 291-298, in: Proceedings of Digital Heritage 2015, Granada, Volume 2; preprint, http://ceipac.ub.edu/biblio/Data/A/0817.pdf

Carlisle P. K., Avramides I., Dalgity A. & Myers D. (2014): The Arches Heritage Inventory and Management System: A Standards-Based Approach to the Management of Cultural Heritage Information. Paper presented at the CIDOC Conference: Access and Understanding – Networking in the Digital Era, Dresden, Germany, 6-11 September 2014. http://archesproject.org/wp-content/uploads/2014/10/I-1_Carlisle_Dalgity_et-al_paper.pdf

Carver G. & Lang M. (2013): Reflections on the rocky road to e-archaeology, pp. 224-236, in: CAA 2012 Southampton, Volume I, Amsterdam University Press, http://dare.uva.nl/cgi/arno/show.cgi?fid=516092

Carver, Geoff (2013): ArcheoInf, the CIDOC-CRM and STELLAR: Workflow, Bottlenecks, and Where do we Go from Here?, pp. 498-508, in: CAA 2012 Southampton, Volume II, Amsterdam University Press, http://dare.uva.nl/cgi/arno/show.cgi?fid=545855

Casarosa V., Manghi P., Mannocci A., Rivero Ruiz E. & Zoppi F. (2014): A Conceptual Model for Inscriptions, pp. 23-40, in: Orlandi S. *et al*. (eds.): Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the First EAGLE International Conference, Paris. Rome: Sapienza Università Editrice, http://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf

Catalogue of Life, http://www.catalogueoflife.org

CATCH Vocabulary and alignment repository demonstrator, http://www.cs.vu.nl/STITCH/repository/

CEIPAC - Centre for the Study of Provincial Interdependence in Classical Antiquity, University of Barcelona, Spain, http://ceipac.ub.edu

Charles V. & Devarenne C. (2014): Europeana enriches its data with the AAT. EDM case study, http://pro.europeana.eu/page/europeana-aat

Charles V., Isaac A., Fernie K. *et al*. (2013): Achieving interoperability between the CARARE schema for monuments and sites and the Europeana Data Model. Proceedings of DC 2013 - International Conference on Dublin Core and Metadata Applications, http://dcevents.dublincore.org/IntConf/dc-2013/paper/view/171/171

Charno M., Jeffrey S., Binding C., Tudhope D. & May K. (2013): From the Slope of Enlightenment to the Plateau of Productivity: Developing Linked Data at the ADS, pp. 216-223, in: CAA 2012 Southampton, Volume I, Amsterdam University Press, http://dare.uva.nl/cgi/arno/show.cgi?fid=516092

Chiarcos C., Lang M. & Verhagen P. (2015): IT-assisted Exploration of Excavation Reports. Using Natural Language Processing in the Archaeological Research Process, pp. 87-93, in: CAA2015 Siena, Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology. Oxford: Archaeopress, http://archaeopress.com/ArchaeopressShop/Public/download.asp?id={77DEDD4E-DE8F-43A4-B115-ABE0BB038DA7}

CIDOC (2012): Statement on Linked Data identifiers for museum objects. CIDOC Annual General Meeting, 2012-06-13, Helsinki, http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/PDF/StatementOnLinkedDataIdentifiersForMuseumObjects.pdf

CIDOC Conceptual Reference Model (CIDOC CRM), http://www.cidoc-crm.org

CIDOC CRM (2015): 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: Overview of CIDOC CRM extensions, http://www.ics.forth.gr/isl/CRMext/

Cimiano P., McCrae J., Rodriguez-Doncel V. *et al*. (2015): Linked Terminology: Applying Linked Data Principles to Terminological Resources, pp. 504-517, in: Proceedings of eLex 2015 - Electronic Lexicography in the 21st century: Linking Lexical Data, Herstmonceux Castle, Sussex, UK, 11-13 August 2015, https://elex.link/elex2015/proceedings/eLex_2015_34_Cimiano+etal.pdf

Cimiano P., McCrae J.P. & Buitelaar P. (2016): Lexicon Model for Ontologies. Final Community Group Report 10 May 2016, https://www.w3.org/2016/05/ontolex/

CLAROS - Classical Art Research Online Services, http://www.clarosnet.org

CLAROS: Data, http://data.clarosnet.org

Consens, Mariano P. (2013): Challenges and Opportunities for the Open Web of Linked Data. WOD'2013 - 2nd International Workshop on Open Data, BNF, Paris (presentation), http://www-etis.ensea.fr/WOD2013/wp-content/uploads/2013/06/Consens-Challenges-and-Opportunities-for-the-Open-Web-of-Linked-Data.pdf

Corcho O., Poveda-Villalón M. & Gómez-Pérez A. (2015): Ontology Engineering in the Era of Linked Data. In: ASIS&T Bulletin 41(4: Special section: Linked Data and the Charm of Weak Semantics), 13-16, http://www.asis.org/Bulletin/Apr-15/Bulletin_AprMay2015.pdf

Coyle, Karen (2012): Linked Data Tools: Connecting on the Web. In: ALA TechSource - Library Technology Reports, 48(4), http://www.alastore.ala.org/detail.aspx?ID=3845

Coyle, Karen (2013): Dublin Core usage in LOD. In: KCoyle weblog, 9 October 2013, http://kcoyle.blogspot.co.at/2013/10/dublin-core-usage-in-lod.html

Creative Commons (CC) licenses, https://creativecommons.org/licenses/

Cripps P. & May K. (2004): To OO or not to OO? Revelations from Ontological Modelling of an Archaeological Information System. In: Proceedings of Computer Applications and Quantitative Methods in Archaeology (CAA), Prato, Italy, 13-17 April 2004, http://proceedings.caaconference.org/files/2004/08_Cripps_May_CAA_2004.pdf

Cripps P., Greenhalgh A., Fellows D., May K. & Robinson D. (2004): Ontological Modelling of the work of the Centre for Archaeology. Technical report, English Heritage - Centre for Archaeology, http://cidoc.ics.forth.gr/docs/Ontological_Modelling_Project_Report%_%20Sep2004.pdf

Cripps, Paul (2014): Colonisation of Britain. In: Geosemantic Technologies for Archaeological Research (GSTAR) weblog, 30 May 2014, http://gstar.archaeogeomancy.net/2014/05/colonisation-of-britain/

Cripps, Paul (2015): Geosemantic Tools for Archaeological Research: GSTAR. Presentation at USW Annual Postgraduate Researchers Presentation Day, 5 May 2015, http://de.slideshare.net/pauljcripps/uswpgr2015-cripps-gstar

Crofts N., Doerr M. & Nyman (2011): Call for Comments - Linked Open Data Recommendation for Museums. CIDOC CRM website, 21 March 2011, http://www.cidoc-crm.org/URIs_and_Linked_Open_Data.html

Cultura Italia: Dati, http://dati.culturaitalia.it

Cuy S., Gerth P. & Förtsch R. (2016): Connecting Cultural Heritage Data: The Syrian Heritage Project in the IT Infrastructure of the German Archaeological Institute, pp. 251-258, in: CAA2015 Siena - Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology. Oxford: Archaeopress, http://archaeopress.com/ArchaeopressShop/Public/download.asp?id={77DEDD4E-DE8F-43A4-B115-ABE0BB038DA7}

D'Andrea, Andrea (2012): Including Links in LinkedData: CIDOC-CRM and the Fourth T. Berners-Lee Rule. VAST 2012 - 13th International Symposium on Virtual Reality, Archaeology, and Cultural Heritage. Brighton, UK, Nov. 19-21, 2012, https://www.academia.edu/4195188/Including_Links_in_Linked_Data_CIDOC-CRM_and_the_Fourth_T._Berners-Lee_Rule

D2R Server: Accessing databases with SPARQL and as Linked Data, http://d2rq.org/d2r-server

D2RQ - Accessing Relational Databases as Virtual RDF Graphs, http://d2rq.org

Damova M. & Dannells D. (2011): Reason-able view of linked data for cultural heritage, pp. 17-24, in: S3T-2011 - Third International Conference on Software, Services and Semantic Technologies. Springer (AISC vol. 101); preprint, https://ontotext.com/documents/publications/2011/S3T-MuseumreasonableView_v7_cameraReady-30Jun.pdf

Damova M., Dannélls D., Enache R., Mateva M. & Ranta A. (2013): Multilingual access to cultural heritage content on the Semantic Web, pp. 107–115, in: Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Sofia, Bulgaria, 8 August 2013, http://www.aclweb.org/anthology/W13-2715

Damova M., Kiryakov A., Simov K. & Petrov S. (2010): Mapping the Central LOD Ontologies to PROTON Upper-Level Ontology, pp. 61-72, in: Proceedings of the 5th International Conference on Ontology Matching, Shanghai, 7 November 2010. CEUR-WS 689, http://ceur-ws.org/Vol-689/om2010_Tpaper6.pdf

DANSlabs: EASY Metadata as Linked Open Data Demo, http://dans-labs.github.io/easy-lod/

DARIAH-DE (2013): Recommendations for Interdisciplinary Interoperability. Project report 3.3.1, V1.0, 15.02.2013, https://dev2.dariah.eu/wiki/download/attachments/14651583/R3.3.1.pdf?version=1&modificationDate=1366904278298&api=v2

DataHub (Open Knowledge Foundation), http://datahub.io

DBpedia (Wikipedia structured information often used in Linked Data projects), http://dbpedia.org

de Boer V. & Leinenga J. (2014): Diepere Maritieme Data. DANS. http://dx.doi.org/10.17026/dans-x8p-mc6a

de Boer V., Van Rossum M., Leinenga J. & Hoekstra R. (2014): Dutch Ships and Sailors Linked Data, pp. 229-244, in: The Semantic Web - ISWC 2014, Springer (LNCS 8796); preprint, http://www.few.vu.nl/~vbr240/publications/deboer_iswc2014_dss_draft.pdf (datasets: http://datahub.io/dataset/dutch-ships-and-sailors)

de Boer V., van Rossum M., Leinenga J. & Hoekstra R. (2015):The Dutch Ships and Sailors Project. In: DHcommons Journal, Issue 1, July 2015, http://dhcommons.org/journal/issue-1/dutch-ships-and-sailors-project

de Boer V., Wielemaker J., van Gent J. *et al*. (2012): Supporting Linked Data Production for Cultural Heritage Institutes: The Amsterdam Museum Case Study. Proceedings of the 9th Extended Semantic Web Conference (ESWC 2012), TPDL conference. Heraklion, Greece. 27-31 May 2012, http://www.few.vu.nl/~vbr240/publications/eswc2012supporting.pdf de Boer V., Wielemaker J., van Gent J. *et al*. (2013): Amsterdam Museum Linked Open Data. In: Semantic Web Journal, 4(3): 237-243, http://www.semantic-web-journal.net/sites/default/files/swj293_2.pdf

De Boer, Victor (2015): Linked Data for Digital History, pp. 5-6, in: SWASH 2016 - 1st Workshop on Semantic Web for Scientific Heritage, Portoroz, Slovenia, 1 June 2015, http://ceur-ws.org/Vol-1364/sw4sh-2015.pdf

Declerck T., Wandl-Vogt E. & Mörth K. (2015): Towards a Pan European Lexicography by Means of Linked (Open) Data, pp. 342-355, in: Proceedings of eLex 2015 - Electronic Lexicography in the 21st century: Linking Lexical Data, Herstmonceux Castle, Sussex, UK, 11-13 August 2015, https://elex.link/elex2015/proceedings/eLex_2015_22_Declerck+etal.pdf

Di Giorgio S., Felicetti A., Martini P. & Masci E. (2016): Dati.CulturaItalia: a Use Case of Publishing Linked Open Data Based on CIDOC-CRM, pp. 44-54, in: Ronzino, Paola (ed.): Extending, Mapping and Focusing the CRM. Proceedings of the EMF-CRM workshop, Poznan, Poland, 17 September 2015, http://ceur-ws.org/Vol-1656/paper4.pdf

Digital Atlas of the Roman Empire (Department of Archaeology and Ancient History, Lund University, Sweden), http://dare.ht.lu.se

Digital Collaboratory for Cultural Dendrochronology - DCCD, http://dendro.dans.knaw.nl; project website: http://vkc.library.uu.nl/vkc/dendrochronology/

Digital Object Identifier System, http://www.doi.org

Dodds L. & Davis I. (2012): Linked Data Patterns. A pattern catalogue for modelling, publishing, and consuming Linked Data (version 2012-05-31), http://patterns.dataincubator.org/book/

Dodds, Leigh *et al*. (2010): Quality Indicators for Linked Data Datasets. Discussion on Semantic-Overflow, 24.06.-13.07.2010, http://answers.semanticweb.com/questions/1072/quality-indicators-for-linked-data-datasets

Doerr M. & Hiebel G. (2013): CRMgeo: Linking the CIDOC CRM to GeoSPARQL through a spatiotemporal refinement. ICS-FORTH/TR-435, April 2013, https://www.ics.forth.gr/tech-reports/2013/2013.TR435_CRMgeo_CIDOC_CRM_GeoSPARQL.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. (2014a): 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. (2014b): The SYNERGY Reference Model of Data Provision and Aggregation. Draft, June 2014, http://www.cidoc-crm.org/docs/SRM_v0.1.pdf

Doerr M., Kritsotaki A. & Boutsika, A. (2011): Factual argumentation - a core model for assertions making. In: Journal on Computing and Cultural Heritage (JOCCH), 3(3), http://dl.acm.org/citation.cfm?id=1921615

Doerr M., Schaller K. & Theodoridou M. (2004): Integration of complementary archaeological sources. In: Niccolucci F. (ed.): Proceedings of the 32nd Computer Applications and Quantitative

Methods in Archaeology Conference, http://proceedings.caaconference.org/files/2004/09_Doerr_et_al_CAA_2004.pdf

Doerr M., Theodoridou M., Aspöck E. & Masur A. (2016): Mapping Archaeological Databases to CIDOC CRM, pp. 443-451, in: CAA-2015 - 43rd Conference on Computer Applications and Quantitative Methods in Archaeology (Siena, April 2015). Oxford: Archaeopress, http://archaeopress.com/ArchaeopressShop/Public/download.asp?id={77DEDD4E-DE8F-43A4-B115-ABE0BB038DA7}

Doerr, Martin (2010): Technological Choices of the ResearchSpace Project. Researchspace.org, August 2010, http://www.researchspace.org/researchspace-concepts/technological-choices-of-the-researchspace-project

Duan S., Kementsietsidis A., Srinivas K. & Udrea O. (2011): Apples and oranges: a comparison of RDF benchmarks and real RDF datasets. In: SIGMOD'11, Athens, Greece, 12–16 June 2011, conference proceedings, pp. 145–156, http://researcher.ibm.com/researcher/files/us-sduan/sigmod2011_RDF_benchmark_duan.pdf

Dublin Core Metadata Element Set, Version 1.1, 2012-06-14, http://dublincore.org/documents/dces/

Dublin Core Metadata Initiative (DCMI) Metadata Terms, http://dublincore.org/documents/dcmi-terms/

Dunsire G., Harper C., Hillmann D. & Phipps J. (2012): Linked Data Vocabulary Management: Infrastructure Support, Data Integration, and Interoperability. In: Information Standards Quarterly, 24(2/3): 4-13, http://www.niso.org/publications/isq/2012/v24no2-3/dunsire/

Dutch Ships and Sailors (Clarin IV project, 4/2013-3/2014), http://dutchshipsandsailors.nl

EAGLE - Europeana Network of Ancient Greek and Latin Epigraphy (EU, ICT-PSP, 4/2013-3/2016), http://www.eagle-network.eu

EAGLE (2015): EAGLE Metadata Model Specification – Second release. Project deliverable D 3.1.2, V1.1, 26 January 2015, http://www.eagle-network.eu/wp-content/uploads/2013/06/EAGLE_D3.1_EAGLE-metadata-model-specification_v1.1.pdf

EAGLE vocabularies (Material, Type of inscription, Execution technique, Object type, Decoration, Dating criteria, State of preservation), http://www.eagle-network.eu/resources/vocabularies/

Eckkrammer F., Feldbacher R. & Eckkrammer T. (2011): CIDOC CRM in Data Management and Data Sharing. Data Sharing between Different Databases, pp. 80-85, in: CAA-2008. 36th Annual Conference of Computer Applications and Quantitative Methods in Archaeology, Budapest; http://proceedings.caaconference.org/files/2008/CD19_Eckkrammer_et_al_CAA2008.pdf

Edelstein J., Galla L., Li-Madeo C., Marden J. Rhonemus A. & Whysel N. (2013a): Linked Open Data for Cultural Heritage: Evolution of an Information Technology. New York: Pratt Institute, Spring 2013, http://www.whysel.com/papers/LIS670-Linked-Open-Data-for-Cultural-Heritage.pdf

Edelstein J., Li-Madeo C., Marden J. & Whysel N. (2013b): Linked Open Data for Cultural Heritage: evolution of an information technology, pp. 107-112, in: SIGDOC'13 - Proceedings of the 31st ACM International Conference on Design of Communication; preprint, http://academiccommons.columbia.edu/catalog/ac:168445

Elliott T. & Gillies S. (2009): Digital Geography and Classics. In: Digital Humanities Quarterly, 3(1), http://www.digitalhumanities.org/dhq/vol/3/1/000031.html

Elliott T. & Jones C. (2014): Moving the Ancient World Online Forward. ISAW Paper 7.6, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

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://dx.doi.org/10.3789/isqv24n2-3.2012.08

Elliott T., Heath S. & Muccigrosso J. (2014): Prologue and Introduction. ISAW Paper 7.1, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

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/

Encoded Archival Description, http://www.loc.gov/ead/

Encyclopedia of Life (EOL), http://www.eol.org

English Heritage Places, DataHub information, http://datahub.io/dataset/englishheritage_places

Entjes, Jeroen A. (2015): Linking Maritime Datasets to Dutch Ships and Sailors Cloud - Case studies on Archangelvaart and Elbing. Master thesis project, https://videbo.files.wordpress.com/2015/08/jeroen_entjes_final_thesis.pdf

Environment Ontology, https://bioportal.bioontology.org/ontologies/ENVO

EpiDoc: Epigraphic Documents in TEI XML, http://epidoc.sf.net

Epigraphic Database Heidelberg, http://edh-www.adw.uni-heidelberg.de

EPNet - Production and Distribution of Food during the Roman Empire: Economic and Political Dynamics (ERC Advanced Grant project, 3/2014-2/2019), http://www.roman-ep.net

Epure E.V., Martín-Rodilla P., Hug C., Deneckère R. & Sanilesi C. (2015): Automatic Process Model Discovery from Textual Methodologies: An Archaeology Case Study. Proceedings of RCIS 2015 - Ninth IEEE International Conference on Research Challenges in Information Science, Athens, Greece, May 2015, https://hal-paris1.archives-ouvertes.fr/hal-01149742/document

Erdman-Thomsen H., Pareja-Lora A. & Nistrup Madsen B. (2016): Term Bases and Linguistic Linked Open Data. TKE 2016 - 12th International conference on Terminology and Knowledge Engineering. Copenhagen Business School, http://openarchive.cbs.dk/handle/10398/9323

Ermilov I., Lehmann J., Martin M. & Auer S. (2016): LODStats: The Data Web Census Dataset. ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, 17-21 October 2016; preprint, https://svn.aksw.org/papers/2016/ISWC_LODStats_Resource_Description/public.pdf

Erp M., Oomen J., Segers R. *et al*. (2011): Automatic heritage metadata enrichment with historic events. Museums and the Web 2011, 6-9 April 2011, Philadelphia, http://www.museumsandtheweb.com/mw2011/papers/automatic_heritage_metadata_enrichment_with_hi

Erxleben F., Günther M., Krötzsch M., Mendez J. & Vrandeci D. (2014): Introducing Wikidata to the Linked Data Web. ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, http://korrekt.org/papers/Wikidata-RDF-export-2014.pdf

EUCLID - Educational Curriculum for the Usage of Linked Data, http://euclid-project.eu

European Coin Find Network (ECFN), http://www.ecfn.fundmuenzen.eu

European Commission, Joinup Portal - Share and reuse interoperability solutions for public administrations, http://joinup.ec.europa.eu

European Language Social Science Thesaurus (ELSST), http://elsst.ukdataservice.ac.uk

European Network of e-Lexicography - ENeL (EU, COST Action, 10/2013-10/2017, http://www.elexicography.eu

European Persistent Identifier Consortium (EPIC), http://www.pidconsortium.eu

Europeana Cloud project (02/2013-01/2015, CIP-ICT-PSP Best Practice Network, http://pro.europeana.eu/web/europeana-cloud

Europeana Data Model (EDM), http://pro.europeana.eu/edm-documentation

Europeana Linked Data, http://labs.europeana.eu/api/linked-open-data/introduction/

Europeana Tech Task Force on a Multilingual and Semantic Enrichment Strategy: final report, 7 April 2014, http://pro.europeana.eu/documents/468623/8b75b054-712e-432b-a0f7-761898e6f60e

EuropeanaConnect (EU, eContent+ project, 5/2009-10/2011), http://www.europeanaconnect.eu

FaBiO - FRBR-aligned Bibliographic Ontology, http://vocab.ox.ac.uk/fabio

Faron-Zucker C., Pajón Leyra I., Poulida K. & Tettamanzi A. (2016): Semantic Categorization of Segments of Ancient and Mediaeval Zoological Texts, pp. 59-68, SWASH 2016 - 2nd Workshop on Semantic Web for Scientific Heritage, Heraklion, Greece, 30 May 2016, http://ceur-ws.org/Vol-1595/paper7.pdf

Felicetti A. & Lorenzini M. (2011): Metadata and tools for integration and preservation of cultural heritage 3D information. 23rd International CIPA Symposium, Prague, Czech Republic, 12-16 September 2011, http://cipa.icomos.org/fileadmin/template/doc/PRAGUE/051.pdf

Felicetti A., Galluccio I., Luddi C., Mancinelli M.L., Scarselli T. & Madonna A.D. (2016): Integrating Terminological Tools and Semantic Archaeological Information: the ICCD RA Schema and Thesaurus, pp. 28-43, in: Ronzino, Paola (ed.): Extending, Mapping and Focusing the CRM. Proceedings of the EMF-CRM workshop, Poznan, Poland, 17 September 2015, http://ceur-ws.org/Vol-1656/paper3.pdf

Felicetti A., Gerth P., Meghini C. & Theodoridou M. (2016): Integrating Heterogeneous Coin Datasets in the Context of Archaeological Research, pp. 13-27, in: Ronzino, Paola (ed.): Extending, Mapping and Focusing the CRM. Proceedings of the EMF-CRM workshop, Poznan, Poland, 17 September 2015, http://ceur-ws.org/Vol-1656/paper2.pdf

Felicetti A., Murano F., Ronzino P. & Niccolucci F. (2016): CIDOC CRM and Epigraphy: a Hermeneutic Challenge, pp. 55-68, in: Ronzino, Paola (ed.): Extending, Mapping and Focusing the CRM. Proceedings of the EMF-CRM workshop, Poznan, Poland, 17 September 2015, http://ceur-ws.org/Vol-1656/paper5.pdf

Felicetti A., Scarselli T., Mancinelli M.L. & Niccolucci F. (2013): Mapping ICCD Archaeological Data to CIDOC-CRM: the RA Schema. In: Alexiev V. *et al*. (eds.): Practical Experiences with CIDOC CRM and its Extensions (CRMEX 2013) Workshop, 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), Valetta, Malta, 26 September 2013, http://ceur-ws.org/Vol-1117/paper2.pdf

Felicetti, Achille (2012): Digital collections of semantically annotated cultural heritage texts. In: Uncommon Culture, Vol. 3, no. 5/6 (2012): 61-64, http://journals.uic.edu/ojs/index.php/UC/article/view/4719/3682

Felle A.E. & Rocco A. (eds., 2016): Off the Beaten Track. Epigraphy at the Borders. Proceedings of the VI EAGLE International Event, 24-25 September 2015, Bari, Italy. Oxford: Archaeopress, http://www.archaeopress.com/ArchaeopressShop/Public/download.asp?id={E7B2AAC6-9986-4C41-9842-6AA93BE7ACD9}

Ferrara A., Nikolov A. & Scharffe F. (2011): Data Linking for the Semantic Web. In: International Journal on the Semantic Web in Information Systems, 7(3): 46-76, manuscript, http://people.kmi.open.ac.uk/andriy/data_linking_for_the_semantic_web.pdf (paper: http://www.igi-global.com/article/data-linking-semantic-web/62562

Ferro N., Munnelly G., Hampson C. & Conlan O. (2013): Fostering Interaction with Cultural Heritage Material via Annotations: The FAST-CAT Way. Proceedings of the 9th Italian Research Conference on Digital Libraries (IRCDL 2013), CCIS Vol.385, http://www.tara.tcd.ie/xmlui/bitstream/handle/2262/67966/fast-cat-IRCDL2013.v2%20copy.pdf;jsessionid=B36C67BA30EC9A76F83C8BDE7A6A03DC?sequence=1

Finto - Finnish thesaurus and ontology service, http://finto.fi/en/

FOAF - Friend-of-a-Friend, http://xmlns.com/foaf/spec/

Forum on Information Standards in Heritage (FISH): http://heritage-standards.org.uk/fish-vocabularies/

Fossilworks, http://fossilworks.org

Free Your Metadata project (iMinds / Ghent University and MaSTIC / Université Libre de Bruxelles) http://freeyourmetadata.org

Freitas A., Curry E., Oliveira J.G. & O'Riain S. (2012): Querying Heterogeneous Datasets on the Linked Data Web: Challenges, Approaches, and Trends. IEEE Internet Computing, 16(1), January/February 2012 24-33, http://www.edwardcurry.org/publications/freitas_IC_12.pdf

Fürber C. & Hepp M. (2010a): Using Semantic Web Resources for Data Quality Management, pp. 211-225, in: Knowledge Engineering and Management by the Masses. Springer: Lecture Notes in Computer Science Volume 6317; preprint, http://www.fuerber.com/publications/Fuerber-Hepp-Using_Semantic_Web_Resources_for_Data_Quality_Management.pdf

Fürber C. & Hepp M. (2010b): Using SPARQL and SPIN for Data Quality Management on the Semantic Web. In: Business Information Systems. Lecture Notes in Business Information Processing, Vol. 47: 35-46, http://www.heppnetz.de/files/fuerber-hepp-sparql-spin-dqm.pdf

Fürber C. & Hepp M. (2011a): SWIQA - A Semantic Web Information Quality Assessment Framework. ECIS 2011 - European Conference on Information Systems, Proceedings, paper 76, http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1075&context=ecis2011

Fürber C. & Hepp M. (2011b): Data Quality Management Vocabulary. V 1.0, 9 October 2011, http://semwebquality.org/dqm-vocabulary/v1/dqm

Fürber C., Hepp M. & Wischnewski M. (2011): Data Quality Constraints Library. V1.1, 28 March 2011, http://semwebquality.org/ontologies/dq-constraints

GBIF (2011): Recommendations for the Use of Knowledge Organisation Systems by GBIF. Released on 4 February 2011. Copenhagen: Global Biodiversity Information Facility, http://www.gbif.org/resource/80656

Geiger C.P. & von Lucke J. (2012): Open Government and (Linked) (Open) (Government) (Data). Free accessible data of the public sector in the context of open government. In: JeDEM - eJournal of eDemocracy and Open Government, 4(2): 265-278, http://www.jedem.org/index.php/jedem/article/download/143/115

GEMET - General Multilingual Environmental Thesaurus (EIONET/European Environment Agency), http://www.eionet.europa.eu/gemet/

Geological Survey of Ireland, http://www.gsi.ie

GeoNames, http://www.geonames.org

GeoSpecies ontology, https://bioportal.bioontology.org/ontologies/GEOSPECIES

German National Library: Linked Data Service, http://dnb.de/EN/lds

Gerth P., Schmidle W. & Cuy S. (2016a): Sculptures in the Semantic Web. Presentation at CAA2016 - 44th Computer Applications and Quantitative Methods in Archaeology Conference, Oslo, Norway, 30 March 2016, http://www.slideshare.net/ariadnenetwork/sculptures-in-the-semantic-web-65237911

Gerth P., Schmidle W. & Cuy S. (2016b): Sculptures in the Semantic Web. In: Proceedings of CAA2016 - 44th Computer Applications and Quantitative Methods in Archaeology Conference, Oslo, Norway, 29 March - 2 April 2016 (paper forthcoming).

Geser, Guntram (2003): A Cultural Heritage Semantic Web Example & Primer, pp. 26-36, in: DigiCULT Thematic Issue 3: Towards a Semantic Web for Heritage Resources. Salzburg, May 2003, http://www.digicult.info/pages/Themiss.php

Geser, Guntram (2004): Assessing the readiness of small heritage institutions for e-culture technologies, pp. 8-13, in: DigiCULT.Info e-Journal, Issue 9, November 2004, http://www.digicult.info/downloads/digicult_info_9.pdf

Geser, Guntram (2009): STERNA Technology Watch Report. A Report on Semantic Approaches for Including Digital Cultural and Bio-Heritage Resources in the European Digital Library Initiative. Salzburg, January 2009, http://www.sterna-net.eu/images/stories/documents/sterna_del.6.5_technology-watch_full-report_20081210.pdf

Geser, Guntram *et al*. (2003): Towards a Semantic Web for Heritage Resources. DigiCULT Thematic Issue 3, May 2003, http://www.digicult.info/downloads/thematic_issue_3_low.pdf

Getty Vocabularies as Linked Open Data, http://www.getty.edu/research/tools/vocabularies/lod/

Getty Vocabularies: LOD, http://vocab.getty.edu

Goddard L. & Byrne G. (2010): Linked Data tools: Semantic Web for the masses. In: First Monday, 15(11), http://firstmonday.org/ojs/index.php/fm/article/view/3120/2633

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

Golden P. & Shaw R. (2016): Nanopublication beyond the sciences: the PeriodO period gazetteer. In: PeerJ Computer Science 2: e44, https://peerj.com/articles/cs-44/

Golub K. & Tudhope D. (2009): Terminology Registry Scoping Study (TRSS): Final report, 3 July 2009, http://www.jisc.ac.uk/media/documents/programmes/sharedservices/trss-report-final.pdf

Golub K., Tudhope D., Zeng M.L. & Žumer M. (2014): Terminology registries for knowledge organization systems: Functionality, use, and attributes. In: Journal of the Association for Information Science & Technology, 65(9): 1901-16, http://dx.doi.org/doi:10.1002/asi.23090

Good B.M. & Wilkinson M.D. (2006): The Life Sciences Semantic Web is Full of Creeps! Briefings in Bioinformatics 2006 7(3):275-286, http://bib.oxfordjournals.org/cgi/content/full/7/3/275?ck=nck#T1

Görz G. & Scholz M. (2012): WissKI: A Virtual Research Environment for Cultural Heritage. In: De Raedt, Luc *et al*. (eds.): ECAI 2012 - 20th European Conference on Artificial Intelligence, Montpellier 27-31 August 2012. Amsterdam: IOS Press; preprint, http://wwwdh.cs.fau.de/IMMD8/staff/Goerz/ecai2012.pdf

Gracy K. & Lambert F. (2014): Who's ready to surf the next wave? A study of perceived challenges to implementing new and revised standards for archival description. In: The American Archivist, 77(1): 96-132, http://americanarchivist.org/doi/abs/10.17723/aarc.77.1.b241071w5r252612

Gracy, Karen F. (2015): Archival description and linked data: a preliminary study of opportunities and implementation challenges. In: Archival Science, 15(3): 239-294, http://link.springer.com/article/10.1007/s10502-014-9216-2

Grassi M., Morbidoni C., Nucci M. *et al*. (2013): Pundit: Augmenting Web Contents with Semantics. Literary and Linguisting Computing, Vol. 28, No. 4, http://dm2e.eu/files/Graasi-et-al.-2013-Pundit-augmenting-web-contents-with-semantics.pdf

Gros, Jean-Sébastien (2016): Ατλας, a Gazetteer Linking Archaeological Collections, pp. 19-24, in: SWASH 2016 - 2nd Workshop on Semantic Web for Scientific Heritage, Heraklion, Greece, 30 May 2016, http://ceur-ws.org/Vol-1595/paper2.pdf

Gruber E. & Smith T.J. (2014): Linked Open Greek Pottery, pp. 205-214, in: CAA 2014 Paris - Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Archaeopress; preprint, https://www.academia.edu/9739936/Linked_Open_Greek_Pottery

Gruber E., Bransbourg G., Heath S. & Meadows A. (2013): Linking Roman Coins: Current Work at the American Numismatic Society, pp. 249-258, in: CAA 2012 Southampton, Volume I. Amsterdam University Press; preprint, https://www.academia.edu/6604014/Linking_Roman_Coins_Current_Work_at_the_American_Numismatic_Society

Gruber E., Gondek R. & Smith T.J. (2015): CAA 2015 Siena, Roundtable – Linked Open Data Applied to Pottery Databases, 1 April 2015, http://2015.caaconference.org/program/roundtables/rt3/

Gruber, Ethan (2016): LOD for Numismatic LAM Integration. Presentation at CAA 2016 Oslo, session "Linked Pasts: Connecting Islands of Content", 30 March 2016, http://de.slideshare.net/ewg118/lod-for-numismatic-lam-integration

Gruntgens M. & Schrade T. (2016): Data repositories in the Humanities and the Semantic Web: modelling, linking, visualising, pp. 53-64, in: Proceedings of WHiSe 2016 - 1st Workshop on Humanities in the Semantic Web, Anissaras, Greece, 29 May 2016, http://ceur-ws.org/Vol-1608/paper-07.pdf

Gueguen G., Marques da Fonseca V.M., Pitti D.V. & Sibille de Grimoüard C. (2013): Toward an International Conceptual Model for Archival Description: A Preliminary Report from the

International Council on Archives' Experts Group on Archival Description. In: The American Archivist, 76(2): 566-582, http://www.ica.org/sites/default/files/EGAD_English.pdf

HADOC – Harmonisation de la production des données culturelles programme (Ministère de la Culture et de la Communication, France), http://www.culturecommunication.gouv.fr/Ressources/Harmonisation-des-donnees-culturelles

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

Hafford, William B. (2014): Linked Open Data and the Ur of the Chaldees Project. ISAW Paper 7.7, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Halb W. & Hausenblas M. (2008): select * where { :I :trust :you }. How to Trust Interlinked Multimedia Data. Proceedings of the International Workshop on Interacting with Multmedia Content in the Social Semantic Web (IMC-SSW 2008), Koblenz, Germany, 3 December 2008, http://ceur-ws.org/Vol-417/paper6.pdf

Hannemann J. & Kett J. (2010): Linked Data for Libraries. IFLA 2010 – World Library and Information Congress, Gothenburg, Sweden, 10-15 August 2010, http://conference.ifla.org/past-wlic/2010/149-hannemann-en.pdf

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

Harpring, Patricia (2016): Art & Architecture Thesaurus. Introduction and Overview. Getty Vocabulay Program, http://www.getty.edu/research/tools/vocabularies/aat_in_depth.pdf

Hart, Glen (2009): Linking to the past, geographically speaking: The Linked Data Web & Historical GIS, http://www.ordnancesurvey.co.uk/oswebsite/partnerships/research/publications/docs/2009/Linking_to_the_Past_GeoS.pdf

Haslhofer B., Momeni E., Gay, M. & Simon R. (2010): Augmenting Europeana Content with Linked Data Resources. Proceedings of the 6th International Conference on Semantic Systems (I-Semantics), Graz, Austria, 1-3 September 2010, http://eprints.cs.univie.ac.at/26/1/ldtc2010_haslhofer_et_al_cr2.pdf

Hasnain A., Sana e Zainab S., Kamdar M.R. *et al*. (2015): A Roadmap for navigating the Life Sciences Linked Open Data Cloud, pp. 97-112, in: Semantic Technology - 4th Joint International Conference, JIST 2014, Chiang Mai, Thailand, 9-11 November 2014. Springer (LNCS 8943); preprint, http://maulik-kamdar.com/wp-content/uploads/2014/10/JIST2014.pdf

Haustein S. & Pleumann J. (2002): Is Participation in the Semantic Web Too Difficult?, pp. 448-453, in: Horrocks I. & Hendler J. (eds.): The Semantic Web - ISWC 2002. Berlin: Springer, http://sfb876.tu-dortmund.de/PublicPublicationFiles/haustein_pleumann_2002b.pdf

Heath T. & Bizer C. (2011): Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. Online: http://linkeddatabook.com/editions/1.0/

Heath, Sebastian (2014): ISAW Papers: Towards a Journal as Linked Open Data. ISAW Paper 7.8, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Heath, Tom (2009): Linked Data? Web of Data? Semantic Web? WTF?, http://tomheath.com/blog/2009/03/linked-data-web-of-data-semantic-web-wtf/

Heath, Tom (2010): Why Carry the Cost of Linked Data? Tom Heath weblog, 16 June 2010, http://tomheath.com/blog/2010/06/why-carry-the-cost-of-linked-data/

Hennicke S., Marlies Olensky M., de Boer V. *et al*. (2011): A data model for cross-domain data representation: The Europeana Data Model in the case of archival and museum data. Proceedings of the 12th International Symposium on Information Science, (ISI 2011). Hildesheim, Germany, March 9-11 2011. http://www.few.vu.nl/~AI.Isaac/publications.php

Hepp, Martin (2007): Possible Ontologies. How Reality Constrains the Development of Relevant Ontologies. In: IEEE Internet Computing, 11(1): 90-96, http://www.heppnetz.de/files/IEEE-IC-PossibleOntologies-published.pdf

Heritage Data - Linked Data Vocabularies for Cultural Heritage, http://www.heritagedata.org

Hirst, Tony (2010): Comments to "Why Carry the Cost of Linked Data?". Tom Heath weblog, 16 June 2010, http://tomheath.com/blog/2010/06/why-carry-the-cost-of-linked-data/

Hodge, Gail (2014): Government Knowledge Organization Systems: Valuing a Public Good, pp. 23-29, in: ASIS&T Bulletin, 40(4), April/May 2014, https://www.asist.org/publications/bulletin/apr-14/

Hoekstra R., Meroño-Peñuela A., Dentler K., Rijpma A., Zijdeman R. & Zandhuis I. (2016): An ecosystem for Linked Humanities Data, pp. 85-96, In: Proceedings of WHiSe 2016 - 1st Workshop on Humanities in the Semantic Web, Anissaras, Greece, 29 May 2016, http://ceur-ws.org/Vol-1608/paper-11.pdf

Hogan A. & Gutierrez C. (2014): Paths towards the Sustainable Consumption of Semantic Data on the Web. AMW 2014 - 8th Alberto Mendelzon Workshop on Foundations of Data Management, Cartagena de Indias, Colombia, 4-6 June 2014, CEUR Workshop Proceedings, http://ceur-ws.org/Vol-1189/paper_7.pdf or http://ciws.cl/media/pdf/amw_2014_hogan.pdf

Hogan A., Harth A., Passant A., Decker S. & Polleres A. (2010): Weaving the pedantic web. In: LDOW 2010 - 3rd International Workshop on Linked Data on the Web, Raleigh, USA, 27 April 2010, http://events.linkeddata.org/ldow2010/papers/ldow2010_paper04.pdf

Hogan A., Umbrich J., Harth A. *et al*. (2012): An empirical survey of Linked Data conformance. In: Web Semantics: Science, Services and Agents on the World Wide Web, Special Issue on 'Dealing with the Messiness of the Web of Data', volume 14, July 2012, http://aidanhogan.com//docs/ldstudy12.pdf

Holmen J. & Ore C. (2010): Deducing Event Chronology in a Cultural Heritage Documentation System. In: Proceedings of the 38th Computer Applications and Quantitative Methods in Archaeology (CAA) Conference, Williamsburg, Virginia, USA, 22-26 March 2009, http://www.edd.uio.no/artiklar/arkeologi/holmen_ore_caa2009.pdf

Holmen J., Ore C. & Eide O. (2004): Documenting Two Histories at Once: Digging into Archaeology. Proceedings of the 30th Computer Applications and Quantitative Methods in Archaeology, vol. 1227 of BAR International Series, Oxford: Archaeopress.

Hong Y., Solanki M., Foxhall L. & Quercia A. (2010): A Framework for Transforming Archaeological Databases to Ontological Datasets. In: Proceedings of the 38th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA). Granada, Spain, April 2010, http://www.tracingnetworks.ac.uk/publications/CAA2010/paper.pdf

Horne, Ryan (2014): Beyond Maps as Images at the Ancient World Mapping Center. ISAW Paper 7.9, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Hoxha J., Rula A. & Ell B. (2011): Towards green linked data. COLD 2011 - 2nd International Workshop on Consuming Linked Data @ ISWC 2011 proceedings, http://ceur-ws.org/Vol-782/HoxhaEtAl_COLD2011.pdf

Huebner, Katherine (2009): How taxonomic revisions affect the interpretation of specimen identification in biological field data. In: mURJ, 4(1): 25-29, http://msurj.mcgill.ca/vol4/iss1/Huebner2009.pdf

Huggett, Jeremy (2012): Promise and Paradox: Accessing Open Data in Archaeology. Proceedings of the Digital Humanities Congress 2012, Edited by C. Mills, M. Pidd & E. Ward, http://www.hrionline.ac.uk/openbook/chapter/dhc2012-huggett

Huijboom N. & Van den Broek T. (2011): Open data: an international comparison of strategies. In: European Journal of ePractice, Issue 12, March/April 2011, pp. 4-15, https://joinup.ec.europa.eu/sites/default/files/76/a7/05/ePractice%20Journal-%20Vol.%2012-March_April%202011.pdf

HUMA-NUM - la très grande infrastructure de recherche des humanités numérique, http://www.huma-num.fr

Hunter J. & Gerber A. (2010): Harvesting community annotations on 3D models of museum artefacts to enhance knowledge, discovery and re-use. In: Journal of Cultural Heritage, 11(1): 81-90, https://www.researchgate.net/search.Search.html?query=Harvesting+community+annotations+on+3D+models+of+museum+artefacts+to+enhance+knowledge%2C+discovery+and+re-use

Hunter J. & Gerber A. (2012): Towards Annotopia - Enabling the Semantic Interoperability of Web-Based Annotations. Future Internet, 4(3): 788-806, http://www.mdpi.com/1999-5903/4/3/788

Hunter J. & Yu C.-H. (2011): Assessing the Value of Semantic Annotation Services for 3D Museum Artefacts. Sustainable Data from Digital Research Conference (SDDR 2011), Melbourne, 13-14 December 2011 (authors' manuscript), http://ses.library.usyd.edu.au/bitstream/2123/7951/1/HunterYu.pdf

Hunter J., Khan I. & Gerber A. (2008): HarVANA - Harvesting Community Tags to Enrich Collection Metadata. Joint Conference on Digital Libraries, JCDL 2008. Pittsburgh, USA, 16-20 June 2008, http://www.itee.uq.edu.au/eresearch/filething/files/get/papers/2008/Hunter_JCDL2008.pdf

Hyland B. & Villazón-Terrazas B. (eds., 2011): Cookbook for Open Government Linked Data. Revised Version, December 2011, https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

Hyland, Bernadette (2010): Preparing for a Linked Data Enterprise, in: Wood, David (ed., 2010): Linking Enterprise Data. Springer, manuscript http://linkeddatadeveloper.com/Projects/Linking-Enterprise-Data/Manuscript/led-hyland.html

Hyvönen E., Ikkala E. & Tuominen J. (2016): Linked Data brokering service for historical places and maps, pp. 39-52, in: Proceedings of WHiSe 2016 - 1st Workshop on Humanities in the Semantic Web, Anissaras, Greece, 29 May 2016, http://ceur-ws.org/Vol-1608/paper-06.pdf

Hyvönen E., Kettula S., Raatikka V. *et al*. (2002). Semantic interoperability on the Web. Case Finnish Museums On-line. In: Hyvönen E. & Klemettinen M. (2002): Towads the Semantic Web and Web Services. Proceedings of the XML Finland 2002 Conference, Helsinki, Finland, 21-22 October 2002, http://www.cs.helsinki.fi/u/eahyvone/xmlfinland2002/ProceedingsXML2002-final.pdf

Hyvönen E., Lindquist T., Törnroos J. & Mäkelä E. (2012): History on the semantic web as linked data: an event gazetteer and timeline for the World War I. Proceedings of CIDOC 2012, Enriching Cultural Heritage, 10–14 June 2012, Helsinki, Finland, http://www.cidoc2012.fi/en/File/1609/hyvonen.pdf

Hyvönen E., Mäkelä E., Kauppinen T. *et al*. (2009a): CultureSampo - Finnish Cultural Heritage Collections on the SemanticWeb 2.0. In Proceedings of the 1st International Symposium on Digital humanities for Japanese Arts and Cultures, Ritsumeikan University, Kyoto, Japan, March 2009, http://www.seco.tkk.fi/publications/2009/hyvonen-et-al-culturesampo-dh-jac-2009.pdf

Hyvönen E., Mäkelä E., Kauppinen T. *et al*. (2009b): CultureSampo: A National Publication System of Cultural Heritage on the Semantic Web 2.0. 6th European Semantic Web Conference, proceedings (Lecture Notes in Computer Science, vol. 5554/2009): 851–856, http://www.seco.tkk.fi/publications/2009/hyvonen-et-al-culsa-demo-eswc-2009.pdf

Hyvönen E., Mäkelä E., Kauppinen T. *et al*. (2009c): Finnish Culture on the Semantic Web 2.0. Thematic Perspectives for the End-user. In: Museums and the Web, volume 2009, http://www.archimuse.com/mw2009/papers/hyvonen/hyvonen.html

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 E., Saarela S. & Viljanen K. (2004): Application of Ontology Techniques to View-Based Semantic Search and Browsing. Proceedings of the 1st European Semantic Web Symposium (Lecture Notes in Computer Science, vol. 2053): 92–106, http://www.seco.tkk.fi/publications/2004/hyvonen-saarela-et-al-application-of-ontology-techniques-2004.pdf

Hyvönen E., Saarela S., Viljanen K. *et al*. (2004): A Cultural Community Portal for Publishing Museum Collections on the Semantic Web. ECAI Workshop on Application of Semantic Web Technologies to Web Communities (CEUR Workshop Proceedings, vol. 107) http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-107/paper8.pdf

Hyvönen E., Tuominen J., Alonen M. & Mäkelä E. (2014): Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. Proceedings of ESWC 2014 Demo and Poster Papers, Springer-Verlag, http://seco.cs.aalto.fi/publications/2014/hyvonen-et-al-ldf-2014.pdf

Hyvönen E., Viljanen K., Tuominen J. & Seppälä K. (2008): Building a National Semantic Web Ontology and Ontology Service Infrastructure - the FinnONTO Approach, pp. 95–109, in: 5th European Semantic Web Conference proceedings (Lecture Notes in Computer Science, vol. 5021), http://www.seco.tkk.fi/publications/2008/hyvonen-et-al-building-2008.pdf

Hyvönen, Eero (2009): Semantic Portals for Cultural Heritage. Handbook on Ontologies. 2nd edition; chapter, http://www.seco.tkk.fi/publications/2009/hyvonen-portals-2009.pdf

Hyvönen, Eero (2012): Publishing and Using Cultural Heritage Linked Data on the Semantic Web. Synthesis Lectures on the Semantic Web: Theory and Technology. Palo Alto: Morgan & Claypool, http://www.seco.tkk.fi/publications/2012/hyvonen-ch-book-2012.pdf

ICOM (2011): ICOM recommendation on Linked Open Data for museums, http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/AGM_2011/LoD_For_Museums%20v1.6.pdf

ICONCLASS as Linked Open Data, http://www.iconclass.org/help/lod

iDAI.gazetteer (German Archaeological Institute), http://gazetteer.dainst.org

Institute for the Study of the Ancient World (ISAW), http://isaw.nyu.edu

Inventaire National du Patrimoine Naturel / National Inventory of Natural Heritage (Muséum national d'Histoire naturelle), http://inpn.mnhn.fr

Inventaires archéozoologiques et archéobotaniques de France - I2AF (Muséum national d'Histoire naturelle), https://inpn.mnhn.fr/espece/inventaire/I100

Irish National Monuments Service monument class list, http://webgis.archaeology.ie/NationalMonuments/WebServiceQuery/Lookup.aspx

ISA - Interoperability Solutions for European Public Administrations (2013): Cookbook for translating relational data models to RDF schemas. Prepared for the ISA programme by PwC EU Services, 27/02/2013, http://ec.europa.eu/isa/documents/cookbook-for-rdf-schemas-v2.pdf

ISA - Interoperability Solutions for European Public Administrations (2012): Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC. Prepared by P. Archer (W3C/ERCIM), S. Goedertier and N. Loutas (PwC EU Services). Project deliverable D7.1.3, December 2012, https://joinup.ec.europa.eu/community/semic/document/10-rules-persistent-uris

Isaac A. & Haslhofer B. (2013): Europeana Linked Open Data – data.europeana.eu. Semantic Web Journal, 4(3): 291-297, http://www.semantic-web-journal.net/system/files/swj297_1.pdf

Isaac A., Clayphan R. & Haslhofer B. (2012): Europeana: Moving to Linked Open Data. Information Standards Quarterly, 2012 Spring/Summer, 24(2/3):34-40, http://www.niso.org/apps/group_public/download.php/9407/IP_Isaac-etal_Europeana_isqv24no2-3.pdf

Isaac A., Waites W., Young J. & Zeng M. (eds., 2011): Library Linked Data Incubator Group: Datasets, value vocabularies, and metadata element sets [W3C Incubator Group Report, October 25, 2011]. http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset/

Isaksen L., Barker E., Kansa E. & Byrne K. (2011): Googling Ancient Places. Proceedings of Digital Humanities 2011 (DH2011), Stanford, CA, June 2011 (online paper), http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-349.xml;query=;brand=default

Isaksen L., Martinez K. & Earl G. (2010b): Interoperate with whom? Archaeology, Formality & the Semantic Web. CAA UK Chapter Meeting, UCL, London, 19-20 February 2010, http://leifuss.files.wordpress.com/2008/08/caauk2010isaksen1.pdf

Isaksen L., Martinez K. & Earl G. (2011): Semantic Technologies in Cultural Heritage. Past, Present and Future. Cultural Heritage & the Semantic Web. British Museum, London (slides: http://leifuss.files.wordpress.com/2008/08/bmisaksenfinalslides.pdf)

Isaksen L., Martinez K., Gibbins N. & Earl G. & Keay S. (2010a): Interoperate With Whom? Formality, Archaeology and the Semantic Web (poster). Web Science Conference 2010 (WebSci10), Raleigh, USA, 26-27 April 2010, http://eprints.soton.ac.uk/150319/

Isaksen L., Martinez K., Gibbins N., Earl G. & Keay S. (2009): Linking Archaeological Data, in: Frischer B. *et al*. (eds.): Making History Interactive. CAA 2009, Williamsburg, Virginia, Archaeopress, Oxford, pp. 130-136, http://proceedings.caaconference.org/files/2009/18_Isaksen_et_al_CAA2009.pdf

Isaksen L., Rainer S., de Soto Cañamares P. & Barker E.T (2016): Pelagios Commons: Decentralizing the Web of Historical Data. Presentation at CAA 2016 Oslo, session "Linked Pasts: Connecting Islands of Content", 30 March 2016 (paper forthcoming)

Isaksen L., Simon R., Barker E. & de Soto Cañamares P. (2014): Pelagios and the emerging graph of ancient world data, pp. 197-201, in: WebSci'14 - Proceedings of the 2014 ACM Conference on Web Science, Indiana University, Bloomington, 23-26 June 2014; preprint, http://www.researchgate.net/publication/266659779_Pelagios_and_the_emerging_graph_of_ancient_world_data

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

Ivanov, Vladimir (2011): The Open Kunstkammer Data Project. In: ERCIM News, 86, July 2011, 43-44, http://ercim-news.ercim.eu/images/stories/EN86/EN86-web.pdf (see also: http://data.kunstkamera.ru)

Jankowski J., Cobos Y., Hausenblas M. & Decker S. (2009): Accessing Cultural Heritage using the Web of Data [CHoWDer - Cultural Heritage on the Web of Data]. In: VAST'09 - 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage 2009, St. Julians, Malta, http://aran.library.nuigalway.ie/xmlui/bitstream/handle/10379/455/VAST2009-CHoWDer.pdf?sequence=1

Janowicz K., Hitzler P., Adams B., Kolas D. & Vardeman C. (2014): Five Stars of Linked Data Vocabulary Use. In: Semantic Web Journal, 5(3): 173-176; http://www.semantic-web-journal.net/content/five-stars-linked-data-vocabulary-use

Janowicz K., Scheider S., Pehle T. & Hart G. (2012): Geospatial Semantics and Linked Spatiotemporal Data - Past, Present, and Future. In: Semantic Web Journal, 3(4): 321-332; http://www.semantic-web-journal.net/sites/default/files/swj330_0.pdf

Janowicz, Krzysztof (2009): The Role of Place for the Spatial Referencing of Heritage Data. Workshop on The Cultural Heritage of Historic European Cities and Public Participatory GIS. University of York, September 2009, http://geog.ucsb.edu/~jano/chwy09_janowicz.pdf

Jansma, Esther (2013): Towards sustainability in dendroarchaeology: the preservation, linkage and reuse of tree-ring data from the cultural and natural heritage in Europe, pp. 169-176, in: Bleicher, Niels *et al.* (eds.): DENDRO - Chronologie - Typologie - Ökologie. Freiburg: Janus (paper available on www.academia.edu)

Jarrett J., Zambanini S., Hüber-Mork R. & Felicetti A. (2011): Coinage, Digitization and the World-Wide Web: Numismatics and the COINS Project. In: New Technologies in Medieval and Renaissance Studies 3, 459–489; preprint, https://www.academia.edu/2147548/Coinage_Digitization_and_the_World-Wide_Web_numismatics_and_the_COINS_Project

Jentzsch A., Cyganiak R. & Bizer C. (2011): State of the LOD Cloud, September 2011, http://lod-cloud.net/state/

Johnson T. & Estlund K. (2014): Recipes for Enhancing Digital Collections with Linked Data. In: Code4Lib Journal, Issue 23, 17 January 2014, http://journal.code4lib.org/articles/9214

Jones S., MacSween A., Jeffrey S., Morris R. & Heyworth M. (2001): From the ground up: The publication of archaeological projects: a user needs survey, http://www.britarch.ac.uk/pubs/puns

Jordal E., Uleberg E. & Hauge B. (2012): Was It Worth It? Experiences with a CIDOC CRM - based Database, pp. 255-260, in: CAA 2011 - Proceedings of the 39th Annual Conference of Computer Applications and Quantitative Methods in Archaeology, Beijing, China, 12-16 April 2011, http://proceedings.caaconference.org/files/2011/28_Jordal_et_al_CAA2011.pdf

JSON - JavaScript Object Notation, http://json.org

JSON-LD - JSON for Linking Data, http://json-ld.org

Kamps, Jaap (2015): When Search becomes Research and Research becomes Search. Keynote presentation at SIGIR'13 - Workshop on Exploration, Navigation and Retrieval of Information in Cultural Heritage (ENRICH) 1 August 2013, Dublin, Ireland, http://de.slideshare.net/jaap.kamps/sigir-workshop-enrich13

Kamura, Tetsuro *et al*. (2011): Building Linked Data for Cultural Information Resources in Japan. Proceedings of Museums and the Web 2011, http://www.museumsandtheweb.com/mw2011/papers/building_linked_data_for_cultural_information_.html

Kansa E. & Bissell A. (2010): Web syndication approaches for sharing primary data in "small science" domains. In: Data Science Journal, Volume 9: 42-53, https://www.jstage.jst.go.jp/article/dsj/9/0/9_009-012/_pdf

Kansa E. & Whitcher-Kansa S. (2011): Enhancing Humanities Research Productivity in a Collaborative Data Sharing Environment. White Paper to the NEH Division of Preservation and Access, 27 June 2011, http://alexandriaarchive.org/wp-content/uploads/2011/09/white_paper_PK_50072.pdf

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, 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 for Digital Curation, 9(1), http://www.ijdc.net/index.php/ijdc/article/view/9.1.57/341

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 (2014a): Open Context and Linked Data. ISAW Paper 7.10, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/kansa/

Kansa, Eric (2014b): Linked Data, Publication, and the Life Cycle of Archaeological Information. Presentation at 8. Deutscher Archäologiekongress, Berlin, 8 October 2014, http://www.ianus-fdz.de/attachments/download/697//06_Kansa_OpenContext.pdf

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

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

Kauppinen T., Baglatzi A. & Keßler C. (2013): Linked Science: Interconnecting Scientific Assets. In: Critchlow T. & Kleese-Van Dam K. (eds.): Data Intensive Science. CRC Press; preprint, http://linkedscience.org/wp-content/uploads/2012/02/linked-science-bookchapter-revised-2011-11-16.pdf

Kerameikos, http://kerameikos.org

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

Kiryakov A., Ognyanoff D., Velkov R., Tashev Z. & Peikov I. (2009): LDSR: Materialized Reason-able View to the Web of Linked Data. In: Proceedings of the 3rd International RuleML-2009 Challenge. Las Vegas, USA, http://ceur-ws.org/Vol-549/paper9.pdf

Kobilarov G., Scott T., Raimond Y. *et al*. (2009): Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections. In: L. Aroyo *et al*. (Eds.): ESWC 2009, LNCS 5554, Berlin and Heidelberg: Springer 2009, pp. 723–737, http://derivadow.files.wordpress.com/2009/06/eswc2009-bbc-dbpedia-2.pdf

Kondert F., Schandl T. & Blumauer A. (2011): Do controlled vocabularies matter? Survey results. Semantic Web Company, Vienna, June 2011, http://www.semantic-web.at/sites/default/files/files/Survey_Do_Controlled_Vocabularies_Matter_2011_June_0.pdf

Kosem I., Jakubiček M., Kallas J. & Krek S. (eds., 2015): Electronic Lexicography in the 21st Century: Linking Lexical Data in the Digital Age. Proceedings of eLex 2015 - Electronic Lexicography in the 21st century: Linking Lexical Data, Herstmonceux Castle, Sussex, UK, 11-13 August 2015, https://elex.link/elex2015/conference-proceedings/

Krueger, Kristi J. (2013): A Case Study of Assertions for the Iron Age and Implications for Temporal Metadata Creation. A Master's Paper for the M.S. in L.S. degree. University of North Carolina at Chapel Hill, April 2013, https://cdr.lib.unc.edu/record/uuid:a8f56c09-954c-45ca-931b-a7fc2bf51dd5

Lana, Maurizio (2014): Geolat: Geography for Latin Literature. ISAW Paper 7.11, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Lang M., Carver F. & Printz S. (2013): Standardised Vocabulary in Archaeological Databases, pp. 468-473, in: CAA 2012 Southampton, Volume II, Amsterdam University Press, http://dare.uva.nl/cgi/arno/show.cgi?fid=545855

Lange, A.G. (ed., 2004): Reference Collections. Foundation for Future Archaeology. Proceedings of the international conference on the European electronic Reference Collection, 12-13 May 2004, ROB, Amersfoort, The Netherlands, http://cultureelerfgoed.nl/sites/default/files/publications/reference-collections-foundation_for_future_archaeology.pdf

LATC - LOD Around The Clock (2012): Final Release of P&C Library. Project deliverable D3.3.1, 29 February 2012, http://latc-project.eu/node/89

LAWD - Linking Ancient World Data ontology, https://github.com/lawdi/LAWD

LAWDI - Linked Ancient World Data Institute (USA, NEH-funded project, 2012-2013), http://wiki.digitalclassicist.org/Linked_Ancient_World_Data_Institute

Le Cornec Rochelois C. & Issac F. (2015): What Terms to Express the Categories of Natural Sciences in the Dictionary of Medieval Scientific French?, pp. 29-42, in: SWASH 2016 - 1st Workshop on

Semantic Web for Scientific Heritage, Portoroz, Slovenia, 1 June 2015, http://ceur-ws.org/Vol-1364/sw4sh-2015.pdf

Le Goff E., Marlet O., Rodier X., Curet S. & Husi P. (2015): The interoperability of the ArSol (Archives du Sol) database: Based on the CIDOC-CRM ontology, pp. 179-186, in: CAA 2014 Paris - Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Paris, France, 22-25 April 2014, Archaeopress, http://www.archaeopress.com/ArchaeopressShop/Public/download.asp?id={5CACE285-4C48-41AE-809E-E98B65C9E4CD}

Ledl A. & Voß J. (2016): Describing Knowledge Organization Systems in BARTOC and JSKOS. In: 12th International Conference on Terminology and Knowledge Engineering (TKE 2016), Copenhagen, 22-24 June 2016; preprint, http://eprints.rclis.org/29366/1/Ledl_Voss_TKE2016_final_version_20160518.pdf

lemon - The Lexicon Model for Ontologies (see also: OntoLex model, Cimiano *et al*. 2016), http://lemon-model.net

LiAM - Linked Archival Metadata project (USA, 10/2012-9/2013, led by Tufts University, Digital Collections and Archives), http://sites.tufts.edu/liam/

Library Linked Data Incubator Group (2011): Datasets, Value Vocabularies, and Metadata Element Sets. W3C Incubator Group Report, 25 October 2011, http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/

Library of Congress: Linked Data Service, http://id.loc.gov

LIDER - LingHub - Linguistic Linked Open Data cloud, http://linghub.lider-project.eu/llod-cloud

LIDER - LingHub (language resources), http://linghub.lider-project.eu

LIDER - Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe (EU, FP7, 11/2013-12/2015, http://www.lider-project.eu

Limp, Fredrick W. (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

Lincoln, Matthew D. (2016): Linked Open Realities: The Joys and Pains of Using LOD for Research. In: Art History and Digital Research weblog, 6 June 2016, http://matthewlincoln.net/2016/06/06/linked-open-realities-the-joys-and-pains-of-using-lod-for-research.html

Linked Ancient World Data: Relating the Past (2016). Panel at Digital Humanities 2016 conference, Kraków, Poland, 11-16 July 2016, http://dh2016.adho.org/abstracts/262

Linked Heritage & Athena (2011): Your terminology as part of the Semantic Web. Recommendations for design and management. November 2011, http://www.linkedheritage.eu/getFile.php?id=244

Linked Heritage (EU, ICT-PSP, 2011-2013), http://www.linkedheritage.eu

Linked Open Vocabularies – LOV (Open Knowledge Foundation), http://lov.okfn.org

linkedarc.net, http://linkedarc.net; datasets, https://datahub.io/dataset/linkedarc

LinkedBrainz - MusicBrainz in RDF and SPARQL, http://linkedbrainz.org

LinkedDataTools - Free tools, information and resources for the semantic web, http://www.linkeddatatools.com

Linking Open Data cloud diagram, http://lod-cloud.net

Liuzzo, Pietro (2016): Mapping Epigraphic Databases to EpiDoc, pp. 149-162, in: Orlandi S. *et al*. (eds.): Digital and Traditional Epigraphy in Context. Proceedings of the Second EAGLE International Conference. Rome, 27-29 January 2016, http://www.eagle-network.eu/wp-content/uploads/2016/04/EAGLE%20D2.6_EAGLE%20Second%20International%20Conference%20Proceedings.pdf

Liuzzo, Pietro M. (2014): The Europeana Network of Ancient Greek and Latin Epigraphy (EAGLE). ISAW Paper 7.12, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

LOCAH - Linked Archives and Linking Lives projects (UK, JISC-funded, 2010-2012, Mimas and UKOL), http://locah.archiveshub.ac.uk

LOD Browser Switch (offers a set of browsers), http://browse.semanticweb.org

LOD2 - Creating Knowledge out of Interlinked Data (2011): State of the Art Analysis. Project deliverable 1.2, 16 January 2011, http://static.lod2.eu/Deliverables/deliverable-1.2.pdf

LOD2 - Creating Knowledge out of Interlinked Data (EU, FP7-ICT, 2010–2014), http://lod2.eu

LOD-LAM, the International LOD in Libraries, Archives, and Museums Summit, http://lodlam.net

LODStats (Agile Knowledge Engineering and Semantic Web Group at University of Leipzig, Germany), http://stats.lod2.eu

MacKay, Camilla (2014): Bryn Mawr Classical Review. ISAW Paper 7.13, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Madsen, Torsten (2004): Classification and archaeological knowledge bases, pp. 35-42, in: Lange, A.G. (ed.): Reference Collections. Foundation for Future Archaeology. Amersfoort, The Netherlands: ROB, http://cultureelerfgoed.nl/sites/default/files/publications/reference-collections-foundation_for_future_archaeology.pdf

Mantegari, Glauco (2009): Cultural heritage on the semantic web: From representation to fruition. Ph.D. dissertation, Università degli Studi di Milano-Bicocca, QUA SI Project, http://boa.unimib.it/bitstream/10281/9184/3/phd_unimib_708063.pdf

Mapping Memory Manager - 3M (facilitates the mapping of databases to the extended CIDOC CRM), Foundation for Research and Technology Hellas, Institute of Computer Science, http://www.ics.forth.gr/isl/3M

Marlet O., Curet S., Rodier X. & Bouchou-Markhoff B. (2016): Using CIDOC CRM for dynamically querying ArSol, a relational database, from the semantic web, pp. 241-249, in: CAA2015 - Keep the Revolution Going: Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology. Oxford: Archaeopress, http://archaeopress.com/ArchaeopressShop/Public/download.asp?id={77DEDD4E-DE8F-43A4-B115-ABE0BB038DA7}

MASA - Mémoire des Archéologues et des Sites Archéologiques, http://masa.hypotheses.org

Maturana R.A., Ortega M. & López-Sola S.(2013): Mismuseos.net: Art After Technology. Putting cultural data to work in a Linked Data platform. In: Proceedings of Veni 2013 - LinkedUp Veni Competition on Linked and Open Data for Education, Geneva, 17 September 2013, http://ceur-ws.org/Vol-1124/linkedup_veni2013_03.pdf

May K., Binding C. & Tudhope D. (2015): Barriers and opportunities for Linked Open Data use in archaeology and cultural heritage. In: Archäologische Informationen, Volume 38, http://journals.ub.uni-heidelberg.de/index.php/arch-inf/article/view/26162/19880

May K., Binding C. &Tudhope, D. (2010): Following a STAR? Shedding more light on semantic technologies for archaeological resources. Computer Applications and Quantitative Methods in Archaeology 2009 (BAR Int Ser 2079), 227-233, http://proceedings.caaconference.org/files/2009/28_May_et_al_CAA2009.pdf

May K., Binding C., Tudhope D. & Jeffrey S. (2011): Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources, pp. 261-272, in: CAA 2011 - Revive the Past. Proceedings of the 39th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Beijing, China, 12-16 April 2011, http://proceedings.caaconference.org/files/2011/29_May_et_al_CAA2011.pdf

May, Keith (2016): The Matrix: Connecting Time and Space with archaeological research questions involving spatio-temporal phenomena and the conceptual relationships between them. Presentation at CAA 2016 Oslo, session “Linked Pasts: Connecting Islands of Content”, 30 March 2016, http://de.slideshare.net/Keith.May/caa-2016-the-matrix-connecting-time-space

Mazzini S. & Ricci F. (2011): EAC-CPF Ontology and linked archival data, pp. 72–81, in: Proceedings of the 1st International Workshop on Semantic Digital Archives, 29 Sept 2011, Berlin, Germany. CEUR Workshop Proceedings, vol. 801, http://ceur-ws.org/Vol-801/paper6.pdf

McCrae J.P. & Cimiano P. (2015): Linghub: a Linked Data based portal supporting the discovery of language resources, pp. 88-91, in: SEMANTiCS2015 - 11th International Conference on Semantic Systems, Proceedings of the Posters and Demos Track, Vienna, Austria, 15-17 September 2015, http://ceur-ws.org/Vol-1481/paper27.pdf

McMichael, A. L. (2014): Byzantine Cappadocia: Small Data and the Dissertation. ISAW Paper 7.14, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Meadows A. & Gruber E. (2014): Coinage and Numismatic Methods. A Case Study of Linking a Discipline. ISAW Paper 7.15, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Meadows, Andrew (2015): Online Coins of the Roman Empire: An Open Resource for Roman Numismatics, December 2015, https://t.co/pKksMjf7qb

Meeks E. & Grossner K. (2012): ORBIS: An Interactive Scholarly Work on the Roman World. In: Journal of Digital Humanities, 1(3), http://journalofdigitalhumanities.org/1-3/orbis-an-interactive-scholarly-work-on-the-roman-world-by-elijah-meeks-and-karl-grossner/

Meroño-Peñuela A., Ashkpour A., Rietveld L., Hoekstra R. & Schlobach S. (2012): Linked Humanities Data: The next frontier? [census data]. In: Proceedings of LISC 2012 - 2nd International Workshop on Linked Science 2012, Boston, 12 November 2012, http://ceur-ws.org/Vol-951/paper3.pdf

Meroño-Peñuela A., Ashkpour A., van Erp M. *et al*. (2014): Semantic Technologies for Historical Research: A Survey. In: Semantic Web Journal, paper 588, http://www.semantic-web-journal.net/system/files/swj588_0.pdf

Meyers, Katy (2014): Exploring an Opportunity to Link the Dead in Ancient Rome. ISAW Paper 7.16, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Michel F., Montagnat J. & Faron-Zucker C. (2013): A survey of RDB to RDF translation approaches and tools. Equipes Modalis/Wimmics. Rapport de Recherche, ISRN I3S/RR, 2013-04-FR, Novembre

2013, https://hal.inria.fr/file/index/docid/903568/filename/Michel_Montagnat_Faron_2013_-_A_survey_of_RDB_to_RDF_translation_approaches_and_tools.pdf

Miller, Paul (2010): Linked Data Horizon Scan. Report commissioned by JISC. January 2010, http://cloudofdata.com/2010/02/final-version-of-linked-data-horizon-scan-now-available-online/

Minadakis N., Marketakis Y., Kondylakis H., Flouris G., Theodoridou M., Doerr M. & de Jong G. (2016): X3ML Framework: an effective suite for supporting data mappings, pp. 1-12, in: Ronzino, Paola (ed.): Extending, Mapping and Focusing the CRM. Proceedings of the EMF-CRM workshop, Poznan, Poland, 17 September 2015, http://ceur-ws.org/Vol-1656/paper1.pdf

MisMuseos.net: DataHub information, http://datahub.io/dataset/mismuseos-gnoss

Missikoff, Oleg (2004): Ontologies as a Reference Framework for the Management of Knowledge in the Archaeological Domain, pp. 35-39, in: Enter the Past. The E-Way Into the Four Dimensions of Cultural Heritage. ArcheoPress; preprint, https://publikationen.uni-tuebingen.de/xmlui/bitstream/handle/10900/60734/02_Missikoff_CAA_2003.pdf?sequence=2&isAllowed=y

Mitchell, Erik T. (2016): The Current State of Linked Data in Libraries, Archives, and Museums. In: ALA TechSource - Library Technology Reports, 52(1), chapter 1, https://journals.ala.org/ltr/article/view/5892/7446

MONDIS - Monument Damage Information System project (Czech Republic), http://www.mondis.cz

MoRe - Metadata & Object Repository aggregator (ATHENA, Digital Curation Unit, Greece), http://more.dcu.gr

Morgan E.L. *et al*. (2014): Linked Archival Metadata: A Guidebook. Version 0.99, 23 April 2014, http://sites.tufts.edu/liam/2014/04/24/version-099/

Morgan, Eric L. (2014): Linked Archival Metadata: Trends and gaps in linked data for archives. LiAM: Linked Archival Metadata, http://sites.tufts.edu/liam/2014/04/23/trends/

Mouromtsev D., Haase P., Cherny E., Pavlov D., Andreev A. & Spiridonova A. (2015): Towards the Russian Linked Culture Cloud: Data Enrichment and Publishing, pp. 637-651, in: The Semantic Web. Latest Advances and New Domains. Springer (LNCP 9088); preprint, http://metaphacts.com/images/Papers/Towards-the-Russian-Linked-Culture-Cloud.pdf

MULTITA - Coudyzer E. & Lheureux B. (2015): Multilingual terminological research (French, Dutch and English) for the development and integration of semantically enriched scientific thesauri (MULTITA). Summary of the research project, 30 January 2015, http://www.belspo.be/belspo/organisation/Publ/pub_ostc/agora/ragLL169sum_en.pdf

MULTITA - Multilingual terminological research (French, Dutch and English) for the development and integration of semantically enriched scientific thesauri (7/2012-12/2014), http://www.belspo.be/belspo/fedra/proj.asp?l=fr&COD=AG/LL/169

Mungall C.J., Torniai C., Gkoutos G.V., Lewis S.E. & Haendel M.A. (2012): Uberon, an integrative multi-species anatomy ontology. Genome Biology 13, R5, http://genomebiology.com/2012/13/1/R5

Murray, William (2014): RAM 3D Web Portal. ISAW Paper 7.17, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Musei Italiani, http://www.linkedopendata.it/datasets/musei

Museums and the Machine-processable Web wiki, edited by Mia Ridge, http://museum-api.pbworks.com/w/page/21933420/Museum%C2%A0APIs

National Museum of Ireland: Artefacts, http://www.museum.ie/en/list/artefacts.aspx

Natural Europe project (EU, ICT-PSP, 10/2010-09/2013), http://www.natural-europe.eu

NCBI Organismal Classification, https://bioportal.bioontology.org/ontologies/NCBITAXON

Ngonga Ngomo A.-C., Auer S., Lehmann J. & Zaveri A. (2014): Introduction to Linked Data and Its Lifecycle on the Web, pp. 1-99, in: Reasoning Web. Reasoning on the Web in the Big Data Era. Proceedings of the 10th International Summer School 2014, Athens, Greece, 8-13 September 2014. Springer (LNCS 8714); preprint, http://jens-lehmann.org/files/2014/reasoning_web_update_linked_data.pdf

Niccolucci F. & Hermon S. (2015): Time, chronology and classification, pp. 265-279, in: Barcelo J.A. & Bogdanovic I. (eds.): Mathematics and Archaeology. CRC Press

Niccolucci F. & Hermon S. (2016): Representing gazetteers and period thesauri in four-dimensional space–time. In: International Journal on Digital Libraries, 17(1): 63-69, http://link.springer.com/article/10.1007/s00799-015-0159-x

Niccolucci F., Hermon S. & Doerr M. (2015): The formal logical foundations of archaeological ontologies, pp. 86-99, in: Barcelo J.A. & Bogdanovic I. (eds.): Mathematics and Archaeology. CRC Press

Nikolov A. & d'Aquin M. (2011): Identifying Relevant Sources for Data Linking using a Semantic Web Index. LDOW2011, Hyderabad, India, 29 March 2011, http://ceur-ws.org/Vol-813/ldow2011-paper10.pdf

Nikolov A., d'Aquin M. & Motta E. (2012): What should I link to? Identifying relevant sources and classes for data linking, pp. 284-299, in: JIST2011 - Joint International Semantic Technology Conference. The Semantic Web. Springer (LNCS 7185); preprint, http://people.kmi.open.ac.uk/andriy/jist2011.pdf

NKOS Task Group of the Dublin Core Metadata Initiative (2015): KOS Types Vocabulary, 2015-10-02, http://wiki.dublincore.org/index.php/NKOS_Vocabularies

Nomisma ontology and numismatics datasets, http://nomisma.org

Nouvel B. & Sinigaglia E. (2014): PACTOLS, un thésaurus pour décrire les ressources documentaires en archéologie. MASA Consortium, weblog, 17 November 2014, http://masa.hypotheses.org/116; slides: https://f.hypotheses.org/wp-content/blogs.dir/1718/files/2014/11/01_PACTOLS_MASA20141013.pdf

Nouvel, Blandine (2015): Des outils d'enrichissement documentaire multilingues pour l'archéologie. MASA weblog, 14 December 2015, http://masa.hypotheses.org/date/2015/12

Nowak K. & Bon B. (2015): medialatinitas.eu. Towards Shallow Integration of Lexical, Textual and Encyclopaedic Resources for Latin, pp. 152-169, in: Proceedings of eLex 2015 - Electronic Lexicography in the 21st century: Linking Lexical Data, Herstmonceux Castle, Sussex, UK, 11-13 August 2015, https://elex.link/elex2015/proceedings/eLex_2015_10_Nowak+Bon.pdf

Nurmikko-Fuller, Terhi (2014): Assessing the Suitability of Existing OWL Ontologies for the Representation of Narrative Structures in Sumerian Literature. ISAW Paper 7.18, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Nußbaumer P. & Haslhofer B. (2007): CIDOC CRM in Action – Experiences and Challenges. Poster at the 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL07), Budapest, http://eprints.cs.univie.ac.at/403/1/cidoc_crm_poster_ecdl2007.pdf

Nußbaumer P., Haslhofer B. & Klas W. (2010): Towards Model Implementation Guidelines for the CIDOC Conceptual Reference Model. Technical Report TR-201. University of Vienna, http://eprints.cs.univie.ac.at/58/

OCLC - Online Computer Library Center: Linked Data, http://oclc.org/developer/develop/linked-data.en.html

Oldman D. & Rahtz S. (2014): Aligning the Academy with the Cultural Heritage Sector through the CIDOC CRM and Semantic Web technology, p. 80, in: CAA 2014 Paris, Book of abstracts, http://caa2014.sciencesconf.org/conference/caa2014/pages/BOACAA_2016.pdf

Oldman D., Doerr M. & Gradmann S. (2015): ZEN and the Art of Linked Data. New Strategies for a Semantic Web of Humanist Knowledge, Chapter 18 in Schreibman S., Siemens R. & Unsworth J. (eds.): A New Companion to Digital Humanities. Blackwell; preprint, https://www.academia.edu/12608990/ZEN_and_the_Art_of_Linked_Data_New_Strategies_for_a_Semantic_Web_of_Humanist_Knowledge

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

Olsson, Carl A. (2016): A Linked (Open) Data hub at the Norwegian Directorate for Cultural Heritage – a case study. Presentation at CAA 2016 Oslo, session “Linked Pasts: Connecting Islands of Content”, 30 March 2016 (paper forthcoming)

Omelayenko, Borys (2008): Porting Cultural Repositories to the Semantic Web, pp. 14-35, in: Kollias S.& Cousins J. (eds.): Semantic Interoperability in the European Digital Library. Proceedings of the First International Workshop, SIEDL 2008, Tenerife, 2 June 2008, http://image.ntua.gr/swamm2006/SIEDLproceedings.pdf

ONKI - Finnish Ontology Library Service, http://onki.fi

Online Coins of the Roman Empire (OCRE), http://numismatics.org/ocre/

ONTOCOM - Ontology Cost Estimation with ONTOCOM, http://ontocom.sti-innsbruck.at

Ontop, platform to query databases as Virtual RDF Graphs using SPARQL (University of Bozen-Bolzano, KRDB research group), http://ontop.inf.unibz.it

Oomen J., Baltussen L.-B. & Van Erp M. (2012): Sharing cultural heritage the linked open data way: why you should sign up. In: Museums and the Web 2012, San Diego, 11-14 April 2012, http://www.museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data

Open Annotation Collaboration, http://www.openannotation.org

Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH), http://www.openarchives.org/pmh/

Open Context: Linked data projects, http://alexandriaarchive.org/projects/linked-data/

Open Data Barometer (international survey of open governmental data), http://opendatabarometer.org

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

OpenRefine, http://openrefine.org

ORBIS - The Stanford Geospatial Network Model of the Roman World, http://orbis.stanford.edu

Ordnance Survey (UK), http://data.ordnancesurvey.co.uk

Orlandi S., Santucci R., Casarosa V. & Liuzzo P.M. (2014): Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the First EAGLE International Conference, Paris, http://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf

PACTOLS - Peuples, Anthroponymes, Chronologie, Toponymes, Oeuvres, Lieux et Sujets (thesaurus), http://frantiq.mom.fr/thesaurus-pactols

Page, Roderic (2009): Semantic Publishing: towards real integration by linking. iPhylo weblog, 20 April 2009, http://iphylo.blogspot.co.at/2009/04/semantic-publishing-towards-real.html

Pan X., Schiffer T., Hecher M. *et al*. (2012a): A scalable repository infrastructure for CH digital object management. In: 18th International Conference on Virtual Systems and Multimedia, Milan, Italy, September 2012, http://havemann.cgv.tugraz.at/Publications/2012_PSHx12__ScalableRepositoryInfrastructureForCHObjectManagement.pdf

Pan X., Schiffer T., Schröttner M. *et al*. (2012b): An enhanced distributed repository for working with 3d assets in cultural heritage. In: 4th International Euro-Mediterranean Conference on Digital Heritage (EuroMed), Limassol, Cyprus, October 2012. Springer LNCS, http://link.springer.com/chapter/10.1007%2F978-3-642-34234-9_35

Parry R., Poole N. & Pratty J. (2008): Semantic Dissonance: Do We Need (and Do We Understand) the Semantic Web? Proceedings of Museums and the Web Conference 2008, http://www.archimuse.com/mw2008/papers/parry/parry.html

PATHS - Personalised Access to Cultural Heritage Spaces (EU, FP7 project, 01/2011-12/2013), http://www.paths-project.eu

Patroumpas K., Alexakis M., Giannopoulos G. & Athanasiou S. (2014): TripleGeo: an ETL Tool for Transforming Geospatial Data into RDF Triples, pp. 275-278, in: Proceedings of the Workshops of the EDBT/ICDT 2014 Joint Conference, Athens, Greece, 28 March 2014, http://ceur-ws.org/Vol-1133/paper-44.pdf

Pearce L. & Schmitz P. (2014): Berkeley Prosopography Services. ISAW Paper 7.19, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Pelagios project, http://commons.pelagios.org

Pelagios: Joining Pelagios, https://github.com/pelagios/pelagios-cookbook/wiki/Joining-Pelagios

Pena Serna S., Schmedt H., Ritz M. & Stork A. (2012): Interactive Semantic Enrichment of 3D Cultural Heritage Collections. In: VAST'12 - The 13th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, Brighton, UK, http://culturalinformatics.org.uk/files/papers/InteractiveSemanticEnrichment2012.pdf

Pena Serna S., Scopigno R., Doerr M. *et al*. (2011): 3D-centred media linking and semantic enrichment through integrated searching, browsing, viewing and annotating. VAST11: 12th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage, Prato, Italy (not openly available online)

PeriodO - Periods, Organized project, http://perio.do

Pett, Daniel (2014a): Linking Portable Antiquities to a wider web. ISAW Paper 7.20, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/pett/

Pett, Daniel (2014b): Making the links to Portable Antiquities Scheme data. In: CAA 2014 Paris, Book of abstracts, p.81, http://f.hypotheses.org/wp-content/blogs.dir/1309/files/2014/04/CAA2014-BOA-S07-20140424.pdf

Pett, Daniel (n.d.): Implementing Linked Data within the Portable Antiquities Scheme, https://www.academia.edu/9347715/Implementing_Linked_Data_within_the_Portable_Antiquities_Scheme

PICO thesaurus (Central Institute for the Union Catalogue - ICCU, Italy, http://purl.org/pico/thesaurus_4.2.0.skos.xml

Pitti D.V., Popovici B.F., Stockting W. & Clavaud F. (2014): Experts Group on Archival Description: Interim Report. Girona 2014: Arxius I Industries Culturals. Girona 2014: Arxius i Indústries Culturals, Girona, Spain, 11-15 October 2014, http://www.girona.cat/web/ica2014/ponents/textos/id56.pdf

Placenames Database of Ireland, http://www.logainm.ie/en/

PlanetData (2012): Conceptual model and best practices for high-quality metadata publishing. Project deliverable D2.1, http://planet-data-wiki.sti2.at/web/File:D2.1.pdf

Pleiades - Gazetteer of the Ancient World, http://pleiades.stoa.org

Poehler, Eric (2014): Pompeii Bibliography and Mapping Resource. ISAW Paper 7.21, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Portable Antiquities Scheme, http://finds.org.uk

Portnoy, David (2014): What Happened to the Semantic Web? September 2014, http://david.portnoy.us/what-happened-to-the-semantic-web/

PricewaterhouseCoopers (2009): Technology Forecast. Spring 2009, http://www.pwc.com/us/en/technology-forecast/spring2009/

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/

Raimond Y., Smethurst M., McParland A. & Lowis C. (2013): Using the Past to Explain the Present: Interlinking Current Affairs with Archives via the Semantic Web, pp. 146-161, in: The Semantic Web – ISWC 2013, 12th International Semantic Web Conference, Sydney, 21-25 October2013, Part II, Springer (LNCS 8219); preprint, http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP260.pdf

Rakhmawati N.A., Umbrich J., Karnstedt M., Hasnain A. & Hausenblas M. (2013): Querying over Federated SPARQL Endpoints|A State of the Art Survey. DERI Technical Report 2013-06-07, June 2013, http://www.deri.ie/sites/default/files/publications/1306.1723v1.pdf

Reinhard, Andrew (2014): Publishing Archaeological Linked Open Data: From Steampunk to Sustainability. ISAW Paper 7.23, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

ReLoad - Repository for Linked Open Archival Data (Italy, 2010-2013, Archivio Centrale dello Stato, Istituto per i Beni culturali dell'Emilia-Romagna and regesta.exe), http://labs.regesta.com/progettoReload/

ReLoad (2013): Project description for LODLAM 2013 summit, http://summit2013.lodlam.net/2012/12/01/challenge-entry-reload-repository-for-linked-open-archival-data/

ResearchSpace - Creating the Cultural Heritage Knowledge Graph project (British Museum), http://www.researchspace.org

Richards J., Tudhope D. & Vlachidis A. (2015): Text Mining in Archaeology: Extracting Information from Archaeological Reports, pp. 240-254, in: Barcelo J. & Bogdanovic I. (eds.): Mathematics in Archaeology. CRC Press; preprint, https://pure.york.ac.uk/portal/en/publications/text-mining-in-archaeology-extracting-information-from-archaeological-reports%28ef5831ea-4a00-4996-b225-ba53cf9019cf%29.html

Richards, Julian (2006): Archaeology, e-publication and the Semantic Web. In: Antiquity, 80(310): 970-979, http://core.ac.uk/download/pdf/50930.pdf

RightField - Semantic data annotation by Stealth, http://www.rightfield.org.uk

Rodriguez Echavarria K., Theodoridou M., Georgis C. *et al*. (2012): Semantically rich 3D documentation for the preservation of tangible heritage. In: VAST'12 - 13th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, Brighton, UK, http://culturalinformatics.org.uk/files/papers/SemanticallyRich3D2012.pdf

Romanello, Matteo (2012): SKOSifying an Archaeological Thesaurus. In: Computers for the Classes weblog, 8 October 2012, https://c4tc.wordpress.com/2012/10/08/skosifying-an-archaeological-thesaurus/

Romanello, Matteo (2014): Mining Citations, Linking Texts. ISAW Paper 7.24, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Ronzino P., Amico N., Felicetti A. & Niccolucci F. (2013): European standards for the documentation of historic buildings and their relationship with CIDOC-CRM, pp. 70-79, in: CRMEX 2013 – Workshop: Practical Experiences with CIDOC CRM and its Extensions, co-located with TPDL 2013, Valetta, Malta, 26 September 2013, http://ceur-ws.org/Vol-1117/CRMEX2013.pdf

Ronzino P., Niccolucci F., Felicetti A. & Doerr M. (2016): CRMba, a CRM extension for the documentation of standing buildings. In: International Journal on Digital Libraries, 17(1): 71-78, http://link.springer.com/article/10.1007%2Fs00799-015-0160-4

Ronzino, Paola (2015): CIDOC CRMba – A CRM extension for building archaeology information modelling. Presentation at CIDOC-CRM SIG, 32nd joint meeting, Oxford University e-Research Centre, 11 February 2015, http://www.cidoc-crm.org/docs/32nd-meeting-presentations/CRMBA_Paola%20Ronzino_32SIG.pdf

Ronzino, Paola (2015): CIDOC CRMba: A CRM extension for buildings archaeology information modelling. Unpublished PhD thesis, The Cyprus Institute, Cyprus, January 2015

Ross S., Ballsun-Stanton B., Sobotkova A. & Crook P. (2015): Building the Bazaar: Enhancing Archaeological Field Recording Through an Open Source Approach, pp. 111-129, in: Wilson A.T. & Edwards B. (eds.): Open Source Archaeology: Ethics and Practice. Walter de Gruyter, https://www.degruyter.com/downloadpdf/books/9783110440171/9783110440171-009/9783110440171-009.xml

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

Ross, Seamus (2003): Position Paper, pp. 7-11, in: DigiCULT Thematic Issue 3: Towards a Semantic Web for Heritage Resources. Salzburg, May 2003, http://www.digicult.info/downloads/thematic_issue_3_low.pdf

Ross, Shawn (2015): Creating Interoperable Digital Datasets: the Federated Archaeological Information Management Systems (FAIMS) Project. Presentation at Mobilizing the Past for a Digital Future: the Potential of Digital Archaeology, Wentworth Institute of Technology, Boston, 27-28 February 2015, http://uwm.edu/mobilizing-the-past/sample-page-2/

Roueché C., Lawrence K. & Lawrence K.F. (2014): Linked Data and Ancient Wisdom. ISAW Paper 7.25, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Sahoo S., Halb W., Hellmann S. *et al*. (2009): A Survey of Current Approaches for Mapping of Relational Databases to RDF. W3C RDB2RDF Incubator Group, W3C, 2009. http://esw.w3.org/Rdb2RdfXG/StateOfTheArt

Samwald, Matthias (2010): Comments to "Why Carry the Cost of Linked Data?". Tom Heath weblog, 17 June 2010, http://tomheath.com/blog/2010/06/why-carry-the-cost-of-linked-data/

Schaible J., Gottron T. & Scherp A. (2014): Extended Description of the Survey on Common Strategies of Vocabulary Reuse in Linked Open Data Modeling. Universität Koblenz-Landau, Arbeitsberichte aus dem Fachbereich Informatik, Nr. 1/2014, http://www.uni-koblenz.de/~fb4reports/2014/2014_01_Arbeitsberichte.pdf

Scheidel, Walter (2015): ORBIS: the Stanford geospatial network model of the Roman world. Princeton/Stanford Working Papers in Classics, May 2015, http://orbis.stanford.edu/assets/Scheidel_64.pdf

Schmachtenberg M., Bizer C. & Paulheim H. (2014a): State of the LOD Cloud 2014, Version 0.4, 30 August 2014, http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

Schmachtenberg M., Bizer C. & Paulheim H. (2014b): Adoption of the Linked Data Best Practices in Different Topical Domains, pp. 245-260, in: The Semantic Web – ISWC 2014. Lecture Notes in Computer Science 8796, http://dws.informatik.uni-mannheim.de/fileadmin/lehrstuehle/ki/pub/SchmachtenbergBizerPaulheim-AdoptionOfLinkedDataBestPractices.pdf

Schröttner M., Havemann S., Theodoridou M. *et al*. (2012): A generic approach for generating cultural heritage metadata. 4th International Euro-Mediterranean Conference on Digital Heritage (EuroMed), Limassol, Cyprus, October 2012, Springer LNCS; https://www.semanticscholar.org/paper/A-Generic-Approach-for-Generating-Cultural-Schr%C3%B6ttner-Havemann/9e8d6f5201f153e4c03e066745967734a8fb5c2c

Sebastian Cuy S., Schmidle W. & Thiery F. (2016): Linking periods: Modeling and utilizing spatio-temporal concepts in the chronOntology project. Presentation at CAA 2016 Oslo, session "Linked Pasts: Connecting Islands of Content", 30 March 2016, https://www.academia.edu/24845165/Linking_periods_Modeling_and_utilizing_spatio-temporal_concepts_in_the_chronOntology_project

Segers R., Van Erp M., van der Meij L. *et al*. (2011): Hacking history: Automatic historical event extraction for enriching cultural heritage multimedia collections. Proceedings of the 6th International Conference on Knowledge Capture (K-CAP'11), http://ceur-ws.org/Vol-779/derive2011_submission_18.pdf

Seifreid, Rebecca (2014): Linked Open Data for the Uninitiated. ISAW Paper 7.26, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Semantic Computing Research Group (SeCo), Aalto University, Finland, http://seco.cs.aalto.fi

Semanticweb.org: List of Semantic Annotation tools, http://semanticweb.org/wiki/Category:Semantic_annotation_tool

Semanticweb.org: Semantic Wiki projects, http://semanticweb.org/wiki/Semantic_wiki_projects

SEMIC - Semantic Interoperability Community, https://joinup.ec.europa.eu/community/semic/description

SEMLIB - Semantic Tools for Digital Libraries (EU FP7-SME project), http://www.semlibproject.eu

SemWebQuality.org (provides information and tools about data quality in Semantic Web architectures), http://semwebquality.org

SENESCHAL - Semantic Enrichment Enabling Sustainability of Archaeological Links (UK AHRC-funded project, 2013-2014), http://hypermedia.research.glam.ac.uk/kos/SENESCHAL/; see also: http://www.heritagedata.org/blog/about-heritage-data/seneschal/

Shadbolt N., Berners-Lee T. & Hall W. (2006): The Semantic Web Revisited. IEEE Intelligent Systems, vol. 21, no. 3, pp. 96-101, http://eprints.soton.ac.uk/262614/1/Semantic_Web_Revisted.pdf

Sibille de Grimoüard, Claire (2014): Archives and Linked Data: Are our tools ready to 'complete the picture'? Girona 2014: Arxius i Indústries Culturals, Girona, Spain, 11-15 October 2014, http://www.girona.cat/web/ica2014/ponents/textos/id9.pdf

Signore, Oreste (2009): Representing knowledge in archaeology: from cataloguing cards to semantic web. In: Archeologia e Calcolatori, no. 20, 111-128, http://soi.cnr.it/archcalc/indice/PDF20/10_Signore.pdf

Simon R., Barker E., de Soto P. & Isaksen L. (2014): Pelagios. ISAW Paper 7.27, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Simon R., Barker E., Isaksen L. & de Soto Cañamares P. (2015): Linking Early Geospatial Documents, One Place at a Time: Annotation of Geographic Documents with Recogito. In: e-Perimetron, 10(2): 49-59, http://oro.open.ac.uk/43613/1/Simon_et_al.pdf

Simon R., Haslhofer B. & Jung J. (2011): Annotations, Tags & Linked Data - Metadata Enrichment in Online Map Collections through Volunteer-Contributed Information. 6th International Workshop on Digital Approaches in Cartographic Heritage The Hague, Netherlands, 7-8 April 2011, http://eprints.cs.univie.ac.at/2849/1/Simon_et_al._-_CartoHeritage_2011.pdf

Simon R., Isaksen L., Barker E. & de Soto Cañamares P. (2016a): Peripleo: a Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time. In: Code4Lib Journal, Issue 31, http://journal.code4lib.org/articles/11144

Simon R., Isaksen L., Barker E. & de Soto Cañamares P. (2016b): The Pleiades Gazetteer and the Pelagios Project. In: Berman M.L., Mostern R. & Southall H. (eds.): Placing Names: Enriching and Integrating Gazetteers. Indiana University Press, http://www.iupress.indiana.edu/product_info.php?cPath=1037_1116_3767&products_id=808056

Simov K. & Kiryakov A. (2015): Accessing Linked Open Data via a Common Ontology, pp. 33-41, In: Proceedings of the Second Workshop on Natural Language Processing and Linked Open Data, Hissar, Bulgaria, 11 September 2015, https://aclweb.org/anthology/W/W15/W15-5506.pdf

Simperl E., Bürger T., Hangl S. Wörgl S. & Popov I. (2012): ONTOCOM: A Reliable Cost Estimation Method for Ontology Development Projects. In: Journal of Web Semantics, Vol. 16, 1-16; preprint, http://www.websemanticsjournal.org/index.php/ps/article/viewFile/320/320

Sinclair, P.A.S. *et al*. (2005): Concept browsing for multimedia retrieval in the SCULPTEUR project. In: Proceedings of the 2nd Annual European Semantic Web Conference, Heraklion, Crete, http://eprints.soton.ac.uk/260913/1/eswc.pdf

SITAR - Sistema Informativo Territoriale Archeologico di Roma, http://www.archeositarproject.it

Skevakis G., Makris K., Arapi P. & Christodoulakis S. (2013): Elevating Natural History Museums' Cultural Collections to the Linked Data Cloud. Proceedings of the 3rd International Workshop on Semantic Digital Archives (SDA), in conjunction with TPDL 2013, http://ceur-ws.org/Vol-1091/paper4.pdf

Smith, Marcus J. (2015): The Digital Archaeological Workflow: A Case Study from Sweden, pp. 215-220, in: CAA 2014 Paris - Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Paris, France, 22-25 April 2014, Archaeopress, http://www.archaeopress.com/ArchaeopressShop/Public/download.asp?id={5CACE285-4C48-41AE-809E-E98B65C9E4CD}

Smith-Yoshimura, Karen (2014a): Linked Data Survey results 1 – Who's doing it. In: Hangingtogether.org OCLC Research weblog, 4 September 2014, http://hangingtogether.org/?p=4137

Smith-Yoshimura, Karen (2014b): Linked Data Survey results 2 – Examples in production. In: Hangingtogether.org OCLC Research weblog, 4 September 2014, http://hangingtogether.org/?p=4147

Smith-Yoshimura, Karen (2014c): Linked Data Survey results 3 – Why and what institutions are consuming. In: Hangingtogether.org OCLC Research weblog, 4 September 2014, http://hangingtogether.org/?p=4155

Smith-Yoshimura, Karen (2014d): Linked Data Survey results 4 – Why and what institutions are publishing. In: Hangingtogether.org OCLC Research weblog, 4 September 2014, http://hangingtogether.org/?p=4167

Smith-Yoshimura, Karen (2014e): Linked Data Survey results 5 – Technical details. In: Hangingtogether.org OCLC Research weblog, 4 September 2014, http://hangingtogether.org/?p=4256

Smith-Yoshimura, Karen (2014f): Linked Data Survey results 6 - Advice from the implementers. In: Hangingtogether.org OCLC Research weblog, 4 September 2014, http://hangingtogether.org/?p=4284

Smith-Yoshimura, Karen (2014g): Linked Data Survey results (results spreadsheet), https://groups.google.com/forum/#!topic/lod-lam/9ZR1FUvPntM

Smith-Yoshimura, Karen (2015): Results of Linked Data Surveys for Implementers. Responses 2014 and 2015 (data sheet), http://oc.lc/0bglX7

Smith-Yoshimura, Karen (2016): Analysis of International Linked Data Survey for Implementers. In: D-Lib Magazine, 22(7/8), http://dx.doi.org/10.1045/july2016-smith-yoshimura

SNAC - Social Networks and Archival Context project (USA, 2010-ongoing, Institute for Advanced Technology in the Humanities, University of Virginia), http://socialarchive.iath.virginia.edu

SNAP - Standards for Networking Ancient Prosopographies (UK, AHRC funded project, 2014-2015), http://snapdrgn.net

Solanki, Monika (2009): Semantic web in Cultural Heritage and Archaeology. W3C Semantic Web, Tracing Networks Workshop 2009, University of Leicester, 13 November 2009, http://de.slideshare.net/nimonika/semantic-web-in-cultural-heritage-and-archaeology

Souza R., Almeida M.B. & Tudhope D. (2010): The KOS spectra: a tentative typology of Knowledge Organization Systems. ISKO 2010 conference, Rome, 23-26 February 2010, http://mba.eci.ufmg.br/downloads/ISKO%20Rome%202010%20submitted.pdf

Souza R., Tudhope D. & Almeida M.B. (2012): Towards a taxonomy of KOS: dimensions for classifying knowledge organization systems. In: Knowledge Organization, 39(3): 179-192; preprint, http://mba.eci.ufmg.br/downloads/Souza_Tudhope_Almeida_-_KOS_Taxonomy.Submitted.pdf

Spampinato D. & Zangara I. (2013): Classical Antiquity and Semantic Content Management on Linked Open Data. In: 1st International Workshop on Collaborative Annotations in Shared Environment: Metadata, Vocabularies and Techniques in the Digital Humanities, Florence, 10 September 2013 (presentation), http://www.cs.unibo.it/dh-case/pdf/Zangara.pdf

Stadler C., Lehmann J., Höffner K. & Auer S. (2012): LinkedGeoData: A Core for a Web of Spatial Open Data. In: Semantic Web Journal, 3(4): 333-354 http://jens-lehmann.org/files/2012/linkedgeodata2.pdf

STAR - Semantic Technologies for Archaeological Resources (UK, AHRC-funded project, 2007-2010), http://hypermedia.research.southwales.ac.uk/kos/star/

STELLAR - Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources project (UK, AHRC-funded project, 2010-2011), http://hypermedia.research.southwales.ac.uk/kos/stellar/

STELLAR Applications (Hypermedia Research Unit, University of South Wales), http://hypermedia.research.southwales.ac.uk/resources/STELLAR-applications/

Stevenson M., Otegi A. *et al*. (2013): Semantic Enrichment of Cultural Heritage Content in PATHS. PATHS project, http://www.paths-project.eu/eng/content/download/5102/38896/file/SemanticEnrichment.pdf

Stevenson, Jane (2011): Putting the Case for Linked Data. LOCAH Project weblog, 12 July 2011, http://locah.archiveshub.ac.uk/2011/07/12/putting-the-case-for-linked-data/

Stevenson, Jane (2012) Linking Lives: Creating An End-User Interface Using Linked Data. In: Information Standards Quarterly, 24(2/3): 14-23, http://dx.doi.org/10.3789/isqv24n2-3.2012.03

Studer R. & Sure Y. (2006): Cost Estimation in Ontology Engineering. IST, Helsinki, November 22, 2006, slide 7, ftp://ftp.cordis.europa.eu/pub/ist/docs/kct/cost-estimation-in-ontology-engineering_en.pdf

Suominen O., Pessala S., Tuominen J. *et al*. (2014): Deploying National Ontology Services: From ONKI to Finto. In: ISWC 2014 - 13th International Semantic Web Conference, Industry Track, Riva del Garda, Italy, http://ceur-ws.org/Vol-1383/paper6.pdf

Swedish National Heritage Board (2014): Lista med lämningstyper och rekommenderad antikvarisk bedömning. Version 4.1, 2014-06-26, http://www.raa.se/app/uploads/2014/07/L%C3%A4mningstypslistan_ver-4_1_20140626.pdf

Swedish Open Cultural Heritage (K-samsök): http://www.ksamsok.se/in-english/

Szabados, Anne-Violaine (2014): From the LIMC Vocabulary to LOD. Current and Expected Uses of the Multilingual Thesaurus TheA, pp. 51-67, in: Orlandi S. *et al*. (2014): Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the First EAGLE International Conference, Paris, http://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf

Szekely P., Knoblock C.A., Yang F. *et al*. (2013): Connecting the Smithsonian American Art Museum to the Linked Data Cloud. ESWC 2013 (LNCS 7882, Springer), 593-607, http://www.isi.edu/~szekely/contents/papers/2013/eswc-2013-saam.pdf

Taylor, Jon (2014): Linked data and the future of cuneiform research at the British Museum. ISAW Paper 7.28, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

TDWG - Biodiversity Information Standards, http://www.tdwg.org

TEI - Text Encoding Initiative, http://www.tei-c.org/index.xml

Thiery F. & Engel T. (2016): The Labeling System: A bottom-up approach for enriched vocabularies in the humanities, pp. 259-268, in: CAA2015 Siena - Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology. Oxford: Archaeopress, http://archaeopress.com/ArchaeopressShop/Public/download.asp?id={77DEDD4E-DE8F-43A4-B115-ABE0BB038DA7}

Thiery, Florian (2014): Linking potter, pots and places: a LOD approach to samian ware. Poster presented at CAA 2014 Paris, https://www.academia.edu/6782320/Linking_potter_pots_and_places_a_LOD_approach_to_samian_ware

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

Tolle K. & Wigg-Wolf D. (2016): How To Move from Relational to 5 Star Linked Open Data – A Numismatic Example, pp. 275-281, in: CAA2015 Siena - Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Volume 1, Oxford: Archaeopress, http://archaeopress.com/ArchaeopressShop/Public/download.asp?id={77DEDD4E-DE8F-43A4-B115-ABE0BB038DA7}

Toms, Elaine G. (2015): Complex Tools for Complex Tasks. In: Proceedings of the First International Workshop on Supporting Complex Search Tasks (SCST 2015), Vienna, Austria, 29 March 2015. CEUR Workshop Proceedings 1338, http://ceur-ws.org/Vol-1338/paper_8.pdf

Tounsi M., Faron Zucker C., Zucker A., Villata S. & Cabrio E. (2015): Studying the History of Pre-Modern Zoology with Linked Data and Vocabularies, pp. 7-14, in: SWASH 2016 - 1st Workshop on Semantic Web for Scientific Heritage, Portoroz, Slovenia, 1 June 2015, http://ceur-ws.org/Vol-1364/sw4sh-2015.pdf

Tree of Life (TOL) project, http://tolweb.org/tree/

Tsonev, Tsoni (2014): Integrating Historical-Geographic Web-Resources. ISAW Paper 7.29, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

Tudhope D., Binding C., Jeffrey S., May K. & Vlachidis A. (2011a): A STELLAR role for knowledge organisation systems in digital archaeology. ASIS&T Bulletin, 37(4): 15-18, http://www.asis.org/Bulletin/Apr-11/AprMay11_Tudhope_etAl.pdf

Tudhope D., Binding C., May K. & Charno M. (2013): Pattern based mapping and extraction via the CRM(-EH), pp. 23-36, in: CRMEX 2013 – Workshop: Practical Experiences with CIDOC CRM and its Extensions, co-located with TPDL 2013, Valetta, Malta, 26 September 2013, http://ceur-ws.org/Vol-1117/CRMEX2013.pdf

Tudhope D., May K., Binding C. & Vlachidis A. (2011b): Connecting archaeological data and grey literature via semantic cross search. Internet Archaeology, Issue 30, http://intarch.ac.uk/journal/issue30/tudhope_index.html

Tzompanaki K. & Doerr M. (2012): Fundamental categories and relationships for intuitive querying CIDOC-CRM based repositories. Technical Report ICS-FORTH/TR-429, April 2012, http://www.cidoc-crm.org/docs/TechnicalReport429_April2012.pdf

UBERON – Uber Anatomy Ontology, http://uberon.org; see also: https://bioportal.bioontology.org/ontologies/UBERON

Unsworth J. (2000): Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This? Symposium on Humanities Computing: formal methods, experimental practice. King's College, London, 13 May 2000, http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html

Unsworth, John (2002): What is Humanities Computing and What is not? In: Forum Computerphilologie, 8 November 2002, http://computerphilologie.uni-muenchen.de/jg02/unsworth.html

van de Sompel H., Lagoze C., Nelson M.L. *et al*. (2009): Adding e-science assets to the data web. Linked Data on the Web (LDOW2009), Madrid, Spain, 20 April 2009, http://events.linkeddata.org/ldow2009/papers/ldow2009_paper8.pdf ; see also arXiv:0906.2135v1 [cs.DL], http://arxiv.org/abs/0906.2135

van der Meij L., Isaac A. & Zinn C. (2010): A web-based repository service for vocabularies and alignments in the cultural heritage domain. Proceedings of the 7th European Semantic Web Conference, Heraklion, Greece, 30 May-3 June 2010, 394–409, http://www.few.vu.nl/~AI.Isaac/papers/STITCH-Repository-ESWC10.pdf

van Erp M., Oomen J., Segers R. *et al*. (2011): Automatic heritage metadata enrichment with historic events. Proceedings of Museums and the Web 2011, http://www.museumsandtheweb.com/mw2011/papers/automatic_heritage_metadata_enrichment_with_hi

van Hooland S. & Verborgh R. (2014): Linked Data for Libraries, Archives and Museums. How to clean, link and publish your metadata. Facet Publishing, http://book.freeyourmetadata.org van Hooland S., De Wilde M., Verborgh R., Steiner T. & Van de Walle R. (2015): Exploring Entity Recognition and Disambiguation for Cultural Heritage Collections? In: Literary and Linguistics Computing, 30(2): 262-279; preprint, http://freeyourmetadata.org/publications/named-entity-recognition.pdf

van Hooland S., Verborgh R. & Van de Walle R. (2012a): Joining the Linked Data Cloud in a Cost-Effective Manner. In: 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., Hercher J., Mannens E. & Van de Walle R. (2012b): Evaluating the success of vocabulary reconciliation for cultural heritage collections. In: Journal of the American Society for Information Science and Technology, Vol. 64: 464–479; authors' paper, May 2012, http://freeyourmetadata.org/publications/freeyourmetadata.pdf

Van Keer, Ellen (2014): Moving from Cross-Collection Integration to Explorations of Linked Data Practices in the Library of Antiquity at the Royal Museums of Art and History, Brussels. ISAW Paper 7.30, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

van Ossenbruggen J., Hildebrand M. & de Boer V. (2011): Interactive vocabulary alignment. TPDL 2011 - International Conference on Theory and Practice of Digital Libraries, Berlin, Germany, 26-28 September 2011, http://semanticweb.cs.vu.nl/lod/tpdl2011/paper.pdf (see also the use case replicability documentation here: http://semanticweb.cs.vu.nl/lod/tpdl2011/)

Vandenbussche P.-Y., Atemezing G.A., Poveda-Villalón M. & Vatant B. (2015): Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web. In: Semantic Web Journal, version 29/09/2015, http://www.semantic-web-journal.net/system/files/swj1178.pdf

Vatant, Bernard (2012): Is your linked data vocabulary 5-star? In: Bernard Vatant weblog, 10 February 2012, http://bvatant.blogspot.fr/2012/02/is-your-linked-data-vocabulary-5-star_9588.html

Vavliakis K.N., Karagiannis G.T. & Mitkas P.A. (2012): Semantic Web in Cultural Heritage after 2020. Workshop on "What will the Semantic Web look like 10 years from now?" held in conjunction with the 11th International Semantic Web Conference 2012 (ISWC 2012), Boston, USA, 11 November 2012, http://stko.geog.ucsb.edu/sw2022/sw2022_paper10.pdf

Vences M., Guayasamin J.M., Miralles A. & De la Riva I. (2013): To name or not to name: Criteria to promote economy of change in Linnaean classification schemes. In: Zootaxa, 3636(2): 201–244, http://biotaxa.org/Zootaxa/article/view/zootaxa.3636.2.1/1556

VIAF - Virtual International Authority File, http://viaf.org

Vici.org - Archaeological Atlas of Antiquity, http://vici.org

Villazón-Terrazas B. & Corcho O. (2011): Methodological Guidelines for Publishing Linked Data. Ontology Engineering Group, Computer Science School, Polytechnic University of Madrid, http://delicias.dia.fi.upm.es/wiki/images/7/7a/07_MGLD.pdf

Vlachidis A. & Tudhope D. (2011): Semantic Annotation for Indexing Archaeological Context: A Prototype Development and Evaluation. In: Metadata and Semantic Research (Communications in Computer and Information Science, Vol. 240): 363-374; preprint, http://hypermedia.research.glam.ac.uk/media/files/documents/2011-10-26/MTSR2011_Vlachidis_A-SemanticAnnoations-Camera_Ready.pdf

Vlachidis A. & Tudhope D. (2013a): Classical Art Semantics Information Extraction: CASIE Pilot Project. Conference of the British Chapter of the International Society for Knowledge Organization (ISKO UK 2013), London, http://www.iskouk.org/conf2013/papers/VlachidisPaper.pdf

Vlachidis A. & Tudhope D. (2013b): The Semantics of Negation Detection in Archaeological Grey Literature, pp. 188-200, in: Garoufallou E. & Greenberg J. (eds.): Metadata and Semantics Research Communications in Computer and Information Science, Vol. 390; preprint, http://hypermedia.research.glam.ac.uk/media/files/documents/2015-04-28/The_Semantics_of_Negation_Detection_Camera_Ready.pdf

Vlachidis A. & Tudhope D. (2015a): A knowledge-based approach to Information Extraction for semantic interoperability in the archaeology domain. In: Journal of the Association for Information Science and Technology, 67(5): 1138-52, http://onlinelibrary.wiley.com/doi/10.1002/asi.23485/abstract

Vlachidis A. & Tudhope D. (2015b): Negation detection and word sense disambiguation in digital archaeology reports for the purposes of semantic annotation. Program: electronic library and information systems, 49(2): 118-134, http://www.emeraldinsight.com/doi/abs/10.1108/PROG-10-2014-0076

Vlachidis A., Binding C., May K. & Tudhope D. (2010): Excavating grey literature: a case study on the rich indexing of archaeological documents via Natural Language Processing techniques and knowledge based resources. In: ASLIB Proceedings, 62(4&5): 466-475; preprint, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.551.1066&rep=rep1&type=pdf

Vlachidis A., Binding C., May K. & Tudhope D. (2013): Automatic Metadata Generation in an Archaeological Digital Library: Semantic Annotation of Grey Literature, pp. 187-202, in: Przepiórkowski, Adam *et al*. (eds.): Computational Linguistics – Studies in Computational Intelligence 458. Springer; preprint, http://hypermedia.research.glam.ac.uk/media/files/documents/2011-11-02/Automatic_Metadata_Generation.pdf

Vlachidis, Andreas (2012): Semantic Indexing via Knowledge Organization Systems: Applying the CIDOC-CRM to Archaeological Grey Literature. PhD Thesis, University of South Wales, http://hypermedia.research.southwales.ac.uk/media/files/documents/2013-07-11/Andreas-Vlachidis_Thesis_print_ready.pdf

VOAF - Vocabulary of a Friend, http://lov.okfn.org/vocommons/voaf/v2.3/

Vocabulary Mapping Framework (VMF), http://www.doi.org/VMF/

Vocabulary Matching Tool (Hypermedia Research Group, University of South Wales, UK), http://heritagedata.org/vocabularyMatchingTool/; source code for local download and installation, https://github.com/cbinding/VocabularyMatchingTool

W3C (2001-2013) Semantic Web Activity, http://www.w3.org/2001/sw/

W3C (2004) Recommendation: Architecture of the World Wide Web (Volume 1), 15 December 2004, http://www.w3.org/TR/webarch/#identification

W3C (2008) Interest Group Note: Cool URIs for the Semantic Web, 3 December 2008, http://www.w3.org/TR/cooluris/

W3C (2008) Working Group Note: Best Practice Recipes for Publishing RDF Vocabularies, 28 August 2008, https://www.w3.org/TR/swbp-vocab-pub/

W3C (2009) Recommendation: Simple Knowledge Organization System (SKOS) - Reference, 18 August 2009, http://www.w3.org/2004/02/skos/

W3C (2011) Interest Group Note: Describing Linked Datasets with the VoID Vocabulary, 3 March 2011, http://www.w3.org/TR/void/

W3C (2012) Recommendation: OWL 2 - Web Ontology Language Document - Overview (Second Edition), 11 December 2012, https://www.w3.org/TR/2012/REC-owl2-overview-20121211/

W3C (2012): OWL - Web Ontology Language – Current status, http://www.w3.org/standards/techs/owl#w3c_all

W3C (2013) Recommendation: SPARQL 1.1 Federated Query, 21 March 2013, http://www.w3.org/TR/sparql11-federated-query/

W3C (2013) Working Group Note: ADMS - Asset Description Metadata Schema, 1 August 2013, http://www.w3.org/TR/2013/NOTE-vocab-adms-20130801/

W3C (2013) Working Group Note: RDFa 1.1 Primer: Rich Structured Data Markup for Web Documents (second edition), 22 August 2013, http://www.w3.org/TR/xhtml-rdfa-primer ; see also: http://rdfa.info

W3C (2013): SPARQL - Current Status, http://www.w3.org/standards/techs/sparql#w3c_all

W3C (2013-ongoing) Data Activity - Building the Web of Data, https://www.w3.org/2013/data/

W3C (2014) Recommendation: DCAT - Data Catalog Vocabulary, 16 January 2014, http://www.w3.org/TR/vocab-dcat/

W3C (2014) Recommendation: RDF 1.1 Concepts and Abstract Syntax, 25 February 2014, https://www.w3.org/TR/rdf11-concepts/

W3C (2014) Recommendation: RDF Schema 1.1, 25 February 2014, http://www.w3.org/TR/rdf-schema/

W3C (2014) Working Group Note: Best Practices for Publishing Linked Data, 9 January 2014, https://www.w3.org/TR/ld-bp/

W3C (2015) Editor's Draft: Data on the Web Best Practices Use Cases & Requirements, 27 March 2015, https://www.w3.org/TR/dwbp-ucr/

W3C (2015): Resource Description Framework (RDF) - Current Status, http://www.w3.org/standards/techs/rdf#w3c_all

W3C website: List of Tagging tools, http://www.w3.org/2001/sw/wiki/Category:Tagging

W3C website: Semantic Web tools (full list): http://www.w3.org/2001/sw/wiki/SemanticWebTools

W3C wiki: Converter to RDF, http://www.w3.org/wiki/ConverterToRdf

W3C wiki: Tools, http://www.w3.org/2001/sw/wiki/Tools

Wallis, Richard (2012): What Is Your Data's Star Rating(s)? Dataliberate.com, 18 January 2012, http://dataliberate.com/2012/01/what-is-your-datas-star-ratings/

Wang S., Isaac A., Schlobach S. *et al.* (2012): Instance-based Semantic Interoperability in the Cultural Heritage. Semantic Web Journal, 3(1), Special Issue on Semantic Web and Reasoning for Cultural Heritage and Digital Libraries, pp. 45-64, http://www.few.vu.nl/~AI.Isaac/publications.php

Wells J.J., Kansa E., Yerka S.J. *et al.* (2014): Web-based discovery and integration of archaeological historic properties inventory data: The Digital Index of North American Archaeology (DINAA). In: Literary and Linguistic Computing, 3(29): 349-360; https://www.academia.edu/11450026/Web-based_discovery_and_integration_of_archaeological_historic_properties_inventory_data_The_Digital_Index_of_North_American_Archaeology_DINAA_

Wester, Jeroen and Nederbragt, Hans (2007): RNA-project: Using things like thesauri and taxonomies in real cases!, pp. 93-99, in: Aroyo, L., Hyvönen, E. and van Ossenbruggen, J. (2007): Cultural Heritage on the Semantic Web. Workshop 9 of the 6th International Semantic Web Conference, Korea, 2007 http://www.cs.vu.nl/~laroyo/CH-SW/ISWC-wp9-proceedings.pdf

Whitcher-Kansa, Sarah (2015): Using Linked Open Data to Improve Data Reuse in Zooarchaeology. In: Ethnobiology Letters, 6(2): 224-231, http://ojs.ethnobiology.org/index.php/ebl/article/view/467/254

Wickett K.M., Isaac A., Doerr M. *et al*. (2014): Representing Cultural Collections in Digital Aggregation and Exchange Environments. In: D-Lib Magazine, 20(5-6), May/June 2014, http://www.dlib.org/dlib/may14/wickett/05wickett.html

Wiljes C., Jahn N., Lier F. *et al*. (2013): Towards Linked Research Data: An Institutional Approach. 3rd Workshop on Semantic Publishing (SePublica), CEUR Workshop Proceedings, Aachen: 27–38, http://ceur-ws.org/Vol-994/paper-03.pdf

Wilson, Scott (2014): Preserving and Curating Software. OSS Watch website, guidance material, 5 November 2014, http://oss-watch.ac.uk/resources/preservation

Wolstencroft K., Owen S., Horridge M. *et al*. (2011): RightField: Embedding ontology annotation in spreadsheets. In: Bioinformatics 27(14): 2021-22, http://bioinformatics.oxfordjournals.org/content/27/14/2021.full

Wolstencroft, Katy (2012): RightField: Semantic Enrichment of Systems Biology Data using Spreadsheets (myGrid, SysMO-DB, University of Manchester). Presentation at IEEE-Escience 2012, Chicago, USA, 11 October 2012, https://seek.sysmo-db.org/presentations/61/download

Wood D., Zaidman M., Ruth L. with Hausenblad M. (2014): Linked Data. Structured Data on the Web. Shelter Island, NY: Manning, http://www.manning.com/dwood/

World Geodetic System 1984 (WGS 84), http://earth-info.nga.mil/GandG/wgs84/

Wright, Holly (2011): Seeing Triple. Archaeology, Field Drawing and the Semantic Web. PhD Dissertation. The University of York, Department of Archaeology, September 2011, http://etheses.whiterose.ac.uk/2194/1/WrightThesis.pdf

Yu C.-H. (2010): Semantic Annotation of 3D Digital Representation of Cultural Artefacts. Bulletin of IEEE Technical Committee on Digital Libraries (TCDL), vol. 6, issue.2, http://www.ieee-tcdl.org/Bulletin/v6n2/Yu/yu.html

Zaino, Jennifer (2013): Art lovers will see there’s more to love with linked data. Semanticweb.com, 21 June 2013, https://semanticweb.com/art-lovers-will-see-theres-more-to-love-with-linked-data_b38088#more-38088

Zaveri A., Rula A., Maurino A., Pietrobon R., Lehmann J. & Auer S. (2013): Quality Assessment for Linked Open Data: A Survey. Semantic Web Journal, 556, http://www.semantic-web-journal.net/system/files/swj556.pdf

Zeng M.L. & Žumer M. (2013): A Metadata Application Profile for KOS Vocabulary Registries. ISKO UK Biennial Conference: Knowledge Organization – pushing the boundaries, London, 8-9 July 2013, http://www.iskouk.org/sites/default/files/ZengPaper_1.pdf

Zeng M.L. & Žumer M. (2015): Networked Knowledge Organization Systems Dublin Core Application Profile (NKOS AP), 2015-10-03, http://nkos.slis.kent.edu/nkos-ap.html

Zhang Y., Ogletree A., Greenberg J. & Rowel C. (2015): Controlled Vocabularies for Scientific Data: Users and Desired Functionalities. In: 2015 Annual Meeting of the Association for Information Science & Technology, St. Louis, USA, 6-10 November 2015; preprint, https://wakespace.lib.wfu.edu/bitstream/handle/10339/57209/zhang-ogletree-greenberg-rowell-controlled-vocabularies-for-scientific-data-preprint.pdf

Zimmermann, Antoine (2010): Ontology recommendation for the data publishers. ORES-2010 - Proceedings of the 1st Workshop on Ontology Repositories and Editors for the Semantic Web, Hersonissos, Crete, Greece, May 31st, 2010, http://ceur-ws.org/Vol-596/paper-12.pdf

ZOOMATHIA: Transmission culturelle des savoirs zoologiques (Antiquité-Moyen Âge): discours et techniques, http://www.cepam.cnrs.fr/zoomathia/

Zuiderwijk A., Jeffery K. & Janssen M. (2012): The potential of metadata for linked open data and its value for users and publishers. In: JeDEM - eJournal of eDemocracy and Open Government, 4(2): 222-244, http://www.jedem.org/index.php/jedem/article/view/138/113