D2.1 First Report on Users' Needs

Version 1.3

April 2014

Grant Agreement number:	313193
Project acronym:	ARIADNE
Project title:	Advanced Research Infrastructure for Archaeological Dataset Networking in Europe
Funding Scheme:	FP7-INFRASTRUCTURES-2012-1
Project co-ordinator:	Franco Niccolucci, PIN Scri - Polo Universitario "Città di Prato"
Tel:	+39 0574 602578
E-mail:	franco.niccolucci@gmail.com
Project website address:	http://www.ariadne-infrastructure.eu

Versions	Nr.	Authors & changes made	Date
Draft	1.0	H. Selhofer (SRFG) – version for first review	11-03-2014
Draft	1.1	G. Geser (SRFG) – amendments & revised conclusions	13-03-2014
	1.2	H. Selhofer (SRFG) – editorial amendments, version for final review	20-03-2014
Final	1.3	G. Geser / H. Selhofer (SRFG) – amendments in response to comments made by internal reviewers	07-04-2014

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

About this document

This document is a contractual deliverable of the ARIADNE research project (D2.1). The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-INFRASTRUCTURES-2012-1) under grant agreement n° 313193.

Partner in charge of the deliverable:	Salzburg Research Forschungsgesellschaft mbH (SRFG)
Authors:	Hannes Selhofer (SRFG) Guntram Geser (SRFG)
Contributor:	Ruth Beusing, DAI
Support (questionnaire design, survey roll-out, comments on survey results)	AIAC – International Association of Classical Archaeology Institute of Archaeology ZRC SAZU, Slovenia INRAP – Institut National de Recherches Archéologiques Préventives Hungarian National Museum National Heritage Protection Centre (HNM NHPC) University of York, Archaeology Data Services (ADS), UK MDR Partners Ltd., UK Data Archiving and Networked Services at the Netherlands Academy of Arts and Sciences (KNAW-DANS) University of Gothenburg, SND OEAW – Austrian Academy of Sciences Athena Research and Innovation Centre, Cultural and Educational Technology Institute (ATHENA RC - CETI), Greece Institute of Heritage Sciences (INCIPIT-CSIC), Spain
Interviewees	We would like to express our sincere gratitude and appreciation to all researchers and data managers who participated in an interview or sent written answers to our questions. The information obtained from these interviews was extremely helpful both for developing the online survey and for putting the results into perspective. The results of the interviews are documented in Section 5 and Annex IV.
Quality review:	H. Wright and J. Richards, ADS

Table of contents

1 Executive summary

1.1 About this deliverable

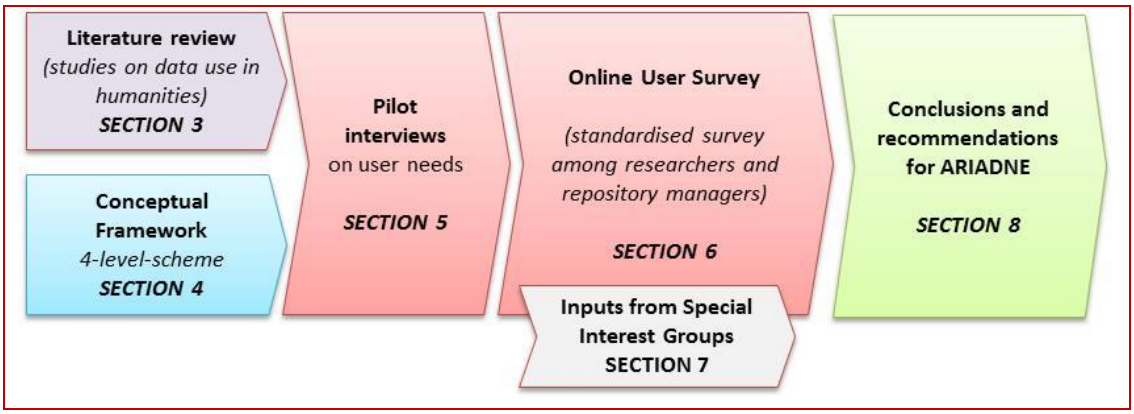
Background and objectives

This document is a deliverable (D2.1) 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 first results of the work carried out in Task 2.1 "Survey of users' needs and community building"; an update and refinement of the results will be presented in D2.2 towards the end of the second project year.

The main objectives of this report are to provide evidence on user requirements of key target groups ("users") of the ARIADNE project with regard to archaeological research data, and to contribute to building a prospective user community (see *Section 2.1*). This information shall support the ARIADNE project in taking informed decisions regarding the specification of the e-infrastructure and services so that the integrated infrastructure is developed in a way that corresponds to perceived and actual research needs. The mandate was to provide evidence on these issues, notably through collecting feed-back from the user communities by way of a user survey.

Methodology

The empirical research that was carried out to provide evidence on user requirements consisted of a literature review, 26 expert interviews (partly carried out as personal interviews, partly as written interviews) and an international online survey among about 700 representatives of the main target communities. In addition, some of the ARIADNE Special Interest Groups provided inputs regarding requirements for specific types of research data. The collected evidence was used to develop a set of recommendations for the ARIADNE project (see *Section 2.2*).



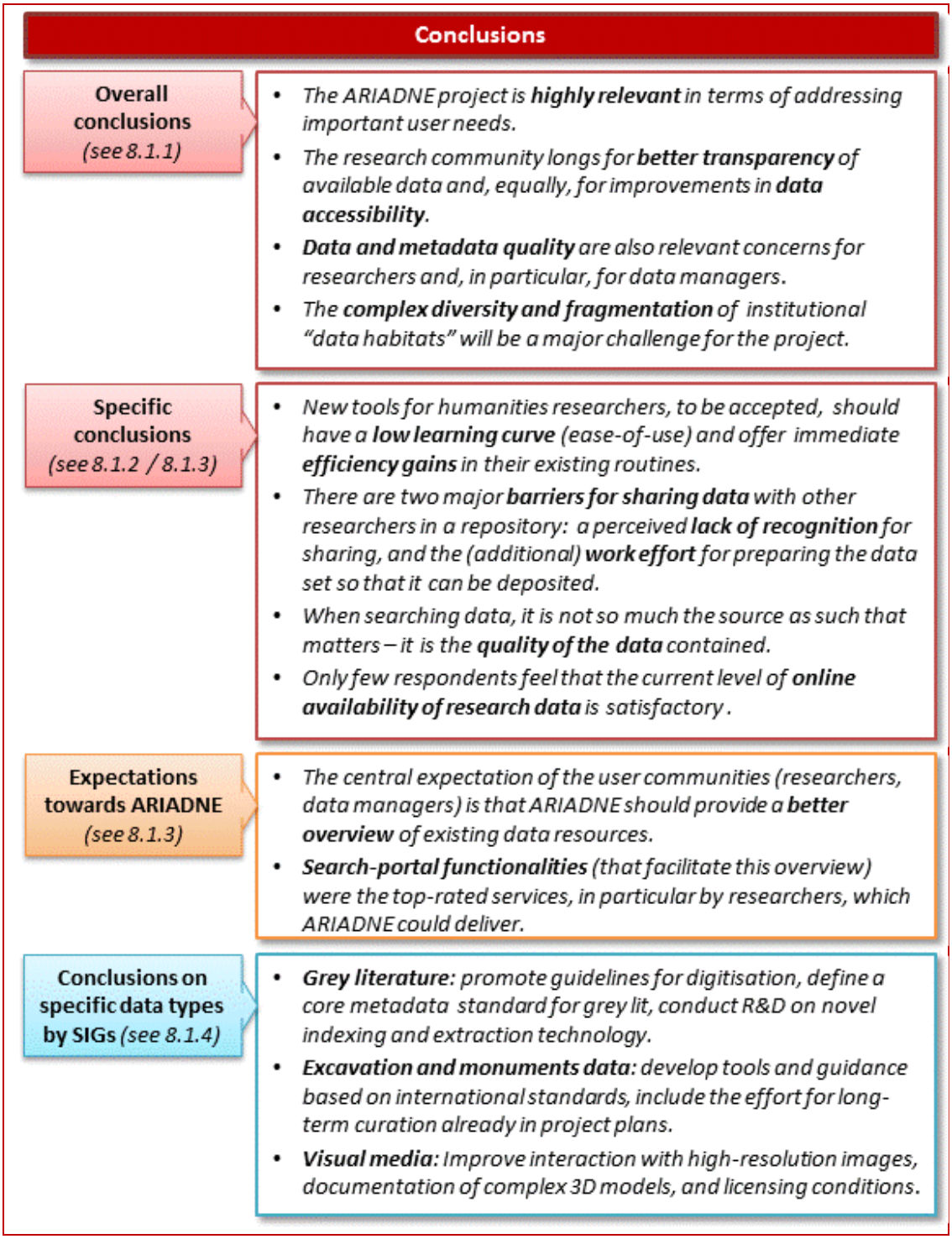
1.2 Conclusions

Conclusions in overview

The analysis of user requirements as documented in this report clearly confirms the **high relevance of the ARIADNE project**. It addresses important user needs with regard to research data which are not well catered for by existing services. More than 60% of the researchers surveyed said they were not or less satisfied with the current situation with regard to major parameters; in particular, they criticised a lack of transparency of available

data, and difficulties in having access to data. Any improvements in these areas would be highly appreciated by the user community.

At the same time, our analysis also confirms the **significant challenge** which ARIADNE is confronted with in its mission to create an e-infrastructure with services that respond to these needs: our evidence documents the enormous degree of fragmentation with regard to potentially relevant data for integration, presented by a complex diversity of institutional data habitats and different types of "repositories". To link these project archives with a common repository will require new workflows (and possibly dedicated staff) which may not be available in many research institutions.



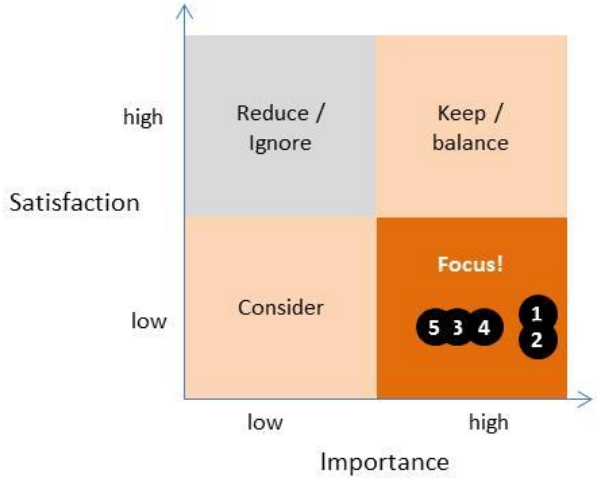
General conclusions on user requirements

The research community expresses, in particular, a need for an improved **transparency** of available research data (it is difficult to know which data actually exists, due to the enormous fragmentation of data resources in the field) and improvements in data **accessibility**. Major barriers with regard to accessibility are costs (e.g. for obtaining licences to use pictures, for subscription fees) and the problem that relevant literature and data is often kept in other places than where it is supposed to be (e.g. in private collections of other researchers). **Data and metadata quality** are further concerns of researchers; any improvements in these fields would be highly welcome.

Essentially, this means that ARIADNE has a broad field of opportunities to create real value for users. While it is clear that the project cannot solve all problems, ARIADNE has a high impact potential if its services can deliver improvement in any of the above mentioned areas (see matrix – all five domains of user requirements are in the segment which suggests focusing on them).

However, in order to take a **strategic decision** on priority areas, and to facilitate the choice and design of technical solutions, a further analysis of specific user requirements in the five domains is required. This will be the main goal for the updated version of this report in D2.2 (see *Section 2.3*), where requirements will be further broken down and explored with lead users among researchers and repository managers.

User requirements of archaeological researchers according to their importance and the satisfaction with the existing situation in a strategy matrix



User needs as shown in the matrix:

(1) = Data transparency

(2) = Data accessibility

(3) = Metadata quality

(4) = Data quality

(5) = International dimension of available data

Conclusions on the fragmentation of the research data landscape

We found in the survey of data managers a complex diversity of "data habitats", comprising a variety of organisational and institutional mechanisms and regimes under which research data is collected, archived and maintained. This includes project-level repositories or databases (e.g. regional or city level, single site, digital corpus of artefacts); single institutes (e.g. research centres, museums); supra-institutional data centres, heritage authorities and related services at county, province or national levels. Moreover, there is also a broad variety of different types of repositories.

The fragmentation is probably further reinforced by project-focused practices in research. The pilot interviews and the survey comments demonstrate that researchers have a predominantly project-centred rather than institutional perspective on data. As a result, the major formats how to organise data are "project archives" (one per excavation site) and "database projects" of small research groups or even a single researcher. The format of "collection" is much less common. Unfortunately, this high degree of fragmentation does

not represent a favourable condition for linking and integration. It will be a critical success factor for the ARIADNE project to develop a mechanism for how data from these fragmented "habitats" can possibly be linked with the e-infrastructure in a feasible, efficient way.

In practice, ARIADNE will probably have to focus on data sets that are already available in existing institutional repositories and national data centres and, on the other hand, promote the flow of currently "dark data" into the repositories or centres. When developing a strategy, ARIADNE will have to consider opportunities for an alignment with close scholarly neighbours such as classics, medieval studies, epigraphy or iconology. They are relevant both as providers and users of data and knowledge resources, as they have produced a multitude of digital materials.

Conclusions from the literature review

Our literature review focused on studies which explored data and tool requirements of humanities researchers in general and (to the extent that literature was available) on archaeologists in particular. In most reports humanities scholars are understood to present a special case, because the type of "data" they are working with (cultural content like texts and images) tends to be quite different than the data used in natural sciences. Archaeologists are closer to natural sciences; both in terms of the data they use and produce, and with regard to the methods they apply in their research.

Studies on **tool requirements of humanities scholars** suggest that they prefer "incremental" (but easy to achieve) improvements over potentially disruptive (but costly) innovations. They state a need for "simple-to-use" tools that facilitate their daily research routines with a low learning curve, i.e. without having to invest a lot of time in getting familiar with the new tools. The focus is thus on efficiency gains and fast return-on-investment.

As for **archaeologists**, their "dream tools" are mostly related to the initial phase of data collection and for presenting project outcomes. The ideal tool should be capable of accommodating existing practices and vocabularies, be highly flexible (i.e. have the ability to customise various things easily), and enable the provision of comprehensive project information. Overall, is seems that archaeologists are rather **reserved towards significant changes** in existing research designs, workflows or vocabulary.

The most **commonly stated specific needs** (with regard to new, improved tools) are searching across distributed resources and filtering hit returns more effectively. Not appreciated, by contrast, are online collections which are pre-culled by others.

Conclusions from the survey and the interviews – the researchers' perspective

Data sources: the online accessibility of research data is not satisfactory. While online publications (in particular with supplemental data) are a very important source of data, only few survey respondents felt that the online availability of research data was satisfactory. Most researchers perceived the following criteria as "very important" for data sources:

- Data transparency: having a good overview of available data – 77%
- Data accessibility: the required data is available in an uncomplicated way – 73%
- Data quality: the available data is complete and well organised – 64%

Types of data used and generated in research projects. The single most important type of data (if measured by the number of researchers for whom it is important) is excavation data (75% said it was "very important" for them). Also very important for a large group of researchers are GIS data, data stemming from material or biological analysis, and data from field surveys. These types of data are also the ones most frequently produced by researchers. However, the feedback indicates that it is not the source as such that matters – it is the quality of the data contained.

Main barriers for sharing data: Whether to share data with others or not is a very important issue for researchers. Survey results and literature on this topic indicate that research data may not only be scattered across different institutional databases, but a good deal of data might not even make it to the institutional database but remain on the computers of the individual researchers. Thus, the publishing and sharing of data in national data archives or international repositories is not yet a common practice. Three main challenges were identified for depositing data in an online repository:

- a lack of recognition for sharing the data;
- the work effort for preparing and documenting the data set so that data can be deposited in an institutional repository;
- (in some cases) a lack of opportunity.

This was also confirmed by the data managers who participated in the survey when asked to assess the barriers for sharing research data from their perspective and presents a big issue for the ARIADNE project, as it presents a barrier that undermines the rationale of the project (see also recommendation on this issue).

Metadata creation at project level: Most researchers seem not to worry as much about metadata as data managers do and, consequently, often do not produce metadata for the various data (data sets) they generate in projects. Researchers are concerned that producing adequate metadata is a considerable additional effort. To allow for effective data sharing, these additional efforts (costs) would have to be covered somehow (e.g. by research grants).

Conclusions from the survey and the interviews – the data managers' perspective

Metadata quality is the major challenge: The major challenge data managers see themselves confronted with in their daily work is ensuring metadata quality. This was by far the most important item out of a list of six challenges. Further relevant challenges are how to manage a rising number of data sets, and changes in the regulatory framework. Technical innovation and user-driven challenges (changes in user requirements, rising number of users) are perceived as much less critical by many data managers. According to several comments received on this question, the sustainability of project-based repositories and costs for operation and further development might be further key issues.

Technology is not the (major) issue: Most managers have a rather stable data management environment, and mainly carry out some upgrading and refinement of services, sometimes in response to external demands (e.g. new regulations). The main themes with regard to important recent technical developments include: database functionality, data standards and exchange protocols, spatial data (GIS), Digital Object Identifiers (DOIs). According to the pilot interviews with data centre managers, there is an increasing demand for DOIs because researchers need to link publications with deposited data.

Three worlds of access rules: About one third of the repositories have an "open access" policy (no registration required) for all, most or at least some data sets, while another third grant access only on request (and not necessarily to all or most data). Furthermore there is a "shadow world" of access restricted for specific communities only. With regard to depositing there is a clear preference of most repositories not to allow uncontrolled deposits by anybody but based on request.

Rising expectations of customers: Data managers observed two important trends regarding the needs of their customers: rising expectations towards the convenience in accessing and downloading data, and with regard to individual service and guidance.

Expectations towards ARIADNE

The main areas where researchers face problems are finding and accessing relevant data, including the usefulness of the available data. Therefore, major expectations and hopes towards the ARIADNE project are that the resulting services can improve the transparency of what is available, the search capability and, possibly, the conditions of access (e.g. promote open access repositories).

Improving transparency: A central expectation is that ARIADNE should provide a broad overview of existing data resources, beyond the partners' resources. The current approach is the ARIADNE Registry (based on the DCAT standard), which could be a stumbling block for many holders of small as well as large and various collections interested in making their resources visible through ARIADNE. In order to create a broad overview another way to "register" many data resources may be required. In general, geo-spatial/GIS/map-based overviews and access may be perceived as particularly suitable.

Capability for cross-searching data repositories was one of the services given a high priority by the respondents. We assume that this is one of the main advantages users of the ARIADNE portal will expect.

Improved conditions of access: This is not a technical requirement but a research policy objective, which the ARIADNE project can support by promoting open access principles as well as leading by example ("walk the talk").

Filtering "useful" and re-useable resources: Concerning data that are accessible online, researchers mentioned that they are sometimes not as useful as they could be, because data is structured in different ways, not up to date, incomplete or lack important details (e.g. how collected or processed). Moreover, a lot of data are not re-usable but "canned content" (such as data tables in pdf documents) or not available under an adequate license.

Portal service portfolio and specific user requirements: Respondents suggested that ARIADNE should establish a new portal for data search. If such a new portal (on top of existing data resources) is established, users will clearly expect an added-value – i.e. it must have other and better features, or provide access to more resources. While an improved overview, cross-searching and filtering of data resources would be quite some progress on the current situation, the specific requirements are not fully clear, however.

Services for repositories and other websites: Thinking beyond individual users, ARIADNE should also be seen as a service for data repositories, other websites and for specific communities of practitioners. ARIADNE might help enrich services of underlying repositories for instance by suggesting (and providing) links to similar or complementary collections or individual items held by other repositories.

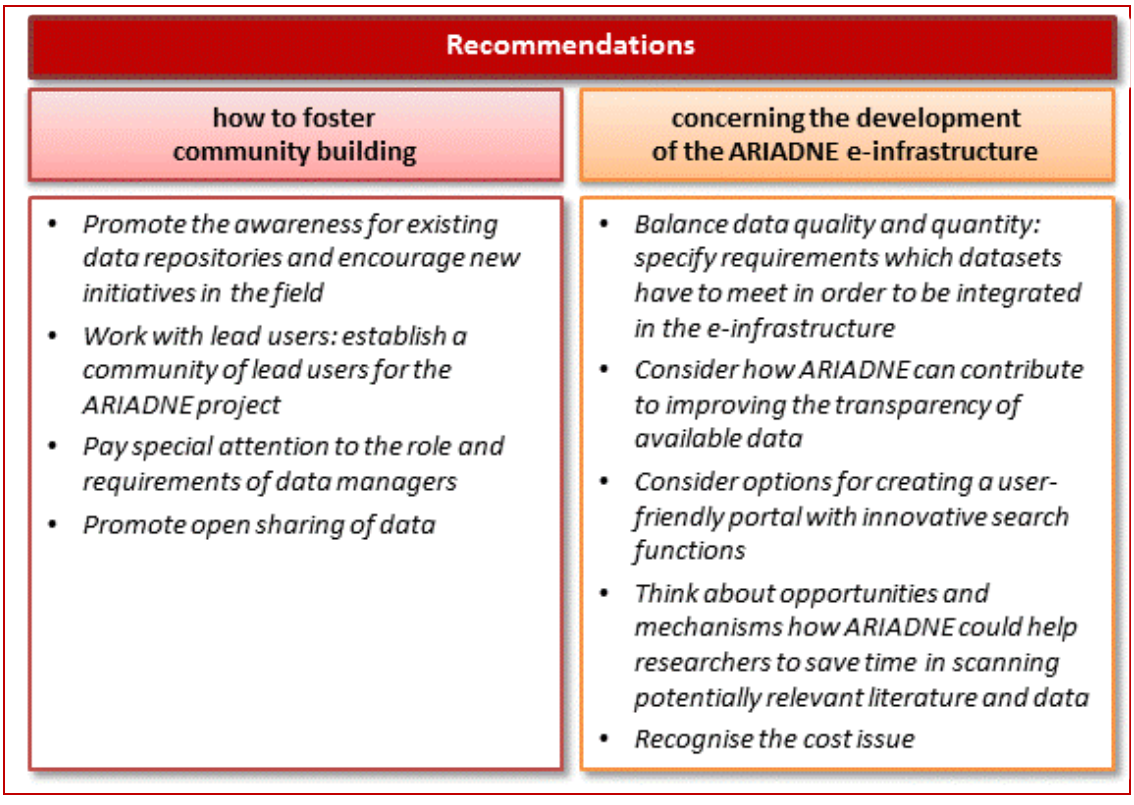
Conclusions and recommendations from the Special Interest Groups

Three ARIADNE Special Interest Groups (SIGs) also provided some initial conclusions on issues and suggested approaches in their fields of interest (see *Section 7*).

- The **SIG on Grey Literature** proposes to promote guidelines for the digitisation of grey literature, the definition of a core metadata standard for grey literature, and to support research on novel indexing and extraction technology in order to facilitate the overview and extraction e.g. of available tabular information contained in literature.
- The **SIG on Excavation and Monuments Data** suggests, inter alia, that tools and guidance based on international standards should be developed, and that the effort for long-term curation of data should be already included in project cost plans.
- The **3D Data and Visualisation SIG** addresses the production and use of various forms of visual representation of archaeological entities, small finds as well as structures or an entire site – an area not explicitly covered by the online survey. The SIG addressed challenges in this field; the main ones are related to standards and interoperability, difficulties of web-based access and interaction with high-quality visual media, and IPR restrictions with regard to the re-use of data and models.

1.3 Recommendations for the ARIADNE project

The following recommendations, which have been derived from the collected empirical evidence, are addressed to the ARIADNE project community. There are two categories, reflecting the key objectives of WP2: recommendations on how to foster community building and recommendations regarding the design, offerings and focus of the future ARIADNE e-infrastructure (see *Section 8.2* for details and suggested activities).



Recommendations with regard to community building

Promote awareness for existing data repositories and encourage new initiatives

The online survey results indicate that most of the existing digital repositories are not yet widely known among the research community, in particular outside the home country of the repository. We recommend therefore that the ARIADNE project should also be used as a platform to raise awareness for digital repositories in general, and encourage initiatives to establish new repositories in the field.

Work with lead users - establish a community of lead users for the ARIADNE project

"Lead users" (cf. van Hippel, 1986), in the case of data resources, are researchers who make intensive use of repositories and data sets in their daily work, and who have therefore specific needs and a genuine interest in developing solutions to these needs. Often, lead users are at the same time early adopters of new technologies and services in their field. We recommend therefore that the ARIADNE project should make every possible effort to identify such lead users and closely work together with them in identifying specific user requirements and developing technical solutions.

Pay special attention to the role and requirements of data managers

Literature suggests that the criteria for the usefulness of a research e-infrastructure should be developed by the prospective user community ("demand pull") rather than from a technological point of view ("technology push"). In particular, the role of data managers in e-infrastructures should be emphasised; they represent a highly important group of stakeholders and should be specifically addressed, both with regard to exploring user requirements and as part of the dissemination and awareness raising activities.

Promote open sharing of data

Initiatives in e-infrastructure and services must address data sharing practices heads on, because most researchers are reluctant to share their data, at least not "open access" and in re-useable form. Research funders increasingly demand data management plans and open sharing of research products. As an integrating activity, ARIADNE is well placed to promote open sharing of data in the archaeology sector in Europe. Going beyond advocacy, this requires advice in matters of data management (targeted at sharing), effective metadata generation, licensing, data citation standards and, overall, trust building and recognition of data sharers.

Recommendations with regard to the development of the ARIADNE e-infrastructure

Balance data quality and quantity: specify requirements which datasets have to meet in order to be integrated	Data and metadata quality (completeness, degree of organisation) are key user requirements with regard to digital repositories. The ARIADNE project will have to carefully consider and specify the quality requirements for specific collections or data sets to be integrated in the e-infrastructure, so that the users regard the resulting services as valuable. In other words, the project needs to think about where and how "to draw the line". These criteria may be different for various types of data. In particular, ARIADNE will have to discuss how to deal with "legacy data".
Consider how ARIADNE can contribute to improving the trans-parency of available data	In our survey, researchers expressed a strong need for having a better overview of available data and thus clearly confirmed the overall rationale for the ARIADNE project. The question is how this lack of transparency due to the enormous fragmentation of the data landscape can be most effectively addressed. We suggest that ARIADNE should consider the means for how to improve the overview of different data sources in Europe.
Consider options for creating a user-friendly portal with innovative search functions	Researchers would greatly appreciate (and probably expect from ARIADNE) portal functionality which makes it more convenient for them to search for archaeological data across different databases. Ideally, such a portal should not only link different data sources, but also offer innovative and more powerful search mechanisms. ARIADNE should take these user requirements with regard to the "front end" of services carefully into consideration.
Consider opportunities and mechanisms for how ARIADNE could help researchers to save time in scanning potentially relevant literature	Researchers mentioned the lack of time to scan relevant literature and data as another major problem in their profession. Even if potentially relevant literature and data resources have been identified and would be accessible, they just do not have the time to review it. While ARIADNE cannot directly solve this basic problem, it might also present an opportunity for the project: any innovative mechanism that allows a quick scan of literature or data (such as offering previews, summaries or samples) would probably be highly welcome and present a real asset for the integrated infrastructure.
Recognise the cost issue	The interviews and the survey confirmed that costs are a major barrier for data access, in particular with regard to digital resources. While ARIADNE, as a meta-infrastructure project, does not have a direct impact on pricing schemes of the underlying institutions, the fact that costs are a concern for many researchers should be carefully considered when developing a business model for offering ARIADNE services. Fees charged to users could severely affect the acceptance and use of these services.

2 Introduction

2.1 Background and objectives

Background

This document is the "First Report on Users' Needs", a contractual deliverable (D2.1) of the EU project ARIADNE ("Advanced Research Infrastructure for Archaeological Dataset Networking in Europe"), which is funded under the European Community's Seventh Framework Programme (FP7-INFRASTRUCTURES-2012-1). The report summarises the results of the work carried out during the first project year in Task 2.1 "Survey of users' needs and community building" of WP 2 "Community Building and Innovation". There will be an update of this report (D2.2 - Second report on users' needs) towards the end of the second project year; a first plan for this update, showing how the analysis of user requirements will be elaborated, is presented in Section 2.3.

Objectives

The main objectives of this report is to provide **evidence of user requirements** of key target groups of the ARIADNE project with regard to archaeological research data, and to contribute to **building a prospective user community**. This information shall support the ARIADNE project in **taking evidence-based, informed decisions regarding the specification of the e-infrastructure and services** so that the integrated infrastructure is developed in a way that corresponds to perceived and actual research needs. This evidence shall be based on feed-back from the user communities, to be collected through a user survey.

The **specific objectives** and goals of the analysis summarised in this document can be derived from the description of Task 2.1 in the "Description of Work" of the ARIADNE project. The specific (operational) objectives relevant for the work presented in this document are:

- to collect feedback from users on the planned data infrastructure and services (as far as its suitability to research practice and needs are concerned);
- to organise and carry out a survey on users' needs exploring the community's perception and reaction to the project;
- to explore (through the survey) perceived needs and expectations of researchers and the degree of their satisfaction with currently available datasets;
- to collect the survey results in a database;
- to contribute to the building of a user community around the ARIADNE integrated infrastructure, in particular by making use of the survey not only as a tool to collect information, but also as a promotional tool to inform about the project;
- to present, at the end of the first project year, the interim results, in particular a snapshot of the database (to be summarized in D2.1).

The description of work also stipulates that a first survey round should involve the partners for fine tuning the survey tools and receive feedback on their suitability. The survey should then be extended to the wider archaeological research community. It was recommended that the survey should seek cooperation with existing archaeological communities such as those of EAA (European Association of Archaeologists) and CAA (Computer Applications to Archaeology conference), and that the survey should be promoted at conferences, in order to effectively reach the target communities.

The above listed objectives have been closely followed and addressed through the work carried out during the first project year in Task 2.1. The results of the various activities, in particular the results of the user survey (a cornerstone of the activities during the first year), are described in this document.

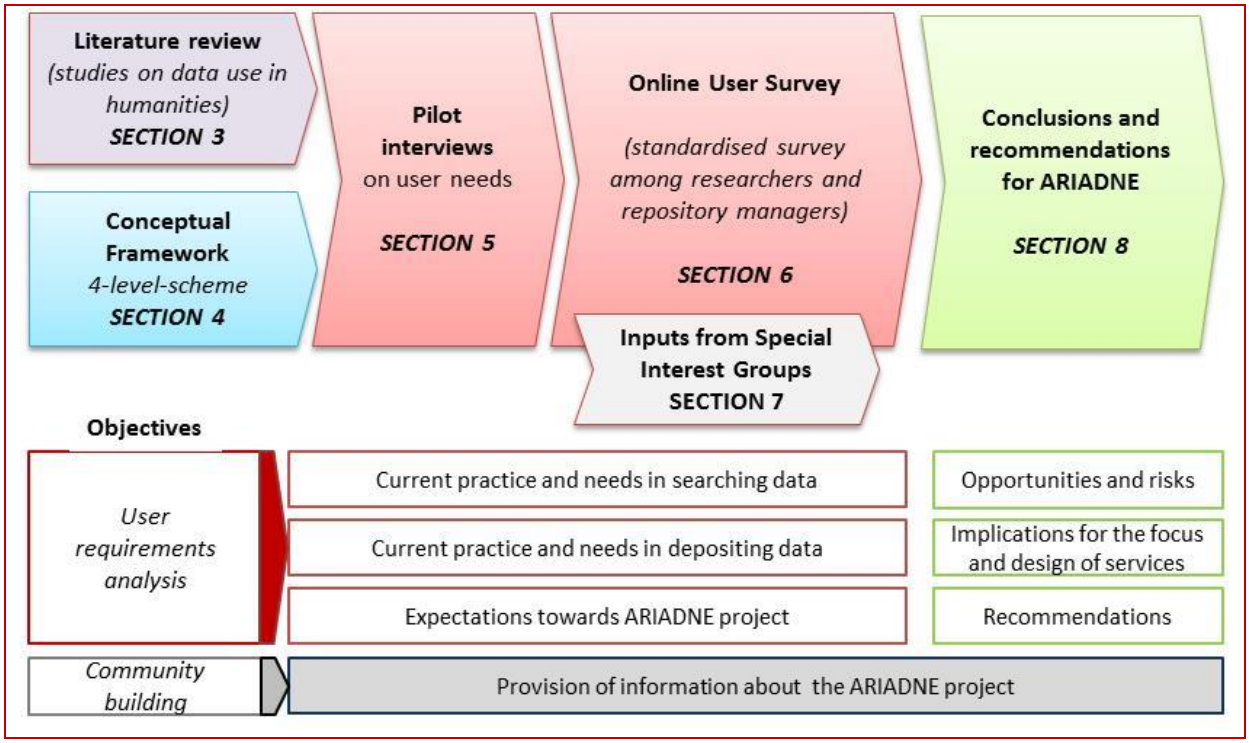
There has been one important deviation from the initial objectives as stated in the description of work, following an agreement on this among the consortium: Initially, it was also mandated that Task 2.1 should "develop a set of forms and questionnaires to analyse and summarize the current organization of datasets". Thus the user survey would have to capture not only user needs, but also collect information about specific datasets used by the respondents. This would be information "about the data structure, quantity and database functionality, including technical information" as well as "the intended and actual use of the dataset and its suitability for research". It was then decided that this information is to be collected by WPs 12 and 13, as this type of information was not compatible with a survey on user needs; it would have been hardly possible to collect this specific information about data sets used and all the details from external users through a survey.

2.2 Methodology

The main building blocks of the methodology that was applied for carrying out the ARIADNE user requirements analysis (with references to the sections where the respective results are presented) are depicted in Figure 2.2-1. The work was conducted in four main phases.

- **Phase I** consisted in the development of a conceptual framework for the definition of the user communities, and in identifying, collecting and analysing relevant existing literature and previous studies (in order not to duplicate efforts as well as for properly preparing the survey).
- In **Phase II**, pilot interviews were conducted with selected researchers and repository managers from among the consortium members in preparation of **Phase III**, the implementation of the Europe-wide ARIADNE Online User Survey. Thus, phases II and III represent the empirical data collection. The collection of evidence was complemented by the work of some of the ARIADNE SIGs which analysed issues and requirements with regard to specific types of research data.
- **Phase IV** focused on the analysis of the survey results, the development of conclusions for the ARIADNE project on this basis and the preparation of this summary report.

Figure 2.2-1: Methodology applied in the analysis of user requirements



In the following, we give some background information on each of the methodological building blocks. For further details on how the work was carried out, in particular about the Online Survey, we refer to the respective sections where the results are presented and the respective Annexes (e.g. with the survey questionnaires).

Literature review

A starting point of the work was to identify, collect and review literature which has explored data requirements of researchers. The specific aim was to identify what is currently known about research infrastructures that are relevant to researchers and other users in the multidisciplinary field of archaeological research, and specifically about their needs and requirements in research infrastructures. The review was guided by a number of specific questions, some of which were of a more generic nature, while others were more specifically focused on the archaeological and related fields of research. Moreover, the perspectives of potential users were considered both on the individual as well as research community levels. The results of this review are summarised in Section 3.

Conceptual framework

In the widest sense, the term "users" refers to the target communities for whom the ARIADNE project sets out to develop an e-infrastructure and services based on this infrastructure. While such a basic definition of "users" is easy to establish (and will be sufficient for many purposes), it is much more challenging to identify and describe the various user communities in a systematic and coherent way. A first step of the work was therefore to develop a framework for mapping the different user communities which the ARIADNE project addresses. We have used a data workflow approach for this purpose. We suggest a framework consisting of four levels of data workflows and the respective data management communities (with a broad understanding of the term "management"). The four levels of the framework are:

- **Research projects** (represented in the survey by researchers);
- **Research institutes** (represented in the survey by directors of institutes);
- **Data centres** and data repositories (represented by repository managers);
- **e-Infrastructures** (not directly represented in the survey).

This four-level framework is described in more detail in Section 4.

Pilot interviews

We conducted 26 pilot interviews with researchers and data managers, mostly representatives of the wider project community. The interviews focused on conditions of data search, production, management and access in the context of their work as well as about advantages they expect from the ARIADNE e-infrastructure and services and other project activities. About half of the interviews were conducted face-to-face, the other half in writing. These pilots had two main goals: (i) to collect first insights in partners' perception of current conditions, and (ii) to identify issues that should be validated and further explored in the Online User Survey. Further information about the set-up of the interviews and the interviewees is provided in the introduction to Section 5.

ARIADNE Online User Survey

The ARIADNE Online User Survey addressed the (international) archaeological research community, including researchers and managers of archaeological data repositories. They were grouped into four categories, in line with the ARIADNE users framework developed. These target groups were addressed with two different questionnaires (one for researches, one for repository managers) – i.e. the user survey consisted effectively of two separate surveys.

The roll-out of the survey to the dispersed population was carried out with support of the ARIADNE consortium, predominantly by leveraging the networks and communities of the various research partners, including the members of major associations and federations. The survey was launched in November 2013. The deadline for answering the questionnaire (as communicated in the invitation e-mails) was December 31st, 2013. The results in this document represent the status of answers received to that date. The final, cleaned sample comprises 692 questionnaires (640 researchers and 52 repository managers). More detailed information about the survey methodology is available in Section 6.1.

Inputs from SIGs

The ARIADNE Special Interest Groups are a mechanism for the project to explore specific issues related to data-related user requirements in more detail. The SIGs have been established under WP 2 (as Task 2.2). Their assignment is to survey the state-of-the-art of a specific theme (typically related to specific types of data), focusing on user requirements and on the strengths and weaknesses of the available infrastructure and tools. On this basis, the SIGs assess the gaps (issues, challenges) and how the ARIADNE project could contribute to addressing the identified challenges. The work of the SIGs thus feeds into the analysis of user requirements. We used the preliminary results of some of the SIGs (which are already more advanced in their work) for this analysis and present their assessment as "case studies" on specific fields in Section 7 (see introduction for further background information):

- a case study on Excavation and Monuments Data;
- a case study on Grey Literature;
- a case study on 3D Data and Visualisation.

The SIGs will be an important mechanism for further addressing specific issues with regard to user requirements in the second project year.

Recommendations

The recommendations in Section 8 are addressing the ARIADNE project; some recommendations are relevant for specific WPs and tasks. They are framed either as "guidelines" suggesting in which direction to go, or as specifically recommended activities. They shall reflect the key objective of this document, namely facilitating the ARIADNE project to "take evidence-based, informed decisions regarding the specification of the e-infrastructure and services so that the integrated infrastructure is developed in a way that corresponds to perceived and actual research needs."

2.3 Outlook on the update (D2.2)

The ARIADNE work plan foresees preparation of an update of this Deliverable ("Second Report on Users' Needs", Deliverable 2.2) after the second project year, but leaves some flexibility in defining which aspects of the deliverable are to be updated or, rather, expanded. The WP2 team proposes that the update should mainly focus on exploring specific user needs in more detail, with the goal to facilitate decisions and choices in the design of ARIADNE services. As a side line to this effort, but not as the primary objective, the results of the online survey should be updated and expanded, notably by further improving the overall representativeness of the sample.

Exploring user requirements in detail: a portal analysis with support of a lead users panel

The main focus of the second report on user requirements should be to analyse detailed, specific user requirements. This report confirms that the ARIADNE project objectives are highly relevant; they address existing user needs and have a high potential to create value for users. However, it was not always possible with this first and broad analysis to explore specific, technical aspects of user requirements in much detail. Therefore, the strategic recommendations that can be made to the ARIADNE project on the basis are limited to general guidelines and directions (see Section 8.2).

The evidence collected with this survey is an excellent baseline, however, for now going into details and exploring specific needs and requirements in terms of their relative importance and satisfaction levels. This is necessary to develop more specific recommendations with regard to options for the design for the ARIADNE services.

As a method to provide this evidence, we suggest setting up a **panel of lead users** and carry out, with their support, an evaluation of existing data repositories or portals. "Lead users" (cf. van Hippel, 1986) are users of a product or service who experience specific needs months or years before the mass market will express the same needs, and who would benefit significantly from obtaining a solution to their needs. Translating this concept to the case of archaeological research and the use of data resources, lead users would be archaeological researchers who make intensive use of (cross-)searchable repositories in their daily work, and who have therefore specific needs and a genuine interest in developing solutions to these needs. Often, lead users are at the same time early adopters of new technologies and services in their field.

We expect that several lead users of digital repositories can be identified among the organisations which are part of the ARIADNE consortium, and from their networks. The panel should consist of about 15-20 lead users. In parallel, in consultation with the consortium, a sample of about 8-10 digital repositories or portals which cover archaeological data will be selected. These will certainly include some of the repositories that were listed already in the online survey and which are also represented in the project. Each lead user would then test and evaluate 2-3 portals. They will receive a questionnaire; in particular, they will be asked to comment on the usability of various features and functions of the portal, to describe their user experience (what did he/she like/dislike?) and to make suggestions for improvements or additional features he/she was missing. We expect that the evaluation of a portal (testing the functionalities, documentation of the experience) will take 1-2 days. SRFG and DAI will then collect the evaluation reports, analyse the various experiences and translate the comments and suggestions received into specific requirements.

These requirements will then be cross-checked and discussed with data managers from online repositories in terms of their feasibility.

In addition, and complementary to this portal evaluation, the **Special Interest Groups** will be involved again with specific assignments (see Chapter 7 for SIG contributions to this reports). The SIGs will be asked to explore and assess specific issues arising from this report within their community of practice, and to deliver opinions and conclusions.

Main tasks:

- establish a panel of 15-20 lead users of digital repositories;
- select a sample of repositories or portals to be evaluated;
- develop guidelines and instructions for lead users, brief them and assist them in their task ("help desk");
- analyse evaluation reports from lead users and prepare synopsis, draw conclusions with regard to user requirements;
- involve data managers in the evaluation, for instance by discussing the suggestions made by lead users with them;
- instruct SIGs, collect and analyse inputs.

Updating the online survey

As for the online survey, the second report will present updated figures which include further responses that have been obtained after the deadline set for inclusion in this report. The survey is still open – this report includes all responses until December 31st, 2013. While it is not to be expected that the general trends and patterns will significantly change, the goal is to further improve the representativeness of the sample. In particular, an effort will be made to increase the number of responses from Germany and other countries that are significantly underrepresented in the sample. The report will present the final figures, but only comment on them if there is a significant change in the evidence as compared to the results presented in this document. In specific cases, it might also be of interest to dig further into the data than was possible for this first report, for instance to analyse whether views and practices differ between various segments of the sample. These analyses will be linked to the specific questions that are posed to the research team in the context of analysing specific user needs (see section above – "going into details").

Main tasks:

- promote the survey in specific countries that have not been effectively reached so far (in particular in Germany);
- collect additional responses and prepare an update of the results (new tables, figures);
- check if any of the existing empirical evidence has significantly changed and (if so) comment and explain;
- prepare break-downs on specific items (if needed to support the analysis of specific user needs);

3 Literature review

According to the survey methodology, first a literature review was carried out to serve three purposes:

- *Knowledge base:* to collect available knowledge about the current offer, demand, and existing issues concerning e-infrastructure and services;
- *User framework:* to support the preparation of a framework of potential user groups of the foreseen ARIADNE e-infrastructure and services;
- *Pilot interviews and online survey:* based on the results of the two previous activities, to help prepare templates for pilot interviews and the online survey on user needs.

Some important aspects of those work items are:

***Knowledge base**:* The overall result is that a lot of knowledge is available already about research e-infrastructures and services for the Arts & Humanities (A&H). But specifically for Archaeology the knowledge base is rather thin. In various surveys aimed at representativeness Archaeology is lumped together with many other disciplines, A&H and social sciences, or the A&H alone, which is still a large and diverse set of disciplines. Nevertheless, we could identify some studies with more or less direct relevance to archaeology.

***User framework**:* The framework has been developed based on the foreseen architecture of the ARIADNE e-infrastructure and services, but from the perspective of different data management communities (research groups, institutional repository managers, national data centres and, typically, international subject- or domain-based repositories). Some of results of the literature review were helpful to understand problem areas in this setup, in particular, data management and sharing practices at the base level of research groups.

***Pilot interviews and online survey**:* In the preparation of the templates for the pilot interviews and, subsequently, the online survey on user needs, the Task 2.1 team combined own knowledge in archaeological data practices (particularly, of the German Archaeological Institute) with some of the findings of the literature review. For example, the strong focus of the interviews and online survey on data sharing practices is a direct response to the observation of the literature review that such practices currently are the "hottest" topic of e-infrastructure and service development.

The sections that follow present the results of the literature review, referencing the most relevant literature. The first section addresses definitions of, and expectations from, research e-infrastructures. This is followed by highlighting key topics in the literature, e.g. "common ground" of developers and users, the need for "embedding" e-infrastructures in research practices, and data sharing as the "hottest" topic.

Next we summarise available surveys on research infrastructures which had a particular focus on the humanities and distinguish different types of e-infrastructures. We also considered as relevant exploration of what distinguishes ICT-supported research in archaeology from the so called "digital humanities". Finally, we summarise a number of available studies on current digital practices of humanities scholars, including archaeologists and classicists. At the bottom of each section, results highlights are added with reference to ARIADNE tasks for which they are relevant.

Results of some study work on repository typologies and archaeological repositories are not included in this chapter but rather in section 4.3.5.

3.1 Contexts and definitions of research e-infrastructures

Research Infrastructures (RIs), both physical (e.g. laboratories, telescopes, research vessels) and multi-tier e-infrastructures require large investments by international partnerships. In Europe, the European Strategy Forum on Research Infrastructures (ESFRI) was launched in 2002 to support a coherent approach to policy-making on pan-European infrastructures, which are perceived as one of the pillars of the European Research Area.

Research e-infrastructures or "cyberinfrastructure" (the term used in the USA) are understood as important drivers of innovative scientific research. In recent decades, research has become increasingly collaborative, distributed and data-intensive, and there is the expectation that with large integrated datasets or "big data", e-infrastructure and tools new scientific questions can be tackled in international and interdisciplinary collaboration (IWGDD 2009; Riding the Wave 2010).

Highly influential publications on research e-infrastructures have been the ESFRI "roadmap" reports (ESFRI 2006, 2008, 2010), and the reports of the US National Science Foundation (NSF) Advisory Panel on Cyberinfrastructure (NSF 2003, 2007). Specifically for the Humanities sector the "Our Cultural Commonwealth" report of the Commission on Cyberinfrastructure for the Humanities & Social Sciences of the American Council of Learned Societies (ACLS 2006) and a recent report of the European Science Foundation (2011) can be mentioned.

Selected definitions

The most widely used definitions of e-research infrastructures probably are the ones suggested by the European Strategy Forum and the NSF Advisory Panel on Cyberinfrastructure. The first ESFRI roadmap report (2006) defines research infrastructures generically, but emphasises the ICT elements:

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

The NSF definition of "cyberinfrastructure" (2007) reads:

"Cyberinfrastructure integrates hardware for computing, data and networks, digitally enabled sensors, observatories and experimental facilities, and an interoperable suite of software and middleware services and tools. Investments in interdisciplinary teams and cyberinfrastructure professionals with expertise in algorithm development, system operations, and applications development are also essential to exploit the full power of cyberinfrastructure to create, disseminate, and preserve scientific data, information, and knowledge." (NSF 2007: 5)

Also often cited is the following definition: *"e-Science is not a new scientific discipline in its own right: e-Science is shorthand for the set of tools and technologies required to support collaborative, networked science. The entire e-Science infrastructure is intended to empower scientists to do their research in faster, better and different ways."* (Hey & Hey 2006: 3)

Notably this definition centres on the empowerment of scientists, the actual users of RIs, rather than the technical implementation of e-research infrastructures. The definition of the NSF Advisory Panel (2007) also refers to required investment in "interdisciplinary teams and cyberinfrastructure professionals", while the ESFRI 2006 definition mentions, although only in passing, "including the associated human resources". Those meant by these notes are mainly technical developers and IT and data management personnel of RIs.

In the literature on research e-infrastructures of course many other definitions can be found. One definition which fits particularly well with the focus of this survey has been coined by Sheila Anderson (Centre for e-Research, King's College London) in the e-science scoping report for the UK Arts and Humanities e-Science Initiative: *"e-Science for the arts and humanities is the development and deployment of a networked infrastructure and culture through which resources – be they processing power, data, expertise, or person power – can be shared in a secure environment, and in which new research questions will arise, new forms of collaboration can emerge, and new and advanced methodologies explored."* (Anderson 2007: 6)

Difficulties of RIs for the Humanities

It is worth noting that the ESFRI (2006) report emphasises the particular difficulty of providing e-infrastructure and resources for the humanities and social sciences: *"The complexity of the record of human cultures – a record that is multilingual, historically specific, geographically dispersed, and often highly ambiguous in meaning – makes digitisation difficult and expensive. (…) Data, information and knowledge are scattered in space and divided by language, cultural, economic, legal, and institutional barriers."* (ESFRI 2006: 20) Consequently the report perceived as a priority to provide RIs that allow for discovering and accessing available and newly created digital resources, i.e. not RIs centred on data processing.

Concerning the definition and suggestions of the NSF Advisory Panel, in the humanities headquarters these have been perceived as over-emphasising the role of technological infrastructure. Therefore the "Our Cultural Commonwealth" report of the Commission on Cyberinfrastructure for the Humanities & Social Sciences (ACLS 2006) stressed the specificity of the humanities and interpretive social sciences and noted several constraints that must be overcome to enable cyberinfrastructure to play a significant role. These include "insufficient training, outdated policies, unsatisfactory tools, incomplete resources, and inadequate access". Thus the constraints are not seen as primarily technological but, instead, cultural, economic, legal, and institutional. Green & Roy (2008) confirm this assessment, explain requirements with reference to humanities lead users of ICT, and conclude: *"One of the biggest issues surrounding cyberinfrastructure and the liberal arts is that, overall, a major cultural shift in both the conceptualization and the practice of scholarship is required to take full advantage of what is being offered."*

Results highlights (incl. references to project tasks)

Researchers in the humanities perceive current definitions of research e-infrastructure as overly technology-centred ("technology-push" model). One background to this may be different views of the role of technology in research, especially in comparison to the natural sciences and engineering disciplines. However, more important arguably is the perceived difficulty of adopting e-infrastructure and tools, "insufficient training, outdated policies, unsatisfactory tools, incomplete resources, and inadequate access" (ACLS 2006), for instance.

The topic of research e-infrastructures is directly connected with perceived new opportunities of advancing research and innovation. Innovation is one of the core goals of ARIADNE, however some related notions of innovation, through "data-driven" and "big data" research, for instance, will need to be scrutinized. – *Relevant for Task 2.4: Innovation Agenda and Action Plan.*

Besides sustained funding, research e-infrastructures and services also require a skilled force of "human resources", i.e. professionals in research data management and reliable operation of various ICT systems and tools. In the archaeology sector, such professionals may not be readily available, but require capacity building, training and career opportunities for such professionals. – *Relevant for Task 4.5: Good Practices and Task 4.5: Guides to Good Practices, and WP5: Transnational Access and Training.*

3.2 Key topics in research e-infrastructure development

3.2.1 Disciplinary embeddedness of e-research practices

It is crucial always to bear in mind that research e-infrastructures are not primarily about technology but research practices that are supported by ICT environments and tools. Such practices are part of the "culture" of scientific disciplines which can have a higher or lower affinity for using technologies. For example technical products created by researchers (e.g. ICT tools, databases, web-based publications, etc.) may be ranked high or low on the hierarchy of valuable scientific contributions.

In this respect considerable differences are assumed between the humanities and the natural sciences. In the natural sciences the importance of technology-based research and, even, technology-drive in research are fully acknowledged. In most humanities disciplines this is not yet part of the "epistemic culture" (Knorr 1999), i.e. the methodologies which are used to produce and communicate knowledge (e.g. a research database vs. a scholarly monograph).

Lower affinity has domain-specific reasons which impact on the willingness and capability to adopt and use e-research tools. The reasons must be taken account of in the development of useful e-infrastructure, because such an infrastructure is not just a technical system to be designed. Rather it must fit with the culture and social frameworks of the research community, which comprise institutional, organizational, legal and other non-technical aspects. This has been called the "human infrastructure" or the required "embeddedness" of e-infrastructure in the research culture and practices. (Borgman 2007; Lee *et al*. 2006; Procter *et al*. 2013)

Embeddedness (also sometimes called "domestication") of e-infrastructure has both technical and socio-cultural aspects. Technical aspects include the installed base (i.e. what is already in use) and its transparency (i.e. that it invisibly supports tasks), which is recognised when it does not work properly or breaks. Socio-cultural aspects concern the learning of e-infrastructure supported practices, their conduct in the day-to-day work of the researchers, and the related values and norms of collaboration and sharing of tools and resources. Proposed tools and services will fail or be less useful than originally anticipated if they do not become embedded in the research practices. Hence the criteria for fit and usefulness of intended e-infrastructure should come from the research community not technologists.

To illustrate the complexity of the matter, the Enabling Wider Uptake of e-Infrastructure Services (e-Uptake) project studied the inhibitors researchers face in taking up e-infrastructures (Voss *et al*. 2009 and 2010). Some 50 researchers from a range of disciplines and institutions were interviewed about their engagement with e-infrastructures. More than 250 issues in e-infrastructure adoption and about 80 enabling factors were identified. Therefore adoption should be conceived of as a process of "co-evolution" in which technical opportunities are successively evaluated and appropriated by the research community, promoting evolutionary rather than radical innovation.

Concerns and warnings expressed about potential limited adoption of e-infrastructures have meanwhile arrived in technology-centred quarters. While initial roadmaps for research e-infrastructure development have presented a "technology-push" perspective, the socio-cultural requirements for adoption are now acknowledged. For example, the European E-Infrastructure Forum (EEF) of Grid and high performance computing providers in a report on ESFRI project requirements writes:

"There are two main constituencies which play a crucial role in the development of scientific data infrastructures: one that uses data-intensive methods and another that creates these methods. So far, most of the roadmaps/white papers/visions that are circulating have been produced or heavily influenced by members of the first constituency with very little involvement by the second constituency. Consequently, these reports mainly describe the application requirements that must be met by the future data infrastructures. The challenges that the researchers of the second constituency

have to overcome in order to make feasible the building of the next generation of data infrastructures are generally ignored." (GRDI2020, 2010: 10)

Results highlights (incl. references to project tasks)
It is crucial to fully take into account the research culture of archaeologists within which e-infrastructure and services play, or rather, are intended to play a significant role. Activities targeted at "embedding" ARIADNE's e-infrastructure and services in the research practices of archaeologists may be required in order to ensure their wide uptake and use. – *Relevant for the overall project management, WP1, i.e. definition of specific targeted actions.*
We assume a rather high affinity of archaeologists for ICT-supported research compared to other humanities, though the criteria for research process and research product can be different. The actual disciplinary rewards for contributions such as web-based publications, databases, ICT tools/software, etc. in comparison to other products (especially the scholarly monograph) may be a point of concern. – *Relevant for the overall project management, WP1 [e.g. measures aimed at raising the recognition of digital products), and Task 4.5: Good Practices and Task 5.6: Guides to Good Practices.*

3.2.2 "Common ground" of developers and users

There has been much debate about how to adjust e-infrastructures as first used by research communities within the natural sciences and engineering disciplines for disciplines of the humanities and interpretive social sciences, and if such e-infrastructures are adequate for them anyway. Specific requirements of the humanities have been noted (e.g. working with cultural content rather than numeric data) and the need to ensure the development of e-infrastructures, resources and tools which are appropriate for humanities research were emphasised (Anderson *et al*. 2010; Borgman 2009; Svensson 2010; Wouters & Beaulieu 2006).

In this context, special attention has been given to user–designer relations and the different perspectives of technical developers and humanities researchers. Indeed, a "common ground" or "methodological commons" between developers and intended users cannot be assumed. This has been observed not only in the development of e-infrastructures for the humanities, but also in disciplines and specialties such as astronomy or marine sciences (Gray & Szalay 2004; Lee, Bietz & Thayer 2010).

Ideally, researchers would take the lead in defining what kind of e-infrastructures, tools and services might fit their requirements and, thereby, develop a sense of ownership. But for "common ground" to emerge, it is necessary that the researchers and technical developers work together closely and develop a common understanding of the objectives, design options, implementation, and future use and overall operation. The intellectual contributions of both parties must be recognised (for example, see Bradley 2012 on experiences in the development of humanities computing at King's College London).

It is not sufficient involving ethnographers to study in detail research practices and workflows of the researchers and explain to the technical experts what the researchers might need. However, studies of ethnographers have detected existing issues in the "common ground". One important issue is that technical developers, researchers and data managers can have considerably different objectives (Lee *et al*. 2010):

- Researchers want to tackle research questions and tools for this need to be fit to purpose and user-friendly.

- Developers may be more interested in trying out new technical methods, where researchers' "use cases" serve to demonstrate possibilities (with the development possibly only reaching the "prototype" stage);
- Categories in the middle ground, who are not researchers or developers (i.e. staff responsible for IT and data management) are in a difficult position as they typically have no clear path of career advancement.

Therefore e-infrastructure projects should ask how collaboration among the actors can be organized so that a learning process takes place which supports the creation of useful outcomes, including important side effects such as a leveraging of the role of data managers.

Results highlights (incl. references to project tasks)
"Common ground", i.e. close cooperation of researchers, data managers and technology experts has been identified as crucial for the successful development of research e-infrastructures. There is need for a clear, common understanding of the objectives, design options, implementation, future use and overall operation. The criteria for fit and usefulness should come from the research community, not technologists. – *Relevant for the technical work packages WP12 and WP13, with regard to the overall design and implementation, and other WPs, e.g. concerning semantic annotation and linking tools.*
Special attention is also required for the opportunities of data managers in e-infrastructure and service development. Their role should be emphasised, but also appropriate recognition and career paths offered. – *Relevant for all partners and stakeholders with substantial data management and service demand; data centres and repositories within and related to ARIADNE might provide advice on appropriate measures.*

3.2.3 The many faces of "data"

The difficulty with the simple word "data" is that it means very different things to actors in the field of e-research infrastructure: *"For some, data is first and foremost a question of things: samples, specimens, collections. For others, data is what comes out of a model – or perhaps the model itself. Data may be tactile, visual, textual, numeric, tabular, classificatory, statistical. Data may be an intermediate outcome, a step on the road to higher-order products of science (publications, patents, etc.). Or data may be the product itself. Where a discipline or research project fits within this spectrum will have enormous consequences for its positioning vis-à-vis cyberinfrastructure. This specificity alone guarantees that cyberinfrastructure should and assuredly never will be a singular or unified thing"* (Edwards *et al*. 2007: 31).

In archaeology a broad notion of "data" is used, comprising anything that can be studied and provide evidence for or against a hypothesis. Among others, this includes archival records, museum specimens, new survey and excavation data, laboratory analysis of finds with various tools. Focused on the material record of past human activity, textual data only plays a significant role for periods where such data exists (including epigraphy, for instance). As Stuart Dunn (2011: 95) elaborates: *"Texts occupy an important place in archaeological research, chiefly in the form of so-called gray literature reports of excavations, which are often the only extant records of those excavations, along with secondary literature and publications. But the bulk of primary archaeological excavation data comes in the form of numeric, graphic, statistical, and formal descriptions of the material record. This mass of digital evidence is geographically distributed, fuzzy, incomplete, inconsistent, and difficult or impossible to access."*

Concerning digital textual sources it may be worth noting that many humanities scholars need "analogue to digital, one language to another, text to data" conversion before they can start working

with the data (Crane, Babeu & Bamman 2007: 4). The most important distinction of types of data is between observational and experimental data, because of their implications (cf. Thessen & Patterson 2011):

- ***Observational data**:* The term observation is used for many methods of data collection, but most of them share an important characteristic: The observations are made at a certain place, at a certain time, and under not fully controllable conditions. They cannot be replicated, only confirmed by further observations. Therefore the data are unique and should in principle be preserved. In archaeology, excavations present a special case: the excavation destroys what can be observed, hence, it is the responsibility of archaeologists to take special care for the data that is produced, and share it with other researchers to fully exploit its value.
- ***Experimental data**:* Experimental data are collected under conditions where the researchers can fully or at least sufficiently control and change the conditions under which the expression of a phenomenon occurs. Typically such research is carried out in a laboratory but in-field experimentation is also possible, albeit without full control over conditions. Experiments can be replicated, however this requires availing of the same environment, source material, facilities (e.g. equipment, including calibration, etc.), and ingredients (e.g. chemicals).

Some characterizations often used in the literature are "big data" versus "small data", and "dark data":

"Big data" research domains benefit from the automation of data capture, with various sensors (e.g. earth and environmental sciences) or "high-throughput" technology like the sequencing instruments used in genetics/genomics. In terms of volume: 1 Petabyte (PB) is 1000 Terabytes (TB), 1 TB is 1000 Gigabytes (GB) and 1 GB is 1000 Megabytes (MB). "Big Data" science is understood to start at the level of Terabytes. The European Bioinformatics Institute (EBI) of the European Molecular Biology Laboratory (EMBL) expanded its storage capacity from some 100 TB in 2007 to over 10,000 TB (10 PB) in 2010. (ELIXIR 2011) The Institute expects a required ten-fold increase in capability until 2020 to also manage data from the next generation sequencing machines. One investigation into data volumes in archaeology did not find much "big data" at project level: there were some large datasets acquired from third parties (e.g. LIDAR and maritime survey data) while 3D laser scanning was considered as a core data driver in archaeology (Austin & Mitcham 2007).

"Small data" research domains are characterised by individual observations in the field and the data is mostly "hand-collected", i.e. researchers filling in templates and Excel sheets. While much of such documentation is produced, it is small volume and forms the so called "long tail" of scientific data. As noted in a recent report of the Australian archaeological e-infrastructure initiative FAIMS: *"Worldwide, archaeology suffers from 'small science' data problems that inhibit the production and dissemination of high-quality, compatible data: diverse and idiosyncratic datasets, customised methodologies and recording systems, lack of core data standards, and limited budgets, among others"* (Ross *et al*. 2013: 107).

"Dark data" is scattered and difficult to locate data, and typically it is also difficult to prepare for aggregation and sharing (Heidorn 2008). For example, it has been estimated that less than 1% of the ecological data ever collected is readily discoverable and accessible (Reichman *et al*. 2011).

An often used characterization of some fields of research is ***"data-driven"***, which typically appears in the context of big data and with the notion that mining such data can allow for unexpected new insights. Though, it is also used in the more general sense that research projects produce ever more data and hence need e-infrastructure for managing it effectively.

Results highlights (incl. references to project tasks)

Archaeologists have a broad notion of "data" and, actually, produce a large variety of observational and experimental data and data-based products. The objective of enabling semantically integrated access to this variety of data in sources which are highly scattered across Europe is challenging.

Archaeological data typically are not "big data", but rather "small" and "dark" data resources which are difficult to manage and prepare for aggregation and sharing. ARIADNE will on the one hand have to focus on data that is already available in existing institutional repositories or national data centres and, on the other hand, promote the flow of currently "dark" data into such repositories or centres. However, in many countries archaeological data centres do not exist and the state of institutional repositories may not be optimal (e.g. lack of Open Archive Initiative compliant systems). ARIADNE can play a significant role in promoting state-of-the-art centres and repositories and the flow of currently "dark data" into them.

Relevant for the overall project management, WP1, i.e. definition of specific targeted actions; data centres or repositories within and related to ARIADNE might provide advice on appropriate measures.

3.2.4 Sharing of research data – the "hottest" topic

Sharing of research data is the "hottest" topic in studies on user needs and requirements in research e-infrastructure, repositories and services, as they will not thrive if the researchers do not make their data available. But most researchers show considerable resistance to do so.

All research policy statements about research e-infrastructure, investment in such infrastructure and repositories rest on the assumption that researchers are willing and able to share their data in an open and trustful manner. In a rather simplistic analogy data are seen as the "raw oil" of science and e-infrastructures as the pipelines. However the tanks (repositories) may not be filled (Nelson 2009; Shotton 2011; Strasser 2013). Therefore data and the perceived tensions the request for data sharing occasions are "the front line of cyberinfrastructure development" (Edwards *et al*. 2007: 31).

General situation

A survey by PARSE.Insight (2009) with 1202 respondents from different research domains and countries found that the most used forms of storing and managing data were a computer at work (81%), a portable storage carrier (66%), a server of the organisation (59%) or a computer at home (51%). Only 20% stated that they store data/datasets in an archive. 14% store it in an archive of the research organisation, and 6% in an external repository of the discipline or research domain. The primary notion of publishing data was through publications, i.e. in the form of summary tables, charts, etc. that present research results. A rather high percentage of the respondents (15%) said that they also submit data to a journal, i.e. as supplemental material like oversized data tables, images from laboratory experiments, audio-visual material, 3D rendering files, etc.

The scientific journal Science polled their peer reviewers about the availability and use of data (Science 2011). They received about 1700 responses representing input from an international and multi-disciplinary group of researchers. Asked about "Where do you archive most of the data generated in your lab or for your research", 50.2% said in the lab, 38.5% on a university server, 7.6% in a repository of the research community, 3.2% "other", and 0.5% that it is not stored. Most of the respondents (80.3%) thought that they do not have sufficient funding available for data curation. However, 38.3% said that they store the data permanently, 17.9% over 10 years, 26.8% 5-10 years, 16.1% 1-5 years, 0.3% below one year, and 0.6% that the data is discarded promptly.

In summary, the two relatively large surveys indicate that across all disciplines only between 6-8% of the researchers deposit datasets in an external archive of the discipline/research domain. The most

common environment for storing, managing and re-using data is the lab and/or individual working environment, down to PCs and portable storage carriers. The category "server" is probably best understood as a file server of the research organisation behind a firewall and with restricted access for defined groups of registered users (cf. the results of IANUS survey reported in section 3.6.7).

Reasons for the low level of open access data sharing

The actual practice of researchers does not comply with what advocates of data curation and data sharing, including ever more research funding organisations, would like the researchers to do. Arguments put forward by advocates include the need to enable replication and validation of research results ("the scientific method"), to prevent potential loss and avoid redundant generation of data, the value of the data for alternative uses and, particularly, "public ownership" of publicly funded research (e.g. OECD 2003 and 2007). Though, these arguments often do no resonate with researchers (Borgman 2010).

The low level of provision for open access data has many reasons, such as lack of decisive deposit mandates, lack of appropriate data archives, and the widespread understanding that the data is owned by the research group, as an asset to be exploited further. On top of this come strong barriers even if researchers are willing to share research data. Such barriers have been identified in many studies of which the studies conducted or commissioned by the UK Research Information Network (RIN) merit highlighting (Pryor 2009; RIN & Key Perspectives 2008; RIN & British Library 2010; RIN & NESTA 2010). Among the core barriers are:

- Little academic reward for the development and curation of databases;
- Priority of publications rather than data sharing;
- Concerns that data could be scooped, misused or misinterpreted;
- Issues of copyright, confidential and sensitive data;
- Required additional effort for providing shareable data, including formatting, metadata creation, licensing.

Overall there are more barriers than incentives for open access sharing of reusable data. Therefore data is primarily shared directly between trusted colleagues within the research community. Where open data recommendations and guidelines have been issued, actual provision is routinely ignored or under-performed. This makes the existing cases of open data sharing all the more valuable and exemplary.

However, such examples can typically be found only were special conditions apply, for example, if there are no or only few commercial interests involved (e.g. astronomy or oceanography), where the data are basic and produced in increasing volumes (e.g. genetics), or where research consortia are mandated to carry out large-scale data collection on a regular basis.

One factor that makes researchers more inclined to share research data is if the data can be captured and processed automatically (Pritchard *et al*. 2005). For example, with high-throughput DNA sequencing technology, masses of data are produced and in need of analysis and interpretation; in the field of structural genomics over 90% of deposited structures are not yet described in literature (Ellrott *et al*. 2010). Hence a high degree of automation of data creation in a discipline makes open data sharing more likely than if the data is "hand-collected". This reduces the perception of high personal effort invested and thereby the sense of data ownership.

Results highlights (incl. references to project tasks)

Initiatives in e-infrastructure and services must address data sharing practices head-on, because most researchers are reluctant to share their data, at least not "open access" and in re-useable form. This often goes hand in hand with data management practices with a high risk of data loss. Research funders increasingly demand data management plans and open sharing of research products, though the impact will take quite some time and could be lower than expected, e.g. if the mandates are tooth-less concerning open data formats.

As an Integrating Activity, ARIADNE is well placed to promote open sharing of data in the archaeology sector in Europe. Going beyond advocacy, this requires advice and support in matters of data management (targeted at sharing), effective metadata generation, licensing, data citation standards (e.g. DataCite) and, overall, trust building and recognition of data sharers.

Relevant for the overall project management, WP1; WP4: Good Practices; WP5: Coordination of Transnational Access and Training.

The question of data licensing for re-use has been dealt with in Deliverable D.3.3 – Report on data sharing policies.

3.3 Surveys on e-research infrastructure with a particular focus on humanities

The ESFRI and NSF reports mentioned in section 3.1 have stimulated various surveys and studies to identify relevant existing and potential new research infrastructures in different disciplines, including the humanities. Below we briefly highlight the results of three surveys on European projects: the first covers a wide range of disciplines while the other two have a particular focus on the humanities.

Trends in European Research Infrastructures

This survey was conducted by the Research Infrastructures unit of the European Commission's DG Research and the European Science Foundation (EC & ESF 2007). 598 infrastructures of European relevance, both physical and virtual, have been selected, including 64 within the humanities sector. The survey used the distinction between "single-site", "distributed" and "virtual" RIs suggested in the first ESFRI report (2006).

The survey identified a much higher percentage of virtual infrastructures with a European dimension in the humanities than in other sectors, 24.6% compared to 12% in the total, though with 41.2% the percentage was much higher for virtual RIs in the Social Sciences. The survey includes information on the age, costs, funding, and visitors/usage of the RIs. With regard to visitors, most infrastructures are used on-site: across all RIs 60% reported less than 10% and some no remote users at all. Though of the 64 Humanities RIs 57% had over 50% remote users.

Accelerating Transition to Virtual Research Organisations in Social Science (AVROSS)

The AVROSS online survey focused on virtual infrastructure projects and received input on 218 projects which have been carried out in Europe and beyond (AVROSS 2007). The survey includes some interesting results for archaeology: Respondents in this sector appeared relatively more experienced in ICT projects, having started in the 1990s and they were involved in projects more often than other respondents.

While most projects used communication and collaboration tools (80%) and distributed data (77%), two elements were most present in archaeological projects: Innovative data collection methods with 68.8% (25.3% in all projects), and 3D objects and virtual environments with 41.7% (15.7% in all projects). Though the archaeological projects typically were not large-scale, the average initial

funding was just 148,000 € and the average duration 18 months, compared to 335,000 € and 36 months in the total sample.

Survey on Infrastructural Research Facilities and Practices for the Humanities in Europe

This survey was conducted by the Humanities in the European Research Area project (HERA-NET 2006) and covered 405 "infrastructural research facilities" (IRF), both physical and virtual. The survey provides interesting insights into the composition of available IFRs for the humanities:

Four main areas of research supported by IFRs were distinguished (share in the total in %): Cultural Heritage: 51%; Languages and Linguistics: 29%; Media Studies (understood to focus on mass media): 9%; Digital Media (including preservation of such media): 11%.

Sixty per cent of the IFRs were digital resources and "experimental facilities". Digital resources made up the bulk of this (92.2%), including: Databases; Digitised texts, objects, artefacts, etc.; Language and linguistic corpora; Metadata linking of existing databases; Historical cartography, atlases, etc. Experimental facilities only accounted for 7.8% of the 405 reported research facilities, including: Technical instruments (5.1%), Audio-visual laboratories (2%) and Virtual Laboratories (0.7%).

With regard to access, 64% of the digital resources were freely accessible via the Internet, 6% required membership or subscription, 20% a special appointment with the host institution (e.g. due to sensitive information, specific technical requirements, etc.); 10% of the access modalities were classified as "other".

Asked if the host institution had developed any methodological and/or analysis tools to improve the researcher's use of the research infrastructure/resources, 56% answered "no". Consequently the study suggested that "the humanities need more clearly to define how their research infrastructures can be used more directly as a methodological tool for their scholars".

The most important issues concerning the further development of IFRs were seen as copyright and free access, better coordination among funding agencies and stable, long-term financial arrangements. With regard to the four fields of humanities research studied (see above), the respondents perceived a particular need for new research infrastructures in the field of Cultural Heritage. Concerning content and services, most respondents wanted to see more digitisation of multimedia resources, standardisation of metadata, and ability for cross-domain federated search.

A survey on archaeological research infrastructures is needed

The surveys summarised above allow the understanding of some specifics for humanities research infrastructures (physical and virtual) and small-scale virtual/ICT infrastructure projects in archaeology. However the insights are limited because archaeology is lumped together with many other humanities disciplines (or humanities and social sciences) that are quite different in terms of research methods and content/data.

We could not identify a similar quantitative survey specifically for archaeological research e-infrastructures. Such a survey would indeed be a major undertaking, because archaeology is a multi-disciplinary field of research that spans several domains and specialties of the humanities (among others, cultural anthropology, classics, ancient/historical geography, medieval studies, epigraphy, …). Moreover the field includes the application of various natural sciences methods to archaeological research questions (e.g. physical, chemical and biological sciences). Investigations could aim at identifying where and how intersections between archaeology and other humanities research communities might be strengthened by e-infrastructure and services; or, concerning the natural sciences, how results of the existing laboratories, might be integrated.

Results highlights (incl. references to project tasks)
The available large surveys (AVROSS 2007; EC & ESF 2007; HERA-NET 2006) allow for a general understanding of the landscape of research infrastructures. Their usefulness for ARIADNE is limited, because archaeology has been lumped together with many domains of research, which is also the case if "only" the arts & humanities are addressed. Missing is a survey on archaeological research (e-)infrastructures in Europe at the level of institutions and research communities. This would be a major undertaking because archaeology is a multi-disciplinary field of research that spans several domains and specialties within the humanities and includes the application of various methods developed within the natural sciences. However, ARIADNE could select a few areas where an overview would be beneficial. *The Scientific Data SIG is considering creating an inventory of relevant laboratories and organizations and involving them in data standardisation. Other SIGs might follow with similar surveys to enable targeted activities. – Relevant for Task 2.2: Special Interest Groups.*

3.4 Types of research e-infrastructures

Research Infrastructures (RIs) include major scientific equipment and sets of instruments, e.g. laboratories, remote sensing instruments, a research vessel etc., and knowledge-based resources such as scientific databases as well as content collections held by libraries, archives and museums. They require technologies for data collection, storage, management, access, processing and sharing, as well as research communication and collaboration tools. This e-infrastructure should allow easy and controlled access to integrated digital resources and tools, and enable research collaboration across geographical, disciplinary and organisational boundaries.

The ESFRI definition of research infrastructures distinguishes between "single-site", "distributed" and "virtual" RIs and emphasises that each one requires information systems for data management and access. Furthermore, both the ESFRI and NSF Cyberinfrastructure Panel definitions mention IT infrastructures based on Grid, middleware and various software applications for working with research data (e.g. data processing with Grid technology). Notably, in the case of "virtual" RIs the IT component can be said to be the (e-)infrastructure, while "distributed" and "single-sited" RIs will depend on available IT infrastructure to a greater or lesser degree.

Different forms of research e-infrastructures have been realized with different technologies, depending mainly on the purposes. Below we briefly describe these forms, starting with Virtual Research Environments which are the broadest category of RIs.

Virtual Research Environments (VREs)

VREs is an umbrella category of research environments that centre on e-research functions which are selected and combined according to the needs of specific research communities. The combination can be centred more on networking and information exchange, on collaborative digital collection formation and usage, or to offer programming and data processing tools (Bos *et al*. 2007; Carusi & Reimer 2010). To provide a couple of examples of VREs in archaeology: BoneCommons promotes communication and collaboration among the zooarchaeological community; Virtual Research Environment for Archaeology (VERA) is a collaborative working environment of the Silchester Town Life Project. The large-scale project needed to enhance the flow of information from excavation and data analysis to the publication of research results (Rains 2011); Alison Babeu (2011) provides many examples of VREs which have been developed for "digital classicists".

Approaches using Web 2.0

The recent trend in using Web 2.0 tools and services for VREs in archaeology, e.g. for research networking and knowledge exchange, is discussed in a recent publication on "Archaeology 2.0" (Kansa *et al*. 2011). Web 2.0 approaches are perceived as more bottom-up, flexible and arguably more inclusive than large-scale e-infrastructure development. However, it is understood that Web 2.0 tools and services are mainly fit for networking, knowledge exchange and informal distribution of information, while long-term collaborative projects and the building and curation of digital collections require other technological systems.

Eric Kansa notes: *"While Web 2.0's impact is far reaching, it does seem to have limits. Web 2.0 platforms and services mainly facilitate informal communications among archaeologists. Web 2.0 systems are simple to use, fast, and geared to content that requires relatively minimal investment to create. Archaeologists tend not to use Web 2.0 platforms as the primary dissemination channel for forms of content that take a great deal of effort and expertise to create. In this light, data sets and sophisticated scholarly manuscripts see less circulation in Web 2.0 channels."* (Kansa 2011: 5; cf. Dunn 2011).

Indeed, substantial surveys on how researchers use Web 2.0 tools (e.g. weblogs and various "social media") found that such tools are mainly an (informal) supplement to established channels of research communication and actively used only by a small segment of researchers. More frequent users can be found among those who are engaged in collaborative research activities beyond institutional boundaries. Surprisingly maybe, young scholars are not among the avid users of Web 2.0 tools, the result of considerable scepticism of their value in the research sector (Procter *et al*. 2010; RIN 2010). One international survey also found that across all participants, social media users were 1.27 times more likely to be found in the arts & humanities and social sciences, and 0.67 times less likely in biosciences and health (UCL & Emerald 2010).

Distributed Computing Infrastructure (DCI)

A Distributed Computing Infrastructure (DC) provides a Scientific Gateway to Grid and/or Cloud resources and software applications for working with research data, e.g. data processing, storage and transfer. User groups can share data resources and computing applications ("virtual machines") and, thereby, form a virtual research community. Leading promoters of DCIs are the European Grid Infrastructure (EGI) community, European Middleware Initiative (EMI) and European Desktop Grid Initiative (EDGI). DCIs have been built in the first place in response to the need of some natural and engineering sciences for massive networking and processing of data, typically based on Grid infrastructure and functionality.

The DCI providers have sought to expand their user base, not least to legitimate the high investments in DCIs, Grid and other high-performance computing. Recently, in response to the market development and user demand, also Cloud services have been included in the DCI offer (Curtis+Cartwright 2010; EGI-InSPIRE 2011; e-IRG 2012). But the results in some disciplines, especially in the humanities and social sciences, have been limited. Research groups in these disciplines seldom deploy Grid-based datasets and processing, though may be attracted by Cloud services for other tasks. The major impediments in archaeology arguably are the diverse and complex types of datasets, lack of consistent data structures; incomplete, isolated and often not openly available data sources (cf. Hedge 2009).

Consequently, other environments based on less heavy technologies than DCIs have been adopted, for example, Web-based platforms with modules for content management, collaboration and other functionality (e.g. the open source system Drupal), Wiki and other technologies. Web GIS has been adopted particularly by archaeologists, but also historians (Gregory & Ell 2008) and other humanities researchers.

Data Infrastructures

A Data Infrastructure centres on making distributed research data accessible and (re-)useable which involves preparation, registration, networking and exchange of data, including the required protocols and standards, for example controlled vocabularies (thesauri, ontologies). Such an infrastructure can be generic, i.e. used across many domains (e.g. spatial data infrastructure), or focused on aggregating and making accessible data for one or several related disciplines.

The ARIADNE e-infrastructure is basically a data infrastructure for archaeology. Chapter 4 contains a schematic representation of the foreseen setup in terms of entities and actors involved, e.g. institutional repositories, data centres and subject/domain-based repositories. Technically the setup would follow a three-tier model: core services for data networking and access to distributed and heterogeneous data resources; reference services (e.g. classification systems, thesauri, gazetteers, etc.) which support semantic interoperability; and a number of tools which are relevant for broader segments of users.

Results highlights (incl. references to project tasks)
Beside Data Infrastructures like ARIADNE there are different forms of research e-infrastructures including Distributed Computing Infrastructure (i.e. Scientific Gateways to Grid/Cloud resources), VREs for specific research communities and, more recently, various Web 2.0 approaches. ARIADNE does not foresee offering VREs, however could benefit from having a good overview of existing VREs that are relevant for tasks in specific areas of archaeological work. Such VREs might focus on work with particular types of data (e.g. 3D models) or be useful for groups of subject specialists (e.g. a VRE developed for biodiversity research might be of interest to archaeobotanists). VREs are potential users as well as providers of the data sources ARIADNE aims to integrate. *Relevant for Task 2.2 Special Interest Groups, and technical work packages, WP13 and others, with regard to the interoperability of ARIADNE services with VREs.*

3.5 Digital archaeology and/vs. "digital humanities"

This section outlines some differences and relations between ICT-supported archaeological research and a much wider field of studies that are subsumed under the label "digital humanities". In a recent compilation of papers on "Archaeology 2.0", Eric C. Kansa writes: *"Archaeology is an inherently multidisciplinary enterprise, with one foot in the humanities and interpretive social sciences and another in the natural sciences. As such, case studies in digital archaeology can help illuminate changing patterns in scholarly communications across a wide array of disciplinary contexts."* (Kansa 2011: 2) The statement raises some important questions: Can we distinguish "digital archaeology" from other "digital humanities", for example, based on different contexts of work, practices and data/content sources? Which ones are the closest neighbours? Might they provide relevant digital resources to, and be served by, the ARIADNE e-infrastructure and services?

Digital humanities

The "digital humanities" comprise as core ICT-supported studies of corpora of texts (e.g. corpus linguistics), archival documents (e.g. historical research), images, music and film (e.g. art history, musicology, film, media and other cultural studies). Concerning the research community, a large segment gathers in the Alliance of Digital Humanities Organizations (ADHO), which brought together the Association for Literary and Linguistic Computing, the Association for Computers and the Humanities, and the Society for Digital Humanities. ADHO also oversees the major annual Digital Humanities Conference and the journal Digital Humanities Quarterly (since 2007). Moreover,

CenterNet was established in 2007 as an international network of about 100 digital humanities centres in many countries around the world.

The change from the initial "humanities computing" to the "digital humanities" label about ten years ago *"was meant to signal that the field had emerged from the low-prestige status of a support service into a genuinely intellectual endeavour with its own professional practices, rigorous standards, and exciting theoretical explorations"* (Berry 2011: 2). Subsequently the "digital humanities" community promoted integrating the use of digital collections, software tools and web-based collaboration in all activities of humanities scholars. For example, already in 2008 the arts-humanities.net of the UK-based Methods Network had documented hundreds of digital arts and humanities projects (mostly in the UK), including information on the digital resources created and methods and tools used in the research. This activity has recently been taken over by the CenterNet's DHCommons initiative "to create one large-scale discovery and review publication for digital humanities projects".

While this field of research is booming, a study of the Research Information Network (RIN) perceived an overall limited uptake of e-research practices by humanities scholars. The study notes: *"There is much talk about developments in 'digital humanities', but the practices and concepts are not as yet well-established or fully-embedded as standard features of scholarly practice."* (RIN 2011: 8) Other observers voice similar opinions. According to David Robey (2013) the types of practices and methods used by digital humanities scholars "still tend to be a minority interest", however he mentions "archaeology or some parts of linguistics" as exceptions.

Katherine Hayles (2011), based on interviews with 17 humanities researchers (at different stages of their careers) and her own experience in the field, estimates that about 10% of all humanities scholars are "seriously engaged with digital technologies", with higher shares among the younger ranks and graduate students. She assumes that in 10 to 15 years, when the "seriously engaged" demographic are in higher positions, the recognition of digital research practices and works (e.g. Web-based publications instead of the scholarly monograph) will change. However, at present print is very much alive also in the digital humanities with ever more compilations of papers rolled out (e.g. Berry 2012; Burdick *et al*. 2012; Gold 2012; Terras *et al*. 2013; Thaller 2012; Warwick *et al*. 2012).

Digital archaeology

Concerning contexts of work, the difference between "digital humanities" and "digital archaeology" is rather clear. "Digital humanities" studies are mainly an activity of academics with few external constraints except getting access to identified relevant content. In contrast, work in archaeology is conducted in contexts which include national heritage agencies, local government curators, infrastructure development companies, and commercial archaeology services (contract archaeologists). The most obvious example is that excavation is subject to regulations, requires permits, entails reporting duties, etc. While most researchers in the humanities and certainly "digital humanities" would not agree with the image of the "lone scholar", in archaeology research teams are common practice. Notably, in excavation but also for other types of work these are interdisciplinary teams involving excavators or surveyors, experts in particular finds, laboratory-based researchers, and data management and other staff.

According to a study by Jeremy Huggett, the relations between ICT-supported archaeological research and the "digital humanities" are rather limited: *"From an archaeological perspective, a relationship between Digital Archaeology and Digital Humanities is largely absent and the evidence suggests that each is peripheral with respect to the other"* (Huggett 2012: 1). He also notes that "digital archaeology" does not need the "digital humanities" for legitimacy or support because archaeological research builds on its own arsenal of methods and tools. One example Huggett analyses is the use of GIS by archaeologists compared to which "DH [digital humanities] applications

of GIS can seem very limited, even simplistic, to archaeological eyes in that they often seem to focus on interactive hypermedia visualisation with little use of GIS analytical tools".

Even better examples might have been the data capture, processing and visualization tools archaeological researchers are using (e.g. photogrammetry, terrestrial laser scanning, 3D models of objects, sites and landscapes, etc.). The adoption of ever more sophisticated digital data tools and the analysis of finds with natural sciences equipment and techniques has certainly contributed to the strong affinity of archaeological researchers also for ICT-supported research (cf. the results of the AVROSS survey).

One of the most important differences between archaeology and other humanities is the research content: scholars in the "digital humanities" analyse mainly cultural content produced by others, e.g. literary texts, paintings, photographs, music, films, media, etc. In contrast, archaeologists produce most of their data themselves, e.g. all data produced in surveys, excavations, laboratory measurements and analysis of physical and biological finds, etc. Some of the latter is carried out by specialised laboratories which serve archaeologists among other clients (e.g. mass spectrometry, electron microscopy, sequencing technology for ancient DNA analysis). Furthermore, data not produced but used by archaeologists (if affordable) are airborne or satellite remote sensing and imaging data.

The main overlaps with "digital humanities" can be found where humanities scholars work with various texts (e.g. epigraphy, manuscripts, etc.) and evidence drawn from material remains from earlier periods like classical antiquity, medieval times, and more recent periods (e.g. historical archaeology). The overlap is particularly clear between archaeological research and classical studies, and also the adoption of information technologies in specialties within classics is particularly high (e.g. digital editions of Greek and Roman texts, epigraphy, ancient world geography, etc.). Alison Babeu (2011) provides an excellent overview of "digital classics".

Results highlights (incl. references to project tasks)

Archaeology is a special field within the humanities with regard to its contexts of work and level of data production, rather than predominantly analysing existing cultural content. The specificities of the sector are also present in the digital realm so that few relations to the so called "digital humanities" can be found. However, there are some closely related domains and specialties such as cultural anthropology, classics, ancient/historical geography, medieval studies, epigraphy, iconology and others.

A multitude of directories, catalogues, bibliographies, reference collections, text and image corpora, digital editions have been produced and could be relevant also for archaeological researchers. This domain has not been fully considered yet by ARIADNE with regard to relevant resources. – *Relevant for the overall project management [WP1] with regard to overall decisions on the data portfolio, and Task 2.2 concerning a potential Artefact Data SIG.*

It is worth noting that relevant resources archaeologists might want to find and access with ARIADNE services also include airborne or satellite remote sensing and imaging data or specific natural sciences data (e.g. ancient DNA analysis of humans, animal and plants). – *Relevant for Task 2.2, Remote Sensing & Spatial Data SIG, Scientific Data SIG and others interested in specific data.*

3.6 User studies on e-research and data sharing practices

A formula often found in reports, papers and presentations is that a "one size fits all" approach to research e-infrastructures for the diversity of scientific disciplines will not work. Concerning technical requirements, typically a set of generic categories of requirements have been defined by major actors in the e-infrastructure arena, e.g. the inter-governmental policy body e-Infrastructures Reflection Group (e-IRG 2005) and interest groups like the European E-Infrastructure Forum of grid and high performance computing providers (EEF 2010).

The goal of the overall mainly technology-focused research e-infrastructure policy is to prevent a "many-headed beast" development of the e-infrastructure landscape in Europe (eResearch2020 2010); the objective is integrated e-infrastructures (e-IRG 2013; GRDI2020 2012). Consequently projects have been funded that aim to consolidate the e-infrastructure "ecology" across several initiatives (e.g. EUDAT). Cost-reduction, shared resources and services, and potential for cross-disciplinary research are highlighted in this context. The perspectives on e-infrastructure at the level of research institutions and individual users may or may not be the same.

3.6.1 Focus, perspectives and methods of studies

As part of the ARIADNE user needs survey we searched, collected and scanned a number of available surveys and studies on digital practices and related needs of researchers, including some on humanities scholars. In the scanning we found some repeating patterns of focus, perspective and study design & methods:

- Overall there is a strong thematic focus on data sharing (the "hottest" topic, cf. Section 3.2.4).
- Concerning the study design and methods, some projects cover a variety of disciplines, mainly to identify differences and commonalities in e-research needs and practices, others on a small sample or only one research area. Most studies use qualitative methods as is common in the social sciences; in particular case studies, questionnaires and interviews.
- Some studies involved ethnographic methods such as observing how researchers actually carry out their research, i.e. participant observation (e.g. Anderson & Carlson 2006), which result in "thick description" (Geertz 1993) of the research culture and practices. Such descriptions provide an information-rich analysis of how researchers carry out their research with attention to important aspects of the context of work (e.g. organizational matters, forms of collaboration, etc.).
- Also used are tools for researchers to describe how they actually work with data and what they understand as a shareable dataset (e.g. Data Curation Profiles).
- Notably bibliometric and novel scientometric methods are rarely used. One example is a bibliometric analysis of relations among humanities disciplines which shows the close relation of archaeology with classical studies as well as with applied chemistry, a category which covers various forms of material analysis (Leydesdorff *et al*. 2011). However, such methods can hardly be used for the analysis of data sharing and re-use practices. Here the expectation is that increasing usage of data citation standards (e.g. DataCite) will allow for revealing interesting patterns as well as rewarding data sharers.

We found some surveys and studies addressing humanities scholars in general as well as in specific disciplines and specialties, including archaeologists and classicists. Concerning needs and requirements, in addition to basic ones, which may be rather common across the humanities disciplines, we also tried to identify those which may be more specific to archaeologists. However, because archaeology is a multi-disciplinary enterprise there seems to be no short cut to "the" needs

and requirements of archaeologists in e-infrastructure. We start the presentation with a summary of what various studies and reports assume to be specific characteristics and requirements of humanities scholars.

3.6.2 Specific characteristics and requirements of humanities scholars

In most reports humanities scholars are understood to present a special situation, because of the "data" (cultural content) they are working with, the expectations they have of tools, and the assumed average technical skills of the scholars. Below we summarise some characteristics and related requirements thought to be specific (or more important) for humanities scholars and their resources (e.g. Brockman *et al*. 2001; ACLS 2006; RIN 2011). The summarization is for humanities researchers in general, the characteristics and requirements of archaeologists may be somewhat different.

User characteristics & expectations

- Novelty is not a relevant criterion for humanities scholars, but usefulness is;
- They will use a tool or service when it fills an existing need, or simplifies or accelerates a task essential to their work;
- Facilitation, simplification, etc. of an existing practice must come before creating or suggesting new ones;
- A low learning curve and ease of use; if a tool requires the acquisition of too many new skills or is too complex to use, it will not be used;
- Individual research tasks often play a greater role than collaborative work.

This means that e-infrastructures, tools and services for humanities scholars must be useful for their tasks, and easy to learn and use for technically less well versed researchers (e.g. compared to researchers in the natural sciences and engineering using Grid computing). This includes "conservative" attitudes of humanities scholars making a deliberate choice not to engage with tools perceived as not immediately useful (Wouters 2007).

Data/content

- Mostly small and heterogeneous data/content (i.e. not "big data", but data that is difficult to combine into a single datasets for advanced computing);
- A large stock of analogue content (e.g. texts that need to be OCR-ed);
- Widely distributed data/content, i.e. high demand for networking of resources;
- Not particularly sensitive data as found in some social sciences or the health sector (i.e. no need for fine-grained authentication and authorization systems);
- High importance of semantics (cultural meaning, including historic change, different languages, etc.).

This means that e-infrastructure, tools and services for humanities scholars must be capable of providing access to scattered data/content that is small in volume (for the particular research task) but highly context-sensitive (i.e. requires human evaluation of meaning and relevance).

Further characteristics and criteria

- Mostly small projects in terms of budget, size of research group, etc.;
- Affordable e-infrastructure and software tools (notably, compared to other infrastructures on the ESFRI Roadmap the costs for DARIAH are minimal).

3.6.3 "Reinventing Research?" – Practices and problems of humanities scholars (RIN, 2011)

This "Reinventing Research?" study of the UK Research Information Network (RIN 2011) aimed at identifying specific issues and needs humanities scholars perceive when working with digital resources and tools. The study involved a total 54 researchers through semi-structured interviews and focus group discussions. The methods were used in the context of six different cases and contexts of digital practices. These practices included the use of specific databases (Old Bailey Online; Digital Image Archive of Medieval Music - DIAMM), use of digital resources within "traditional" Humanities departments (University of Birmingham – English Department; UCL - Philosophy Department), in a specific field of research (Corpus Linguistics), and in a collaborative project of UK and Dutch researchers (The Digital Republic of Letters).

Thus, no one case related to an archaeological field of research. However the results provide some insights into information practices of humanities scholars that involve digital resources which may be quite similar for archaeologists or classicists working with archival texts and images.

The main findings of the study in this respect are (cf. RIN 2011: 6-8):

- *Variety of sources used:* The study shows that humanities researchers "are engaging with a wide range of digital resources and services, alongside printed and manuscript texts and images". Though the research work in the surveyed use cases mainly focused on texts and images held by libraries and archives.
- *Insufficient standardisation of online resources:* An important issue is the current lack of standardisation and inconsistent interfaces and functionality across different online resources which "make for delays in research, repetitive searching, and limitation on researchers' ability to draw connections and relationships between different resources".
- *Selectivity and accuracy of resources:* While humanities researchers appreciate digital resources they are also often concerned about their selectivity, i.e. incomplete digitisation and access, accuracy of provenance information.
- *Linking currently disconnected resources:* According to the study "disconnected resources proved to be the largest barrier to accessing information"; therefore the key challenge (requirement) of humanities researchers is overcoming existing difficulties in data linking, i.e. "to improve their ability – with user-friendly tools and methods - to link data housed in different archives".
- *Sustainability of resources and tools:* Scholars worry about the long-term sustainability of digital resources and tools, i.e. that their investment in learning how to use and probably contribute to their development may be lost. They are also aware that regular scholarly input and technical maintenance is required to retain the usefulness of digital resources.
- *Technologies used:* The scholars use tools and resources that fit with their styles of research and are reluctant to use something just because it is new. The study identified "only limited uptake of even simple, freely-available tools for data management and sharing. Rather, they manage and store information on their desktops and laptops, and share it with others via email."
- *Adoption of advanced technologies:* The study also found "little evidence as yet of their taking full advantage of the possibilities of more advanced tools for text-mining, grid or cloud computing, or the semantic web". The authors of the study thought that most humanities researchers "are not moving from less complex information uses to more complex ones, but are broadening their information ecosystems".

- *Dissemination:* The dominant channels are the long-established ones such as journal articles, conferences and workshops, monographs and book chapters. Channels such as weblogs or other social media were used rarely, and the researchers expressed doubts about the quality of the content that is published in such media.
- *Sharing of content/data:* The study found that for humanities scholars the sharing of resources between researchers "is not such a prominent challenge as it is in the many of the sciences".

Overall the study confirmed the general assumptions about less advanced "digital humanities" scholars, but made clear their concerns and reasons for their sceptical position.

3.6.4 RePAH humanities portals survey (2006)

The RePAH - User Requirements Analysis for Portals in the Arts and Humanities project (2005-2006) was commissioned by the UK Arts and Humanities Research Council to evaluate the relevant branches of the Resource Discovery Network (RND), Humbul and Artifact, as well as the Arts and Humanities Data Service (AHDS). The context was the rework of the RND into Intute (Williams 2006), and therefore precarious situation of the AHDS. The RePAH project surveyed researchers' needs for information portals with focus groups, interviews, one online survey (128 respondents) and a Delphi exercise (Brown *et al*. 2006). In parallel the Log Analysis of Internet Resources in the Arts and Humanities study analysed web server log data of the portals (Warwick *et al*. 2006 and 2008).

One key result of the RePAH study was that most humanities scholars did not appreciate online material pre-culled by others. Rather they "readily accept that individual resource discovery is fundamental to their research". Thus, they want to identify and select what is valuable for their research themselves, and their interest would typically be "in the particular, or the anomalous" (Brown *et al*. 2006: 8).

This does not mean that humanities scholars are searching in the wild. Rather the identified preferences are related to what has been called "chaining", i.e. following references found in publications of other scholars or between online resources. One archaeologist noted: *"As real specialist I wonder whether a portal is good for us, because we're good at seeking out the specialist information that we need, and filtering ourselves, so you'd need a hundred experts on the different aspects of archaeology to be there to provide the right stuff from a portal".* (Brown *et al*. 2006: 190)

The RePAH online survey found that among the Classics, ancient history and archaeological scholars Google, the Humbul Humanities Hub and PERSEUS were the most popular among many other online resources mentioned. Though, the scholars were frustrated by the lack of sophistication of search engines, and the current generation of portals were not perceived as a solution for the needs of humanities scholars. This was addressed by "demonstrators" of research tools of the future which would tie in with services offered by portals. Functions and features that were evaluated as relevant by humanities scholars were (Brown *et al*. 2006: 11):

- A personally-managed research environment, e.g. searching the Web and their own hard-drive in an integrated way, selected RSS news feeds that by-passed personal e-mail accounts;
- Better resource discovery tools, including searching across distributed databases and filtering the quality of hit returns;
- Greater control over digital resources and workflows, e.g. more developed bookmarking features, personal editing features, and an automated copyright management system.

The scholars were less excited about online tools for real-time communication and collaboration such as instant relay chat and Grid videoconferencing with integrated computer applications. *"The picture that emerged is of researchers who find asynchronous and largely mono-media*

communication channels such as email, web pages and telephone quite satisfactory. (…) However most respondents declared themselves happy to collaborate at the basic level of sharing the sources they used." (Brown *et al*. 2006: 11)

The RePAH report also provides another lesson: It lauded the archaeology sector as outstanding among the arts & humanities with regard to taking user-needs into account in developing its portal services. Between 1998 and 2005 the sector had carried out user-needs surveying and evaluation work which "enabled the archaeology community to define its needs, and to see them met, in a way that is unmatched in the rest of the Arts and Humanities sector". The reference for the match was the HEIRPORT (Historical Environment Information Resources Portal) managed by HEIRNET, described as "the major portal provision for archaeologists" and the "a model for other disciplines in the Arts and Humanities to follow". (Brown *et al*. 2006: 94-95) Today, the HEIRNET website is still online, but not updated since about 2009 and the portal search function does not work (http://www.britarch.ac.uk/HEIRNET, section portal, 9.3.2014).

It may also be noteworthy that the Arts and Humanities Data Service (AHDS) ceased in 2008 and of the five AHDS archives only the Archaeology Data Service remained; Intute (until 2006 the Resource Discovery Network) ceased in 2011. Both AHDS and Intute gave up after cut of funding by the Joint Information Systems Committee (JISC).

3.6.5 "To Share or not to Share" – Classics (RIN, 2008)

The study "To Share or not to Share" (RIN 2008) was conducted by Key Perspectives Ltd. for the Research Information Network (UK). As with other studies commissioned by RIN, the motivation was to inform research funders and institutions about policies and measures that seem appropriate given the specificity of different disciplines and subject-areas, as well as different kinds of data.

The study mainly used interviews with over 100 researchers to collect and analyse information on data-related attitudes and practices of researchers in eight very different research areas: Classics, astronomy, chemical crystallography, genomics, systems biology, climate science, social and public health sciences, and the UK rural economy and land use programme.

Below we only summarise some results from classics researchers (15 interviews) who were active in art history, epigraphy, prosopography, ceramics and other specialties. The following results are of interest to ARIADNE (RIN 2008, Annex: 1-8):

- *Background:* The study notes that classics have a history of database development and electronic dissemination of data for nearly 30 years. Researchers tend to be well-informed of the issues surrounding data publication.
- *Search for relevant data/content:* By and large classicists find most data/content they need using physical and web-based aids within libraries, museums and archives, based on the knowledge of the niche in which they specialise. They also look for available information from other scholars that might be incorporated in own collections/editions. All researchers expect to be acknowledged by others who link to or use content items they have provided.
- *Own products:* Classicists quite often create datasets/databases which are used as a basis for print products (e.g. scholarly articles or monographs, sometimes with a digital supplement) as well as online products. The "raw data" is usually not shared, because the final products (e.g. lexica, catalogues, annotated texts/images) have undergone significant editing and interpretative analysis. In some cases links to other sources (e.g. online publications, thesauri/terminology, geo-spatial data) are also used to augment own online collections.
- *Data publication:* Classicists have an established culture of data sharing and take pride in publishing their products on institutional or their own websites to be acknowledged by peers.

Many of the datasets are intended to be freely and widely available, fairly cited and referenced by users. Digital formats are appreciated because they allow for easy updating of online collections and editions, periodically, or as soon as new information is available.

- *Interoperability, VREs:* Some classicists are also exploring options for interoperability of their resources with those of other scholars and using Virtual Research Environments for collaborative work.
- *Metadata:* Those who produce datasets generally appreciate the role of metadata, have created metadata, and would upgrade it if specific funds were available.
- *Archiving:* Classicists archive data when requested by funding grants and want the integrity of their datasets to be protected for the long term.

3.6.6 Archaeologists' scholarly communication (UK Berkeley, 2010)

From 2005 to 2010, The Center for Studies in Higher Education (eScholarship programme) at the University of California Berkeley (USA) carried out an exploration of faculty values, practices and needs related to scholarly communication and publication in seven disciplines (Harley *et al*. 2010a). The disciplines comprise archaeology, astrophysics, biology, economics, history, music, and political science. In total 160 interviews across 45, mostly elite American research institutes were conducted.

The results for scholars in archaeology (25 interviews) are available in a 135 page "case study". They are arranged according to topics which are introduced briefly followed by text with many quotes from the interview transcripts (Harley *et al*. 2010b). We summarise here some of the results that are of interest to ARIADNE:

- *Data management practices:* Concerning surveys and excavations, the practices appeared to be defined by varied criteria, including the nature and scope of the site/area, the training and interests of the lead researcher, the methods and types of data involved, and the influence of various other actors (e.g., funding bodies, university departments, governments, local authorities, museums).
- *Pre-publication:* Most scholars keep data and work-in-progress close to their chest before publishing any final results. For in-progress scholarly communication, conferences are an important vehicle for dissemination and publication. Overall, informal networks are the most important means of sharing early ideas and receiving feedback on drafts of work.
- *The scholarly monograph:* Publication practices are heavily determined by considerations of tenure and promotion, with the scholarly monograph as the major element (besides papers in core, high-impact journals and other recognized serials, etc.).
- *Publication in digital formats:* There is a desire for new publication formats that would support presenting various multimedia content and databases, though scholars worry about the peer-review of such publications. Mainly well-established tenured researchers are experimenting with novel forms of online communication and publication both individually (e.g. working papers published on personal websites) as well as within on-going projects.
- *Data sharing:* "Data sharing (particularly the open dissemination of data not presented in archival publication) is constrained by a variety of factors, including stakeholder interests, the sensitivity of archaeological sites, fear of poaching, concerns about the 'messiness' of data, and the lack of common data standards" (p.70); not surprisingly, most archaeologists perceived open access sharing of data difficult to accomplish.
- *Requirements for open publication of research data:* The study authors thought that "experiments in 'radical data sharing' have the potential to open up swathes of archaeological

> data for reuse", though this required: "Additional funding, clear statements on intellectual property and data reuse, centralized repositories, and funder-led guidelines for data curation are important for preserving more work generated in the field. Making more archaeological data accessible in a centralized manner may also raise questions about evaluating the contributions of multiple scholars to final publications based on 'curated' data, rather than original fieldwork" (pp. 31-32).

The study centred on archaeologists working in the context of American universities and other institutions but would be likely be similar in Europe.

3.6.7 IANUS Research Data Centre – stakeholder survey (2013)

The IANUS Forschungdatenzentrum (Research Data Centre) project, funded by the German Research Foundation, is meant to be a collaboration of all stakeholders in the sector of "Altertums-wissenschaften" in Germany (subject domains see below). The data centre is being set up with the task to bundle, archive and provide online access to research data of the sector. From May to October 2013, the IANUS project carried out an online survey to evaluate existing demands and expectations of the community – research institutes, associations, heritage authorities and others – towards the services of the data centre. About 300 German institutions were invited to take part in the survey; it was also advertised via various channels, including social media.

Fully or sufficiently filled questionnaires were received from 243 respondents, mostly from universities (108), other institutes, associations, etc. (77), museums (20) and heritage authority departments (18). With regard to the subject domains, Archaeology in general (45), Classical Archaeology (45), Archaeology of the Ancient World (46), Prehistory (13) and Near Eastern Archaeology (12) were most present.

The preliminary results (plain survey results) were published in January 2014 and are available online from the IANUS wiki (IANUS 2014); the publication of the final analysis is expected for April 2014. The main results of the preliminary analysis, provided by the German Archaeological Institute, include:

Most participants held positions at universities (45%), followed by other research institutions (32%), museums (8%), cultural heritage departments (7%), and various others (8%). The relevance of digital research data was estimated to be "very high" or "high" by 75% of the participants; the majority were quite familiar with the handling of digital data, 68% rated themselves "very skilled" or "skilled".

A vast majority declared that they store their data on the hard drives of their own computers, and a minority delivered them to central servers, computing centres or professional archives. Only 16 (out of 240) share their data via a web-based service, and typically digital data is only provided to external researchers on individual request.

Asked about any obligation to store and make their data available in a certain way, 49% denied any obligation or did not know about any, and 51% were subject to such an obligation. Asked where the data is required to be stored, 178 (of 240) participants answered that the data was to be held within their own institute (81) or even by the researcher (96).

On top of IANUS' list of tasks was expected to be provision of research data online, followed by long-term archiving. Concerning the long term curation of data, most of the participants valued "confirmability of data and results", "long term access" and "reuse of data". Less important considerations were: "visibility of own research", "possibility of citation" and "reduction of printing costs".

3.6.8 Profiling data curation practices (DARIAH, eIUS, Data Curation Profiles)

The profiling of data practices could be of particular interest to ARIADNE, in order to support not only generic uses of the e-infrastructure and services but also domain-specific uses and workflows. Profiling goes beyond schemes and shallow descriptions of so called "scholarly primitives" (Unsworth 2000), e.g. searching, collecting, comparing, annotating, referring, etc. (Palmer *et al*. 2009 provide a taxonomy of such primitives).

Classical studies facing digital research infrastructures

In the "Preparing DARIAH" project (2008-2010) researchers from ARIADNE partner Athena (Digital Curation Unit) developed a model of scholarly research activity based on "scholarly primitives" and conducted interviews with scholars in classical studies on research practices and related needs (e.g. usability of available digital resources and tools). Twenty-four semi-structured interviews were carried out with researchers at different career stages, e-research early adopters as well as "laggards".

The study found that researchers in classical studies using text-based research – considered as the mainstay of "classical" research – also engage in various other forms of research such as historical geography. Hence the need for various other digital resources and tools DARIAH might provide (Benardou *et al*. 2010; Bernardou 2011a/b).

eIUS – e-Infrastructure Use Cases and Service Usage Models

The eIUS project (2007-2009) was carried out by University of Oxford's e-Research Centre and Computing Services and the National Centre for e-Social Science (NCeSS) under a grant of the JISC e-Infrastructure Programme (UK). eIUS together with the e-Uptake and Engage projects formed a cluster of JISC-funded projects.

eIUS studied current and planned usage of e-infrastructure and perceived barriers to adoption across several disciplines. The project used the following methods (eIUS 2009):

- *Experience Reports*: concrete examples of the use of existing e-infrastructures by individuals or groups of researchers;
- *Use Cases*: derived from the experience reports to provide non-technical case histories of how users are currently interacting (or intend to interact) with e-infrastructure;
- *Domain Models and Service Usage Models*: describing patterns or combinations of e-infrastructure services required to fulfil the specific use cases.

Experience reports and use cases were produced for various disciplines, for example, earth sciences, engineering, computational biochemistry, bioinformatics, epidemiology, human geography, applied econometrics, corpus linguistics and archaeology. The work on archaeology centred on the Silchester Town Life Project. The experience report, a protocol (7 pages) with project team members; a video (5.5 minutes); and the use case description (4 pages) are available (eIUS 2009b-d). The results are related to the Silchester Project's Virtual Research Environment for Archaeology (VERA) project (Rains 2011).

Data Curation Profiles

The Data Curation Profiles project (2008-2011) was carried out by researchers at the Purdue University (Distributed Data Curation Center) and University of Illinois at Urbana-Champaign (Graduate School of Library and Information Science). The project investigated which data researchers are willing to share, when, and with whom, and their requirements for sharing the data through an institutional repository (Cragin *et al*. 2009; Witt *et al*. 2009; Data Curation Profiles 2011; Wright *et al*. 2013).

Digital curation profiles describe important aspects such as data forms and stages, tools used, value of data, copyrights, etc. The profiles were created based on the analysis of interviews and worksheets (filled in by the researchers); each profile also presents a specific dataset. Profiles have been created for cases in different fields such as carbonate sedimentology, soil ecology and plant genomics. The most important aspect is that the profiles describe how researchers work with data and what they understand as a shareable dataset (in terms of formats, elaboration, etc.). Such profiles can be a valuable resource of information for repository management and e-infrastructures, taking account of how researchers in particular fields of research would like to make datasets available. Issues in the curation of different kinds of data and their implications for re-use have also been studied by Carlson and Anderson (2007) for four very different cases.

We think that data curation profiles for particular archaeological research data could also be an interesting approach for ARIADNE. Profiles might be produced by researchers and annotated with suggestions for minimum information standards (about the investigation for which the data was produced), discovery metadata, licenses, etc. in order to enhance the potential re-use of the data. Such profiles might also be published as "data papers" describing datasets that fulfil the defined criteria and are deposited in an open access repository. An e-journal that welcomes such papers is the *Journal of Open Archaeology Data* (started 2012), *Internet Archaeology* recently has initiated a series of peer-reviewed data papers and others might follow.

3.6.9 Archaeologists' "dream tools" (Open Context, 2009)

In a project related to the development of the Open Context repository and services, researchers of the Alexandria Archive Institute (Kansa & Kansa 2009 and 2011) asked archaeologists about their "dream tool" in the context of working with primary data. When publishing the results, the project had collected responses of 45 people through interviews and workshops. As expected, the responses were quite diverse, though some common needs could be identified (Open Context 2009):

- Above all, the "dream tools" had to help work toward two ends: comprehensiveness (or at least sufficient depth of information) and efficiency (i.e. faster conduct of tasks).
- Related to comprehensiveness, there was great interest in accessing relevant literature, including grey literature, which was a particular concern of the respondents.
- One third of the respondents wanted tools that link published project information with other types of content (such as images, databases and grey literature), including integration beyond the project level.
- Also important was linking information from many sites, especially a map interface to be able to evaluate what information is available for a region.
- Furthermore field tools for recording and integrating excavation data easily, i.e. tools that help organise and integrate content of a single project from the start. This was envisaged to be a "closed" system for a team but the results might also be made searchable later by others not involved in the project.

The following things were mentioned only by a few respondents:

- Archiving data (a task, however, which is perceived to come at the end of a project);
- Community contribution (only in cases that would involve a discipline-specific forum for discussing e.g. artefacts and methods);
- Citation or worries about proper attribution of their research data.

The final project report also notes that managers of large data sets did not want to be over-burdened with metadata and data documentation, and across all participants "little motivation or interest in having researchers 'markup' their own data to align these data with more general Web or semantic

standards". Such alignment the researchers perceived as disconnected from their immediate needs as well as outside of their area of expertise. (Kansa & Kansa 2011: 5-6)

3.6.10 Creating e-research tools for archaeologists (FAIMS, 2013)

Federated Archaeological Information Management Systems (FAIMS) is an initiative to develop an e-infrastructure, tools and services for the archaeology sector in Australia. Already well advanced in terms of community building and technological development, FAIMS has also invested much effort on capturing and analysing user needs and requirements (Ross *et al*. 2013: 111-114):

One activity was an online survey (128 respondents) which, among other results, found:

- *Most desired tools:* Asked to rank a list of new tools, the most desired were:
  - Laptop and desktop applications for data entry – 92%,
  - Data analysis tools – 91%,
  - Mobile device applications for data capture/entry in the field – 87%,
  - Tools facilitating online data publication – 81%.
- *Facilitation of initial data entry:* Removing the "double-entry problem" was identified as the clear priority, because for primary data collection were used:
  - Paper – 84%,
  - Laptop/desktop – 70%,
  - iPad or Android mobile device – 14%,
  - PDA – 5%.

Another activity was a large stocktaking workshop that involved 80 archaeologists, associated researchers and software developers. Some results that might be of particular interest to ARIADNE include:

- *Flexibility:* Overall the participants' main concern was that any applications offered to them must be very flexible and capable of accommodating their existing practices, vocabularies, etc.
- *Customization:* The participants wanted to have the ability to customise data schemata and interfaces easily.
- *Shared standards:* There was little enthusiasm for adopting shared data standards or terminology; for example, to adopt other vocabularies or to record an agreed set of attributes about excavation contexts or artefacts.
- *Data compatibility:* Applications that promote (but do not enforce) good practice towards data compatibility were appreciated in principle, if they did not require significant adjustment of existing research designs, workflows or terminologies.

These results made the FAIMS team rethink their approach to interoperability, which was initially planned to build around a stable (if extensible) core of data standards, data schemata and user interfaces.

3.6.11 Scoping archaeological e-science needs (King's College London, 2006)

The UK Arts and Humanities Research Council (AHRC) sponsored a project aimed at scoping the needs of researchers in e-science, i.e. research practices and products based on advanced technologies. The background was that an E-Science Programme focused on "grand challenges" was available in the UK since 2001 but generally not well understood and used by scholars in the arts and humanities. The scoping exercise was managed by a researcher of the Centre for e-Research at King's

College London, with support by acknowledged researchers in several disciplines, including archaeology (project documentation: AHDS 2006; Anderson 2007).

For each discipline, an expert seminar with 10-12 researchers was run, prepared, moderated and reported by the lead researcher. Below we summarise some results of the expert seminar (Kilbride 2006):

- *"Bigger" is not the answer:* The expert group agreed that doing what some archaeologists were already doing at a bigger scale was not the answer to the e-science challenge. Examples were: remote sensing, imaging, surface modelling, volumetric modelling, numeric simulation modelling, geo-spatial and geo-temporal processing and text mining.
- *Archaeological e-science "way-finding" initiatives:* The expert group felt that "way-finding" was needed to help archaeologists along the road of e-science development, and suggested the following initiatives:
  - Studies of large scale data gathering and delivery;
  - Data mining to extend data sharing beyond metadata;
  - Ontology development and testing as a support to data mining and integration;
  - Development of simple data exposure and integration tools;
  - A review of web services and their relation to e-science;
  - Reconnaissance of e-science tools for archaeological sciences;
  - Transition from research frameworks to grand challenges.
- *Going beyond resource discovery:* The group also agreed that one challenge was unlocking the great potential of the many data sets and tools that had already been produced. Resource discovery was seen as important but to exploit existing potential more sophisticated search and retrieval would be required.
- *Preparing and linking for computation:* There was consensus that a lot of work was needed to make archaeological datasets fit for advanced computational processing. Given appropriate datasets, persistent identifiers and deep and dependable linking between datasets and tools was required beyond file level to items within files.
- *Generic principles:* The expert group also considered and recommended 12 principles for e-science activities in archaeology, e.g. that they should focus on the needs of archaeology; be recognised as fundamental research; assessed on the basis of archaeological values; pushing the boundaries in the use of ICT in research; involve the whole community or sector (not only the academic segment).

3.6.12 Summary and results highlights

Below we summarise the content of the previous sections and highlight points which are relevant for ARIADNE project WPs and tasks.

Data/content sharing

Informal contacts and networks play a major role in gaining access to the data of colleagues which are not already used for publication. (cf. 3.6.6) Humanities scholars in general seem willing to share some data with other researchers, which may have to do with the fact that the "data" is mostly cultural content, typically from archives, libraries and museums. But the "raw data" such as a database that is used for the creation of a scholarly edition is usually not shared. The scholars wish to be acknowledged, cited and referenced by peers for their published works (cf. 3.6.3).

The IANUS stakeholder survey confirms that researchers are not yet prepared to make data available beyond the project or institutional level, i.e. to researchers not involved in projects (cf. 3.2.4 and 3.6.7). Data is predominantly stored on own computers, only a minority provide data to a central server, professional archive or computing centre. With regard to existing obligations to make data available, 178 of 240 respondents (74%) could comply by holding it and making it available in their institute (81) or at the level of individual researchers (96). Only 16 out of 240 respondents share data via a web-based service. Mostly some data is provided to external researchers only upon individual requests.

Results highlights
These results add to those highlighted in section 3.2.4. Especially the importance of informal relations and networks for gaining access to data which has not been published. In social terms the demand for "open access" probably can be understood as the wish to get access based on common, formally established mechanisms.

Profiling researchers data curation practices

Projects have in various ways sought to capture the need of researchers for tools and services which might support and possibly enhance their research practices. Typically use cases towards envisaged services have been produced based on interviews, focus groups, etc. (e.g. in the eIUS project). One field of practice which might be enhanced is the preparation of shareable data or datasets. The provision of data and adequate metadata to a repository is typically not an immediate objective of researchers when producing and managing project data. But the route to sharable data starts at this level. Profiling researchers' practices in this area aims to understand in sufficient detail how they prepare data (e.g. to underpin a publication) and what they understand as shareable data, beyond just presenting summary tables, charts, etc. in a publication. This has been explored by the Data Curation Profiles project in view of enhancing researchers' data sharing through repositories (cf. 3.6.8). Profiles like the ones produced in this project might also be relevant for ARIADNE.

Results highlights (incl. references to project tasks)
The path to sharable data starts at the level of data curation within projects. Curation profiles elaborated by researchers based on a defined scheme could be annotated with suggestions for minimum information standards (about the investigation in which the data was produced), discovery metadata, licenses, etc. in order to enhance the re-usability of the data. The profiles could be evaluated and published as "data papers" (i.e. a journal publication), which describe datasets that fulfil the defined criteria and are deposited in an open access repository. Probably also a service might be conceived that proposes templates for different types of data, licensing options, etc. *Relevant for WP4, Tasks 4.5/4.6: Good Practices; training activities under WP5, and probably for Task 2.2: Digital Archaeology SIG and WP17: Innovation in Archaeological Research Methodology (e.g. with regard to data curation workflows in projects).*

Humanities scholars' common requirements for tools and services

We certainly can assume that humanities researchers across the board welcome tools and services which they perceive as useful, i.e. in some way simplify or accelerate a task essential to their work. But a low learning curve and ease of use is required for adoption by a wider user community. The usefulness of a suggested new tool or service is evaluated with regard to the individual or group of "scholarly primitives" it aims to support (searching, collecting, comparing, annotating, etc.) and what researchers are already using. To invest time and energy in learning how to use a tool effectively, researchers need to be convinced that it will not become obsolete rapidly (cf. 3.6.2 and 3.6.3).

Archaeologists produce most of their data themselves whereas other humanities researchers predominantly work with cultural content, historical or contemporary, produced by others (cf. section 3.5). The "scholarly primitives" of archaeologists are roughly the same if they can search, collect and work with relevant archival and new material for "their" site as well as many sites (i.e. in comparative research for broad synthesis or meta-analysis). This will include records of heritage authorities, museum specimens, historical texts, inscriptions, drawings, maps, etc., with the latter probably already worked on by philologists, epigraphers and other humanities scholars. For "dream tools" for established forms of research and novel ones for advanced e-research see below.

What all humanities researchers wish to do is searching across distributed resources and more effectively filtering hit returns. What most humanities scholars, including archaeologists, do not appreciate are online collections pre-culled by others like library staff and other subject-experts not acknowledged in their niche of study (cf. 3.6.4, especially the quoted statement of one archaeologist).

Like other humanities researchers, archaeologists complain about inconsistent interfaces and functionality across different online resources (which make it time-consuming to spot relevant material), insufficient provenance information, and lack of capability to establish relations between scattered and heterogeneous resources. What scholars like philologists, epigraphers, iconographers and others might need more than archaeologists are online tools for editing, annotating, linking and updating a scholarly edition of a corpus of cultural content, on their own or institutional website (cf. 3.6.3).

Results highlights (incl. references to project tasks)

a) Common basic requirements of humanities researchers with regard to novel tools and services:

- Useful tools/services – should in some way simplify or accelerate essential tasks;
- Better than what is already in use;
- Low learning curve and ease of use;
- Affordable and sustainable;

b) Most common needs:

- Searching across distributed resources;
- Filtering hit returns more effectively;
- Not appreciated: online collections pre-culled by others;

Relevant for Task 2.1 and Task 12.2, definition and specification of "most wanted" tools and services.

Differences: For the same tasks ("scholarly primitives") the needs of archaeologists and other humanities researchers will be quite the same. Considerable differences can be found where different research methods are applied to produce different products (e.g. an archaeological field survey vs. a scholarly edition of inscriptions).

Relevant for Task 2.2: Digital Archaeology SIG, and WP17: Innovation in Archaeological Research Methodology.

Archaeologists' "dream tools" for established forms of research

Archaeologists' seem to seek improvements especially in the initial phase of the data lifecycle (field recording, data entry) and when publishing project results online (cf. 3.6.9 and 3.6.10). They do not like significant changes in existing research designs, workflows or vocabulary. Achievement of data compatibility and interoperability are a concern, though the willingness to adopt unfamiliar standards is generally low.

Results highlights (incl. references to project tasks)

Characteristics and functions of "dream tools":

a) General criteria:

- Capable of accommodating existing practices, vocabularies, etc.;
- Highly flexible, ability to customise various things easily (e.g. data schemata, interfaces);
- Enabling provision of comprehensive project information.

b) Initial data capture and entry:

- Field tools and mobile applications for data recording;
- Efficiency gains, e.g. removing the "double-entry problem" (getting rid of paper templates).

c) Online publication:

- Linking project information with various other accessible data/content (e.g. grey literature, images, databases), also beyond the project level;
- Integration of information from many sites, especially map interfaces (e.g. what information is available for a region).

d) Not appreciated ("nightmares"):

- Adoption of unfamiliar (but predefined) data standards, schemata, vocabulary, user interfaces;
- Required markup of data to align it with more general Web or semantic standards (perceived as disconnected from immediate needs, and outside of practitioners' area of expertise).

a: Relevant for Task 2.1 and Task 12.2, definition and specification of "most wanted" tools and services.

b: Not directly relevant for ARIADNE.

c/d: Highly relevant for WP3, WP14 and WP15.

Way-finding towards novel forms of archaeological e-research

ARIADNE e-infrastructure and services might support first steps in advanced e-research practices, for example, aggregation, integration and mining of datasets. As suggested by an archaeological e-science scoping exercise, this would require going beyond resource discovery across distributed datasets (cf. 3.6.11).

Results highlights (incl. references to project tasks)

Overall assumptions about archaeological e-science challenges:

- Unlocking the potential of archaeological datasets for e-science methods, which are understood to use large datasets and advanced computational processing;
- Achieving a higher level of datasets integration fit for processing.

ARIADNE services might provide/support:

- Identification of available datasets which could be relevant;
- Data aggregation, search and delivery (e.g. via web services);
- Links to relevant tools/workbenches (e.g. data exposure and integration tools);
- Deep and dependable linking between datasets and tools (beyond file level);
- Ontologies for data mining, integration, analysis.

Probably relevant for WPs 14, 15, 16 and 17

4 Conceptual framework: users and user requirements

4.1 Working definition of key terms used

This report describes user needs with regard to data in archaeological research. As a starting point, it is necessary to specify our understanding of some essential terms such as "users", "data" and "needs". All of these terms are widely used in everyday language in different contexts; one could therefore assume that their meaning is intuitively clear. Unfortunately, terms which are used in many different contexts tend to be vague when applied to a specific context. In this section, we outline our understanding of these key terms as it applies to the ARIADNE project and to this report. In particular, we suggest a framework for describing the user communities of the ARIADNE project and how they relate to each other in a systematic way (see Section 4.2).

"Stakeholders"

With "stakeholders" of the ARIADNE project, we refer to projects, institutions, companies or other entities that have an interest in the project (for instance because they are affected by the project outcomes), and/or entities that may have a (positive or negative) impact on project completion. This includes internal stakeholders that are actively involved in the project (the members of the project consortium, the project sponsors) and external stakeholders. External stakeholders include, in particular but not only, the targeted users of the services that will be produced (e.g. researchers, research institutions), archaeological data repositories (e.g. if contributing to the e-infrastructure to be developed), technology providers (whose tools are needed to establish the e-infrastructure) and related initiatives pursuing similar objectives.

Sometimes, when mapping out the stakeholders, a distinction is made between primary stakeholders and secondary stakeholders. This distinction tends to be more adequate for mapping corporate stakeholders, however, than for mapping project stakeholders. Transferring the concept to a project context,

- **Primary stakeholders** are those who engage in direct (economic) transactions with the project: the project "customers" (i.e. the targeted users), the "suppliers" (the consortium members, the funding organisation, the technology providers).
- **Secondary stakeholders** are those who are affected by the actions of the project or can affect its actions, without being directly engaged in economic transactions with it, such as other projects and initiatives working on similar topics, the archaeological research community in the wider sense, specialised media).

Stakeholder analysis is an important early task in almost any larger project. It involves identifying and grouping the relevant stakeholders, analysing their position towards the project, and establishing an action plan of recommended activities to address specific stakeholder communities. In ARIADNE, the stakeholder analysis has been dealt with as part of Task 2.3 – Liaison with related initiatives. For this particular report and project deliverable, the detailed results of the stakeholder analysis are not required. In this report, the focus is on a specific segment of stakeholders: the potential users of the ARIADNE e-infrastructure and their requirements.

"Users"

"Users" is a central term in this report. The term is often used, in the widest sense, to describe the target communities for whom the ARIADNE project sets out to develop an e-infrastructure and services based on this infrastructure. The "users" in a project context (at least to a large extent) are the equivalent to "customers" in a business context.

The difference between the terms "users" and "stakeholders" is simple: users of the ARIADNE project are an important segment of the project stakeholders. In other words: all users are stakeholders, but not all stakeholders are necessarily users (of the project results).

While such a basic definition of "users" is easy to establish (and will be sufficient for many purposes), it is much more challenging to identify and describe the various user communities in a systematic and coherent way. The difficulties result, inter alia, from the following three factors:

- **Institutional vs. individual users**: "Users" can be framed and looked at from an institutional perspective (e.g. research institutes) or from an individual perspective (researchers).
- **"Customers" vs. suppliers of data**: "using" a data centre can either mean searching and possibly downloading existing data that has been produced by others and is available at the data centre (e.g. when conducting research for a project), or it can mean depositing new data which the user has produced with his/her research project. To deal with these two important use cases, we have adopted a work-flow perspective on data management communities which takes a holistic view and encompasses both cases.
- **Researchers vs. repositories as users:** the ARIADNE project addresses two basic user communities: the research community that carries out research projects and generates data with these projects (researchers, institutes); and the archaeological data centres and repositories (institutional, domain-specific, international) where data can be deposited and which thus act as intermediaries for sharing data among the research community. Both are users, but the requirements and expectations they have towards the ARIADNE project can be quite different.

Based on these general considerations, we have developed a four-level scheme of users of the ARIADNE project (see Section 4.3) which describes the various user communities and how they relate to each other in more detail. This conceptual framework is based on a work-flow perspective of the research process (data search –> data generation –> data depositing and sharing, see Section 4.2); it describes the main individual and institutional players involved in this process.

Needs, requirements vs. tools and solutions to address them

The terms "needs" and "requirements" are used synonymously in this report. This is a slightly different use of the terms than in engineering and software projects, where **needs analysis** often "sits alongside requirements analysis and focuses on the human elements of the requirements".[1] The broader process of **requirements analysis**, from a software engineering perspective, "encompasses those tasks that go into determining the needs or conditions to meet for a new or altered product, taking account of the possibly conflicting requirements of the various stakeholders, analysing, documenting, validating and managing software or system requirements."[2]

For the purpose of this report, we adopt a modified version of the engineering-focused definition of requirements analysis.[3] Here, user requirements (or user needs) describe those conditions, services or features (whether existent or non-existent) which the various user communities desire in order to be able to effectively and efficiently carry out their professional activities. The goal of the ARIADNE project, from this perspective, is to contribute to better fulfilling user requirements in archaeological research, for instance by enabling new services and tools which meet user requirements in a better way than the existing services and tools.

---

1 see "Needs analysis" on Wikipedia: http://en.wikipedia.org/wiki/Needs_analysis (accessed in Feb. 2014)

2 see "Requirements analysis" on Wikipedia: http://en.wikipedia.org/wiki/Requirements_analysis (accessed in Feb. 2014).

3 for the technical work packages of ARIADNE, and in particular for WPs 12 and 13 (which also identify user requirements, but specifically with regard to the technological concepts) it may probably be more useful to stick to the classical concept of requirements analysis.

We also recommend not confusing *needs* (or *requirements*) and possible *tools* (i.e. *solutions)* to address these needs. While we should think of a need as a "job to get done", the solution is the means to carry out the job. For instance, when somebody wants to hang up a picture in their living room, they may consider using a hammer to put a nail into the wall and then hang the picture on the nail. In this example, the hammer and the nail are, strictly speaking, not a need – but a solution to address the need. The need (the job to get done) is: "I would like to be able to hang up pictures in order to decorate my living room". It may sound somewhat theoretical to introduce this distinction; however, we suggest it is very important not to jump to conclusions about what is the most effective and efficient solution. It also helps to avoid an overly technical perspective (thinking in terms of solutions) in cases where a demand-driven perspective is preferable.

"Data"

The most challenging and potentially conflicting definition is what we understand as "data" (or, rather, what is *not* considered as data in the strict sense). While this is a decision which is ultimately beyond the mandate of WP2 and thus a bit outside the scope of this deliverable, we had to adopt at least an implicit, operational definition for the purpose of the online survey, as most of the questions implicitly or explicitly address data management issues.

The challenge arises, to a large extent, from the wide use of the term "data" in many different contexts, including technical terminology as well as everyday language. This can lead to considerable confusion as to what we mean by "archaeological research data", for instance in interviews with researchers. The following basic distinctions between different types of data can be helpful in this context, if not to sort out all difficulties then at least to be aware of potential pitfalls and misunderstandings.

- **Raw data vs. processed data**. A major distinction is to be made between raw data (i.e. unprocessed data as collected, for example, in field surveys, in labs or through experiments and processed data (for instance aggregations of raw survey data as presented in this report). Processing normally occurs in different stages. In research projects, for instance, the final level of processing (and highest level of aggregation) is typically the final project report which informs about the overall results, but may not contain any of the available raw data on which the conclusions are based.
- **Metadata vs. the 'actual' data**. Another important distinction is whether it is "data about data" (metadata) describing the structure and/or content of data, or whether the discussion is about the actual research data themselves. Both data and metadata are highly relevant for the ARIADNE project and its users.
- **Different types of data**. In archaeology, "data" can mean a lot of different things – including images, texts of different types, GIS and other location-based data, maps or technical data from lab tests.

A critical question for this study of user requirements was how to set the focus with regard to "data" considered. During the interviews with researchers, we noted they tend to think of "archaeological data" in a rather broad way. They would not only think of raw data, but include highly processed data (such as project reports and grey literature). Given the importance of processed (aggregated) data for many research purposes, we opted for a rather wide and unspecific working definition and did not discard any type of data or any stage of processing from the analysis. Essentially, this means that the term "data" as used in this report is very close to the concept of "information". Some argue that "information" is a type of processed data, i.e. data enriched with interpretation and context; in this document, the terms are not clearly separated. The study explored the requirements of users with regard to research data and information irrespective of the degree of processing or aggregation, and the type of content described by the data.

4.2 A workflow perspective on data management communities and their requirements

As outlined in the introduction, the ARIADNE user framework we have developed for this study is based on a data workflow approach. We distinguish between four levels of data workflows and the respective data management communities (with a broad understanding of the term "management"). This approach takes account of the fact that there are further direct users of the ARIADNE infrastructure, services and tools beyond the research community.

We understand that ARIADNE does not intend to replace any existing infrastructures, services or tools but provide "integrated services" on top of them, thereby making currently isolated data resources and services better accessible and useable for various research purposes.

The data workflow scheme represents an "up-stream model" where:

1. data is produced and managed by research projects or in the context of other work carried out on archaeological sites and objects, e.g. heritage management (Level 1),
2. the content and data or datasets together with metadata is deposited in institutional repositories (Level 2) or data centres or subject- and domain-based repositories (Level 3),
3. and the metadata from several repositories is collected (e.g. harvested) into a common metadata pool, and search and other services are provided based on the metadata (Level 3). Furthermore on this level there can also be special services that support some workflows such as controlled vocabulary services, for instance.

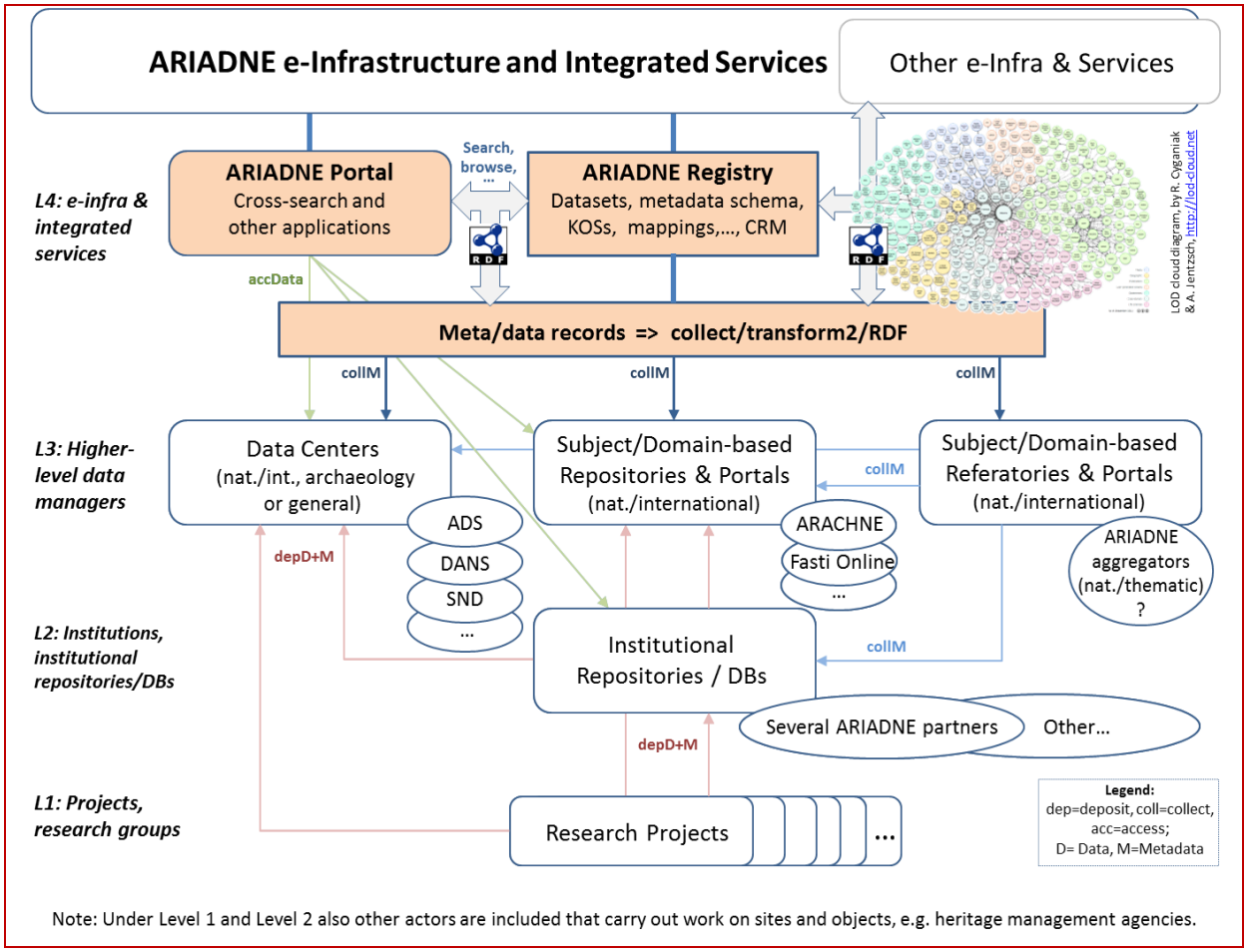
The ARIADNE e-infrastructure and integrated services operate at a higher level, Level 3 upwards, and constitute a further layer (Level 4). Services at these levels represent the integrating function of ARIADNE, e.g. enable cross-search and other services for the archaeological research community.

However, there may also be a need for services "down-stream", in support of other aggregators and service provides (e.g. subject-based portals), institutional repositories and research projects. Base level data management communities (Level 1: research or other projects, Level 2: institutional repositories) often operate rather self-sufficiently, i.e. do not necessarily have "up-streaming" of their data as a priority. Therefore providing special services for them may help "open up" and promote wider sharing of more archaeological data.

4.3 The four levels scheme of users

For dealing with stakeholders and analysing user requirements, we propose a four-level user framework. This distinguishes four institutional levels of relevant user communities. Figure 4.3-1 depicts the major elements of this framework which are explained and described in the following sections.

Figure 4.3-1: Conceptual framework for the analysis of ARIADNE stakeholders and users: the four levels scheme



Note: Under Level 1 and Level 2 also other actors are included that carry out work on sites and objects, e.g. heritage management agencies.

4.3.1 Research projects

Research projects constitute the first level in the framework. The foundation of new knowledge, or revision of established hypothesis, is mainly produced in research projects. The basis for this is the research data that is acquired, managed, analysed and interpreted by the researchers. Such a project can be, for example, an excavation with a lead excavator and core team, and related subject experts for certain types of objects (e.g. epigraphy, numismatics, botany, zoology, etc.). Other relevant projects carried out on archaeological sites and objects, for example, include the work of heritage management agencies.

Among the project team members, the project data managers are especially important for ARIADNE. This can be is a researcher from among the team or another team member with a special focus on managing the data that are generated by the project (in particular, in large and longer-term projects). Workflows of research projects comprise study of the state-of-the-art, research design (e.g. sampling strategy), data creation, data management, data analysis and interpretation, and finally publication of research results (usually reports and papers, with supplemental material).

The results generated by the project will, ideally, be deposited in an institutional repository (Level 2) or data centre or subject- and domain-based repository (Level 3). Results could include, for instance, excavation or field survey reports, datasets and other content (e.g. images), including "self-archived" papers, book chapters or monographs. The repository or data centre should (ideally) also be

accessible to other researchers so that the empirical evidence can be scrutinised and the data re-used for further research on similar or alternate research questions.

We understand that the ARIADNE e-infrastructures and services will not support the various workflows, methods and tools which projects use to produce data (e.g. in excavations, field surveys and laboratory analysis). ARIADNE will mainly support the discovery and access to data that has already been collected, used for publications, and deposited in institutional repositories, data centres and subject-based repositories. From that perspective, ARIADNE can be seen as a meta infrastructure.

In this study on user requirements, the research project's level is represented by **individual researchers** who have either been interviewed during the pilot interviews (see Section 5) or who participated in the ARIADNE User Survey (Section 6). The sample of the User Survey includes 586 archaeological researchers.

4.3.2 Research institutions and institutional repositories

Research institutes and research centres are typically the institutional umbrella for carrying out research projects. Projects are often centred on focus areas (e.g. of periods, subjects, methods) and defined by the mission and research programme of the institute. There are also other relevant institutional actors (such as heritage management agencies). For the purpose of identifying user requirements, however, we focus on research institutions.

At such institutions or centres one or more research directors oversee and give advice (e.g. on research design, methods, etc.) to several projects. Furthermore, most research institutes manage an institutional repository where projects can (should) archive publications, data and datasets produced in the projects.

Thus at the institutional level database or repository management is carried out by dedicated IT or data management staff for the content (data) of on-going projects (if such data is deposited regularly) and completed projects. Workflows at this level include, but are not limited to, support in the depositing of Level 1 data (especially with regard to metadata creation), archiving and curation (e.g. migration of data), and local search & retrieval and other services.

If the repository or database is meant to be also accessible to external users, the IT or data management also implements external search and retrieval and other services. The metadata of the repository or database may also be exposed to OAI-PMH harvesters which collect it into a metadata pool of several institutions and provide services on top of it (e.g. search, alert and other services).

Furthermore current research information will be available (i.e. information about researchers, projects, funding, etc.). This information is typically managed in a separate administrative system, which may or may not include links to the research results such as publications and datasets (on current Research Information Systems [CRIS] see: EuroCRIS, http://www.eurocris.org).

According to the data workflow scheme, the core stakeholder categories for ARIADNE at this level are research directors (overseeing several projects) and managers of the local IT infrastructure, institutional databases and repositories, and local services.

In this study, the research institutions level is represented by **directors of research institutes** who have either been interviewed during the pilot interviews (see Section 5) or who participated in the ARIADNE User Survey (Section 6). The sample of the User Survey includes 54 directors.

4.3.3 Data centres, subject- and domain-based aggregators

Alternatively or in addition to institutional repositories, in some countries archaeological data centres have been established. Furthermore in various areas, archaeological research domain or subject-based repositories and portals are available.

The main differences between these entities are: data centres support researchers or research projects of many institutions in a country, while subject or domain based repositories typically are internationally focused, and may comprise content and data from several archaeological subject areas or domains (e.g. prehistory or classical archaeology) or support research specialties like zooarchaeology, epigraphy, etc.

More specifically we distinguish:

- **Data Centres**: In such centres, research projects (researchers, research groups) from different institutions deposit data, datasets and other research output (e.g. research reports, papers, etc.), and the centres have implemented workflows to preserve, curate and allow for access to the data and metadata in the long-term.
- **Subject or domain based repositories**: Such repositories also receive deposits of data and metadata from Level 1 (research projects) and maybe Level 2 (institutional data managers); the workflows are similar to those of institutional repositories and data centres but the provision of access through a portal plays a more important role.
- **Subject or domain based referatories**: Collect only metadata from many institutional and other repositories (mostly via OAI-PMH harvesting), hence, do not hold the research results (i.e. publications, datasets, etc.). Based on the aggregated metadata they provide various services (e.g. search and retrieval, alerts, etc.).

Subject- or domain-based repositories and referatories typically run a portal that provides several user-focused services which can include search (e.g. keyword-based, facetted searching and browsing, "related content"), alerting, bookmarking, myCollection, annotations, etc.

Such services are relevant to research projects (Level 1), e.g. for searching publications on the state-of-the-art in a field of research, methods used, available data, etc. As they allow for searching across content and data from many projects, such portals can also promote research with a wider scope, e.g. cross-regional comparison, interdisciplinary research or other).

The core category of stakeholders here are of course the managers of the repositories and referatories, including portal managers. Under Level 3 we also include providers of special services like controlled vocabulary and terminology services, for instance, which should be included in the ARIADNE "semantic infrastructure".

The complexity with this level is that repositories are "users" and "providers" of services at the same time, depending on the perspective. From an individual researcher's perspective, they are a provider of data services; from the ARIADNE project perspective, however, repositories are both suppliers (of metadata and data) and users (as beneficiaries of the e-infrastructure to be developed).

In this study, Level 3 (data centres and subject- or domain-based aggregators) is represented by **managers of data repositories** whose holdings include, inter alia or exclusively, archaeological research data. The repository managers were either interviewed during the pilot interviews (see Section 5) or participated in the ARIADNE User Survey (Section 6). The sample of the User Survey includes 52 repository managers.

4.3.4 e-Infrastructures and integrated services

The ARIADNE e-infrastructure and integrated services operate from Level 3 upwards, constituting a further layer (Level 4). ARIADNE does not intend to replace any existing infrastructures, services or tools on the other levels (Levels 1–3), but provide integrating functionality and services on top of them, thereby making currently isolated data sources and services better accessible and useable for various purposes of research.

Such functionalities or services may include:

- Registration of Level 2 and 3 repositories (e.g. OAI-PMH targets, SPARQL endpoints, etc. allowing for identification and access based on criteria such as metadata, controlled vocabularies used, exchange formats, licenses;
- Registration of metadata schemas and KOS including metadata cross-walks, KOS mappings, etc., providing a "semantic infrastructure" for the sector;
- Services for Level 3 aggregators and service providers (such as metadata de-duplication, indexing, annotation);
- ARIADNE portal with various features for Level 1 users: e.g. search and browsing, visualization of networks among research groups (e.g. based on co-authoring, co-citation, etc.).

Going beyond ARIADNE, co-operation with developers and providers of other e-infrastructures and services will be established aimed at enabling interoperability, e.g. DARIAH, CENDARI, Pelagios and many others, including closely related domains of research.

This level of the user framework (e-Infrastructures and integrated services) was not directly represented in the interviews and surveys, as this level describes a provider of relevant services rather than the users. (This report is about user requirements and actually addresses the provider).

4.3.5 Repository typologies and archaeological repositories

Some notes seem appropriate concerning the typology of repositories we use in the four levels scheme. There is an extensive literature on digital repositories, including various studies and surveys (e.g. Adamick & Reznik-Zellen 2010; Armbruster & Romary 2010; Burns *et al*. 2013; COAR 2013; e-SciDR 2008; Marcial & Hemminger 2010; SURF 2008). We mined the literature to first understand different types of digital repositories and secondly, to identify some available repositories beyond the ARIADNE partnership.

Repositories typologies

There are different kinds of scholarly and other repositories that can be distinguished according to criteria of scope. Most widely used is the distinction *Institutional* vs. *Subject-based* repository, i.e. a repository of an individual institution (only comprising content of affiliated researchers) versus a repository of a discipline, domain or even specific sub-field of research where researchers make accessible publications on relevant topics. In a study on different types of repositories, Adamick & Reznik-Zellen (2010a) found that institutional repositories receive the lion's share of attention, while subject-based repositories, especially the smaller ones, are "under-studied and under-represented" in the literature.

The Directory of Open Access Repositories (OpenDOAR) distinguishes between institutional (or departmental), disciplinary (cross-institutional, subject based), and governmental repositories. Furthermore, OpenDOAR identifies aggregating archives, which receive digital content and metadata

from several subsidiary repositories. Some also receive direct deposits by many individuals who have no other fitting place, e.g. Zenodo (offered by CERN, "enabling researchers to share and preserve any research outputs in any size, any format and from any science").

Importantly, OpenDOAR considers only repositories which hold data. A further category sometimes used in the literature is "referatories" which collect metadata from several repositories and provide services based on the aggregated metadata pool. They do not hold content but only refer and link to it. This term is useful to make clear the distinction. For the ARIADNE users framework we decided to use the basic categories of institutional repositories and subject- or domain-based repositories, and included the important additional category of referatories.

A sub-classification applied specifically to subject-based repositories is *Multi-disciplinary*, *Inter-disciplinary*, and *Single-subject* (Adamick & Reznik-Zellen 2010b). Here *Multi-disciplinary* means a non-integrative array of disciplines and *Interdisciplinary* a field where traditional academic boundaries are crossed or blended. *Single-subject* of course means a repository dedicated to a single subject (e.g. the Archaeology Data Service, UK).

Still another classification is *Research*, *Subject-based*, *Institutional*, and *National* (Armbruster & Romary 2010). Here *Research* stands for a repository that offers services which support research tasks of research communities beyond general deposit and search functionality. The authors emphasise this category and would like to see two shifts in the current focus of repositories that aim to serve research communities:

"With regard to content, a well-defined and high quality corpus is essential. This implies that repository services are likely to be most successful when constructed with the user and reader uppermost in mind. With regard to service, high value to specific scholarly communities is essential. This implies that repositories are likely to be most useful to scholars when they offer dedicated services supporting the production of new knowledge" (Armbruster & Romary 2010: 1).

This resonates with the main result of the "Supporting Research" project (2010-2011) that found a strong dissatisfaction of researchers with repository services of universities which invite them to deposit research output. Such repositories are often seen more in the service of the institution rather than research (MacColl & Jubb 2011: 3-4).

Archaeological repositories

In the literature scanning we found little specifically on archaeological digital repositories, except of publications of the archaeological data centres and services Open Context (Alexandria Archive Institute, USA), tDAR (Digital Antiquity Consortium, USA), Mappa (University of Pisa, Italy), and of course the ARIADNE partners ADS (UK) and DANS (Netherlands). The most active promoters are the Open Context developers and researchers (cf. Open Context bibliography; on tDAR see Kintigh & Altschul 2010, and on Mappa see Anichini *et al*. 2012).

The best available source for searching repositories, the Directory of Open Access Repositories (OpenDOAR), was not helpful for spotting archaeological repositories. OpenDOAR currently holds information about 2603 repositories around the world. The directory does not expect that an "open access repository" makes all content available without access control, but some should be. Typically this is a collection of publications, learning resources, multimedia (or a mix thereof).

Most of the repositories covered are institutional repositories (82.6%), followed by disciplinary or subject-based repositories (10.9%), while aggregating and governmental repositories make up only 3.7% and 2.8% respectively (figures as of 12 March 2014). The repositories of the larger institutions like universities are "multidisciplinary", i.e. hold content of all or most subjects (1548 repositories). The bulk of the content is publications. 626 (24%) of the repositories hold also multimedia and audio-visual materials, but only 108 contain datasets (about 4%).

212 of the 2603 repositories (about 8%) have "History and Archaeology" among their subjects. The pattern of content across those repositories is somewhat different than in the total, while also only few (6) hold datasets. 109 of the repositories (51.4%) are located in Europe. In most, "History and Archaeology" is but one among several subjects. Repositories with only this subject mainly hold historical content (e.g. Biblioteca Virtual de Aragón or Ergani - Historical Archive of Aegean).

The few repositories with archaeology, anthropology or classics material are: Archaeology Data Service (UK), Acropolis Educational Resources Repository (Acropolis restoration service, Greece), Parthenon Frieze Repository (Greece), Digital Library of Polish Institute of Anthropology, Propylaeum-DOK (Heidelberg University Library), Riksantikvarens vitenarkiv (Directorate for Cultural Heritage, Norway), DIGIMOM (Maison de l'Orient et de la Méditerranée, France). Only one repository for an open access archaeological journal is present, the JIIA Eprints Repository (249 items) of the Journal of Intercultural and Interdisciplinary Archaeology. All contain mainly publications, except the Archaeology Data Service which is the only data repository in the sample.

Also the Registry of Open Access Repositories (ROAR) allows for searching repositories. ROAR offers more detailed categories like "Archaeology", "History of Civilization", "Ancient History" and "The Greco-Roman World", but they only return 12 results. These included two mentioned above (Acropolis Educational Resources and Parthenon Frieze Repository) and other relevant ones like Human Origins (University of Southampton, UK), the Encyclopedia of Iranian Architectural History (IranShahrPedia), Faculty Scholarship at the Claremont Colleges (California) and the Historical Philological Journal of the National Academy of Sciences of Armenia. The other results were various repositories of learning resources, publications and thesis (e.g. in Ireland, Ukraine, Peru, El Salvador).

Fragmentation of the archaeological research data landscape

Although but one general indicator, OpenDOAR's small coverage of repositories with a special focus on archaeology supports the assumption of a highly fragmented landscape of archaeological research material. Major factors that lead to this fragmentation may be the perspective and forms of data organisation in the field of archaeological research. The pre-dominant perspective seems to be project-oriented rather than institutional. In terms of data organisation, this could mean that "project archives" (one per excavation site) and "database" projects of small groups or even one researcher are the most frequent forms; "collections" as found in the libraries and museums sector would be less common.

Organising this into a common repository requires a system, workflows and dedicated staff which may not be available in most research institutions. In short, we think that the forms of data organisation in archaeological research will probably require an investigation that goes beyond the concept of institutional repositories. A fragmented situation of data organisation at the base level of archaeological institutes could mean that it is necessary to promote a more "collections" style organisation of data sets (i.e. with metadata that are harvestable based on the Open Archive Initiative protocol). An alternative might be to promote data centres, which are not yet available in many countries, but might be the most effective solution in terms of capacity and cost.

5 Pilot interviews

Introduction

In the pilot interviews project partners were asked about conditions of data search, production, management and access in the context of their work as well as about advantages they expect from the ARIADNE e-infrastructure and services (under development) and other project activities. The interviews were conducted to gain first insights in partners' perception of current conditions and for preparing the online survey.

Most questions from the interview templates have been included in the online survey, extended and adapted for the quantitative data collection. However, the qualitative results of the interviews can also stand on their own, as a record of observations and opinions on various matters of data management, search and access in the sector.

26 representatives from the ARIADNE partners and the wider project community were interviewed or provided written input in the interview templates. Two different templates were prepared, one for researchers and data managers of projects (level 1 and 2 of the ARIADNE users framework), and one for managers of data centres (level 3).

Most of the interviews (18) involved senior researchers (incl. heads of research groups, project directors) and a few directors of research institutes or departments as well as young researchers (PhD candidate, research assistant). The participants work at institutes in 10 countries, Austria, Cyprus, Greece, Hungary, Ireland, Italy, Netherlands, Romania, Sweden and the United Kingdom, with a particularly strong contribution of colleagues in different positions at The Discovery Programme (Ireland) and ÖAW (Austria).

Eight interviews were conducted with managers of data centres, repositories and service providers, including Archaeology Data Services (ADS), UK; Data Archiving and Networked Services (DANS), Netherlands; Swedish National Data Service (SND); ARACHNE, Germany; Open Context (Alexandria Archive Institute), USA, and three service providers in Germany.

The two sections that follow summarise the interviews with members of the two groups (more detailed summary reports are included in the Annexes).

5.1 Researchers, project data managers, and directors of institutes

This section summarises the 18 interviews with researchers and project data managers (the larger part of the interviews) and directors of research institutes and programmes. The list of interviewees and a more detailed summarization of the interviews (including a large number of referenced statements) is available in Annex IV.

1 – Search for relevant data, main data sources used

***Literature**:* The main source the researchers most often mentioned were publications "for traditional literature review", e.g. scholarly monographs, books, conference proceedings, print and electronic journals, and unpublished reports of excavation work, field surveys etc. Online search plays a major role, though a lot of required literature and other documents are not available online or need to be found via different websites. For getting access to unpublished literature as well as "raw data" personal contacts were perceived as crucial.

***Data**:* There is not one main type of data but very different things researchers are searching for: Records of monuments authorities and services; published results of excavations, field surveys, laboratory analysis; museum deposits and specimens (e.g. "physical anthropology, faunal remains,

lithic raw material"), and specialist databases (e.g. WODAN - Archaeological Wood & Charcoal Database). The data looked for is needed for the planning of own research, as a basis or reference for products (e.g. 3D reconstructions), or for building and integration of databases (e.g. in the Late Iron Age and Roman Ireland – LIARI project).

***Exemplary statements**:* "own archive and that of others"; "unpublished archives of individual practitioners and archaeological companies"; "the personal contact matters very much"; "amphorae people search for online amphorae databases, coins for numismatic ones, etc."; "relevant details of excavations, finds, new radiocarbon dates, etc. (…) available scientific data such as dates and isotope analysis"; "images, texts, etc. for 3D reconstructions at detail".

2 – Importance of digital repositories and online databases

***Level of importance**:* Several researchers perceived digital repositories to be very important (e.g. "invaluable sources"), quite often referring to excavation results but also various other data.

***Level of satisfaction**:* However, the availability was described as unsatisfactory both in general as well as for the researchers' particular fields of the study. Assessments for example included "very few and dispersed" or "no digital data repositories in the field". Interestingly, subject-based repositories relevant for researchers in specific fields of study were mentioned only by archaeobotanists. Databases of research projects were often "only accessible by the participating researchers" (this topic is addressed in more detail in a section below).

***Exemplary statement**:* "There are not many repositories; if there are, they do not contain much information or relevant data; access problems: access is either limited to certain people or only specific data and information can be accessed, especially those with information on unpublished excavations."

3 – Types of data searched for and generated in projects

The interviewees were asked what types of data they are searching for, which ones they or their research groups are producing, as well as about the overall availability, including own data. Some categories of data were suggested in the interview template. The overview below is a highly condensed version of a table in Annex IV which contains more details and statements. Broad categories of data were used, two of them "data for model-based computing, simulation, etc." and "results of data mining" to trial their relevance for the online survey. In the online survey both as well as more detailed categories for the other types of data were included.

	Importance	**Own production (research group or individually)**	**Availability (selected comments)**
Excavation data	Generally "very important" (also "crucial" or "essential")	For a large number of the interviewees (10) the main or an important part of the work (8 did not produce such data)	"in most cases not available"; "there are very few such data published entirely"
Field survey data	Often "very important" (also "crucial" or "essential") or just "important"; though not for archaeobotanists (e.g. "mostly no ecofacts retrieved")	For eight an important part of the work (group or individually), seven did not mention field survey data, and for three, including two botanists such data was not relevant	"only when publishing a survey", "there are very few complete data on the topic"

Laboratory measurements & analysis data	Mostly considered as "very important" or "important"	For most (11) an important part of the research work; others did not produce or use such data (or only sometimes)	"these are usually published in scientific articles", "we have a database which is only accessible to members of the institute"
Data for model-based computing, simulation, etc. (i.e. "in silico" research	Mixed results: Some "important" (or even "essential"), some "not used often" (e.g. "some-times models") or "not important"	Only two mentioned the production of such data (nine did not or only seldom use such data); relevant types of data included LiDAR and raster and vector data from GIS analysis	"still in its infancy"; "essential and produced in-house with our own specialists"
Results of data mining for identify-ing patterns or interesting outliers	Mixed results, "very important", "important" or not relevant and not used	Not often addressed in the interviews; some did not produce or use such data; examples of use were "detailed distributional analysis", "settlement patterns", "regional or landscape projects"	"practically non-existent so far online"; "very important, if such results were easily accessible"

Other research data mentioned

"images, maps and texts", "bibliographical research, historical data, maps, etc.", "artefact studies"; "mostly 3D models from 3D survey and 3D reconstruction"; "geomorphology, geology and vegetation data"; "14C dating and carbon and nitrogen isotope analysis".

4 – Gaps and problems when searching or accessing data

The main areas where researchers faced problems were finding and accessing relevant data, including the quality and usefulness of the data. Often personal contacts rather than formally established mechanisms enabled access to the data. Not often mentioned were issues of metadata, language, costs or legal matter.

Availability and quality of metadata: Only few researchers mentioned lack of or insufficient quality of metadata (e.g. "lack of metadata in datasets").

Not accessible or only in a very restrictive form: Most comments related to the accessibility of data. Data appeared as difficult to find, not available online, and if online difficult to access (e.g. one has to apply). Two selected statements of many: "Biggest issue: to find relevant data – we do not really know what is actually available"; "there are databases, but they are not accessible online or via a formal way".

Importance of personal contacts: Personal contacts are important to gain access to unpublished data; "often access is based on informal matters: you need to have the right contacts, you need to be established".

Lack of data quality and usefulness: Concerning data that are accessible online, several researchers mentioned a lack of usefulness because the data are structured in different ways, not up to date, incomplete or lack important details. For example: "the main problem is the variability and inhomogeneity of data content and structure"; "incomplete datasets; online databases that aren't kept up to date (this is a big problem)"; "lack of details on how data was collected – it is difficult to assess the quality of data published online".

***Language**:* Only a few researchers mentioned that other languages are an issue for accessing and using data and datasets. Some languages were perceived as difficult (e.g. Bulgarian, Hungarian, Russian), but "basic data (tables) are mostly understandable"; "mostly not immediately relevant. If raw data (…) such as measurements and Latin names".

***Costs**:* Were perceived as an issue only by a few researchers and mainly concerning journals.

5 – Storage and management of data in collaborative projects

The typical practice was storage and management on the computer of individual researchers and on a restricted access server of the research institute or unit. None of the interviewees mentioned deposit of data in an open access institutional or external, subject- or domain-based repository.

6 – Responsibility for storing, archiving and maintaining data

Most researchers thought that storing, archiving and maintaining the data of a research project is the responsibility of each individual researcher or the team. But some mentioned one researcher, e.g. "the scientist in charge of the project", "principal surveyor" or "excavation director". A data management plan was mentioned only by one researcher.

7 – Forms and extent of data publication

Data is published in various forms, but mostly in traditional formats such as research reports (grey literature), proceedings, journal articles, book chapters or monographs (i.e. as summary tables, charts, etc.). Two researchers mentioned monographs with data tables and other supplemental material on a CD/DVD or on the project website. The research data is seldom published.

Among the mentioned barriers to making data available online were legacy technology, lack of metadata, and additional effort for turning datasets into web publications. One researcher thought: "The underlying institutional data sets are not being published or made available publicly. They are an asset for the research organisation and are therefore not being shared with external communities."

8 – Percentage of data deposited in (open access) digital repositories

The accessibility of data for researchers not involved in a project is rather limited. Some statements were: "For people not involved in a project it's zero"; "fairly close to zero"; "none is made available at the moment"; "no data from the project has yet been deposited in an openly accessible digital repository"; "basic project data is accessible to researchers on request".

There are projects that make all or a considerable part of their information and data openly available, or aim to do so in the near future. However, often only basic information or data in closed formats is provided. For example: "All our archaeological survey data is published online through maps (location of site and general description)"; "The publication represents about 90% of the data. Formats are pdfs, the user interface allows searching individual graves. No download of tables possible."

9 – Technical or other changed conditions that would ease data search and access

Most of the researchers suggested implementation of open access principles and technical improvements that would allow easier searching and access. For example: "Large open databases, easily accessed, well managed (…) Would also make it easier for me to enter my material or parts of my material there." Among the technical suggestions were: "A coherent method of data publication, accessibility (intelligent interfaces) and transparency of the process data has been created."

10 – Technical or other changed conditions that would ease deposit of data in a digital repository

The responses included availability, setting up a repository, ease of data provision, and specific technical suggestions: "An online database template that is easy to use (and allows you to deposit data in a simple, straightforward way)"; "Would be important to be able to export an own cloud instance with the metadata included into a digital repository (with a metadata conversion tool like MINT)." Furthermore non-technical conditions (mainly funding) and incentives were mentioned (e.g. "incentive has to come to share data, e.g. from funding bodies").

11 – Expected benefits from ARIADNE infrastructure and services

Most responses related to common approaches (e.g. data structures, metadata), open access, ease of search and access, better access to specific data – particularly to leverage comparative research, and fostering of collaboration on the national and international levels.

***Common approaches**:* "a common approach for the presentation of archaeological research data (data structure, …)", "… standardization. If metadata sheets are produced in a similar way, language barriers can be overcome, too"; "a thematic or methodological approach, rather than just combining data in an arbitrary way".

Ease of documentation and metadata production: "simple-to-use tools for documenting and sharing"; "a metadata ingestion and management tool (…) that can be easily installed in our server to organize the workflow during the production phases".

Promote open access to high quality data: "raising awareness on the need to share high quality data"; "opening existing data to free access"; "could be a big step to remove barriers to open access (…) make it open and international".

***Ease of search and access**:* "facilitate a better knowledge of existing datasets, improved access conditions"; "user-friendly queries and interfaces"; "cross-searching data repositories".

***Enable comparative research**:* Quite often emphasised, e.g. "make results, sites and evidence comparable"; "more potential for quantitative comparisons"; "that comparative data sets be made available for furthering and developing advanced research in archaeology".

Fostering collaboration: "the research between different institutes (with different specialisations and their own databases) would benefit greatly"; "access to a wider geographical datasets will in time help facilitate cross collaboration and enhance funding opportunities".

Only one interviewee mentioned specific technology: "ARIADNE should have a strong GIS support".

5.2 Managers of data centres

This section summarises three interviews with managers of data centres in Europe, one interview each with the Archaeology Data Service (ADS, United Kingdom), the Data Archiving and Networked Services (DANS, Netherlands), and the Swedish National Data Service (SND, Sweden). A more detailed summarization of the interviews (including a large number of referenced statements) is included as Annex IV. The data centres have similar missions which include support for research and learning by offering deposit and curation of, and sustained access to, digital resources, training and advice in related matters, and collaboration in national and international projects. The main user groups are researchers and research students of universities, other research institutes and governmental agencies, as well as commercial users (i.e. contract archaeologists). As the data centres provide services to the research community, in the summary below we sometimes relate results of the interviews with the data mangers to insights gained from the ones with researchers.

1 – Main critical issues and challenges in fulfilling the centre's mission

***Relationship with depositors and users*:** As clearly evidenced in the interviews with researchers, depositing data in an external repository is an uncommon practice for most archaeologists across Europe. Data is mainly presented as summary tables, charts, visualizations, etc. in publications, while the re-useable data is not shared. This practice will only change if researchers are obliged (e.g. according to the terms of research grants) to deposit their data in a community archive. This has been the case with ADS in the UK and DANS in the Netherlands (in Sweden publicly funded researchers are not yet obliged to deposit data at SND).

Open access: The data repositories are generally in favour of open access, but have to take account of various conditions that may not allow this (e.g. DANS: "open if possible, protected if necessary").

Make clear the added value: Data repositories are but one element in the research environment of archaeologists (and, at present, missing in many countries). They must make clear their specific role and added value (in comparison to libraries, for instance), and become "embedded in research practice".

***Metadata**:* While researchers generally worry little about metadata, for data repositories good metadata is one of the most critical issues; e.g. "Quality of the metadata – we put a lot of effort into this, because data without metadata is meaningless. Also try to explain to researchers why it is important that they should invest the effort to provide good metadata together with the data they deposit in DANS." "Managing depositors is the greatest challenge – getting them to provide metadata for comprehensive datasets."

Service enhancement: Depositors and users value services (ideally, personalised), but they must be effective and affordable (e.g. automating services as much as possible).

Trustworthiness and reliability: The data centres aim to prove that depositors can fully trust their preservation and curation work (e.g. Data Seal of Approval; ISO 16363:2012 Audit and Certification of Trustworthy Digital Repositories).

2 – Most demanding technical or other issues

The data managers did not perceive technical issues as a major challenge, also not "big data". Some specifics of archaeological data and user needs were noted (e.g. importance of GIS data, mapping services, etc.). For SND also the rapid growth of the organisation because of the extension of its mandate was mentioned.

Generally standardised and secure processes for handling registered data were seen as most important. One recent development mentioned was Linked Data, i.e. establishing links between related content ("you have to think in new ways how to organise your archive").

3 – Trends in user needs

The managers' views of current trends were fairly similar:

- Main user needs are an easy way to store data, to search and retrieve data at one place, and to refer to it consistently;
- The amount of data that is deposited is growing (e.g. GIS data as a growth area);
- There is an increasing expectation of open access by research funders, research communities, and the interested public;
- The researchers increasingly need to underpin publications with data, which means that the ties between repositories and publishers are strengthened;
- Directly related to the above, more requests for Digital Object Identifiers (DOIs), Persistent IDentifiers (PIDs) for linking publications and data;
- Users are looking for advanced search capability, for example, searching and browsing data across repositories, e.g. data for the same subject contained in project archives and data-sets;
- More open and more linked data (i.e. between publications and datasets, and among sets of data) will impact on the life cycle of data, e.g. re-use of data in new, integrated datasets.

4 – Measures implemented in response to emerging needs

Major measures taken or foreseen concern the relationship with institutional and commercial user groups, user involvement, and enhancement of internal capacity and services:

***Relationships with universities**:* This relationship was seen as particularly important, not least because universities also often establish own repositories. Offers include "outsourcing" the long-term preservation and access to the data centre and having an institutional view on all archived archaeological data of university researchers and projects (ADS, DANS).

***Integration of current research information**:* Such information about research projects, researchers involved, etc. is becoming increasingly important, for example, in the context of research funding & reporting, institutional and individual research profiles (mentioned by DANS).

Enhancement of tools to ease deposit and decrease costs: Making it easier and cheaper to deposit small to medium sized research archives (e.g. output of commercial archaeology companies) – ADS; also DANS aims to improve functionality and cost control.

Advanced services: Allowing users to select, extract and download data items from data sets, and possibly also enabling the running of calculations on datasets online (considered as future services by SND).

***User involvement**:* A user panel that provides feedback on specific issues has been established by DANS; also a review system is being developed so that users can rate data quality.

5 – Relevance of ARIADNE for own service development

The data managers emphasised the sharing of expertise in ICT for archaeology, enhancement of collections and services, and benefits of common standards and interoperability:

***Shared expertise**:* Access to expertise in information sciences (ARIADNE technical partners), and application of novel technologies to the specific needs of archaeologists and repositories.

***Enhanced data collections and services**:* Enhancing collections and developing new services based on integrating XML schemas, controlled vocabularies, etc.; also new tools for data curation and providing access to special types of data were mentioned.

***Common standards and interoperability**:* For the users, an international infrastructure of data repositories based on common standards and services would make it easier to discover and access

relevant material across national borders. This would also allow easier promotion of the usefulness of high-quality metadata by the repositories.

***Leverage of visibility**:* Higher recognition of the work of both the national data repositories and the researchers who make data available.

6 – Most important ARIADNE service for the data centre

Asked about the single, most important ARIADNE service for the enhancement of own services the managers thought of

- Improved integration of data mining (ADS);
- Use of common vocabularies (thesauri) and open data approaches (DANS);
- Data discovery services to find data wherever it is, irrespective of different languages (SND).

7 – Other important e-infrastructure and services projects

The managers considered the European projects DARIAH, CESSDA, Europeana and E-Cloud (Europeana Cloud) as important to collaborate with; but also to engage with international e-infrastructures outside Europe. This was seen as important concerning common standards and tools as well as for making available data visible.

6 The ARIADNE Online User Survey

6.1 Methodology

The ARIADNE Online User Survey was conducted as an open inquiry. Anybody with the URL of the website with the questionnaire was able to participate. The survey was implemented by researchers from Salzburg Research GmbH on the online survey platform SurveyGizmo (www.surveygizmo.com).

Survey population

The survey population was defined as the (international) archaeological research community, including both researchers and managers of archaeological data repositories. They were grouped into four categories, in line with the ARIADNE users framework developed by the project:

(1) Archaeological researchers (academic and independent);

(2) Directors of archaeological research institutes (or heads of major departments);

(3) Managers or directors of institutional data repositories;

(4) Managers of international, subject- or domain-based repositories.

Questionnaire development

These target groups were addressed with two different questionnaires (i.e. two surveys were conducted in parallel):

- a questionnaire for target groups (1) and (2) (with minor differences in the wording of how the respondent was addressed in the question, but no difference in the substance or structure) – results are presented in "Part I – Archaeological researchers";
- a questionnaire for target groups (3) and (4) without any difference – results are presented in "Part II – Managers of data repositories".

An important source for the development of the questionnaires, besides literature, were the results of the **pilot interviews** conducted prior to the survey with archaeologists and repository managers from among the consortium (see Section 5). The information obtained from these interviews helped the research team, in particular, to develop and expand the list of potential response items in standardised questions (such as challenges experienced with regard to data search, data sources and types of data used or produced, expectations towards ARIADNE). Several items which turned out to be relevant were identified in this way. The survey team developed a draft version of the questionnaire on this basis and sent it out for consultation within the consortium. Several partners suggested modifications, further questions and additional response items. The consolidated questionnaire, reflecting the proposed modifications, was then programmed for the online survey. The structure of the questionnaires is described in the introduction to the results sections (see Sections 6.2 and 6.3).

Survey roll-out and field work

The roll-out of the survey to the dispersed population was carried out with support of the ARIADNE consortium, predominantly by leveraging the networks and communities of the various research partners, including the members of major associations and federations(see Annex I for details). In addition, Salzburg Research informed the members of various archaeological mailing lists. The first estimate is that the various invitations to participate in the survey reached several thousand representatives of the target population and thus potential participants.

The survey was launched in **November 2013**. Invitations were sent out over a period of about two weeks. The deadline for answering the questionnaire (as communicated in the invitation e-mails) was

December 31st, 2013.[4] The results in this document represent the status of answers received to that date.

Response rate

The **initial (gross) sample** comprises all respondents who opened the survey website and answered question A.1 (on their professional status) with one of the four options (a)–(d), i.e. their profession complied with the definition of the survey population. These were about 880 people. In total, about 1200 users have opened the survey starting page in their web-browser by clicking on the link contained in the invitation e-mails that were sent out. Of those, about 1000 answered (at least) the first question A.1 about their professional background, 200 users left the start page without having answered any question.

The (cleaned) **net sample** consists of all respondents from the gross sample who fulfil the following conditions:

- They have answered at least two of the survey questions apart from questions about their professional background in Module A (i.e. questions of Modules B-E).
- The pattern and content of their answers suggest that they have deliberately answered the questions ("cleaned net sample").[5]

The final, cleaned net sample comprises 692 questionnaires (see Table). Most of the questionnaires from the gross sample that have been excluded as part of the data cleaning process were excluded because they did not contain a sufficient amount of answers.

Table 6.1-1: Gross and net sample of the ARIADNE Online User Survey (based on answer to Question A.1)

A.1	Professional activity (category)	Gross sample	Action	Net sample
(a)	Archaeological researchers	725	→ 790 passed on to Questionnaire 1	586
(b)	Director of a research institute	65		54
(c)	Institutional repository managers	78	→ 87 passed on to Questionnaire 2	47
(d)	International & domain-based repository managers	13		5
	TOTAL GROSS SAMPLE	**881**	**TOTAL NET SAMPLE**	**692**
(e)	Other	120	→ survey terminated because not member of target population	n.a.
	No answer on A.1 (user quit the survey before answering the first question)	204		n.a.

Partially completed questionnaires

It is common in online surveys that not all respondents fill in all questions, but quit during the survey. Answers from partly completed questionnaires have been used, provided they meet the conditions outlined above. In total, about 53% of the questionnaires received were partial (i.e. the respondents terminated the survey before having reached the last question) and 47% of the questionnaires were complete (i.e. all questions have at least been clicked on and looked at by the respondents). "Complete" does not mean, however, that every single question has been actually answered –

[4] The questionnaire is still open. Further responses coming in after the deadline for the first wave will be used for the update on survey results (to be reported in Deliverable 2.2).

[5] The data set is cleaned from obviously meaningless responses (for instance respondents who have consistently ticked the first answer option only). These are relatively few questionnaires, however.

interviewees were given the option to move on to the next screen without having answered all questions on the previous page.

Difficulties experienced

The field work ran smoothly overall, except for a few respondents (about 5) from the A.1 categories (a) and (b) who reported technical difficulties in answering question D.2 (they could not proceed to question D.3 after having answered question D.2). The survey team then exchanged this question in the online questionnaire, and no further difficulties were reported hence.

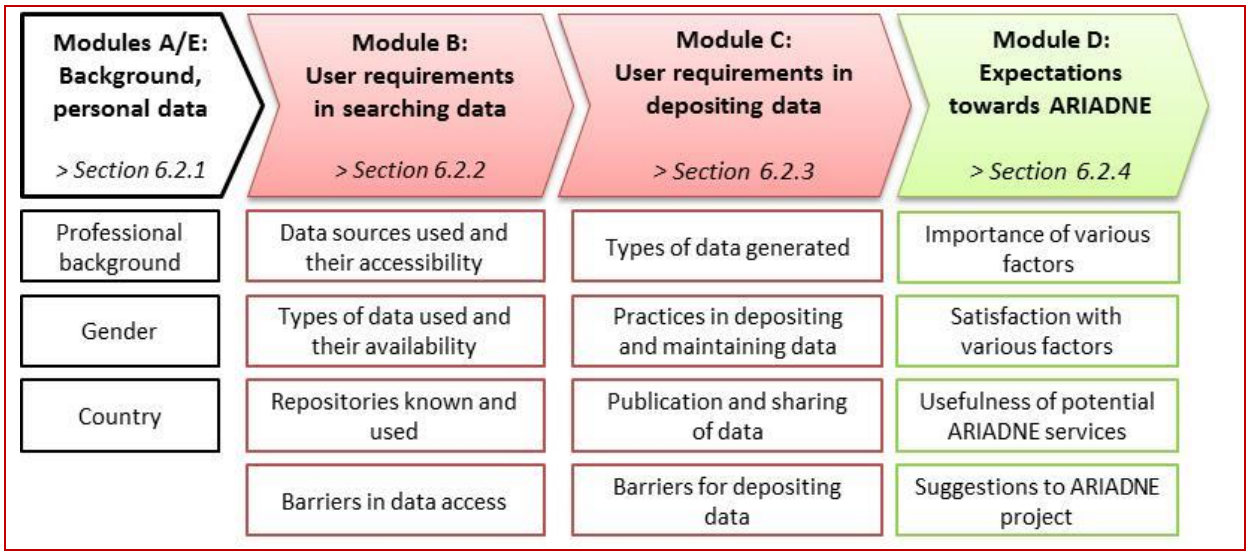
Some respondents made use of the free comments boxes for making specific comments about the survey; their comments and suggestions are included in the report.

6.2 Results - Part I: Archaeological researchers and directors of research institutes

Structure of the questionnaire

This part of the survey addressed archaeological researchers and directors of research institutes. The survey structure reflects a simplified work flow of any typical research project in which researchers are both "users" of (existing) data and "producers" of (new) data (see Section 3.2). After a few questions about the professional background of the respondent (to have information about the sample structure and analyse the impact of various factors), the first module of the survey addressed current research practices and challenges in searching data, i.e. the "data user" role of the researcher. The second module then explored their practices, requirements and needs in depositing the data which they are producing with their research activities. Finally, respondents were asked about their expectations towards the ARIADNE project (see Figure 6.2-1).

Figure 6.2-1: Structure of the ARIADNE User Survey of researchers

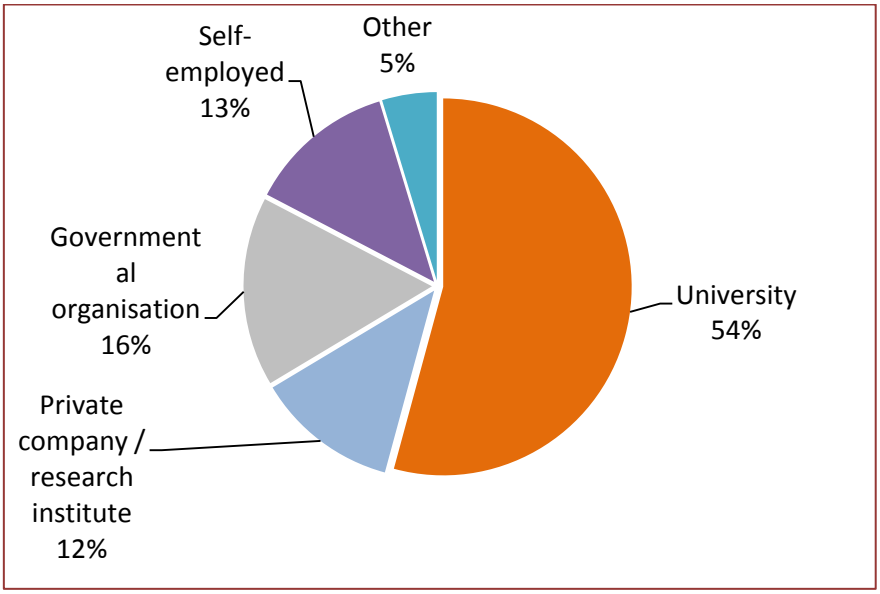


6.2.1 Sample structure

Professional affiliation and responsibility

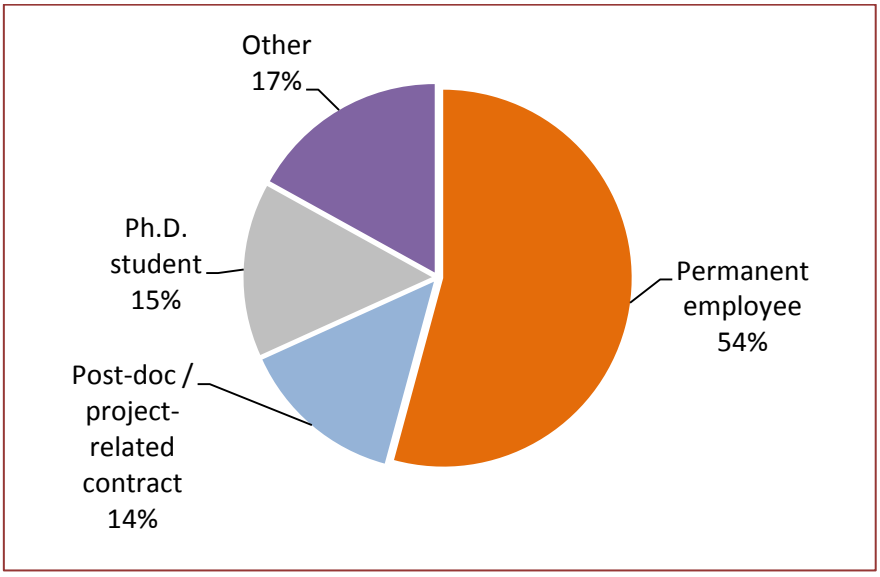
The respondents were first asked which type of research organisation they were working for. More than half of them are university researchers, 12% are researchers at private companies or institutes, 16% work for governmental organisations and 13% are self-employed researchers (see Figure 6.2-2). Although the sample structure is to a large part determined by how effectively the various communities were reached by invitations to participate in the survey (see Annex I for details), and without having any statistical figures about the actual spread across the various types of organisations, the figures suggest a fairly representative distribution of the sample. A lot of archaeological researchers are actually affiliated with universities.

Figure 6.2-2: Question A.2 – "What type of research organisation are you working for?"



N = 640 (excl. no answer)

Figure 6.2-3: Question A.3 – "What describes best your current position in your research institute?"

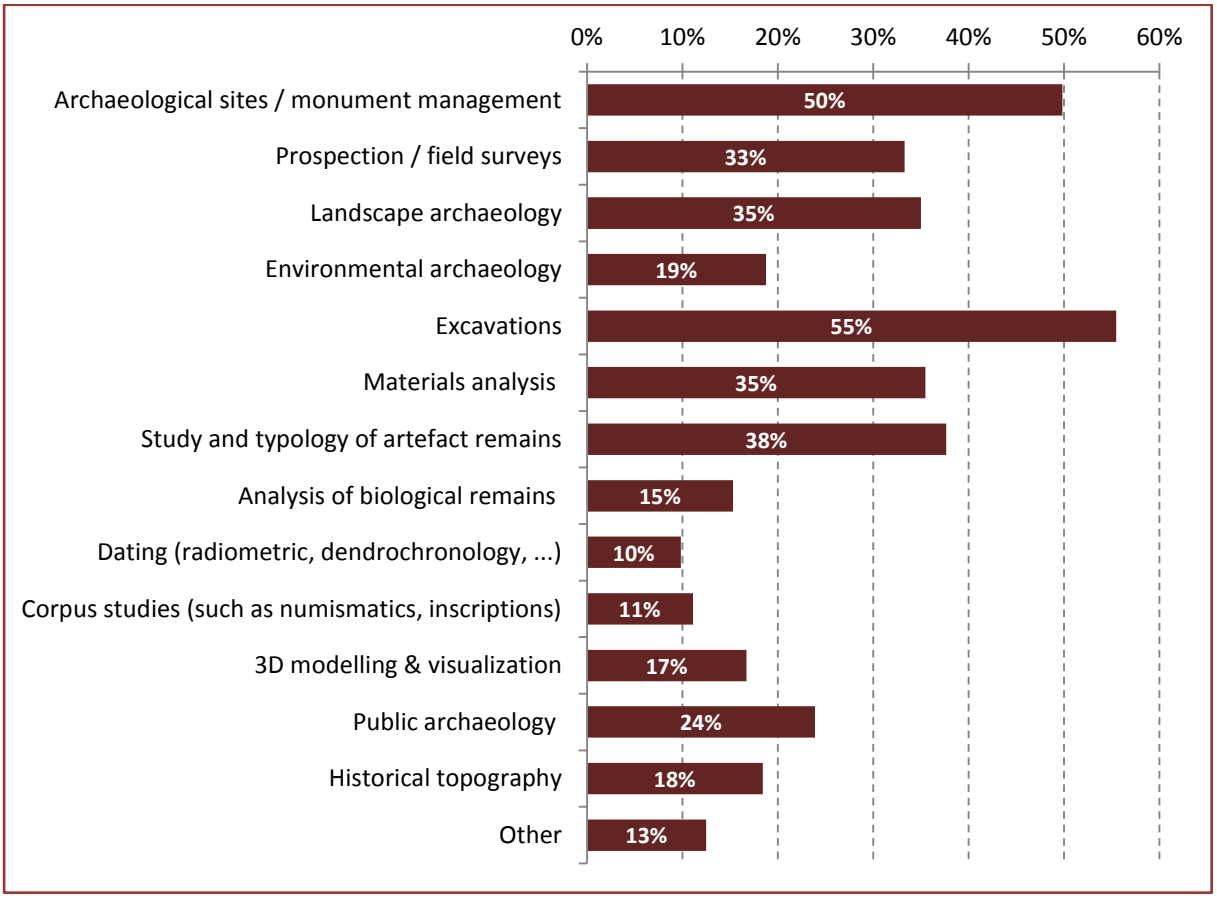


N = 640 (excl. no answer)

Respondents were then asked to describe their current position in their institute (A.3) and their main research focus (A.4). Again, the responses indicate that the sample could be fairly representative for the target population. More than half of the respondents said they were permanent employees, about 15% held temporary work contracts (post-doc or project-related) and 15% of the respondents were Ph.D. students. This leaves a remaining 17% with other positions that are not adequately described by these categories (see Figure 6.2-4).

A majority of the researchers interviewed said that excavation and/or the management of archaeological sites and monuments were among their research priorities (see question A.4). Other major research activities quoted by at least a third of the respondents as a focus of their work were field surveys, landscape archaeology, material analysis and the study and typology of artefact remains. Other activities represent more specialist research methods or disciplines, such as the analysis of radiocarbon, dendrochronology or other dating data (10%), corpus studies (11%) or the analysis of biological remains (15%). Admittedly, the list of these research "activities" does not represent any established canon, but includes a mix of methods and disciplines; moreover, some categories are not sharply differentiated from each other. However, the answers provide a good-enough picture of the research work in which the survey respondents (and, thus, probably the ARIDNE community of practitioners) is actively involved. The low response rates for some of the activities may have an implication when taking a decision on which type of data the ARIADNE services should focus on, although data should not be excluded only because they are relevant only for a minority of the practitioners in the field.

Figure 6.2-4: Question A.4 – "Please describe briefly your research focus." (multiple answers)



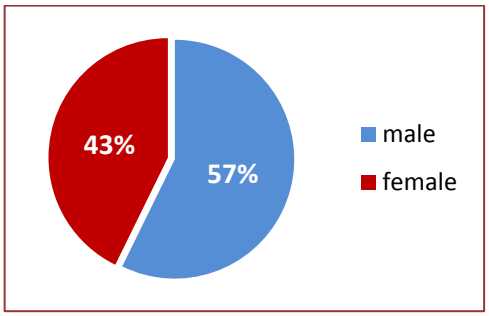
N = 640 (excl. no answer)

Gender structure

With a share of 43% of all respondents, female researchers are well represented in the sample. This figure corresponds almost perfectly with the results of an empirical study conducted by the "Discovering the Archaeologists of Europe" project in 2007/2008. This survey found that the distribution of female and male archaeologists was 46% and 54%, respectively.[6] The result is based on a total of 9,109 archaeologists from twelve European countries. The sample structure of the ARIADNE project also reflect that the share of women in social sciences and humanities is generally high as compared to natural sciences – in some disciplines even above 50% (specific figures for archaeology were not available). According to German statistics, for example, the share of female researchers in social sciences (here including archaeology) is about 55%.[7]

Figure 6.2-5: Gender distribution of the survey sample

	E.2	#	%
a	male	276	57%
b	female	206	43%



N = 482 (excl. no response)

Geographic distribution

More than 80% of the respondents work in one of the EU member states. Most of the other respondents were researchers from non-European countries (12%), in particular from the United States (42 out of 60). Among the EU member states, five countries are particularly well represented in the survey: the UK (79 respondents), France (51), Italy (47), Slovenia (38) and the Netherlands (35). This reflects, on the one hand, the larger proportion of researchers in the large countries (in the case of the UK, France and Italy). On the other hand, however, the sample is to some extent biased in this regard by the effectiveness of the various promotion activities as undertaken by the consortium members. The high share of researchers from Slovenia is clearly an "artefact" demonstrating that the outreach to this community was extremely successful. Similarly, the good response rates from the Netherlands, but also from Austria, Greece, Hungary and Ireland were facilitated by having research partners from these countries in the consortium who helped to promote the survey.

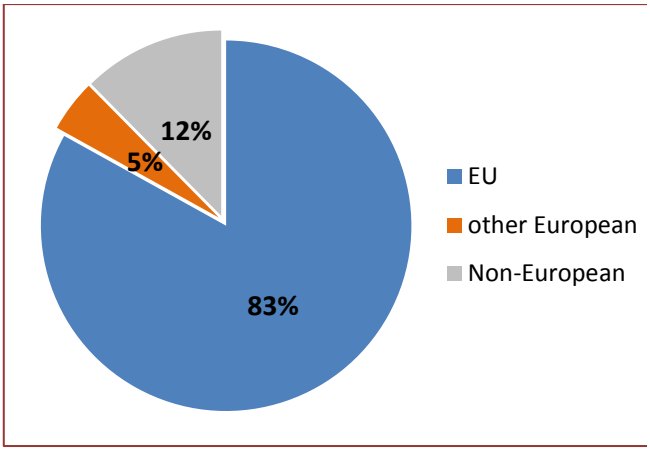
A weakness in the geographic distribution is the lack of German researchers among the sample (only 15). Given the response rates from the UK and France, one would have expected a higher figure of responses from Germany. Quite in contrast to the Slovenian example, the communication measures to promote the survey in Germany have not been as effective. The survey team will make a special effort to obtain responses from Germany during 2014 and present these in the updated results in Deliverable 2.2.

[6] Aitchison, Kenneth (2009): Discovering the Archaeologists of Europe: Transnational Report. Institute for Archaeologists, January 2009, http://www.discovering-archaeologists.eu/DISCO_Transnational_Report.pdf.

[7] See "Berufe im Spiegel der Statistik", IAB, 2010, quoted in: "Liste von Frauenanteilen in der Berufswelt", Wikipedia: http://de.wikipedia.org/wiki/Liste_von_Frauenanteilen_in_der_Berufswelt

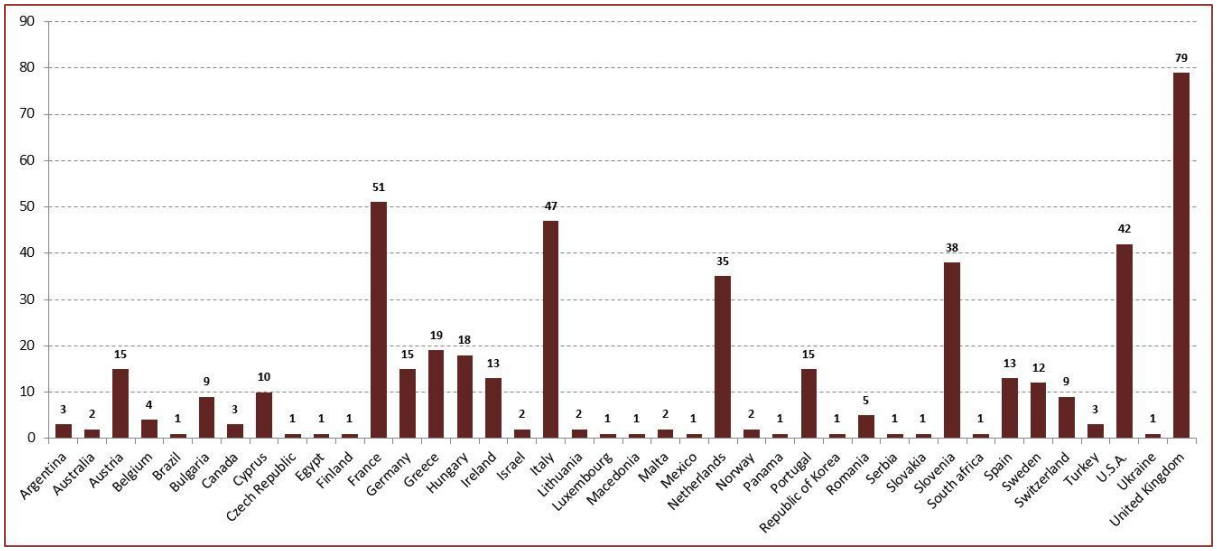
Figure 6.2-6a: Geographic distribution of the survey sample

	E.1	#	%
a	EU	400	83%
b	Other European	22	5%
c	Non-European	60	12%



N = 484 (excl. no response)

Detailed distribution *



N = 482

Although the geographic distribution of the responses is not fully representative in that sense, there is no evidence that this could severely undermine the validity of the results as a whole. The interviews with individual researchers (see Section 5) and the literature reviewed have not indicated that this would be a major factor. We will look into this aspect in more detail, however, in Deliverable 2.2 (update on user requirements) and analyse in detail the differences with regard to requirements for research data in accordance to the geographic base of a researcher.

6.2.2 Importance of different sources and types of data

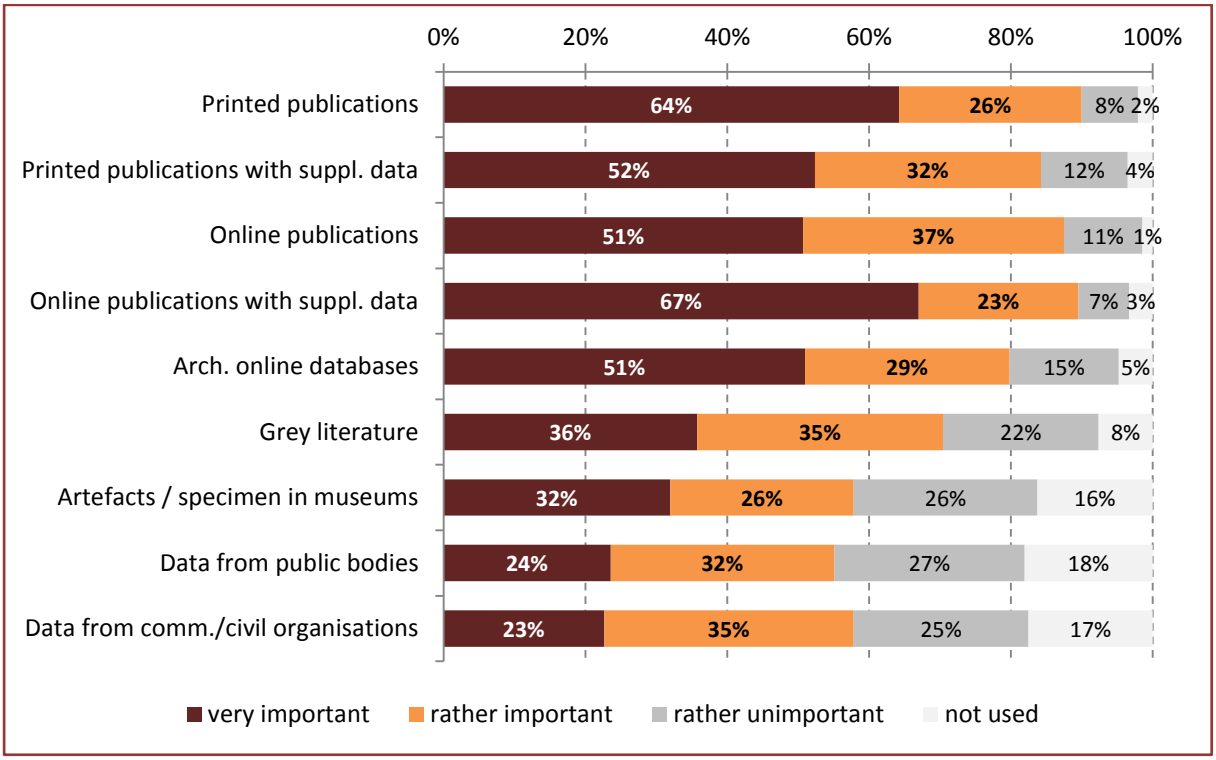
The first module of the survey addressed current research practices and challenges in searching data, i.e. the role of the researcher as a "data user".

Importance of data sources

Researchers were first asked about the importance of different sources for collecting data during the research process. The goal of this question was to identify the relative importance of digital sources (such as online repositories) compared to traditional sources such as printed publications. They were presented with a list of nine types of sources and asked to rate the importance. The main finding, quite expectedly, is that *there is no single most important source* – researchers in the digital era need to be flexible and make use of all kinds of data sources, depending on where the required data are available. It is not the source as such that matters – it is the quality of the data contained. This was also confirmed by the comments received from respondents (e.g. "*All sources are important and must be first-hand*").

This can be demonstrated by a direct comparison of "printed publications" vs. "online publications": both are highly relevant – about 90% of the respondents said (for both sources) that they are "very important" or "rather important" for collecting data (see Figure 6.2-7). Similarly, printed and online publications with supplemental data were rated as very or rather important by a vast majority. About 80% also said that specialised archaeological databases are important sources. Each of the proposed data sources mattered for at least 50% of the respondents. There is another (important) source which is not mentioned in the list but which was proposed in the comments: museum image databases.

Figure 6.2-7: Question B.1 – "When working on research projects and searching for data: how important are the following sources for you and your research group for collecting data?"



N = 543-579 (depending on number of respondents without answer)

		++	+	-	--	
	B.1	very important	rather important	rather unimport.	not used	N
a	Printed publications	372	149	46	12	579
b	Printed publications with supplemental data	299	182	70	20	571
c	Online publications	288	209	63	8	568
d	Online publications with suppl. data	384	129	41	19	573
e	Arch. online databases	288	163	87	27	565
f	Grey literature	202	196	124	43	565
g	Artefacts & specimen in museums	179	145	145	91	560
h	Data from public bodies	129	173	147	99	548
i	Data from comm./civil organisations	123	191	134	95	543

Admittedly, it was a challenge to pose this question in a way that would accurately deliver the intended information, since the term "data" is very broad and it may have different connotations for people. The risk is that some respondents misunderstood the question in the way that they responded about the *general importance* of e.g. printed publications for their research (which is, obviously, high), rather than about the *specific importance for collecting data*. The notion of data that was underlying this question (also with regard to the ARIADNE objectives) was one of structured, comprehensive datasets (tables, images) rather than an occasional image or table in a book. However, this also depends on the use case. For some types of research, large structured data sets available in some specialised repositories are pivotal; for other research projects, such large data sets may not exist, and researchers depend on whatever they can extract from conventional literature and project reports.

The respondents made some interesting comments and suggestions which helped to better assess the results. One respondent claimed that grey literature could be a more important source for extracting data if the availability could be improved. Another respondent made a similar comment, arguing that it is not the type of source ("all are important"), but the availability that matters. These remarks present a clear use case and opportunity for ARIADNE – it seems that there is a clear demand for an improved availability of (international) archaeological data sources.

Respondents also confirmed in their comments the need for good online databases, as they greatly facilitate access to data ("*If I can't find it online it's much harder to include or refer to in my work*"; "*Online databases that do not exist are not important to me, but they would be if they did*."). This indicates a general dilemma with this question: the current importance vs. the potential of a certain data source. Something which barely exists is not important in current practices, but could become very important if it was created, and in fact is missed by researchers. This is encouraging feed-back from the community for the ARIADNE objectives.

Comments from respondents:

- *"All sources are important and must be first-hand; then you have to evaluate them critically."*
- *"Given the lack of data repositories, their value is currently underrated. "Gray lit" would play a more important role if it were more readily available."*
- *"Use of 'grey literature' useful (and acknowledged in text and bibliography) but as my focus is mainly county-wide, I have access to such material in Herts Archives and Local History Archives. Were my focus wider, then 'grey literature' from elsewhere could be vital, assuming I observe the same rules as listed in the brackets above."*
- *"Curiously "grey lit" for the province is now mostly available as digital text through "data provided by government administration" and is now more searchable (and much more comprehensive) than most published data. In the local case, the 'grey' lit has also been peer-*

reviewed through the permitting process, as the regulator's reviews are often more comprehensive than normal peer review."

- *"I would use all the 4-rated resources if they were accessible (they aren't, effectively)."*
- *"If I can't find it online it's much harder to include or refer to in my work. Online coverage of archaeology journals is patchy at my institution, and many published reports or monographs are not available online. For "local" (UK) excavations I can access published, printed monographs etc., though sometimes with difficulty, but I suspect that I miss a lot of useful European information. Unpublished research reports and grey literature are also problematic, I would make more use of them if they were more accessible, but it's sometimes important to know that a search for information has been comprehensive at a certain level (e.g. published papers), not just "everything I could find" which might include some grey literature I could get at easily, but not other equally valid reports I didn't know about or couldn't get."*
- *"Online databases that do not exist are not important to me, but they would be if they did. Therefore, I wasn't sure how to answer the questions about databases."*
- *"I seem to be spending a lot of time with the archives of archaeologists who are elderly or deceased to be sure their data is represented in my publications and to find a way to scan it and make it accessible to other scholars. In some cases the heirs have not known what to do with the records and in other cases the sponsoring university is overwhelmed by such collections."*
- *"I largely work on projects that aggregate content for other researchers and the results here reflect the kinds of data we aggregate."*
- *"Archaeological online databases are very important, but they often prove to be very out of date, not very comprehensive, or are difficult to search for specialist data (even at the simple level of searching for report titles). For remote sensing data, LiDAR is becoming more important."*
- *"Not sure where museum picture databases fall in here - have thus added extra field"*
- *"TNA, County Record Offices and NMR also invaluable"*
- *"It is impossible to rank the different types of resource, all contribute. The only reason for scoring 'artefacts & biological specimens' more lowly is that colleagues tend to deal with them."*
- *"It's hard to evaluate how important an archaeological online database would be for my research, since none exist as yet, for the time and place on which my research focuses."*
- *"Since my research focuses on Neolithic sites in Ireland, the UK and France, access to online data is crucial, as I live in Los Angeles. Unfortunately, a good deal of the information I would like to use is only accessible for a fee, and since I have no budget for such, I've had to rely solely on data and images from free sources or on that provided to me by organizations such as The Discovery Programme, RIA etc."*
- *"I have only recently set up as a sole trader and my uses here are as anticipated based on former job in local authority archaeology service."*
- *"Online databases are in my case mainly epigraphic and iconographic."*
- *"Online databases would be more important if there were better and more comprehensive data available for a more comprehensive range of sites. Unpublished research reports would be more important if there were a better way to find them."*
- *"All is important ...if it is available"*

- *"The importance of sources comes because most of the information available is in publications available online; it is the most common source for me. There is almost no or not much information in my area in Archaeological online databases or repositories or "grey literature", although I think that could be critical."*
- *"Information (site description and interpretation), accompanying data tables and methodological papers are very important whether published in printed format or online. We do not have subscriptions to many online publications so it can be no easier to get to these than printed texts."*
- *"Database and repositories usually don't have the materials' information I need"*
- *"being able to correspond directly with authors of reports in which data are described is also important, e.g., to help clarify details of identifications if illustrations, measurements etc. are not provided"*
- *"It is very difficult to answer some of these questions, because the online documentation is really poor and in some cases non-existent. The answers therefore outline a desired situation, rather than a real situation."*
- *"Unpublished research reports and "grey literature" would be of great help, but in Italy it is almost impossible to access them."*
- *"Most important are publications available online, also data about prehistoric environment and agriculture"*
- *"it depends from the quality of the data source"*

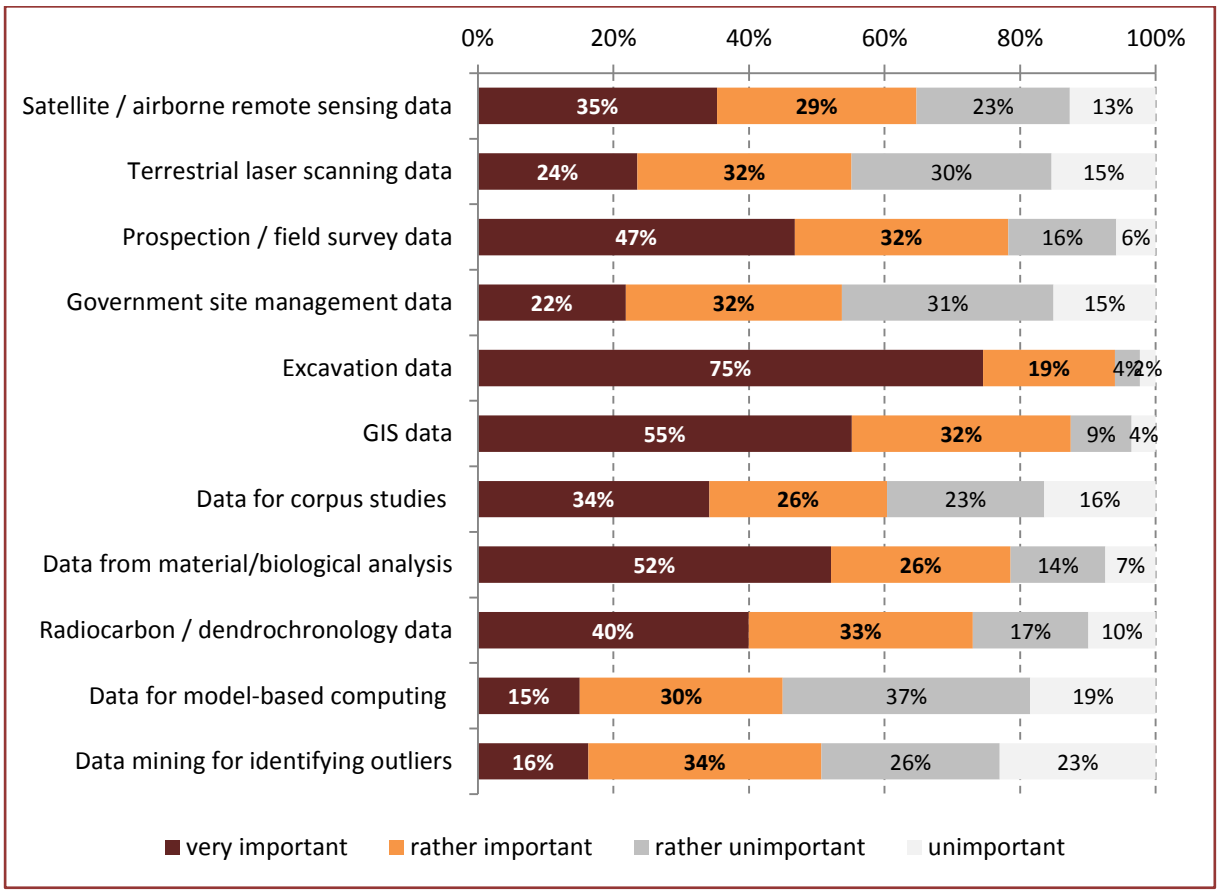
Importance of different types of data

The respondents were then asked to rate the importance of different types of data. With a view to the development of potential ARIADNE services, it was hoped that the responses could provide some guidance e.g. when it comes to setting priorities and focusing on specific data to be integrated. The challenge in posing this question was to come up with a "classification" of different categories of data which archaeological researchers use and produce in their work. The list finally submitted to the respondents for evaluation was drawn up in consultation with the research partners among the consortium.

The single most important type of data if measured by the number of researchers for whom they are important is excavation data. Seventy-five percent of the respondents said that excavation data was "very important" for them to carry out their research projects. Also very important for a large group of researchers (about 50% each) are GIS data, data stemming from material or biological analysis, and data from field surveys. This is not to say that the other types of data are not relevant; quite the contrary, they all have their users; in most cases, at least 50% of the respondents said that they were at least "rather important" (see Figure 6.2-8).

Again, the comments made by individual respondents are very useful as they put results into perspective. For instance, similarly to data sources, it was remarked that specific types of data were seen to have significant potential for future research (and could, therefore, become very important); however, they are difficult to collect and obtain at present, at least for some researchers ("*The items I placed as unimportant are actually important, we just don't have the means for them, and as such are not able to use in our research*.") This again indicates that there is an access problem.

Figure 6.2-8: Question B.2 – "How important are the following types of data for you and your research group in preparing and carrying out your projects?"



N = 521-592 per item (depending on number of respondents without answer)

		++	+	-	--	
	B.2	very important	rather important	rather unimport.	un-important	N
a	Satellite & airborne remote sensing data	209	174	134	75	592
b	Terrestrial laser scanning data	138	185	173	90	586
c	Prospection & field survey data	273	184	93	34	584
d	Government site management data	126	184	180	87	577
e	Excavation data	445	116	22	14	597
f	GIS data	326	191	53	21	591
g	Data for corpus studies	199	153	135	96	583
h	Data from material & biological analysis	309	157	83	44	593
i	Radiocarbon & dendrochronology data	237	196	101	59	593
j	Data for model-based computing	87	173	211	107	578
k	Data mining for identifying outliers	85	179	137	120	521

Comments from respondents:

- *"Information Technology provides reliable and real data as far as plans of buildings are concerned; but then you must understand their meaning and process correctly the data, otherwise they are useless."*

- *"Datamining is currently undervalued due to a lack of repositories."*
- *"One single source of data (e.g., airborne LiDAR) can fulfil many of your categories ....."*
- *"For those marked (3), it is more that we have limited opportunity to explore these rather than believing the data to be unimportant."*
- *"Isotope, residue and DNA analyses are increasingly seen as important but their use is still limited in commercial archaeology"*
- *"there is not a great deal of data mining useful to my preferred area of research - industrial heritage and archaeology"*
- *"The items I placed as unimportant are actually important, we just don't have the means for them, and as such are not able to use in our research"*
- *"All the data sets, marked here as rather important or unimportant (terrestrial laser scanning; C14 and so forth; simulations and data mining) appear to be rarely accessible (or totally lacking), which affects the options to include such info in the project preparation stage."*
- *"All of the following questions are answered based on one project that I am currently involved in. Apart from that I have no other experience with such research projects."*
- *"I wish I could use radiocarbon more often!"*
- *"Not much available in the way of data mining - it may be important in the future"*

Online availability of different types of data

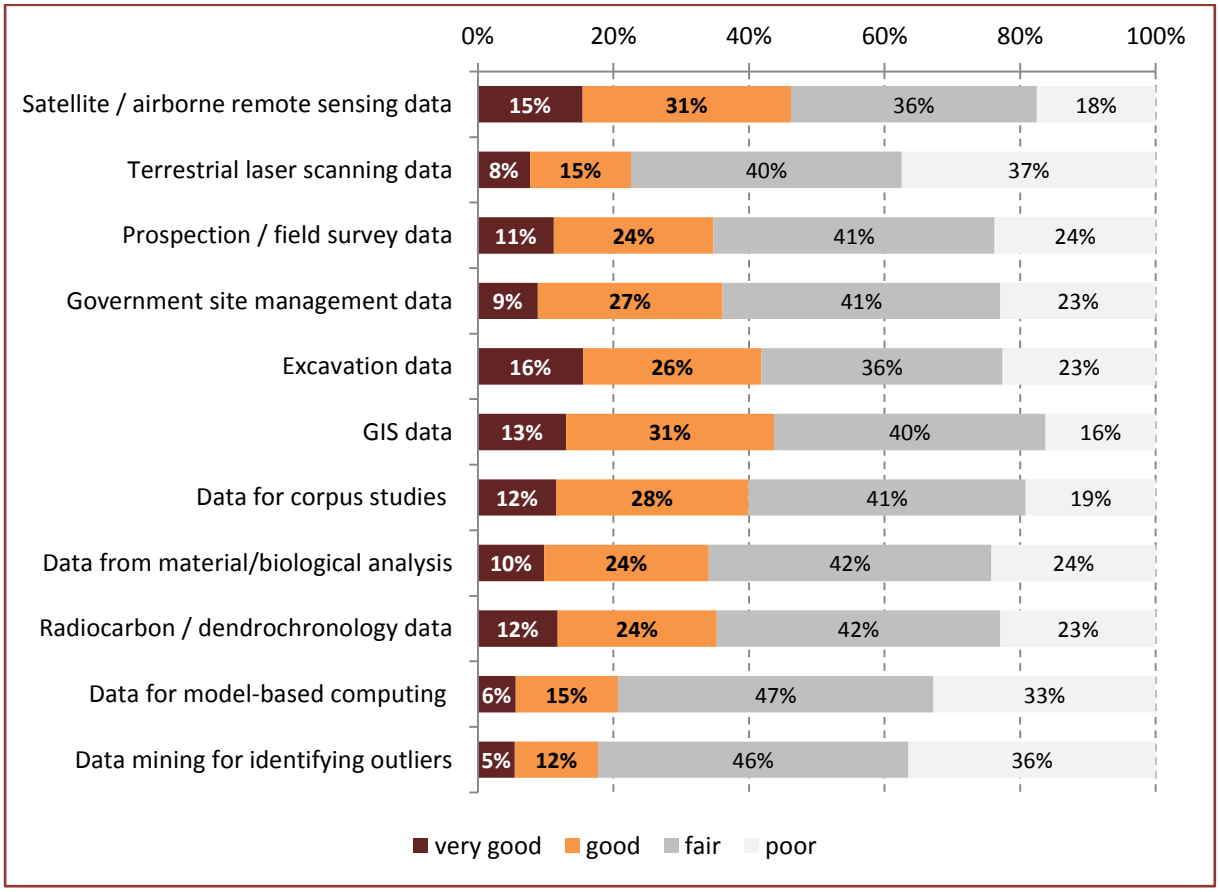
As a follow-up question to the importance, respondents were asked to rate the online accessibility of these types of data. Only few respondents feel that the online availability of research data is fully satisfactory. For any type of data, only a minority of researchers (typically 5-10%) feel that the accessibility is "very good", and less than 50% said that access was at least "good" (see Figure 6.2-9). The online availability of data appears to be most advanced for satellite and airborne remote sensing data, excavations data and GIS data. As in other research domains in the humanities, there are several challenges to making research data available online in a comprehensive way, practically all along the work flow. One challenge is the huge effort for the digitisation of material which is not yet available in a digital format (such as printed documents, physical objects, non-digital images). This is something which cannot be facilitated by the ARIADNE project. However, the ARIADNE project could probably provide some support related to other segments of the "digitisation work flow".

Interestingly, this question produced a lot of comments from respondents, which are again very helpful to put the answers into perspective. Clearly, as some respondents commented, they could not assess the availability of resources which they actually do not need or use in their research. This was a shortcoming in the way the question was posed and may have led to results being lower than they would be if only the actual users of specific data would have been asked.

An interesting comment with relevance for ARIADNE services was the appreciation of making available online digital scans of ancient texts and plans (as is made by the ARACHNE database). As stated above, the project cannot deal with the scanning of documents; but it can possibly improve the availability of existing scanned documents to a wider group of researchers.

As a side notice, a learning point from the comments is that it is not only the technology that matters in digitisation, but also the "*human factor*". One researcher admitted frankly that he/she was just not experienced in online research and that potentially available data would therefore be difficult for them to find and obtain ("*I am only a recent internet user so am not able to comment on accessibility of many of the data sources. I am still learning!").*

Figure 6.2-9: Question B.3 – "How would you rate the online accessibility of these types of data?"



N = 497-525 (Base: all respondents who declared in B.2 that at least one type of data was "very important" or "rather important" for them)

		++	+	-	--	
	B.3	very good	rather good	fair	poor	N
a	Satellite & airborne remote sensing data	81	161	190	92	524
b	Terrestrial laser scanning data	40	77	207	194	518
c	Prospection & field survey data	59	124	218	125	526
d	Government site management data	45	139	209	117	510
e	Excavation data	84	142	192	122	540
f	GIS data	68	160	209	85	522
g	Data for corpus studies	59	145	209	98	511
h	Data from material & biological analysis	51	126	217	126	520
i	Radiocarbon & dendrochronology data	61	122	217	119	519
j	Data for model-based computing	28	76	234	165	503
k	Data mining for identifying outliers	27	61	227	181	496

Comments from respondents:

- *"In Italy Government data about surveys and excavations usually are not available on paper, and even less on-line. I found extremely useful websites as Arachne with scans of ancient text and plans: one of the best things in Internet."*
- *"Most of my material study objects origins from determination sites, like the Portable Antiquity Scheme (PAS), muntenbodemvondsten.nl or bodemvondstenwereld.nl. These are all non-published data elements."*
- *"Excavation data accessibility fair to poor - slightly better but patchy."*
- *"I've left fair that data that I've not had cause to try and find."*
- *"I really don't know how accessible a lot of these data sets are and wouldn't have chosen don't know if it was an option"*
- *"As a self-employed pottery specialist my access to on-line journals and a good deal of on-line data is restricted because of the stranglehold that academic libraries have over such sources. I can afford to buy books and excavation reports but not to subscribe to the many journals that I require to stay up-to-date in my field. As I have no academic affiliation or means of obtaining such affiliation, I am unable to use academic libraries. With regard to grey literature and unpublished data I cannot always get access to what I need because of the limits placed on access by the interpretation of the copyright laws adopted by my local HER effectively placing what should be public records beyond practical use. Data provided by my clients varies widely from the utterly useless to the excellent - depending largely on their model of practice and the extent to which it follows MAP II."*
- *"I don't know the situation for material I don't use, so have not rated them"*
- *"In almost every case there is a very mixed picture with some excellent online resources while others are completely inaccessible. answers may also reflect my ignorance in some areas what is a rapidly changing field (mostly for the better)"*
- *"I have compiled a c14 data set for sites in my main area of work (Victoria Australia) https://docs.google.com/spreadsheet/ccc?key=0AtcLyWT58-K8dEhyR2NOX0g2bU0tX0VUOHpId3A3aGc&usp=drive_web&authkey=COGm0PsB&authkey=COGm0PsB"*
- *"If data is available it is not uploaded in raw-data format, only available as PDF, not as excel or GIS. Specific for archaeological data it tends to be available in MAPinfo format, which is not an open access format, ESRI shapefile would be better. Community Policy maps of Archaeology are not available in Archis, and sometimes poorly available on community websites and it is even harder to find if you have the valid version in temporal sense. Moreover rules connected to the policy zones are not clear of vague or even hard to find."*
- *"This varies widely depending on the region for which you are trying to obtain data. For example, online accessibility of satellite data and local government planning data for the UK is very good while the online accessibility of similar data for Italy is generally fair to poor depending on the region."*
- *"GIS data and maps, and remote sensing data are often "accessible" online in various interactive mapping portals, but they are not very often downloadable for use in GIS software - or are available only at an exorbitant fee (e.g. townland boundary data or road network data from the Irish Ordnance Survey, which despite being publicly funded, charges outrageous fees for access to its data and often under extremely restrictive usage licences)."*
- *"I use some of those types of data too rarely to be able to reply to all..."*

- *"Much of the material my research uses is in the form of C14 dates, site excavation data and materials data (specifically zooarchaeological data such as quantification of species, skeletal element). Much of the material available online is in the form of online articles, which do present some data, but understandably only that data that is directly relevant to the discussions in the paper-very few actually make all the data available, making broad comparisons between multiple sites very difficult."*
- *"Note that my research is not conducted in the EU."*
- *"Excavation data published online at the archaeology companies' websites are easily accessible nowadays. But this only goes back to the ca year 2005. Nearly all the older excavation data (report text, documentation, GIS-data, photos) is hard to access and to overview (archives, libraries, but no joint database)."*
- *"I wasn't aware that the bottom 4 categories were even accessible online"*
- *"access very variable though - site by site"*
- *"Don't know might be a good idea- plus ability to take out an earlier entry and leave a button blank"*
- *"Not enough experience of these to comment"*
- *"have not tried to access data on these topics as such. I normally find what I need in the literature or by asking colleagues directly. there should be the option to answer "don't know" or "n/a"*
- *"I am only a recent internet user so am not able to comment on accessibility of many of the data sources. I am still learning!"*

6.2.3 Awareness and use of data repositories

Awareness of digital data repositories

A specific goal of the Online User Survey was to explore to what extent the community is aware of existing online repositories with relevant archaeological holdings, and to what extent researchers make use of these repositories. In consultation with the ARIADNE consortium members, the following repositories were selected as the most important and relevant ones (also with regard to the envisaged ARIADNE services):

- The **Archaeology Data Service** (ADS) was established in September 1996, as one of five discipline-based service providers within the Arts and Humanities Data Service (AHDS). The ADS developed from a successful bid to the AHDS made by a consortium of university Departments of Archaeology and the Council for British Archaeology, led by the University of York.. It supports research, learning and teaching with freely available digital resources. It also promotes good practice in the use of digital data in archaeology. ADS is a member of the Europeana Network (http://archaeologydataservice.ac.uk).
- **Arachne** is the central object database of the German Archaeological Institute (DAI) and the Archaeological Institute of the University of Cologne. Arachne is a freely available internet research tool for archaeologists. It provides access to hundreds of thousands of datasets (http://arachne.uni-koeln.de/drupal).
- **Artefacts**, the Online Encyclopedia of Archaeological Small Finds, describes itself as "not a finished product, but an evolving programme reflecting the work of a group of researchers" (http://artefacts.mom.fr). Essentially, it is a freely available internet database which reflects part of a working base to which many researchers have contributed.

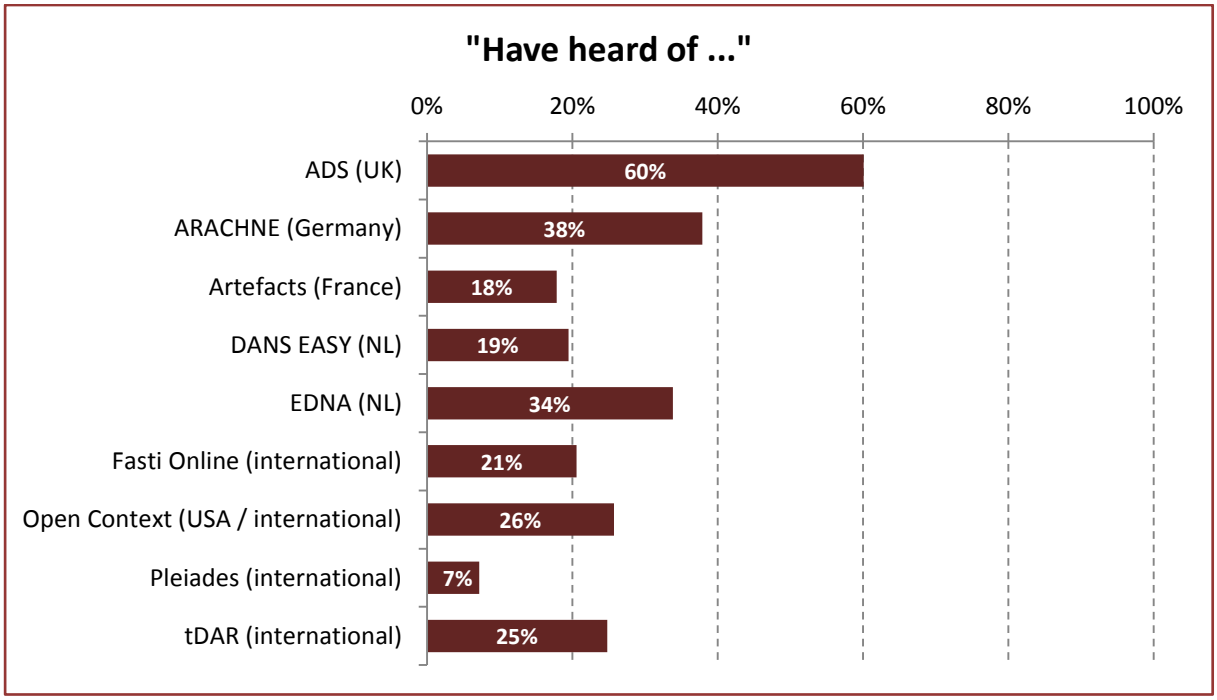
- **Data Archiving and Networked Services (DANS)** is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO). It provides sustained access to digital research data in different disciplines, with a focus on e-humanities and social sciences. **EASY** is the online archiving system of DANS. It offers access to thousands of datasets in the humanities, the social sciences and other disciplines. EASY can also be used for the online deposition of research data. DANS stores data in a permanent and sustainable manner, according to the guidelines of the international Data Seal of Approval (https://easy.dans.knaw.nl/ui/home).
- The **e-depot for Dutch archaeology (EDNA)** is accommodated at DANS. The e-depot stores the digital files from the research data of Dutch archaeologists. In 2014, the e-depot holds over 21,500 archaeological data sets (including about 18,500 publications and 3,000 larger datasets with photographs, GIS and data tables). New data are added daily.
- **Fasti Online** is a project of the International Association of Classical Archaeology (AIAC) and the Center for the Study of Ancient Italy of the University of Texas at Austin (CSAI) (http://www.fastionline.org). It provides a database of excavations since 2000, and a record in English and in the local language for each excavation season. Fasti is the digital successor of the earlier publication "Fasti Archaeologici" (published from 1946 to 1987 by AIAC).
- **Open Context** is a publishing service maintained by the Alexandria Archive Institute, a US non-profit organisation. Open Context reviews, edits, and publishes archaeological research data and archives data with university-backed repositories, including the California Digital Library (which archives the Open Context data). Open Context can complement and enhance conventional publications through comprehensive dissemination and preservation of rich digital data and media (http://opencontext.org).
- **Pleiades** is a gazetteer for ancient world studies operated by NYU's Institute for the Study of the Ancient World and supported by the US National Endowment for the Humanities. It is derived originally from the Barrington Atlas of the Greek and Roman World and continually adds new resources (http://datahub.io/dataset/pleiades).
- The **Digital Archaeological Record (tDAR)** is an international digital repository for the digital records of archaeological investigations. tDAR operates under the organisational umbrella of Digital Antiquity, a multi-institutional organisation that has been explicitly designed to ensure the long-term financial, technical, and social sustainability of tDAR (https://www.tdar.org).

While this sample of digital data repositories is not exhaustive, it represents a solid base for analysing the awareness and use of such repositories by practitioners across Europe. The UK-focused ADS was found to be the best known repository of all – 60% of the respondents said they had heard of ADS. The German ARACHNE and the Dutch EDNA databases were also quite well known (by about 35% of the respondents each). About a quarter of the respondents had heard of the US-based, international repositories Open Context and tDAR (see Figure 6.2-10). The results are certainly influenced by the geographic composition of the survey sample with a relatively high share of respondents from the UK and the Netherlands.

The relatively low level of awareness among researchers for any single repository (except ADS) raises some issues, however, for the development of ARIADNE services. On the one hand, the results suggest that most digital repositories still have, to a large extent, a national use context and user base, even if they are accessible for the international research community. The knowledge about these repositories is limited outside their own country; in the case of international repositories, awareness appears to be limited to specific research domains or practices. This presents a potential challenge for ARIADNE, but – at the same time – an opportunity. If datasets from (some of) these repositories can be effectively linked, it could boost the "cross-border use" of repositories and, ulti-

mately, the awareness of holdings in other repositories which have previously not been on the radar of researchers, and possibly enable new research practices with a stronger international dimension.

Figure 6.2-10: Question B.4 – "Which of the digital data archives for archaeologists have you already heard of?"



N = 536

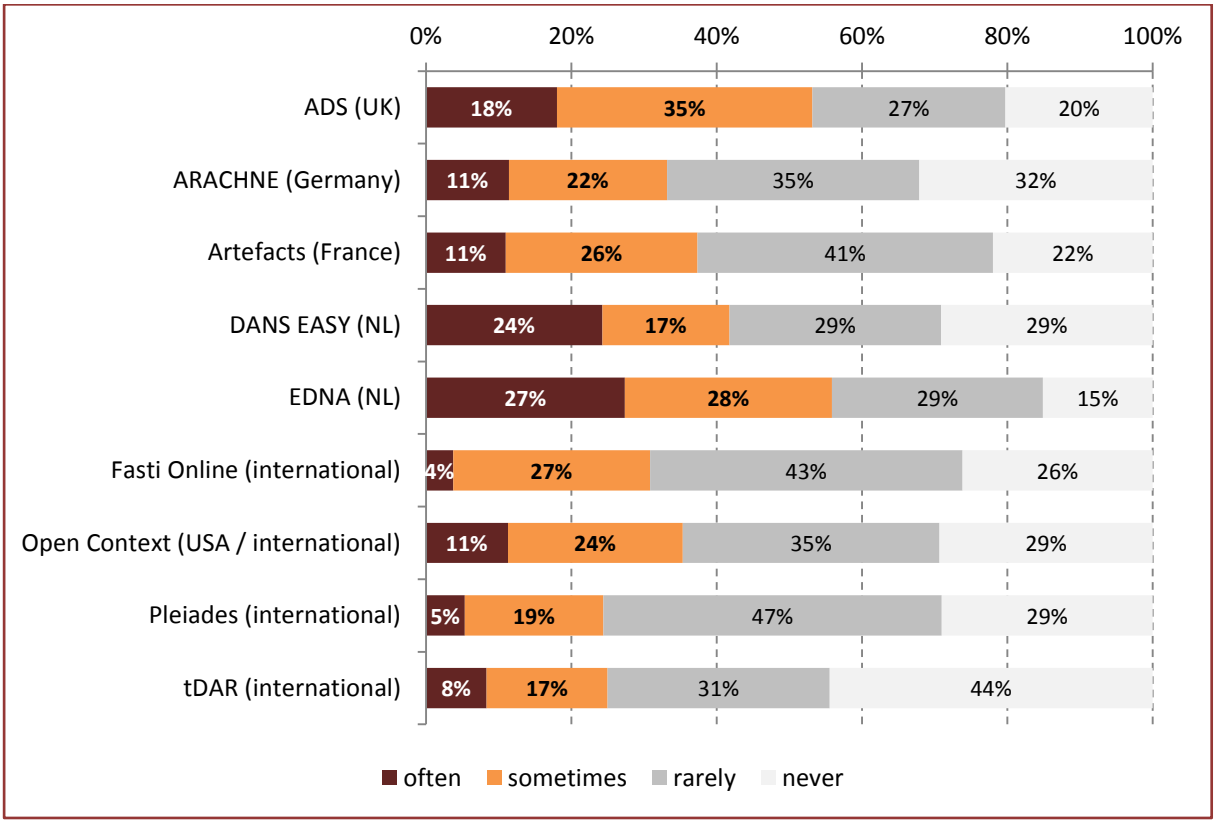
	B.4 Have heard of …		N
a	ADS (UK)	324	536
b	ARACHNE (Germany)	203	536
c	Artefacts (France)	95	536
d	DANS EASY (NL)	106	536
e	EDNA (NL)	184	536
f	Fasti Online (international)	112	536
g	Open Context (USA/international)	140	536
h	Pleiades (international)	39	536
i	tDAR (international)	135	536

Use of digital data repositories

Those respondents that had stated they were aware of a specific repository were then asked whether they actually used it, and to what extent (often/sometimes/rarely/never). ADS (UK) and EDNA (NL) are not only the best known repositories; they are also those which have the highest percentage of actual users. More than half of all respondents who were aware of these two databases said that they use them at least sometimes. Most of the other databases are (often or sometimes) used by about 30-40% of those researchers that actually know them.

Figure 6.2-11 shows the results for a different baseline: here, it is assumed that respondents who were not aware of a specific repository are not using it (they are here included in the "never" category, in contrast to the chart above). This breakdown of the figures shows clearly that ADS and EDNA are the only repositories in the sample which are, across Europe, used in a regular way by a significant share of the respondents in the sample (33% and 19% respectively).

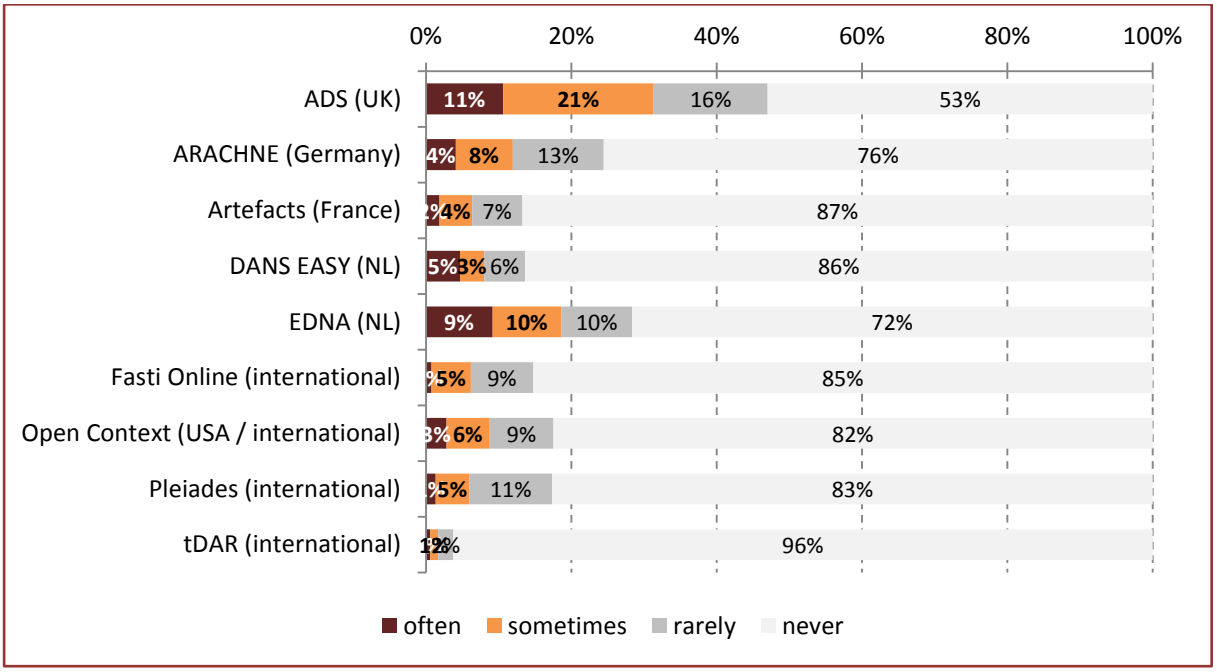
Figure 6.2-11: Question B.5 – "Are you using these digital data archives for your own research (and, if so, how often)?"



N = 36-316 per item (Base: all respondents that have heard of the respective data repository)

		++	+	-	--	
	B.5	often	some-times	rarely	never	N
a	ADS (UK)	57	111	84	64	316
b	ARACHNE (Germany)	22	42	67	62	193
c	Artefacts (France)	10	24	37	20	91
d	DANS EASY (NL)	25	18	30	30	103
e	EDNA (NL)	49	51	52	27	179
f	Fasti Online (international)	4	29	46	28	107
g	Open Context (USA/international)	15	32	47	39	133
h	Pleiades (international)	7	25	61	38	131
i	tDAR (international)	3	6	11	16	36

Figure 6.2-12: Question B.5 – "Are you using these digital data archives for your own research (and, if so, how often)?" – results for base "all respondents"



N = 536 (Base: all respondents having answered question B.4 about the awareness)

6.2.4 Challenges in searching data

The next questions in the survey aimed to explore the challenges which researchers are confronted with in their daily research routines when searching for data. The challenges can be grouped into two main categories:

- Challenges related to a **lack of transparency** on what is actually available: difficulties in finding out what data is actually available (where, how, in which format). This challenge could, in economic terms, be framed as a "market failure" stemming from high transaction costs. The researcher has no practical chance to identify and use relevant data sources (which would actually be available) because the effort for identifying and locating the source would be unreasonably high.
- Challenges in having **access to data**: A researcher may have identified relevant data he/she would like to use, but faces difficulties in getting hold of the data sets. The main difficulties in accessing (theoretically available) data as identified in the pilot interviews and subsequently explored in this survey are linked to the following reasons:
    - Restricted access rules (e.g. due to legal issues, data privately owned);
    - Cost issues;
    - Language barriers.

All of these issues were mentioned, to a different degree, in interviews conducted prior to the survey. The goal for the Online Survey was therefore to "calibrate" the various challenges in terms of their relative importance. This has implications for the ARIADNE project, which focuses primarily on improving the situation with regard to the first challenge – the lack of "market transparency". And indeed, the results confirm that the project addresses a relevant challenge.

General assessment of data transparency and access

In the first question of this block, interviewees were confronted with four statements which paraphrase the above mentioned challenges. They were to what extent they agreed or disagreed with these statements (see Figure 6.2-13). Two statements (no. 1 and 4) concerned transparency issues, the other two statements access-related issues (cost, restricted access).

With regard to **data transparency**, almost 90% of the respondents agreed (fully or partially) that they often did "not know what is available, because research data are stored in so many different places and databases", almost 60% agreed fully with this statement. It was thus by far the statement with the highest level of agreement. This statement presents, essentially, a key rationale for the ARIADNE project. The responses thus confirm empirically the validity of the ARIADNE objective to improve the overview of available research data by linking scattered resources. The outstanding importance of this particular statement was also confirmed by a comment from one of the respondents: "*The last point is the key one here.*"

There was another statement that addressed the transparency dimension, but addressed the issue from a reverse angle: respondents were asked if they agreed that they could find all or most of the data they needed in libraries, archives and museums. The goal was to see if there is an actual need to create or improve online resources. In a way, this question can also be seen as a validation of the views on the general availability of data. The results are, again, encouraging for all those engaging themselves in the development of digital repositories. About 45% of the respondents disagreed (partially or completely) with the statement. This represents a significant group which would probably welcome new, improved offers and services which facilitates their search for data.

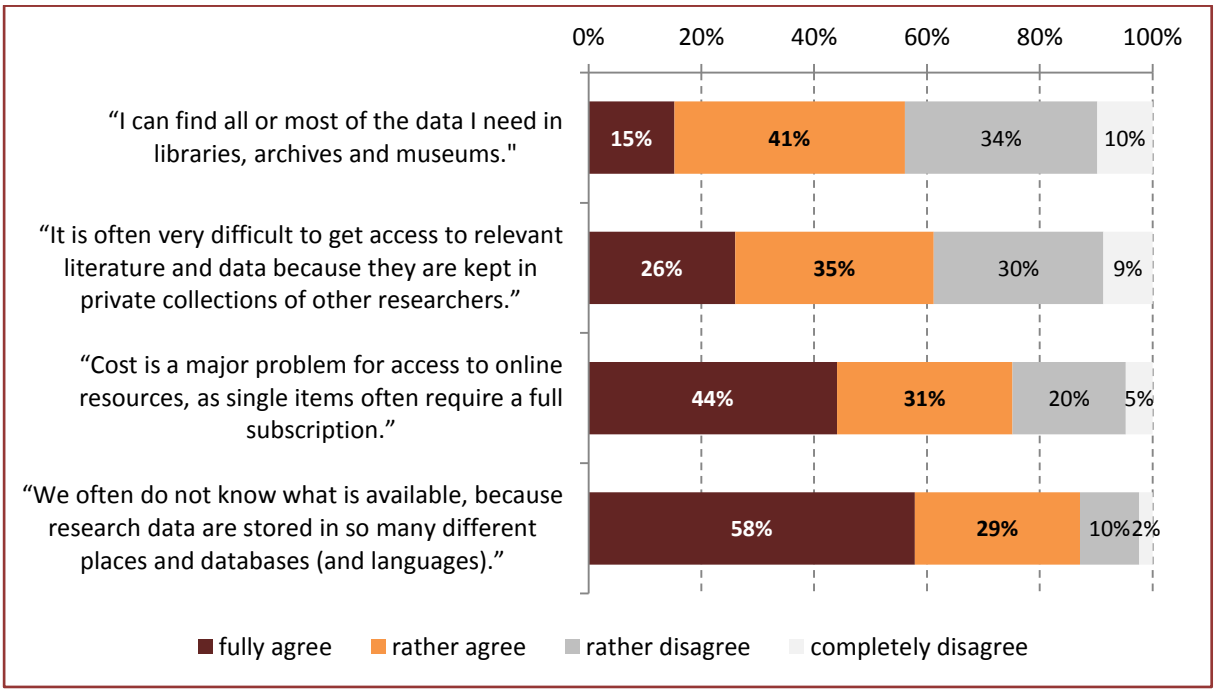
The results also imply that slightly more than half of the researchers surveyed appear to be mostly satisfied with the conventional data sources. However, this needs to be put into perspective. Only 15% said they fully agreed with the statement (compare this figure with the nearly 60% of the statement on not knowing what is actually available). Furthermore, it can be assumed that not all researchers have the same level of awareness regarding new methods for searching and accessing research data. Therefore, it can be expected that even from the group of researchers that are (at least partially) satisfied with the current situation, many may appreciate and actually use additional services once they become available. One of the comments shows that it can even be cumbersome to actually *collect* the data and material which has been identified, as this would be too time consuming: "*Find out that the data exists, yes - have time and money to get to all libraries and archives, no.*" This statement highlights the challenge that "data" can mean a lot of things in archaeology, not only digital data which can simply be downloaded from a database. The connotation which archaeological researchers have with "data" is much broader – a lot of materials would have to be looked at directly in their respective libraries or archives. The interesting question here is to what extent digitisation of such sources may help to overcome the barrier of having to go to the library or archive – and where the limits of digitisation lie.

Concerning another issue with regard to **data access**, the survey confirmed that **costs** can be a relevant barrier, in particular with regard to digital resources. Nearly three quarters of the respondents agreed (fully or partially) that cost was "a major problem for access to online resources", because "single items often require a full subscription". More than 40% of the respondents fully agreed with this proposition. Examples and illustrations of this problem were given in the individual comments to this question (e.g. "*It's frustrating when a museum asks for 50€ for each photo of an object*"; "*As an independent researcher I often feel 'priced out' of important discussions in my field.*"). The ARIADNE project will most probably not have a direct impact on this aspect, at least not in the short term. ARIADNE presents a meta-infrastructure and will not affect the commercial conditions for obtaining specified data sets (which are determined by the actual owner of the IPR). However, the fact that a majority of researchers experience cost as a major problem (for

accessing digital resources) should be carefully considered in case ARIADNE attempts charging fees for using its services, as it could severely affect their acceptance and use.

About 60% of the respondents also agreed (fully or partially) that "it is often very difficult to get access to relevant literature and data because they are kept in private collections of other researchers". Thus, this issue is also highly relevant, even if not as prominently quoted as the cost issue. The issue was also mentioned and illustrated during the pilot interviews by some of the researchers. They explained that they were often confronted with the situation that literature or other data sources which should theoretically be available in a library or museum, were actually located in the private collections of colleagues, for instance as part of a collection related to a specific research project. In practice, it would then often be very difficult to get access to these pieces, and depend on individual or institutional contacts.

Figure 6.2-13: Question B.7 – "Based on your own experience: do you agree with the following statements?"



N = 585-592 per item (depending on number of respondents without answer)

		++	+	-	--	
	B.7	fully agree	rather agree	rather disagree	fully disagree	N
a	"I can find all or most of the data I need in libraries, archives and museums."	90	242	202	58	592
b	"It is often very difficult to get access to relevant literature and data because they are kept in private collections of other researchers."	152	206	176	51	585
c	"Cost is a major problem for access to online resources, as single items often require a full subscription."	259	183	118	28	588
d	"We often do not know what is available, because research data are stored in so many different places and databases (and languages)."	342	173	62	14	591

Comments from respondents:

- *"Sometimes is easier to find new sources on Google Books than in institutional websites; probably it is a question of tags."*
- *"It's also frustrating when for instance a museum asks for 50,- euro per photo of an object."*
- *"See previous comment about grey literature. "Balkanization" of data hasn't been a major problem here; but getting archaeological data in a useful form can be difficult, for instance when tables are stored as simple text."*
- *"For all I know, some French or German researchers have pinpointed Herts as an exemplar of Early Saxon settlement and the means by which that came about....but I have never known of the European counterparts to ADS, nor what research projects might be underway there."*
- *"As an independent researcher I often feel 'priced out' of important discussions in my field."*
- *"What is urgently needed is access to academic libraries with good collections of internet journals and clarification of the laws around copyright for local government civil servants so that access to digital copies of full site reports and other data is possible."*
- *"Even if I feel confident that I have successfully mined the data of libraries, archives, and museums, I now always suspect that there is more out there than I do not yet have the tools to access."*
- *"Q2: I am thinking mostly of grey literature from developer-funded excavations, which often you can access by asking the relevant person or organisation it is just that it takes time and can be difficult to get hold of quickly. General: Universities in the UK seem rather unaware of the research potential of nationwide Historic Environment Records (HERs)."*
- *"Data in EDNA is not synchronised with Archis. The 'Rapportmeldingen' in Archis should all be available in EDNA and vice versa. In the current situation you have to check both archives, to see if there is a public report available.*
- *"Libraries, archives, museum catalogues and indexes are seldom available online."*
- *"If the comments to statement (N 2 and N 3) involve both ethical and archaeological marketing issues on national and international level, there is little room for doubt, that visibility of research data online N4 is still rather low. In this respect ARIADNE mission to integrate and raise interoperability seems as a logical solution."*
- *"The museum which holds one of my planned core research assemblages, this year introduced commercial-rate research fees for research students, without warning. I am now effectively "locked out" of looking at this material, as the costs have not been included in my budget and would be prohibitive given the amount and type of research that I need to perform. I am having trouble identifying a suitable alternative assemblage and this may result in my PhD being abandoned. The fact that this museum seems unable to locate the majority of the assemblage I need (despite repeatedly assuring me that they know where it is) is now a side issue."*
- *"Find out that the data exists, yes - have time and money to get to all libraries or archives, no."*
- *"The last point is the key one here-"*
- *"I have answered with my personal view although being member of a research institution I have access to an enormous range of online resources. Yet I believe cost is a major problem for most people."*
- *"I disagreed with the first statement in that while I could (in theory) get all the info I need from libraries, archives and museums, the costs and logistics of accessing all the repositories that I would need to get to renders it impossible."*

Specific barriers experienced

The next question elaborated on the four general statements discussed above. Respondents were presented with a number of items that might present a barrier for their research and asked to assess the importance of each item. The items represent specific aspects of transparency and access. In summary, all items were found to be relevant; except for two items, a majority of the respondents said they were either "very important" or "rather important" (see Figure 6.2-14). On the whole, the results confirmed the picture as obtained from the introductory statements. Three issues in particular emerged as major barriers to data access (with more than 70% of respondents saying these were very or rather important):

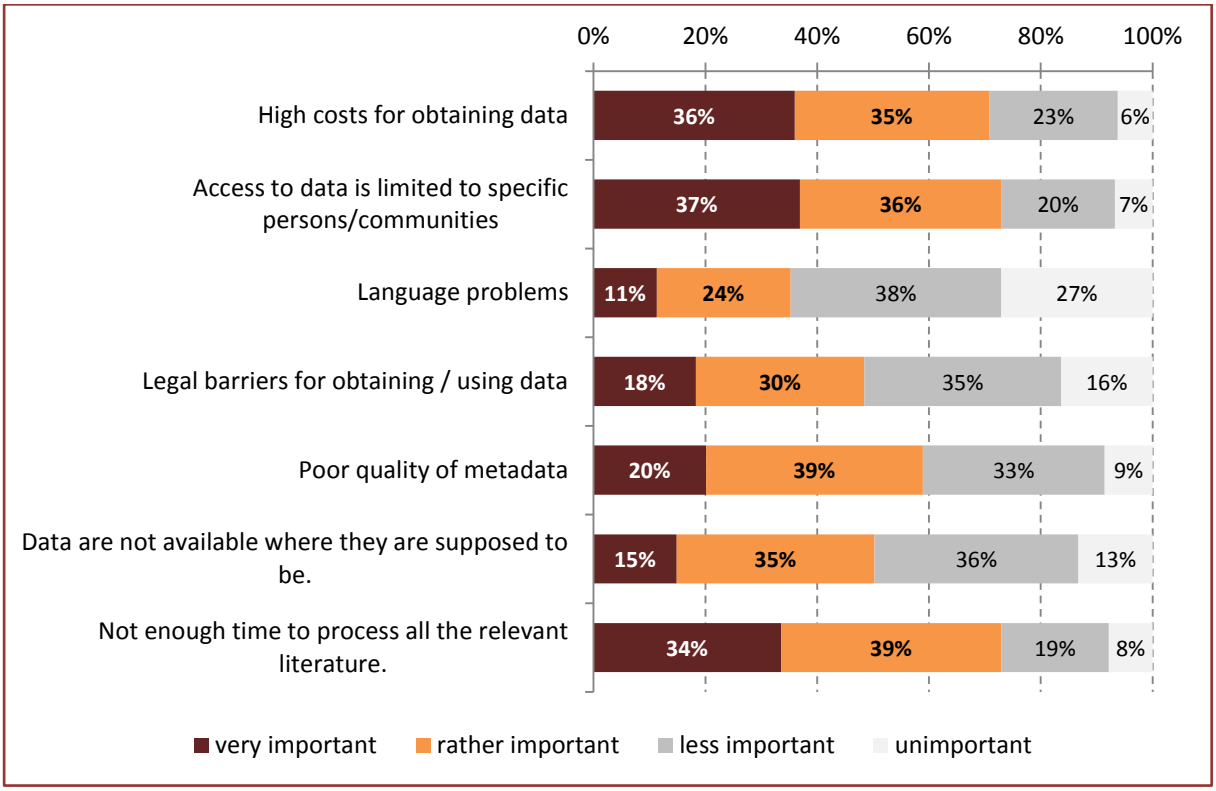
- **Costs** (for obtaining data);
- **Restricted access** (limited to specific persons or communities);
- **Lack of time** to process relevant literature.

The dominance of the **cost** issue is confirmed by the comments made by respondents on this question. Many of the comments made concern this item (see quotes below). Costs do not only incur from access fees, but also when researchers have to travel in order to view the holdings of libraries and archives ("*The cost of travel and research time spent finding archives is significant, especially when access is not guaranteed, and when archives may be scattered and unpublicized*.").

About 60% of the respondents also confirmed that the poor **quality of metadata** was an important problem. This is definitely an area where the ARIADNE project aims to contribute to improvements. Linking data from different sources implies almost by definition that some work must be done on metadata. There was one comment on this particular item, arguing that metadata was, in particular, a problem for data from older projects: "*Metadata is getting better and better lately. But for older data (excavations) there is almost no good data to find. No GIS, and usually only a report.*" It might be worthwhile for one of the ARIADNE Special Interest Groups to look into this issue in more detail, in particular with regard to the implications which this may have for integrating data in the ARIADNE services to be developed.

"**Language problems**" are apparently not seen as a major issue. Only about a third of the respondents felt that it was an important issue. A slight caveat here is that the survey sample had a relatively high share of respondents from the English speaking community (UK, USA); as much of the international literature and data is available in English language, English speaking researchers may be in a preferred situation. However, by and large, the results indicate that the "language problem" should probably not be overrated. If research activities are international, researchers can be expected to speak English; if research is national or regional, sources in local language may be sufficient. A preliminary, tentative conclusion for the ARIADNE project is that the project should probably not be too concerned about the language dimension of the services it is going to develop (see also the findings about expectations towards ARIADNE services, Section 6.2.6).

Figure 6.2-14: Question B.8 – "To what extent have you been confronted with the following barriers in your research when trying to access data? Please rate the importance of these barriers."



N = 565-578 per item (depending on number of respondents without answer)

		++	+	-	--	
	B.8	very important	rather important	rather unimport.	un-important	N
a	High costs for obtaining data	208	201	133	36	578
b	Access to data is limited to specific persons/communities	213	208	117	39	577
c	Language problems	65	138	217	156	576
d	Legal barriers for obtaining/using data	104	171	200	93	568
e	Poor quality of metadata	115	222	186	49	572
f	Data are not available where they are supposed to be.	84	200	206	75	565
g	Not enough time to process all the relevant literature.	192	226	110	45	573

Comments from respondents:

- *"The cost of pictures for publication is very high, especially when academic publications are below 500 copies."*
- *"Rarity of some old books; some are available online, but many are not. Dissertations are also a problem."*
- *"I can imagine for professionally paid researchers the above answers would be different. "*
- *"Many books and journals are not reviewed due to JSTOR and similar charges. Alternate access (e.g., by travelling to university library) is also often cost-prohibitive."*

- *"I am retired, aged 73; no income except small pension-so much of the material on the web is beyond my pocket and this must then affect the quality of my final study adversely. But should the work of the amateur or retired researcher be so compromised?"*
- *"Q 5: Online resources are improving but it is often difficult to screen for environmental archaeology reports."*
- *"Language problems are soluble!"*
- *"Data provided in proprietary formats, or only as an online service rather than downloadable files."*
- *"keywords and tags in databases are sometimes reliable, often times missing or unreliable (such as misnamed items, or misinterpreted scenes of craftwork); for international or cross-database networks, tags would need to be more unified and translated to be really useful"*
- *"I have had trouble accessing materials held at public museums because my affiliation was with a foreign university (even though I happened to be of the nationality of the museums)."*
- *"Limited budget for each project!"*
- *"The cost of travel and research time spent finding archives is significant, especially when access is not guaranteed, and when archives may be scattered and unpublicized."*
- *"Metadata is getting better and better lately. But for older data (excavations) there is almost no good data to find. No GIS, and usually only a report."*
- *"Whilst affiliated to a university, I have the advantage of free access to journals, but this will become a problem when I am no longer a post-grad."*
- *"Also not respecting copyright laws regarding publications of excavations and materials, i.e. the hoarding of materials by some researchers, making it impossible for others to examine them, even if they no longer have any copyright rights to the material.*"

6.2.5 Research practices and challenges in depositing data

After having explored the practices and challenges in searching for *existing data*, the next module of the survey explored how researchers deposit (and possibly) share their own data, i.e. the *new data* they have produced themselves with their research projects. Again, the survey focused on current practices, requirements and challenges in this context. In some of the questions, the items proposed were similar or the same as in the corresponding question in the module on searching for data, for instance in the first question on the types of data which researchers typically generate with their research projects.

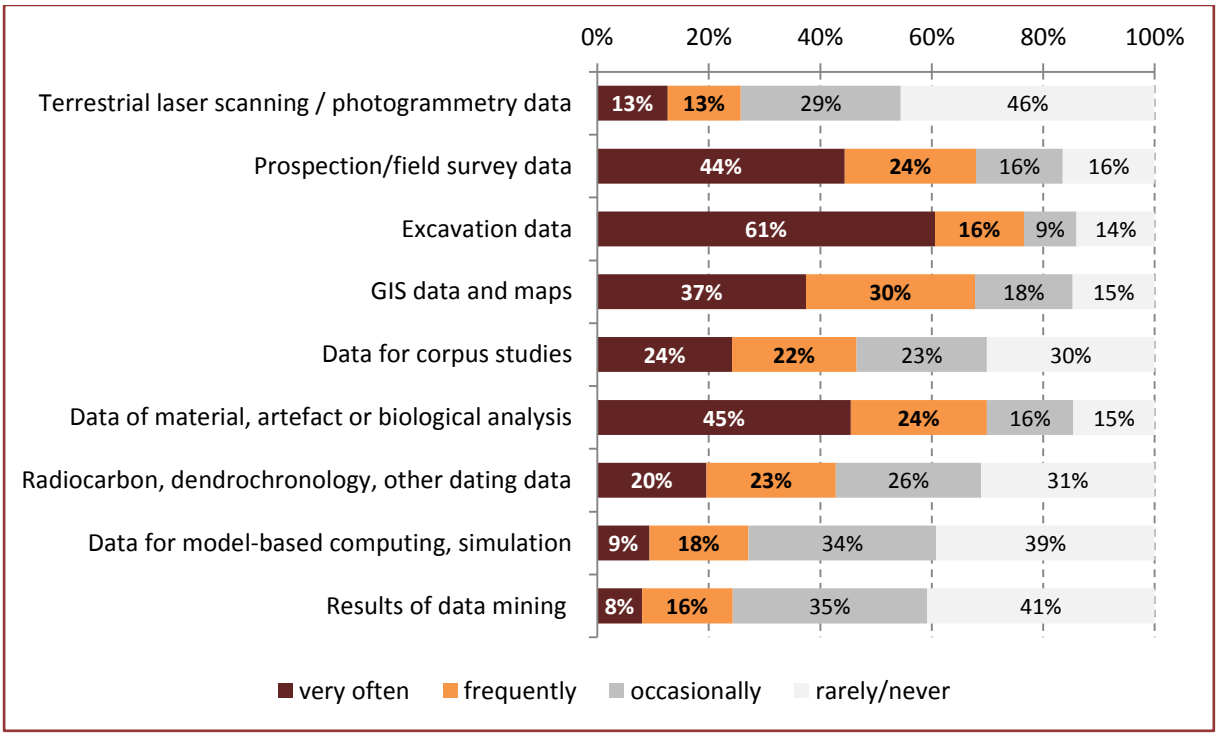
Types of data generated

Unsurprisingly, the answers mirror quite closely the picture obtained from the corresponding question on the importance of various types of data for research purposes. It can be assumed that a researcher who generates, for example, predominantly excavation data will also be interested in the same type of data and therefore attribute "high importance" to excavation data.

In fact, **excavation data** was not only the single most important type of data during the search phase (75% of the respondents said that excavation data were "very important" for them), but is also the data most widely produced by the community. Close to 80% of the respondents said that excavation data were "very often" or at least "frequently" a result of their research activities (see Figure 6.2-15).

About 70% of the researchers interviewed said that their projects generated very often or frequently GIS data, data of material or biological analysis, and data from field surveys are also results. Again, these are the same types that were also ranked as second most important in the module on searching for data. This is not to say that the other types of data are not relevant, of course. For almost all types of data proposed, at least 50% of the respondents said that their projects produced them at least occasionally. For all types of data, at least 20% of the respondents said that these data were generated frequently or very often.

Figure 6.2-15: Question C.1 – "What type of data do you (or does your institute, in case of directors) typically generate with your research projects? Please rate the suggested data in terms of their importance."



N = 536-563 per item (depending on number of respondents without answer)

		++	+	-	--	
	C.1	very often	frequently	occasion-ally	rarely or never	N
a	Terrestrial laser scanning and photogrammetry data	68	71	155	246	540
b	Prospection and field survey data	245	130	86	91	552
c	Excavation data	341	90	53	79	563
d	GIS data and maps	209	169	98	82	558
e	Data for corpus studies	134	124	130	167	555
f	Data of material, artefact or biological analysis	255	137	87	82	561
g	Radiocarbon, dendrochronology, other dating data	108	128	144	172	552
h	Data for model-based computing, simulation	51	97	184	214	546
i	Results of data mining	43	87	187	219	536

Comments from respondents

- *"I am creating a data base and a critical review of previous sources and publications concerning my field of study."*
- *"Would do more, but the state services ("Landesaemter") rarely require more than the bare minimum."*
- *"We tend to be early-adopters of technology in our company, and provide services to clients that rely especially on GIS-based research (often involving data not previously considered for GIS approaches), so most of the columns are 1 where this would not be typical of competitors or local academic institutions."*
- *"Come back to me once my article is complete: apart from photos of artefacts, I am not yet in a position to define what will be in the final mix."*
- *"We have used, or tested, terrestrial laser scanning occasionally but it is still not often used in our projects."*

Storing and depositing data

Following the work-flow-based approach of this survey, the next questions explored how the data produced in research projects was typically deposited and, possibly, shared with the project community. As there are many possible practices (and combinations thereof), it was difficult to simplify all the possible practices into a standardised question. Four basic practices of depositing data were proposed, and respondents were asked to say whether they were applied in their institute in most projects, many, a few, or not at all. Multiple answers were allowed.

The answers indicate a fundamental problem for bringing together archaeological research data: data may not only be scattered across different institutional databases, but a good deal of data might not even make it to the institutional database but remain on the computers of the individual researcher. Nearly half of the respondents said that they store data on their own computer in all or most projects, and another 25% that this was the case in many projects. The figures are similar but slightly lower for storing and depositing data in shared project archives on servers of the institute.

A possible reading of these figures is that, typically, during a project, researchers store data locally on their computers and upload (e.g. as a backup, or to share with colleagues) specific data to the institutional server. This practice was also described in pilot interviews conducted for the online

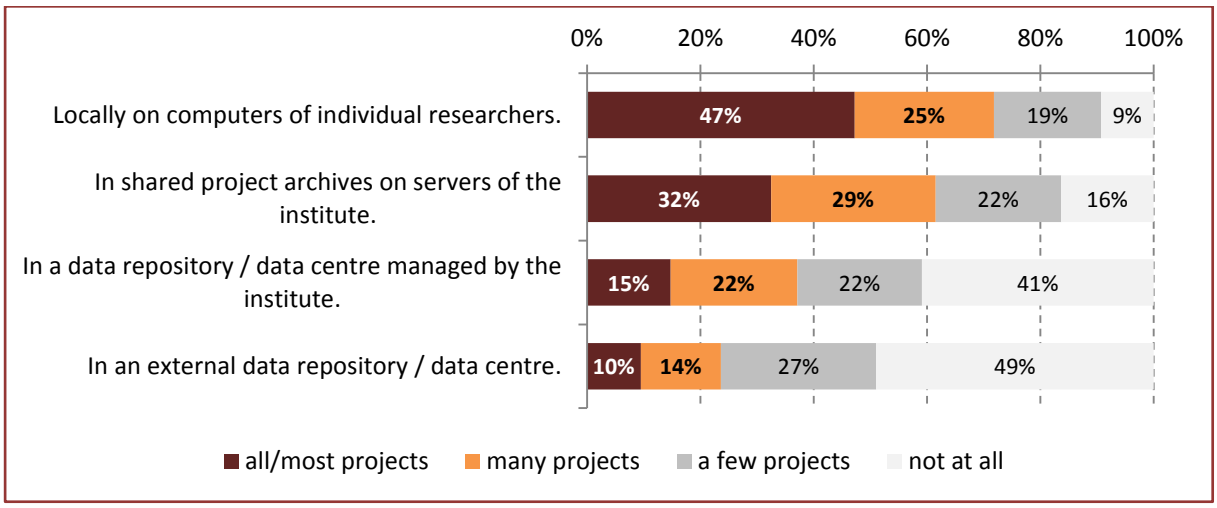
survey. After the end of the project, the documentation (including the resulting data) is then typically deposited in shared project archives on servers of the institute, and/or in a data repository managed by the institute. However, a significant share of the respondents do not deposit data in any shared archive or repository. There are, theoretically, two main reasons for keeping research data only on the local computer: either there is no opportunity or no incentive to deposit data other than storing it locally.

- **Lack of opportunity**: 60% of the respondents said that their organisation does not have an institutional content repository that is managed by dedicated staff (see next section). In particular, self-employed archaeologists (accounting for about 13% of the sample) are unlikely to deposit data in any kind of institutional database, as they are not affiliated with a research organisation. For many of them, the core "IT infrastructure" will mainly consist in their own computer or laptop (and, hopefully, a solution for archiving and back-ups). The same may apply to researchers working on their Ph.D. thesis (14% of the sample), even if their research is conducted within the broader organisational framework of a research organisation. The actual Ph.D. project may not necessarily be part of the institutional work, and therefore they are likely to store related data on their own computer.
- **Lack of incentive**: Data sharing is not necessarily a rewarding activity from the perspective of the individual researcher (see also next section). Researchers, if not explicitly requested to do so, may have a preference for keeping their data to themselves rather than depositing and thus sharing them.

Even if the results of most projects will not make it into a supra-institutional database, the use of such external repositories or data centres is not uncommon in the field. Twenty five percent of the respondents said that they deposited research data in such repositories for many projects or even most projects; and 27% said that the results of a few projects were shared in that way. Thus, more than half of the respondents in the sample confirmed that, at least occasionally, project results would be deposited in external repositories such as those introduced in Section 6.2.3. This is somewhat in contrast to the rather low level of awareness and use of these repositories that was found in the questions about them, and an issue that should be further investigated for the update of this report during the second project year.

The results are broadly comparable to a stakeholder survey carried out by the IANUS Research Data Centre initiative in 2013 among about 240 German institutions. In this survey, a vast majority of the respondents declared that they store their data on hard drives or computers, only a minority delivered them to central servers, computing centres or professional archives. Only 16 out of the 240 shared their data via a web service (see Section 3.6.7 for further details on this survey).

Figure 6.2-16: Question C.2 – "How are these data stored and deposited? Are the following practices typical for all or most projects, many projects, a few projects, or not at all?"



N = 525-561 per item (depending on number of respondents without answer)

		++	+	-	--	
	C.2	all or most projects	many projects	a few projects	not at all	N
a	Locally on computers of researchers.	265	138	106	52	561
b	In shared project archives on servers of the institute.	179	160	122	90	551
c	In an institutional data repository.	78	118	116	216	528
d	In an external data repository.	50	74	144	257	525

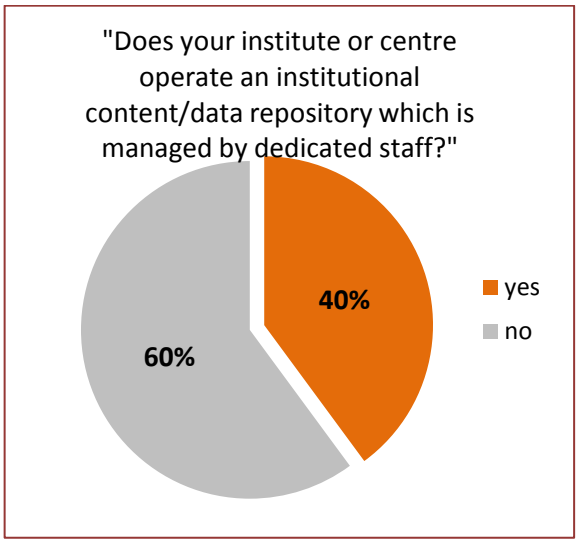
Operation of institutional repositories

As already indicated, 40% of the respondents worked in a research organisation that operates an institutional repository which is managed by dedicated staff. Admittedly, this question leaves room for speculation, as different people may have different understandings of the concept of an "institutional repository". It is, unfortunately, hardly possible to use extensive definitions for such constructs within an online survey. In particular, there may be a misunderstanding of whether a centrally managed server already constitutes a repository or not (in our understanding, it does not).

But even when accepting this lack of sharpness in the definition, the answers indicate that a majority of researchers works in smaller research entities (if affiliated with an organisation at all) which do not provide a repository. This could present an opportunity for supra-national repositories, as it presents an opportunity for researchers to share data and promote their own work (provided that they have an incentive for doing so). On the other hand, if research organisations maintain their own repository, it is often restricted to internal use, and there may be little incentive to "open" it to external users or motivate researchers from the organisation to share data by providing them to supra-institutional repositories. The know-how as represented by the research data in the institutional repository is an asset for the organisation that is not to be disclosed (this was confirmed in pilot interviews for this survey).

Figure 6.2-17: Question C.3– "Does your institute or centre operate an institutional content/data repository which is managed by dedicated staff?"

	C.3	#	%
a	yes	225	40%
b	no	344	60%



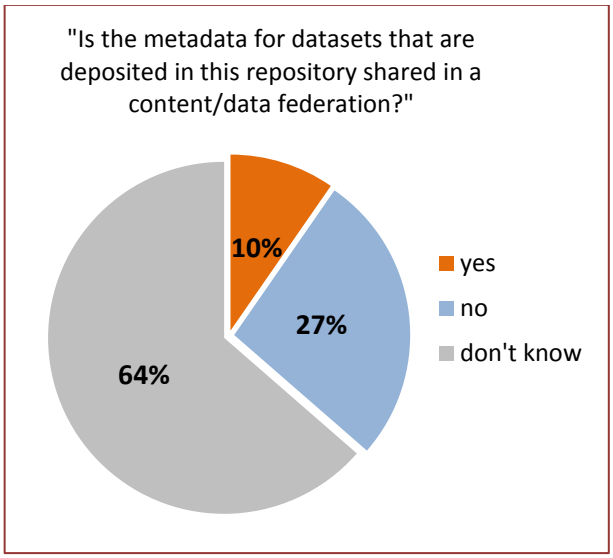
N = 569

Sharing of metadata

Those researchers who had responded that their organisation operates an institutional repository were asked whether the metadata for data sets deposited in this repository were shared in a data federation. Admittedly, this is quite a complex question for anybody not directly involved in managing data sets, and, expectedly, a large share of the respondents (about two thirds) did not know whether this was the case. Out of those that were aware of it, about a quarter said that metadata were shared in a data federation, and three quarters said it was not. It is speculative whether this ratio can be simply extrapolated to the full sample; but there is no immediate argument why it should be completely different.

Figure 6.2-18: Question C.4 – "Is the metadata for datasets that are deposited in this repository shared in a content/data federation (for instance through providing an OAI-PMH target)?"

	C.4	#	%
a	yes	22	10%
b	no	61	27%
c	don't know	145	64%



N = 228 (Base: C.3 = "yes" or no answer)

Responsibility for maintaining the data after project completion

The respondents were then asked who was typically responsible for maintaining the data after the completion of a research project. As to be expected (and as to be recommended), in most cases, this is still part of the responsibility of the project manager (54%). About a quarter of the respondents said that there was a dedicated member of the research team (other than the project manager) who manages data after the completion, and about 20% said that neither of the two options applied.

While this question only focused on maintenance, from a procedural view, the preparation of the data for depositing them in an institutional database or external repository should also be considered as a stage. In a pilot interview, a senior researcher and project manager explained that the procedure for documenting and archiving research data from a research project in the institutional database would typically involve both the project manager and a research assistant. The project manager would draw up the plan (which data to be archived and how), supervise the procedure and possibly upload the main final report, while the assistant would do the preparatory work (such as arranging that the respective data sets are available in the required format, provide the requested metadata, of prepare the forms to be submitted).

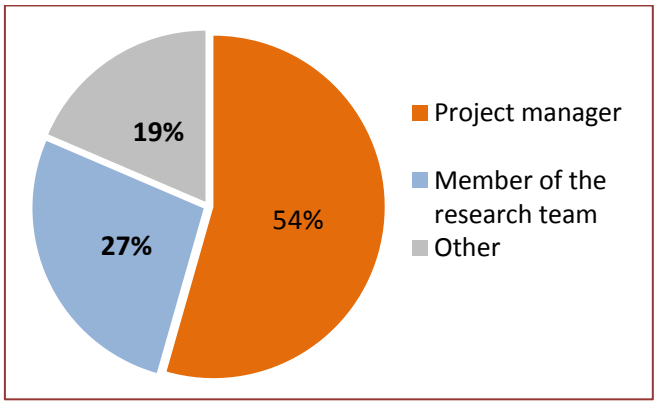
The comments received on this question also indicate that the answer options given were possibly too simplistic (again, often a necessary compromise in consideration of the time constraints of an online survey). For instance, if data are delivered to an archive, museum or county after completion of the project, they are often also maintained by the respective organisation (rather than by the researchers who had conducted the research.

The comments also demonstrate that the situation is quite different in the case of contract research, where results are delivered to the client and then maintained by him: "*Once I have written a report the whole thing goes to the client and I expect it to be made available without restriction to anyone who wants to read or use it - but I have no control over this (…).*"

In summary, archaeological research data are stored, deposited and maintained in many ways after a project has been completed, depending on the context of the research (e.g. academic vs. commercial) and the legal context regarding the data that are generated. It is hardly possible to describe standard procedures.

Figure 6.2-19: Question C.5 – "Who is responsible for maintaining the data after the completion of a research project?"

	C.5	#	%
a	The project manager or team leader	308	54%
b	A member of the research team specifi-cally appointed for this task.	153	27%
c	Other	105	19%



Comments from respondents:

- *"From our voluntary organisation, we publish everything through www.academia.edu. After each digital publication is ready, it's uploaded."*
- *"Ultimately the county or local museum"*

- *"We are hoping to archive our data within the organisation where possible and with the ADS when this is not possible."*
- *"Once I have written a report the whole thing goes to the client and I expect it to be made available without restriction to anyone who wants to read or use it - but I have no control over this and the data usually ends up being restricted for pointless or wholly fictional reasons. Naturally, if someone approaches me directly I will normally make an effort to supply them with the data they need, usually by getting an OK from the project manager or client"*
- *"free-lance contractor research data is copied to the project manager, and sent to archive. Researchers keep the data on their home computers, and lose it if their home computer crashes and is not backed up."*
- *"Some of my projects are museum-sponsored, but there are 'orphaned' projects with no institutional home; if anything happened to the project director, no one is responsible for the material in his/her or my possession."*
- *"I'm only guessing here. I actually have no idea."*
- *"And this is clearly not a sustainable system, as personnel will come and go."*
- *"Not applicable to our society. It is a group responsibility."*
- *"data that should be in commissioning body archive frequently not or is unknown - requests for data especially for work over 10 years old, comes direct, therefore large burden and responsibility for independent workers.*
- *"And other members"*
- *"According to the Antiquities Law, an excavation or survey project archive should be submitted to the National Archaeological Archive. At the moment though, the Archive has no facilities for digital data storage."*
- *"Most of my data is accessioned by the English Heritage Archive, formerly known as the National Monuments Record (Swindon, UK)"*
- *"The archive of the articles and main results of the different research projects in the UNIARQ - University of Lisbon Research unit is to be implemented in the University open repository."*
- *"For us, the scientific community, with a technical support, must be responsible for maintaining and adding the data in an ongoing process."*
- *"En principe le ou les responsables, mais les données sont éparpillées entre les différents chercheurs et on a accès à ces documents que de manière limitée ou par accident."*

Publication of data

The next question in the depositing module addressed how research data are published. Respondents were given a number of options and asked whether this way of publishing archaeological research results was used in most projects, many, a few or none at all. Similarly as for the maintenance of data after the research project, the publication depends clearly on the context of the research. While publication of results is normally a key objective in academic research, it is rarely a goal in contract research.

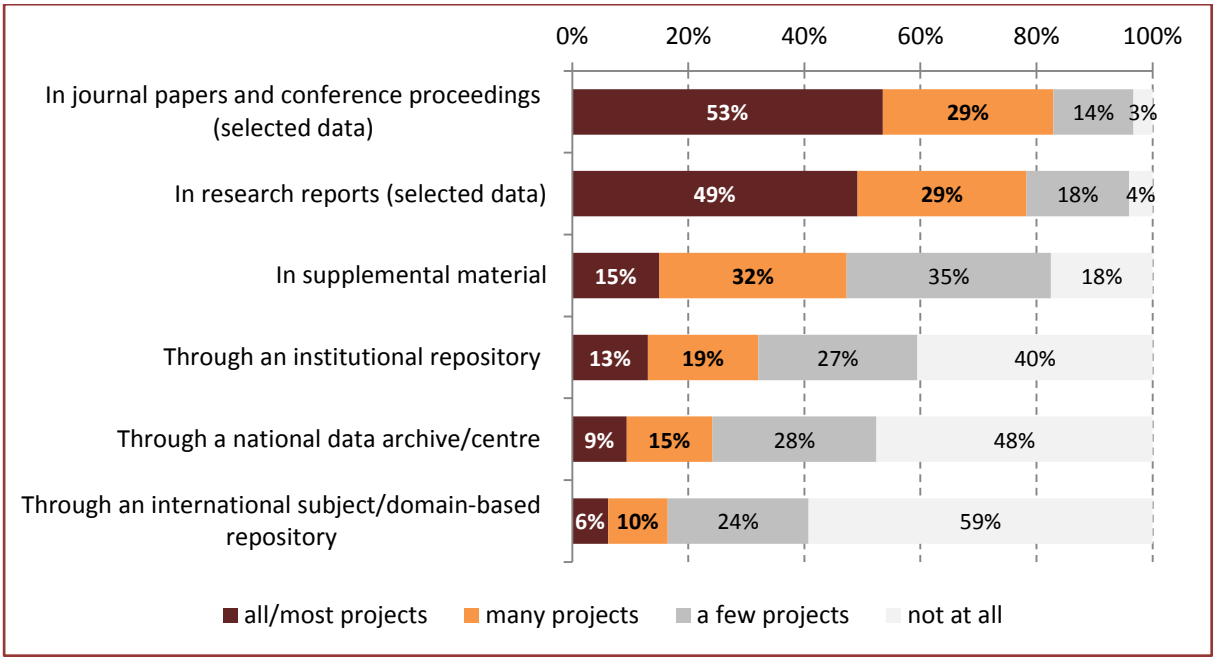
Expectedly, the standard academic approach of presenting the main findings (with selected data) in journal papers, conference proceedings or research reports is by far the most common and frequent way of sharing research results. More than 80% of the respondents said that contributions to **academic journals** and **conference proceedings** were a common means for publishing results, in many or even all research projects they were involved in. Similarly, preparing a standard **research report** which contains a selection of the data generated by the project (for instance selected tables, charts) is standard practice applied in the vast majority of projects.

The publication of extended selections of data (for instance in supplemental material, such as extended tables) is also quite common; close to 50% of the respondents said that this was the case in many or all projects.

When it comes to publishing data in institutional or supra-institutional **repositories** ("publishing" here essentially means agreeing to share the data with others by granting them access to the data), the picture is more nuanced. The percentage of researchers who confirm that their project results are rarely or never published in such repositories goes up significantly. In particular, the publishing and sharing of data in national data archives or international repositories is not yet common practice. Only about 15-25% of the respondents said that they would do so in many of their projects, while 50-60% do not make use of repositories at all.

The reasons are the same as those discussed above. In many cases, researchers may have no incentive to publish the full data set they have generated (even if they have the contractual opportunity to share their data). In other cases, such as contract research, the researchers are not focusing on publication and sharing.

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



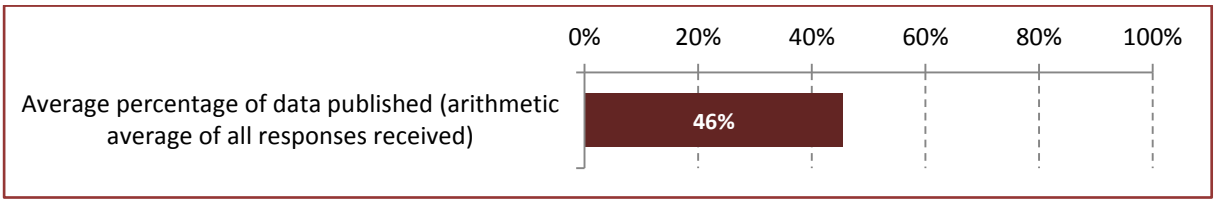
N = 517-547 per item (depending on number of respondents without answer)

		++	+	-	--	
	C.6	all or most projects	many projects	a few projects	not at all	N
a	In journal papers and conference proceedings (selected data)	291	160	75	18	544
b	In research reports (selected data)	265	157	95	22	539
c	In supplemental material	78	168	183	91	520
d	Through an institutional repository	68	99	143	211	521
e	Through a national data archive or centre	49	77	147	248	521
f	Through an international subject- or domain-based repository	32	53	125	306	516

The respondents were also asked to estimate the percentage of their research data which they deposit in a digital database or repository and which can then be accessed by other researchers who were not involved in the project. The average of all answers received from respondents was 46% (see Figure 6.2-21). This figure would imply that researchers, on average, share about half the data they generate in their research projects with others. The percentage is surprisingly high, considering that other evidence in the field rather points to a one-digit figure (i.e. less than 10% of research data is shared – see literature review, Section 3.2.4).

There are several explanations for this apparent discrepancy. First, researchers tend to have a broad notion of what is meant by "data", often including any type of publication (e.g. research reports, journal papers). Thus, if a research report from a project is deposited in an institutional repository, but the detailed data is not, they might consider this project as "deposited" and count it in. In other words, the question inevitably leaves some room for interpretation as to whether something is actually deposited (and shared) or not. Second, even if data is deposited in a certain way, it may actually not be directly re-useable by others (e.g. data tables presented in a PDF document). Again, some may then count it in while others may not, depending on the criteria.

Figure 6.2-21: Question C.7 – "Please estimate the percentage of your research data that is deposited in a digital database/repository and can be accessed by researchers not involved in the project."



The question received a considerable number of comments, which document the importance of different contexts and interests with regard to publishing and sharing data. One of the comments described very well a common approach and position to this issue which was also described in pilot interviews prior to the survey. "*Only selected research data is published. The basic data of an excavation (inventories, site description typology etc. = grey literature) is available to project members only. Other researchers have to ask permission to have access to data that are not their own!"* Academic researchers need to publish (as the academic career depends very much on the publication output), but at the same time, aim to protect their knowledge similarly as a company protects its IPR. Therefore, the detailed data are rather not disclosed and shared, as it represents a potential competitive advantage. It would go beyond this report to embark on the discussion of how to create incentives for sharing data, and on the many trends in this domain; in any case, the responses and comments received in this survey contain some interesting evidence on the conflicts between "sharing" and "protecting" research data.

Comments from respondents:

- *"I published several papers and I made presentations in Conferences because I want to publish some results of my project as a pre-view, while I am preparing the main publication that requires longer time"*
- *"Much of our work is conducted under a permit which requires reports, metadata, and some types of produced data to be submitted to the regulator, who makes them available to professionals and others with a 'need to know'. Theoretically, any bona fide researcher worldwide could access the data; but because it is not open to data crawlers, it won't show up in search engines. Comparatively little is published in traditional paper journals or books."*
- *"Hertfordshire Archaeology is a scholastic publication, produced about every 3 years: it is open to both amateur and professional contributors."*

- *"We are still trying to improve our publication through national sources; perhaps international sources will follow."*
- *"We still hold most of our data which means currently only we have access to it. We hope to deposit it soon but until then its availability is limited to us making it available by post or file sharing site."*
- *"Again, my new projects are not yet at the point of dissemination."*
- *"Most of my reports end up as sections in grey literature, and access to these depends upon the policy of the HER concerned."*
- *"A few years ago, the publications department of my former employer contracted with tDar to put data from the books online."*
- *"Only selected research data is published. The basic data of an excavation (inventories, site description typology etc. = grey literature) is available to project members only. Other researchers have to ask permission to have access to data that not their own! In Hungary archaeologists consider their data as their own even though projects were financed by the government and they carried out their work as employees. Therefore, it is very difficult to access complete site reports of other archaeologists."*
- *"no national data archive apart from the National Library and Australian Archives which have not shown a particular interest in archaeological data"*
- *"We use Edna and the Royal Library at The Hague for this"*
- *"Some of these I do not know if they apply at all?"*
- *"I am not entirely certain how much of the data we produce is made available through an international repository but, as most of it is publically available through our open access web pages and online databases, much of it can be searched, viewed and downloaded by international researchers."*
- *"Our (meta)data is not available through an open access, only the results (research report)."*
- *"We mainly give advice, we are not much producing data. We collect data and report."*
- *"Paper publication of a volume on the excavation or survey, with strat reports, GIS, catalogues etc. published on line, in a University-operated site."*

Barriers for depositing data

The last question in this module addressed barriers with regard to depositing their research data in digital repositories. The goal was to explore whether it is rather a lack of incentives or a lack of opportunity (e.g. due to technical barriers) that keep researchers from sharing their data with others. The respondents were confronted with nine items representing potential barriers and asked to assess the importance of each of them. The results show that all types of barriers are relevant – there is no clear ranking in terms of the number of researchers that are confronted with the respective challenge (see Figure 6.2-22). Three challenges seem to be particularly important, however:

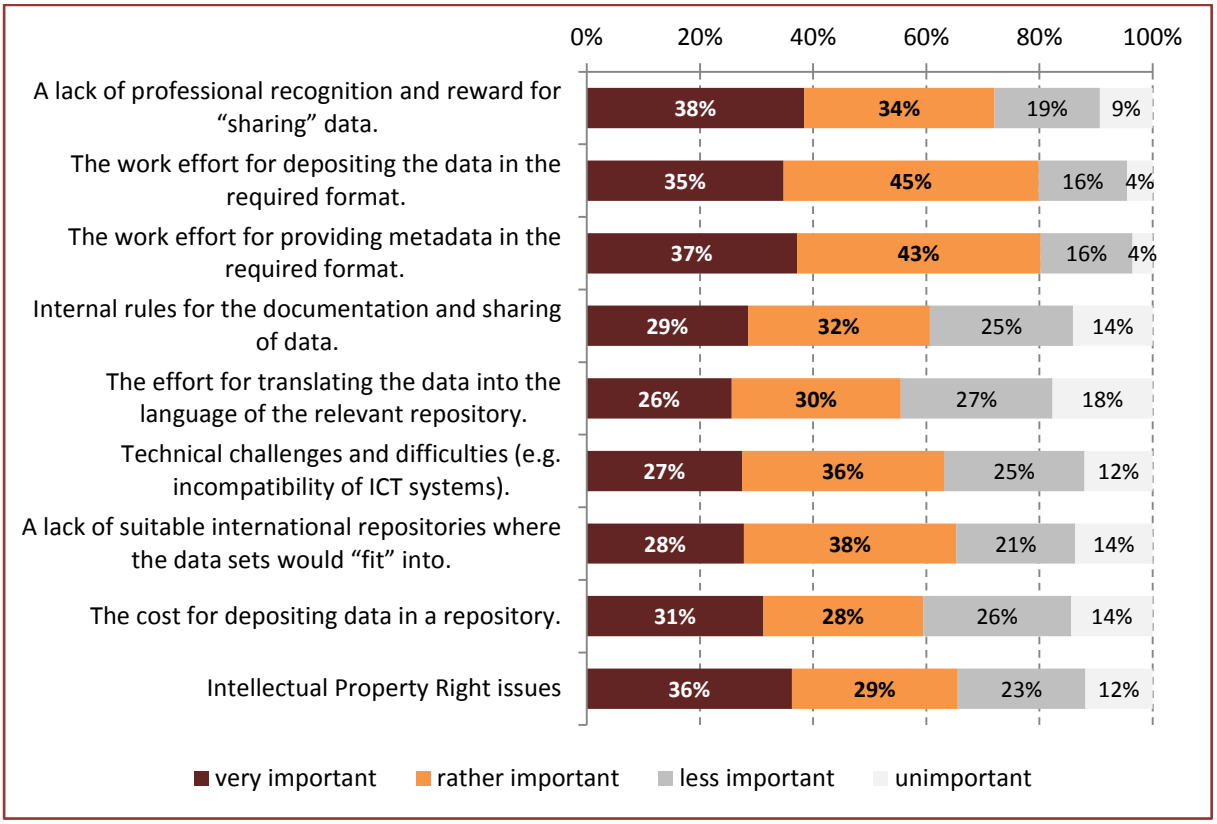
- The **work effort** required for preparing data in a way that it can be deposited in a repository (converting and archiving the data themselves and the metadata in the required format): about 80% of the respondents confirmed that this was a "very important" or "rather important" barrier.
- A **lack of professional recognition** and reward for sharing data with colleagues. About 70% said that this was a very or rather important barrier. A specific aspect here is that sharing may compromise the opportunity to produce further research and publications on the respective data (quote: "*I might use data for another analysis and publish it, before it can be shared with colleagues*").
- Challenges resulting from **Intellectual Property Rights issues** (75% very/rather important), for instance if the organisation conducting the research is not the owner of the results.

All of the other items, including "technical difficulties" and a "lack of suitable international repositories" were also confirmed as important by at least 50% of respondents each.

As with most of the other questions, respondents were given the opportunity to comment on specific aspects or identify further barriers. The comments received (see list after the graph) give some practical examples for the various barriers and help to better understand the organisational contexts. For instance, the comments highlight that reluctance to sharing data can be a cultural issue ("*in … archaeologists consider their data as their own and are not willing to make it available*"), possibly stemming from the risk of not receiving the adequate recognition ("*some scholars copy (or even steal) other researcher's data, from their books or their website without quoting the source, and this on the Internet is a serious problem")*. Comments also provide evidence for cost and time-related challenges ("*for digital repositories that are open to all to add data, the barrier is the cost (in time) not being funded through the project*")

All together, the results again confirm the validity of the ARIADNE project objectives, since the e-infrastructure to be developed by ARIADNE addresses, explicitly or implicitly, most of these barriers. Any improvements the ARIADNE services can make with regard to reducing these barriers should be welcomed by researchers who would, in principle, be willing to share data, but are kept from doing so as it is too time-consuming or complicated. While ARIADNE cannot be expected to have a direct impact on academic reward systems, the project can contribute to improving the technical framework conditions for sharing data in international research by offering an attractive platform. In the long run, these two fundamental pre-conditions for data sharing (professional recognition, technical opportunities) can be expected to reinforce each other. The better the technical framework conditions for linking and sharing data are, the more attractive it will become for researchers to "cooperate" in this way, even if sharing data maybe in conflict with personal academic interests from a short term perspective. ARIADNE can be framed from this perspective as an initiative that aims to support the trend (or even paradigm shift) towards sharing and collaboration in international archaeological research.

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



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

		++	+	-	--	
	C.9	very important	rather important	less important	not important	N
a	A lack of professional recognition and reward for "sharing" data.	194	170	94	47	505
b	The work effort for depositing the data in the required format.	178	231	80	23	512
c	The work effort for providing metadata in the required format.	188	218	82	18	506
d	Internal rules for the documentation and sharing of data.	142	160	126	70	498
e	The effort for translating the data into the language of the relevant repository.	129	150	135	89	503
f	Technical challenges and difficulties (e.g. incompatibility of ICT systems).	137	178	124	60	499
g	A lack of suitable international repositories where the data sets would "fit" into.	136	184	103	67	490
h	The cost for depositing data in a repository.	154	139	129	71	493
i	Intellectual Property Right issues	177	143	110	58	488

Comments from respondents:

- *"Unfortunately some scholars copy (or even steal) other researcher's data, from their books or their website without quoting the source, and this on the Internet is a serious problem."*
- *"Archaeologists generally seem to be afraid of technology, so they aren't very interested in the relevant issues: dissemination, data formatting, etc. Also very proprietary: this is my data, you can't see it; you can only see what I've published."*
- *"Mainly the connectivity to all repositories, including open source. Often publications are available in different places. Now I still have to use google to get most in view."*
- *"Clients may require confidentiality which restricts data access, although this is comparatively rare and, for much of the data that could possibly be included in a report, there are explicit requirements to ignore intellectual property rights. However, much of what we produce these days is GIS shapefiles, grids, geodatabases, 3D models, voxels, etc that can't readily be handled by the government repository."*
- *"For digital repositories that are open to all to add data, the barrier is the cost (in time) not being funded through the project. For digital depositories that are controlled by another person or organisation the barrier is the lack of recognition of our work – i.e. they seem very selective of the projects and the people or organisations they take data from. In the last case there is often a strong bias towards published rather than grey literature work."*
- *"The copyright issues are a barrier but this is something that we need to see an internal strategic change direction to a more OpenGovernment model."*
- *"Much of my data is photographs, interviews and field notes. I find it easy to forget that other researchers would find these things useful"*
- *"Free-lance sub-contractors analyse most of the material from archaeology, since almost all field archaeology is carried out by commercial CRM contractors. They therefore generate almost all the data. There is no centralized repository known for this data. The copyright and intellectual property rights to the analysis are signed over to the CRM contractor by the contract."*
- *"There isn't a coherent policy within my institution (that I know of) - if it were taken for granted that data would be shared at some point and institutional support was automatic, we'd do it"*
- *"All of these together with getting a brain trained in the old ways oriented to all the possibilities and new organizational systems"*
- *"These all strike me as significant issues in many cases."*
- *"In Hungary archaeologists consider their data as their own and are not willing to make it available. They are worried that someone else will publish something out of it before they do."*
- *"No barriers. It is required that the report(s) generated with data acquired from repository sources be submitted to the repository."*
- *"These answers are what I think, not necessarily how these barriers are perceived by those who are most affected (e.g. Intellectual Property is usually seen as a major obstacle or excuse, but that's nonsense). Also, most of the times just laziness and ignorance."*
- *"as a commercial firm, we encounter property right barriers. Our clients are not always interested in 'sharing' results and data."*

- *"barrier is also that I might use data for another analysis and publish it, before it can be shared with colleagues..."*
- *"Research needs: I have been data mining datasets and the effort involved in obtaining the data are considerable. We are still analysing these and will be looking for further research funds to support this work. As such this data forms an important part of the research and there is little incentive for me to deposit the data until I have finished with it, indeed it would be a backwards step. I do recognise this is a circular argument since the reason it took so much effort is the lack of free data depositories!!"*
- *"Archaeological data is a commercially important property, particularly where regional comparative datasets are involved."*
- *"Software and file format compatibility is a big issue for archiving digital materials"*
- *"Many professionals most able to make critical contributions to the distribution of data, i.e. those furthest along in their careers and professional reputation, are reluctant to, or unable to, acquire the digital skill sets needed to distribute that data. Younger researchers, who possess the skill sets needed to disseminate data in the digital realm are discouraged to do so based on established mores which support the reputation of their superiors."*
- *"I work primarily as a contractor for the US government and am not allowed to share data without clearance."*
- *"repositories are not well-known"*
- *"I love to visit libraries and archives! It is an excellent way to avoid sitting too many hours in front of a screen and makes it possible to travel to different libraries and archives national and international (and meeting colleagues). Young colleagues are sitting too isolated in their offices. To put too much on line would make the isolation even harder. You have to discuss the data with colleagues in order to understand it fully!"*

6.2.6 Specific needs and expectations towards ARIADNE

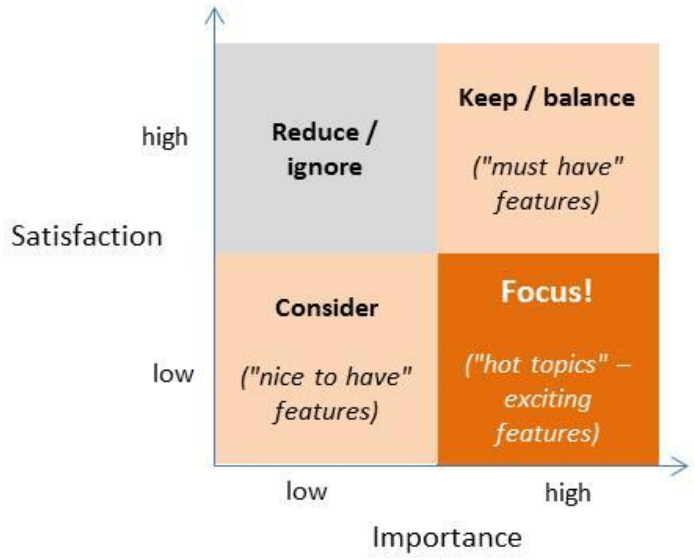
The last module of the questionnaire focused on specific user needs and expectations with regard to the type of services which the ARIADNE project might develop. The results of these questions shall help the project to set priorities in the selection and design of specific features (for instance when deciding on the allocation of resources, or when having to decide between different options).

Importance vs. satisfaction: the user requirements matrix

The first question of this survey module asked respondents to assess a number of potential user requirements (a) in terms of the importance they attribute to each requirement, and (b) in terms of their satisfaction with the current situation.

This is a common approach in user requirements analysis. Ideally, after having analysed the perceived importance and satisfaction with existing solutions, the various user requirements can be placed in one of four quadrants which suggest strategic responses how to priorities them when working on new or improved products and services. (see Figure 6.2-23). The baseline recommendation is that service providers should, in particular, focus on the "hot topics": these are those needs which are important to users, but which are not well catered for by the existing solutions. If they manage to provide useful solutions for those needs, they can make a real difference and achieve excitement among users.

Figure 6.2-23: User requirements matrix – grouping user needs according to their importance and the satisfaction with the existing solutions



The ARIADNE User Survey followed this approach in Question D.1, except that the list of requirements evaluated in that way focused on generic rather than very specific detailed user needs.[8] The restricted use of the approach was due to time constraints (the survey also had to cover other aspects such as current patterns of data use), as well as the situation that some basic decisions about the scope, focus and design of the ARIADNE services are still to be made. From a procedural view, this survey informs the ARIADNE project about *basic* user requirements; the project then has to take some basic decisions about what it is going to develop; on this basis, a second analysis of user requirements which focuses on much more specific user needs (with regard to the services) can be conducted.

[8] When analysing consumer needs with regard to a specific product, the first step is often to develop up a much more comprehensive list of many detailed, specific needs (consisting typically of about 50-100 items), which are then consolidated (e.g. by merging similar items) list a final list of about 20-30 items. These are then assessed in terms of their relative importance and satisfaction, as described above.

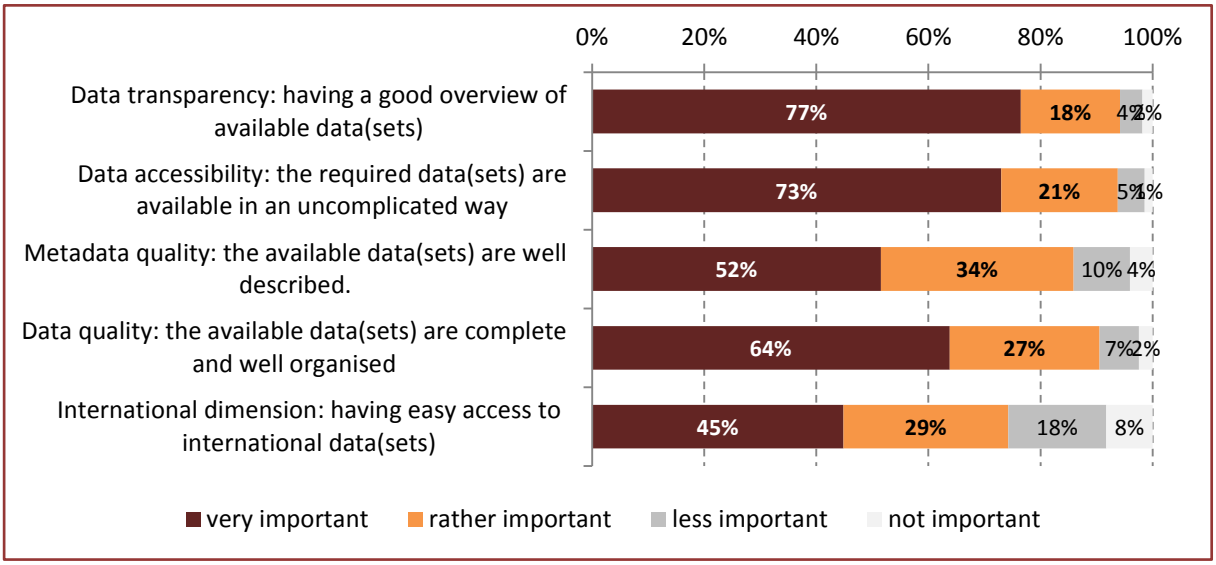
The importance of and satisfaction with various aspects of research data

For these reasons, the following (rather aggregated) user requirements were presented to respondents in the survey, asking them how important they were for conducting their research, and how satisfied they were with the current situation in this regard:

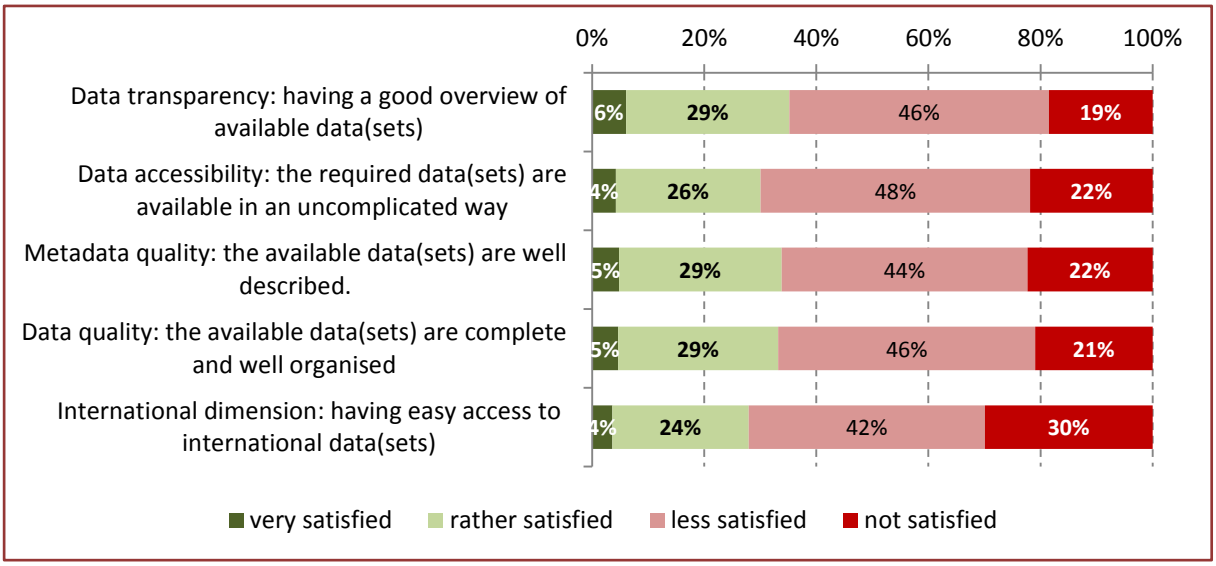
- Data transparency needs: having a good overview of available data(sets)
- Data accessibility needs: the required data(sets) are available in an uncomplicated way
- Metadata quality needs: the available data(sets) are well described
- Data quality needs in general: the available data(sets) are complete and well organised
- The need for an international dimension: having easy access to international data(sets)

Figure 6.2-24: Question D.1 – "Please say how important the following aspects are for you in order to conduct your research, and how satisfied you are with the current situation in this regard."

(a) Importance



(b) Satisfaction



N = 502-506 per item (depending on number of respondents without answer)

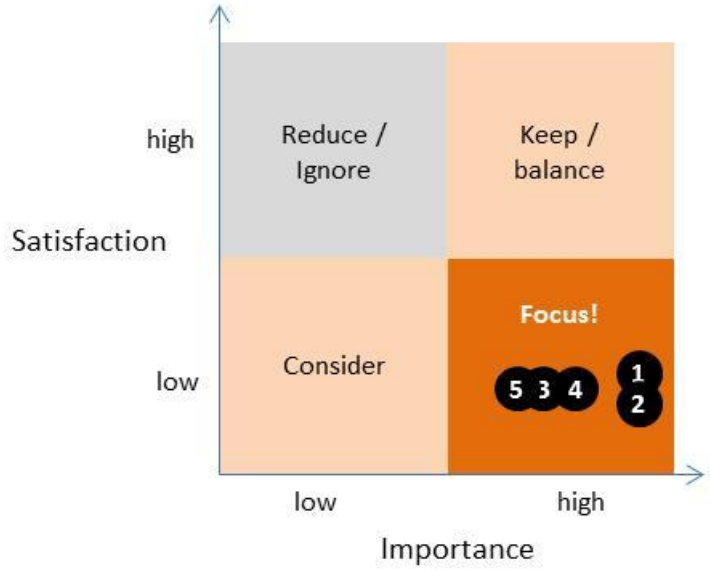
		++	+	-	--	
	D.1 – IMPORTANCE	very important	rather important	less important	not important	N
a	Data transparency: having a good overview of available data(sets)	381	88	20	9	497
b	Data accessibility: the required data(sets) are available in an uncomplicated way	365	104	24	7	499
c	Metadata quality: the available data(sets) are well described.	256	171	50	20	497
d	Data quality: the available data(sets) are complete and well organised	317	133	35	12	497
e	International dimension: having easy access to international data(sets)	223	146	87	41	498
	D.1 – SATISFACTION	very satisfied	rather satisfied	less satisfied	not satisfied	N
a	Data transparency: having a good overview of available data(sets)	30	145	230	92	497
b	Data accessibility: the required data(sets) are available in an uncomplicated way	21	129	240	109	499
c	Metadata quality: the available data(sets) are well described.	24	144	218	111	497
d	Data quality: the available data(sets) are complete and well organised	23	142	228	104	497
e	International dimension: having easy access to international data(sets)	18	121	210	149	498

The results are easy to summarise: all of these items are highly relevant for researchers (see Figure 6.2-24a), and in all of these areas, the researchers see important gaps between what they would ideally expect with regard to these aspects and the actual situation (compare with degrees of satisfaction, Figure 6.2-24b). Two requirements in particular were attributed the highest importance: data transparency and data accessibility. In both cases, about three quarters said that this was very important for their research.

On the one hand, these results are excellent news for the ARIADNE project, as they confirm that **the project rationale is highly relevant and very much to the point.** ARIADNE addresses user needs of the archaeological research community which are confirmed to be highly relevant (important for doing the actual work), while at the same time the satisfaction with the current situation is rather low. For each aspect explored, more than 60% of the respondents said that they were either "not satisfied" or "less satisfied" with the current situation. This implies that any improvement the ARIADNE project manages to achieve with regard to any of these dimensions should be highly appreciated. The "market potential" of ARIADNE services is therefore high.

However, the flipside of the coin is that the items covered by the survey were so highly aggregated. As all items are seen as (nearly) equally important by the target community, the results at this stage cannot provide much information to the project about specific issues and priorities in the definition of the services. When applying the "importance vs. satisfaction" tool as introduced above, it would mean that all of the five user requirements explored fall into the "high priority" quadrant of the matrix (see Figure 6.2-25): needs which are rated as very important and which are not yet properly addressed by the existing solutions in the market. Essentially, it means that ARIADNE has a broad field of opportunities. While it is clear that the project cannot solve all problems in these fields, it indicates the high potential which such initiatives have, given the perceived gap between the existing and expected level of solutions.

Figure 6.2-25: User requirements of archaeological researchers (user needs according to their importance and the satisfaction with the existing situation)



User needs as shown in the matrix:

(1) = Data transparency

(2) = Data accessibility

(3) = Metadata quality

(4) = Data quality

(5) = International dimension

It will be a major objective for the update of this report (after the second project year) to break down these five generic user requirements into more detailed, specific needs and to explore which of the specific needs (within each of the five dimensions) are the most relevant ones (see Section 2.3 – Plan for the update).[9] The time line for the sequence of these analyses was well-planned: the specific results will be available in due time when ARIADNE starts planning and developing the portfolio of services and its specific features (in project years 3 and 4).

Usefulness of potential ARIADNE services for researchers

In the second question of this module, the respondents were given a list of nine potential types of services which the ARIADNE project may consider developing, and asked to assess how useful each of these services would be for their own research. Fortunately, the results for this question are much more nuanced than the ones from the previous question, and they allow some preliminary recommendations for the framing of the ARIADNE services.

The results (see Figure 6.2-26) can be summarised as follows:

- Researchers would greatly appreciate (and probably expect from ARIADNE) a **portal functionality** which makes it more convenient for them to search for archaeological data

[9] D2.2 – Second Report on User Needs (a year after D2.1)

across different databases. Ideally, such a portal should not only link different data sources, but also offer innovative and more powerful search mechanisms.

- **Everything is useful** ("we take anything we can get"): As to be expected, the respondents collectively regarded none of the proposed services as "not useful". For each service, at least 70% stated that it would be "very useful" or "rather useful" to have it. This can be attributed to the pre-screening and selection of the proposed services which partly resulted from interviews with researchers, and to the fact that customers (users) are inclined to take all features of a product as long as it does not increase the price. In retrospect, it might have been useful to ask respondents, in addition to commenting on the usefulness, to also rank the various items (this would have forced them to establish a priority for what would be the most important services). Establishing such a ranking might be a goal for the update on this report.

The two top-rated services (in terms of the share of users saying that this would be "very helpful") are **search portal functionalities** which facilitate supra-institutional data search across the holdings of different databases or institutes. Nearly 80% of the respondents said that such a portal would be very helpful.

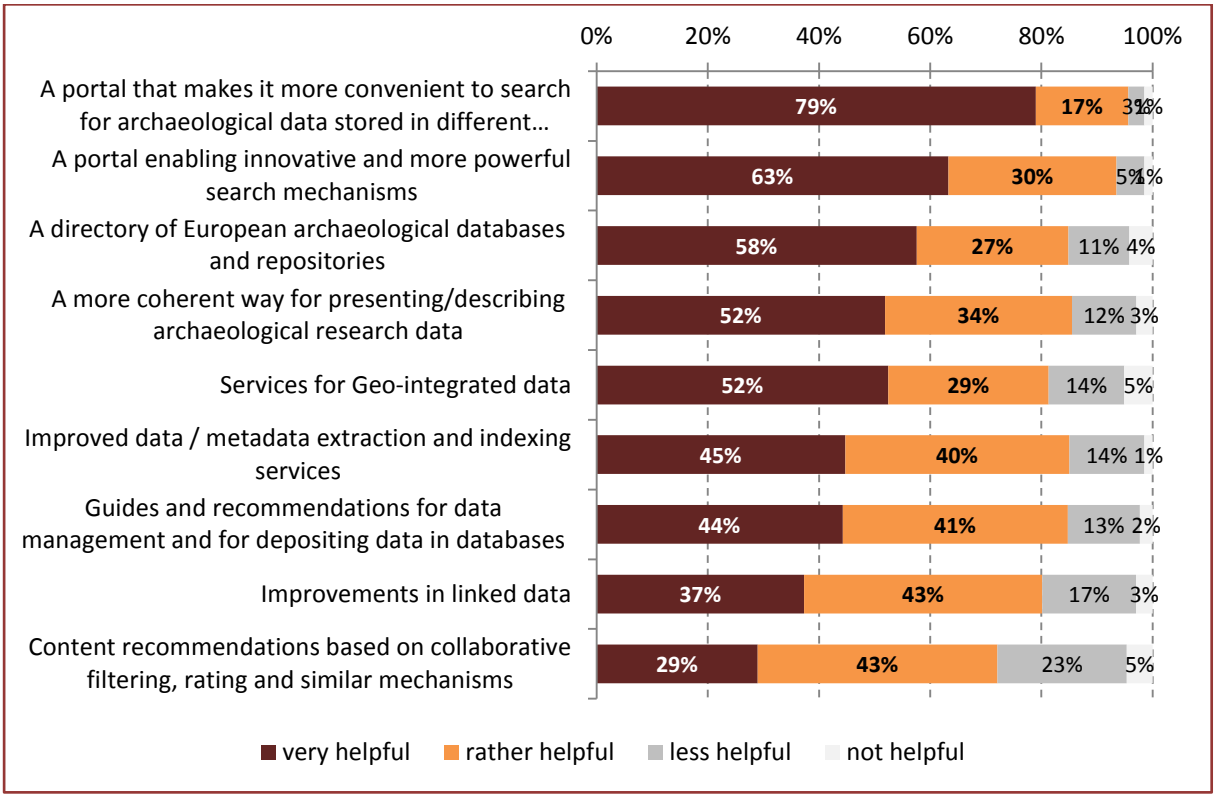
Without questioning that most users would probably greatly appreciate such a portal (if it is well designed and rich in content), it should be considered that the concept of a "web portal" is something which practically all respondents are familiar with. It is **easy-to-understand** and practical, while some of the other concepts and services proposed (such as "improved metadata extraction and indexing services") may be more difficult to understand for many. One of the comments made in the free-text fields supports the assumption that some of the items have probably been framed in a rather technical way and are not properly understood by all ("*Since I cannot access data from repositories, I have not learned the terms. What is metadata?*" "*Linked Data methods: term unclear to me*"). In fact, when looking at the results, one might note a certain correlation between the technical complexities of the concept (service) proposed and the aggregated level of desire for the respective service.

With this caveat in mind, the "collective ranking" of the services in terms of their usefulness is still interesting evidence for the ARIADNE project. For instance, when it comes to the decision of whether the project should focus only or rather on the backend integration of data sets, or whether a smart (front-end) portal with innovative search tools should also be envisaged, the responses obtained in the survey give a clear answer.

The service which came out third (after the two portal functionalities) in terms of having supporters who would find this "very useful" is to have a **directory of European archaeological databases and repositories**. Indirectly, this can also be seen as a portal functionality (if not linking data or metadata directly, it would at least be good to know which databases exist and what kind of data they offer). Moreover, it is also a service addressing the lack of transparency about what actually exists.

Interestingly, the service which received the lowest number of "very useful" ratings (while still being seen as useful by a majority) was the proposal to create a mechanism for using "crowd intelligence" in the research community: having content recommendations based on collaborative filtering or rating mechanisms. Some of the existing repositories (e.g. DANS-EASY) are currently going that way and aim to establish a rating system where actual users rate the quality and usefulness of data contained in the database, similarly to the peer review systems that have become an important feature, for instance, of online hotel reservation systems. Also, the need for improvements in linked data is probably still in an early phase. About a third of the respondents said that this would be very helpful. It might be an issue to be further explored if the comparatively lower appreciation is due to the fact that services are not yet widely developed, and the concept therefore not well understood, or whether it is actually a feature that is just not important to many people.

Figure 6.2-26: Question D.2 – "To what extent would your own research (or the research of your institute) benefit from the following potential services or improvements (as to be provided by ARIADNE)?"



N = 471-481 (depending on number of respondents without answer)

		++	+	-	--	
	D.2	very helpful	rather helpful	less helpful	not helpful	N
a	A portal that makes it more convenient to search for archaeological data stored in different databases	380	80	14	7	481
b	A portal enabling innovative and more powerful search mechanisms	302	144	24	7	477
c	A directory of European archaeological databases and repositories	274	130	52	20	476
d	A more coherent way for presenting/describing archaeological research data	247	160	55	14	476
e	Guides and recommendations for data management and for depositing data in databases	212	194	62	11	479
f	Improved data/metadata extraction and indexing services	212	191	64	7	474
g	Improvements in linked data	177	203	80	14	474
h	Content recommendations based on collaborative filtering, rating and similar mechanisms	137	204	110	22	473
i	Services for Geo-integrated data	247	136	64	24	471

Specific comments and suggestions

Again, quite a lot of respondents made use of the opportunity to propose further services (or post a comment) in a text box after having rated the listed services. Their suggestions are listed below. These comments, as in in other sections of the survey, can be seen as a kind of crowd-sourcing for ideas. They contain a lot of specific needs, but also ideas for solutions as well as information about already existing services that come close to those proposed in the questionnaire. All these comments are "food for thought" – they will help to draw up the more detailed list of needs (and possible solutions) that will then be explored in update of this report (as interesting as some of these ideas from individual respondents may be, it is recommended to explore to what extent they are shared by others).

The comments also show the importance of going into more detail to properly understand user requirements. For example, two comments highlight that a "portal" is only useful for researchers if it comes with an API (Application Programme Interface), which allows the programmatic extraction of metadata. This demonstrates that different user segments have not only different expectations with regard to the type of services to be provided, but also with regard to the depth of the services. Such observations are particularly useful and important, as they indicate the specific needs (behind the general need of getting a better overview of data).

Comments and suggestions from respondents:

- *"A Thesaurus of Terms that creates a common language for scholars, in defining objects or artefacts"*
- *"Harmonisation of metadata: CIDOC-CRM is fine. Concerning "a directory of European archaeological databases and repositories with information about their holdings and conditions for access": That's easy enough to find. "Guides and recommendations for data management and for depositing data in databases": ADS had some guides to good practice."*
- *"Since I cannot access data from repositories, I have not learned the terms. What is metadata?"*
- *"My research is primarily in west and central Asia, but many of the researchers are British and European; I'd like to see geographical expansion!"*
- *"Not sure I understood all of the above. How about an Idiot's Guide to making archaeological data accessible?"*
- *"A portal is OK, but more or less redundant if there isn't an API to go with it."*
- *"Guides and recommendation (protocol) what an entered data (e.g. artefact description, petrographic description etc.) should include in order to avoid incomplete descriptions"*
- *"I do not primarily work in European archaeology."*
- *"Please no new systems; try to use existing programs like access (Microsoft) of the Dutch (but more and more international) Archeolink."*
- *"An API to allow programmatic extraction of metadata."*
- *"All of these services and improvements would be of dramatic benefit."*
- *"recommendations for archiving raw & processed geophysical data"*
- *"A way of searching for 'reports within reports' in professional reports would be extremely useful. For example, through the ADS grey literature library or the Archaeological Investigations Project (AIP) of Bournemouth University, searches can easily be made by title to pull out evaluations, watching briefs, historic building recording reports etc, but it is very difficult to 'see' the specialist reports that often appear as appendices within these reports. Through my role I have been involved in a brief scoping exercise to see whether we could easily*

identify the body of plough zone archaeology (field walking, geochemical analysis, legal metal-detecting etc.) research that has taken place through, and been published by, professional archaeological units. These elements are almost always published or reported within the larger body of work of and overarching project making them hard to find or quantify without prior knowledge that the work had been carried out."

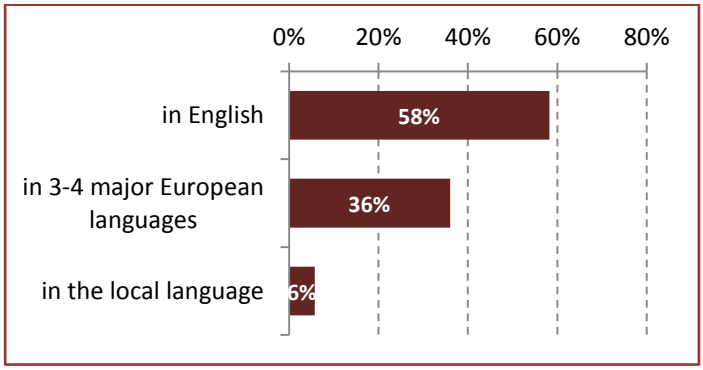
- *"Linked Data methods: term unclear to me"*
- *"The most important is that the data can be found and relevant people, projects and sources contacted and consulted. That is the paramount obstruction that needs to be mitigated for all research, as not knowing what is out there stops any progress in any search. All the rest is essentially a bonus and standardisation is probably unachievable and less of a priority if optimal 'findability' is emphasised. The use of any dataset will require the user to do work to make it suitable and/or fit for their purposes and it is to be expected one has to work through the particulars of the project(s) that generated the data. This initiative should urgently go beyond Europe and hopefully increasingly include commercially obtained data. Enabling worldwide comprehensive map based, geographical searches are of a principal concern!"*
- *"Project by project archive costs are not sustainable - long-term institutional access to national facilities is important. Also, the role of the traditional, paper archive - 50% of our data are born digital and we have no legacy digital data now to manage. The cost barrier is in the digitisation and conservation of paper archive. National regimes in public archaeology also need to be considered – much of Europe is operating a public archaeology system that is essentially non-digital."*
- *"All this is important. But still, the most important thing is the way data is generated, not how it is made available"*

Language requirements for metadata

Respondents were also asked about the language requirements for ARIADNE services, in particular when providing metadata. A majority of the respondents (close to 60%) felt that metadata in English would be sufficient, while about a third argued that data should be available in 3-4 major European languages. Only few said that it was required in the local language. This reflects, on the one hand, the international dimension of research (those researchers engaging in international data search can be expected to have at least a basic command of English), but, on the other hand, also the dominant position of the English research community in the sample (see also Section 6.2.4, Question B.8 – language issues were not seen as an important barrier when searching for data).

Figure 6.2-27: Question D.3 – "Supposing the ARIADNE project manages to integrate metadata from different databases. In which language(s) would this information have to be available so that it is useful for you?"

	D.3	#	%
a	English	282	58%
b	3-4 major European languages	174	36%
c	local language	28	6%



N = 484

Suggestions for the ARIADNE project

The last question of the survey invited respondents to make further suggestions or comments on any of the issues that were raised by the survey in a free-text box. Quite a lot of respondents made use of this opportunity. Their suggestions are listed below. The quotations have not been edited (except for corrections of spelling errors or minor linguistic corrections). The list focuses on those suggestions which have practical relevance for the design of the ARIADNE e-infrastructure; comments which referred predominantly to mere technical issues of the online survey are not quoted. Some positive comments about the ARIADNE initiative as such are included as anecdotal evidence that the initiative is welcome among the target community. In synopsis, the suggestions obtained could be grouped into the following four categories:

Area	Examples
General recommendations	• more effort needs to be spent understanding the data creation process (…) • "the only way to create something useful in this field is ... to make something useful (user-friendly, user-oriented, easy, functional, self-evident)" • "open ALL DATA, give space for storage of data and for GIS"
Specific technical recommendations	• basic structure of ARIADNE should follow the CIDOC-CRM to facilitate multilingual availability of metadata • search words should be made easier and with a number of subcategories to choose from
Comments on challenges that need to be addressed and on risks	• The biggest challenge outside Academia is Paywall access • "The biggest issue I see is ignorance about data archives (still seen as a "technical thing" by most colleagues)" • "… bear in mind the very different motivators and obstacles for archaeologists working in Heritage Management and those working in Research or Educational Institutions …" • "… some European or international aggregators are attempting to create exclusive fiefdoms of big data that lock data sets in with proprietary structures and metadata" • "Data management is very complicated and expensive" • archaeological research data is becoming more and more complex

Comments and suggestions obtained

(E.3) Before closing the questionnaire: would you like to make any further comments on related issues (i.e. the access and management of archaeological research data)?

- *"Multilingual metadata shouldn't be very difficult to do if the basic structure follows the CIDOC-CRM."*
- *"Will become more and more important as data becomes more sophisticated and less possible to: a) print it on paper; b) archive it for posterity"*
- *"Thanks for this ARIADNE initiative; if it can get into operation, it will be a tremendous asset."*
- *"More effort needs to be spent understanding the data creation process. For example, how we capture data that is then made digital through a variety of means, how its quality is ensured, the resource required and the impact on the archaeologists. This requires a great deal of change and we are struggling to come to terms with this. There are efficiency impacts to getting data into digital systems and there are few ways around this but we need to make sure the benefits are made clear and the quality is ensured through effective resourcing."*

- *"Archaeological research data is becoming more and more complex, projects like this are increasingly important"*
- *"Here's an example. I found a type of medieval pilgrim-badge, and thought it would be useful to generate a national catalogue of examples, and an analysis. To scope whether this would be a worthwhile project and apply for funding, I tried to find out approximately how many there were. There are 200+ museums that hold excavation archives; there is no central national database for any artefacts, almost none of the museums have individual databases online, and the largest archaeological archive (in the capital city) assured me they had no examples, despite two catalogues published by that museum-archive service describing their examples. The basic info required to find out whether a topic is worth researching is unavailable: the door on archaeological research is slammed shut at the first attempt to look at other's data."*
- *"If the project succeeds in its aims, perhaps all government-, EU- and developer-funded support for archaeology should be conditional on results being deposited in the appropriate place? Perhaps with embargos for a certain length of time being possible for some types of data on request. In the long run, openness and collaboration should be beneficial for everyone."*
- *"My observation is that in the US there is very little institutional support for archaeology, and even less for archaeobotany. There is no organization with the financial and personnel resources to provide for long-term data storage. Maybe if technology change slows down, the need to constantly upgrade hardware and software will diminish."*
- *"The issues and questions posed have little relevance to the professional practice of archaeology in the U.S. Suggestion: add questions regarding qualifications or certifications of individuals for professional status."*
- *"I appreciate this is of less relevance to Australia, but the possible improvements to archaeological data management elsewhere in the world might trickle down to assess us in the future, and also provide a model to lobby our own archaeological institutes and authorities with"*
- *"The biggest issue I see is ignorance about data archives (still seen as a "technical thing" by most colleagues, rather than a foundational layer for all our work). The availability of guides to good practice is, sadly, not enough."*
- *"Thank you for explicitly including independent researchers in this survey - I have had trouble filling out surveys before as they were just not geared towards including the odd indies!"*
- *"It is commendable that the questionnaire had the option to make additional comments after every question and also that it did not force answers. On the other hand, as an independent archaeologist employed at a university to do non-archaeological work but doing archaeological work with a museum as a volunteer, I was not always sure how to answer."*
- *"The major barrier to independent archaeologists for deposition of data is cost. The will is definitely there."*
- *"Long term sustainability of ADS seems still to be an issue in the UK"*
- *"I think it is important as ARIADNE develops its strategies to bear in mind the very different motivators and obstacles for archaeologists working in Heritage Management and those working in Research or Educational Institutions. In Research and Educational Institutions I see the relatively low reward for making data accessible and managing it well as a major barrier. In Heritage Management the obstacle is more making the data well structured while remaining flexible enough to meet the needs of a wide variety of projects. While it is clear that access and management issues in these different institutional settings are related and impact on one*

another, I suspect the strategies devised by ARIADNE will need to be oriented in different ways to best serve these groups, or be relatively flexible."

- *"'rules' and recommendations for data submission for storage should not be so difficult that independent workers would not have the time or expertise to submit their data. Data should be stored securely with length of intended storage indicated (so that data will not be lost from units that have closed for business)."*
- *"The biggest challenge outside Academia is Paywall access (so called open access) to scientific and academic literature and periodicals. It re-enforces an artificial exclusivity of academia and prevents access to learning and research for the majority of the population. The monopoly of the major publishers must end. I cannot afford to pay EU30-50 *per paper* for hundreds of papers, of EU Thousands for eJournal subscription."*
- *"As an independent researcher, with no academic affiliation and no standing within the archaeological community, there are a number of barriers in getting my research out to the international community. 1) Getting permission to use the data and images used in my papers. 2) That every journal requires articles etc. to be in a different format. 3) Lastly and perhaps most importantly, since my research has been rejected out of hand by the hundreds of archaeologists whom I've contact personally, as it invalidates much if not all of their work with respect to Neolithic petroglyphs, getting my work past a peer review panel is impossible."*
- *"Es importante que incluyan el idioma Español, pues si bien no es un idioma de los "países centrales", sí es uno de los más hablados y existe una comunidad científica muy importante en América Latina que siempre queda excluida."*
- *"Access to archaeological data is currently very poor, patchy and rather ad hoc. If archaeology is to contribute to large scale problems this needs to be addressed. Even a list of databases and sources from a single source would a step in the right direction, there seem to be many of them! I have one such source: http://www.chrono.qub.ac.uk/instar/. Here you will find an archaeobotanical database."*
- *"Data mining: term unclear to me. A serious obstacle to work with data bases in archaeology is the wide array of typological terminology in use for artefact types - often not even uniform in one and the same country in one and the same language."*
- *"In (commercial) archaeology not enough time and money is spent on making data set available. On the contrary, in some cases information (publications, datasets, artefacts, etc.) is deliberately not shared for various reasons. This is caused mainly by the Dutch neoliberal policy and the commercialization of science."*
- *"Pattern very varied. Data in state sector impossible to obtain (i.e., English Heritage)."*
- *"To my view, the only way to create something useful in this field is... to make something useful (user-friendly, user-oriented, easy, functional, self-evident,...). If the costs of going public and open (preparing your data, translating them, being skilled in ontologies, metadata, data models and semantic reasoning, reading a 100 pages manual for learning how to access external data, accessing a web site that is overloaded,...) are higher than the benefits, nobody will go that way!"*
- *"It could be said that some European or international aggregators are attempting to create exclusive fiefdoms of big data that lock data sets in with proprietary structures and metadata for continual large scale funding and maintenance grants. The aim should be facilitating the linking of databases and portals."*
- *"Very good job!"*

- *"Remember the problems that have the Latin American research centres to disseminate data in Spanish and difficult to translate to English or any other language, or to use other language data."*
- *"There is no doubt technically speaking a lot of things are possible, but linking the content of several national databases together, let alone preparing them for data-mining by interested third parties will proof to be a Gargantuan undertaking... I wish the people of ARIADNE lots of success in their effort!"*
- *"I think that this survey should also be made in other languages (German, Italian, French...). I think it is not really easy for people who don't work every day with data management and archiving to understand all the details."*
- *"The main objectives of Ariadne project are important to develop archaeological research in European Union. I hope Ariadne Project achieve some objectives."*
- *"Search words should be made easier and with a number of subcategories to choose from and so on. This because I'm sometimes at a loss of words and I don't want to use the dictionary every time."*
- *"In order to satisfy the economic limitations of management and research of cultural heritage has become more and more complex. Unfortunately, most of the data in either dispersed, unavailable, scarce, payable, has limited access, is available in different languages or is inaccessible in any other way. A database that would provide a more accessible data, would not only contribute to the scientific research but would also help to follow the economic restrictions of applied projects."*
- *"You should include DAACS in your list of databases."*
- *"Thanks for your effort and i hope something good will come out of this project :D*
- *The ARIADNE project addresses major issues of archaeological data. Many archaeologists are waiting for the results of this project."*
- *"Data management is very complicated and expensive. No one at my institution is paid to deal with it and it often falls on unpaid volunteers to put it together, Researchers often don't want to share data."*
- *"We would welcome a wider and more accessible range of data, in the UK this is restricted by copyright restrictions both for map data and for air photographs. Greater pan-EU agreements to encourage greater open access for such data would be most welcome."*
- *"Before any excavation I took part in the reconnaissance, preparation and mark of the field, which will be performed digging. During the excavation, I'm always ready to various weather conditions and also very paying attention to detail and careful editing of technical documentation for a better presentation to the public."*
- *"Even services providing content that we might not find directly relevant to our work establish standards and methods we can point to as models and to make the case for improvements to our data curation"*
- *"Researchers should have and publish their own results and not summarize the results from others!"*
- *"Very interesting project. Is there any way to further collaborate with it?"*
- *"We generate, in public archaeology, a lot of archaeological survey data. We wish to make this fully available, but doing so has long term costs, for which we need financial support."*

- *"There is a tension between wanting to make my research data available and also having to 'keep something back' for the next project. Attracting research funding is a crucial part of my job, and I am under pressure to continually bring money in. This conflicts with my ethical desire to make research data freely available, particularly when data collection and analysis have been publicly funded."*
- *"I think data is overrated in the field in which I work. What matters are ideas."*
- *"Open ALL DATA, Give space for storage of Data and For GIS. Open right for antique maps. Announce who make what and what is made."*
- *"Main problems are : - the costs of online access to scientific literature (our institute is not subscribing online-access to most journals and the depositories of the publishers) - the poor quality and doubtful quality of many open access depositories and online-publishers (a lot of "predatory publishers" are developing without reliable control of scientific content - the protection of archaeological sites and objects when archaeological data and maps are freely available by online-depositories, which can also be used by people destroying sites, their scientific context and stealing archaeological objects (detectorists)."*
- *"too much time to complete data bases to put them on open access."*
- *"Given the tremendous variability of archaeological research data standardization should be type and/or context etc. specific."*
- *"This is a splendid initiative - and I hope that it will reap the fruits we all wish!"*
- *"One of the greatest hurdles we face, in my opinion, is the reluctance to share archaeological data with the wider scholarly community. So often this results in information being lost."*
- *"Most info for getting data requires a payment of fees."*
- *"ARIADNE is a great opportunity."*
- *"Avec le développement de l'archéologie préventive et l'explosion du nombre de découvertes, la gestion de bases de données est devenu un des enjeux majeurs de la recherche archéologique."*

6.3 Results - Part II: Managers of data repositories

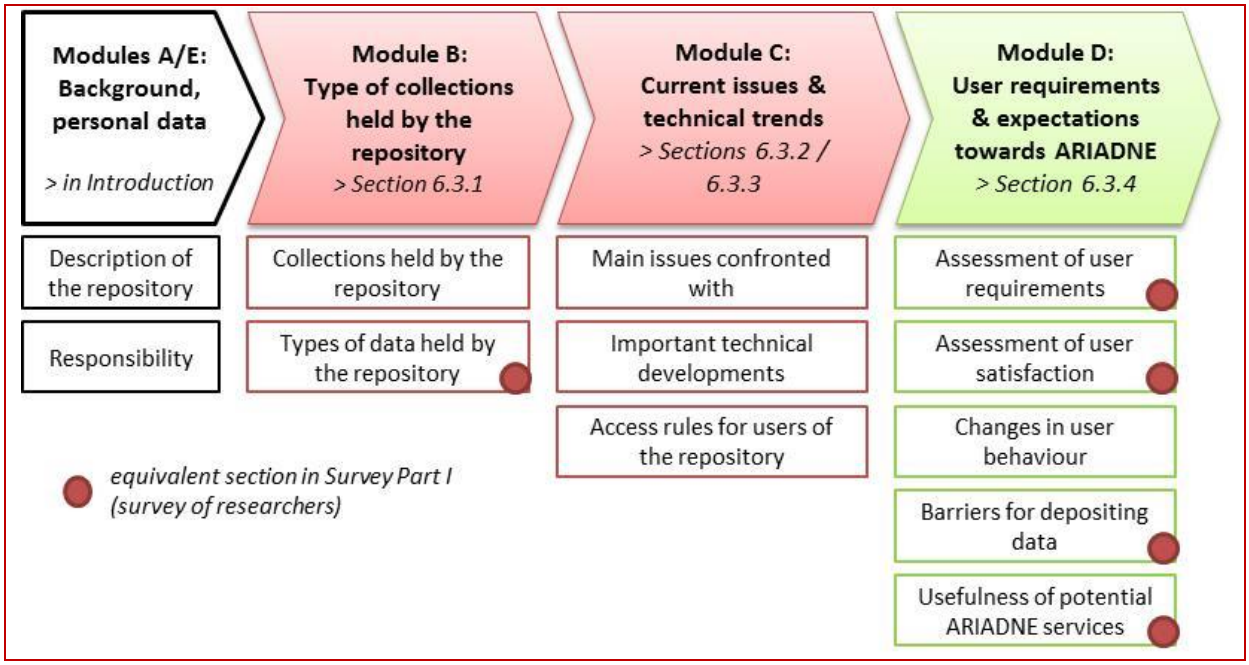
The second target group of the survey were managers of data repositories or data centres. These are institutions or initiatives which collect and maintain databases which cover (at least inter alia) archaeological research data. An overview of the different types of repositories that were reached with the survey is given in Section 6.3.1. All data repositories are important intermediaries (and, to some extent, gatekeepers) for making research data available to the community at large.

The term "managers of data repositories" is necessarily a simplification. The professional responsibilities of the respondents have within their organisations can differ, but typically involve activities such as database administration, maintenance or development. Many of the respondents described their position with terms such as "curator", "database administrator", "project manager" or "data manager" (see Section 6.3.1). The sample of the ARIADNE survey includes, in total, 52 representatives from this target community. Most of them represent institutional data repositories (47), but the sample also includes five representatives of domain or subject-based repositories which hold contents from several archaeological subject areas or institutes. Due to the small sample size, the percentages in the figures and tables should be regarded as indicative, as the statistical confidence interval is considerable. For example, a result of "45%" should rather be read as "about half of".

Structure of the questionnaire

For this target group, a specific questionnaire was developed which differs considerably from the questionnaire for researchers (Part I). Some of the questions, however, let respondents choose from the same items as an equivalent question in the researchers' part of the survey (see red dots in Figure 6.3-1). This makes it possible to compare responses from two perspectives (the "user perspective" of the researchers and the "curator perspective" of the repository managers). The first part of the survey focused on obtaining information about the repositories reached with the survey (including a brief free text description and information about the types of collections and data held). The second part explored current issues the repository managers see themselves confronted with and which technical trends they consider as particularly important. Finally, the main part of the survey asked for their assessment of the needs of their "customers", how these needs are changing and what they expect from the ARIADNE project.

Figure 6.3-1: Structure of the ARIADNE User Survey of researchers



6.3.1 Data repositories represented and responsibility of the respondents

Data repositories represented in the sample

The respondents were asked to briefly describe (in a free text field) the data centre or repository they represent and their own role. The following list contains an overview of the repositories that are represented in the survey, sorted by country (for reasons of confidentiality, the roles of the respondents are not disclosed). The responses account for 41 out of 52 respondents. The remaining 11 respondents either have not given any information, or the information does not provide a clear picture of what kind of institution they represent.

The sample of known entities comprises project repositories/databases (e.g. regional or city level, single site, digital corpus of artefacts, etc.); single institutes (research centres, museums and other); supra-institutional data centres, and heritage authorities and related services at county, province or national levels.

The overview shows that the sample is dominated by responses from the UK. Other countries represented with several responses are Sweden, Italy and The Netherlands. The sample focus on these countries does not only reflect the structure of the ARIADNE consortia, but also the fact that these countries taking a leading position in this field. There are also a few responses from non-European countries (USA, Australia). Thus, while not claiming to have a fully representative sample (which is nearly impossible for a small survey among 52 respondents), the institutions covered by the survey are a good representation of the variety of organisations and initiatives in this field.

Figure 6.3-2: Data repositories and initiatives represented in the sample (descriptions as given by the respondents, partly complemented by information from the quoted websites)

Country	Repositories
Greece	The Documentation Office of the Byzantine Museum, in Athens, is the central point for the management and organization of the digital information preserved and created inside the museum. The DO is responsible for the digital curation of the archaeological collections (more than 12,000 artefacts), the historical and photographic archives (more than 20.000 artefacts) as well as the conservation archives (more than 5000 documents).
Italy	The MOD (Mappa Open Data www.mappaproject.org/mod) is an open archaeological data archive for Italian archaeology. It's managed by the University of Pisa - Mappa Lab. The MOD contains raw data and grey literature from professional and research archaeology (we theoretically reject this distinction because we consider the archaeological practice always as a research activity).
	MiBACT/SSBAR/SITAR Project repository at http://sitar.archeoroma.beniculturali.it
	archaeological WebGIS called AIS, Archaeological Information System, with databases http://laboratoriobagolini.it/ais/
Netherlands	National Scientific archive where the e-depot for Dutch Archaeologists is based: https://easy.dans.knaw.nl/ui/home
	Archaeological finds and documentation of archaeological surveys in the Province of North-Brabant (The Netherlands), the archaeological surveys of five municipalities (Eindhoven, Helmond, Breda, 's-Hertogenbosch and Bergen op Zoom) excluded. URL: http://www.brabant.nl/dossiers/dossiers-op-thema/cultuur/cultuur-toen/grootschalige-erfgoedcomplexen/archeologie/provinciaal-depot-bodemvondsten.aspx
Slovenia	Libera is a database incorporating primarily archaeological literature. It encompasses the time ranging from the 5/6th century till the 10/11th century in the region of Europe, the Near East and North Africa. It consists only of publications in the library of the Institute of Archaeology (Scientific Research Center at the Slovene Academy of Sciences and Arts) in Ljubljana, Slovenia. In this moment it has over 40 000 records. http://zrcalo1.zrc-sazu.si/libera/lang_en/predstavitev.htm

Sweden	I am a dendrochronology software developer doing also dendrochronology research together with my wife
	Database for the Bolin Centre for Climate Research, http://bolin.su.se/data/
	Swedish National Data Service (SND)
	National aggregator for ancient monuments, protected buildings, photographs, and museum collections (http://www.ksamsok.se/in-english) and closely associated with national register for ancient monuments (http://www.fmis.raa.se/cocoon/fornsok/search.html) and protected buildings (http://www.bebyggelseregistret.raa.se/bbr2/sok/search.raa). Leading development of linked open data at the National Heritage Board and forming a node in Swedish cultural heritage sector.
	Jonkopings county museum with archives and archaeology items from the area
Switzerland	Archives of the Swiss School of archaeology in Greece, mainly datasets from 50 years of excavations at Eretria, Euboea www.unil.ch/esag
	The Swiss Data- and Service Center for research data in Humanities is a national pilot project for a nationwide data repository for humanities
UK	City of York Council Historic Environment Record www.york.gov.uk
	UK (Gloucestershire) historic environment record
	Bath & North East Somerset Historic Environment Record: a SQL database of archaeological Sites & Monuments, Events, and Finds plus a physical reports archive for works carried out since 1996 (when the former county of Avon split into 4 unitary authorities, including BANES) http://www.bathnes.gov.uk/services/tourism-and-heritage/archaeology/sites-and-monuments-record
	The database and GIS mapping that I manage is the Wiltshire and Swindon Historic Environment Record. It records archaeological and historical site information for the administrative areas of Wiltshire Council and Swindon Borough Council.
	Our collections contains all know material recovered from Fishbourne Roman Palace, West Sussex, and the area of 1 mile surrounding the site since 1960, including associated paper and photographic archive.
	Lancashire County Council's Historic Environment Record http://www.lancashire.gov.uk/corporate/web/index.asp?siteid=4398&pageid=19836
	Cambridgeshire HER http://www.cambridgeshire.gov.uk/leisure/archaeology/record/
	Herefordshire Historic Environment Record
	Gwynedd Historic Environment Record
	University of Aberdeen Museums include archaeological collections from Scotland, Latin America, Egypt and the Mediterranean. Active collecting is restricted to archaeological material from North-East Scotland acquired through the Treasure Trove procedure.
	Surrey Historic Environment Record
	Historic Scotland manage a service to provide up to date information on the protected historic environment on behalf of the Scottish Government. http://data.historic-scotland.gov.uk/
	Manx National Heritage Library & Archive Service. The collections include archives of a number of archaeological excavations that have been conducted on the Isle of Man. http://www.manxnationalheritage.im/what-we-do/our-collections/library-archives/
	The records of the Archaeological Survey of Ireland, National Monuments Service, Department of Arts, Heritage and the Gaeltacht, underpin the statutory list of protected monuments under Section 12 of the National Monuments Acts 1930 to 2004. http://webgis.archaeology.ie/NationalMonuments/FlexViewer/
	Historic Environment Record (HER) for Buckinghamshire. See http://www.buckscc.gov.uk/leisure-and-culture/archaeology/historic-environment-records/
	Humber SMR (serving Hull and the East Riding of Yorkshire)
	http://www.brighton-hove-rpml.org.uk/Pages/home.aspx

	A museum service with substantial archaeological collections
	Digital Corpus of Argonne Ware sequences stamps
Australia	A digital repository for the archiving and redistribution of the data outputs of research projects: http://datacommons.anu,edu.au (for the service) http://itservices.anu.edu.au/projects/completed/anu-data-commons/ (for the background and documentation)
	Federated Archaeological Information Management Systems Project, http://www.fedarch.org The FAIMS Project is a $950k eResearch Tools project, funded by the National eResearch Collaboration Tools and Resources (NeCTAR) program. NeCTAR is an Australian Government program to build new infrastructure for Australian researchers, conducted as part of the Super Science initiative and financed by the Education Investment Fund.
USA	A municipal repository in the United States.
(unknown)	Central repository and database (GIS) of research reports (grey literature), database on known sites (and GIS)
	Mostly digital archive, database of past (and present) archaeological research projects (digitised reports and mapped research areas), database of known sites (digitizing), database of lidar and aerial photographs data
	I created and manage an archaeological catalogue relational database. I have created databases within my institution that link all of the project and artefact information.
	Collections Management database, digital images of artworks, database for archives

Professional role of the respondents

As indicated in the introduction, the professional responsibilities the respondents have in their organisations can differ, but typically involve activities such as database administration, maintenance or development. Many described their position and professional responsibility either as "curator", "database administrator", "project manager" or "data manager" (or similar). Some representative examples how the respondents describe their role are the following:[10]

- *"I manage this record - including data entry, answering enquiries, database enhancement"*
- *"I am the curator - managing the preservation, documentation and accessibility of the collection."*
- *"Database Administrator and principle Archivist. My role also includes community outreach and liaison with educational institutions."*
- *"Head of (…), so responsible for all aspects acquisition, care and access."*
- *"Database coordinator"*
- *"Project manager/architect"*
- *"Manager of the Archive"*
- *"Responsible for managing the database and computing systems, the physical archive/library."*
- *"I manage the library & archive service for (…)"*
- *"I'm a manager of the repository; my functions concern with data storage and 'educational' activities"*
- *"Head of data analysis department. I established the databases (in digital form) and protocols regarding the (mostly digital) repository were started from scratch."*
- *"Archives management, digital repositories management, database creation and maintenance"*

[10] For confidentiality reasons, the organisational affiliation of the quoted responses is not disclosed.

- *"I am the head of the centre, and also responsible for the software architecture used"*
- *"I am the author of the database structure, I make the selection of the bibliographical records which may be entered in the database, I organise the input of the bibliographical data, I am indexing the records with the keywords according to my own standard."*
- *"Management of the museum digital collections"*

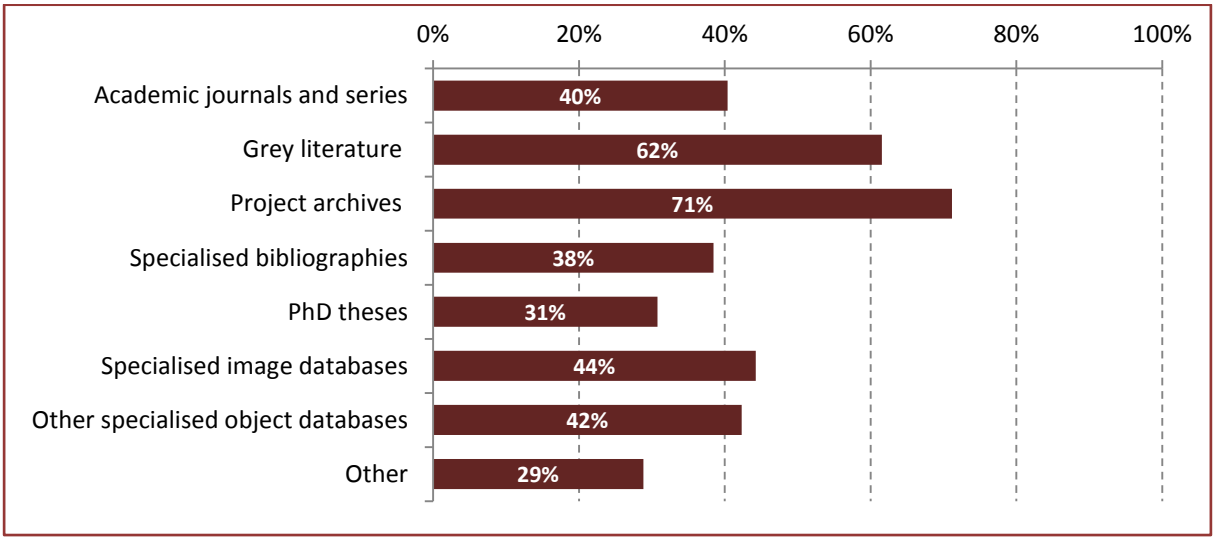
Considering these descriptions as given by the respondents, the term "repository manager" seems to be an adequate generalisation to refer to this group as a whole. We will use it in the following if referring to the whole community.

6.3.2 Collections and type of data held by the data repositories

Collections held and their relative importance

The repository managers were then asked about the types of collections maintained by their institute. They were given a list of eight generic types of data collections and asked whether these were part of the repository's offer. The results (see Figure 6.3-3) show that the portfolios of the various repositories covered by the survey differ considerably, even in terms of such basic types of collections. Project archives and grey literature seem to be the only assets which are essential components of most portfolios (70% and 60% respectively), while all the other asset classes were represented in less than 50% of the repositories. On the other hand, all types of collections are relevant, being represented in more than a quarter of the repositories each.

Figure 6.3-3: Question B.1 – "Which of the following collections does your data repository hold?"



N = 52

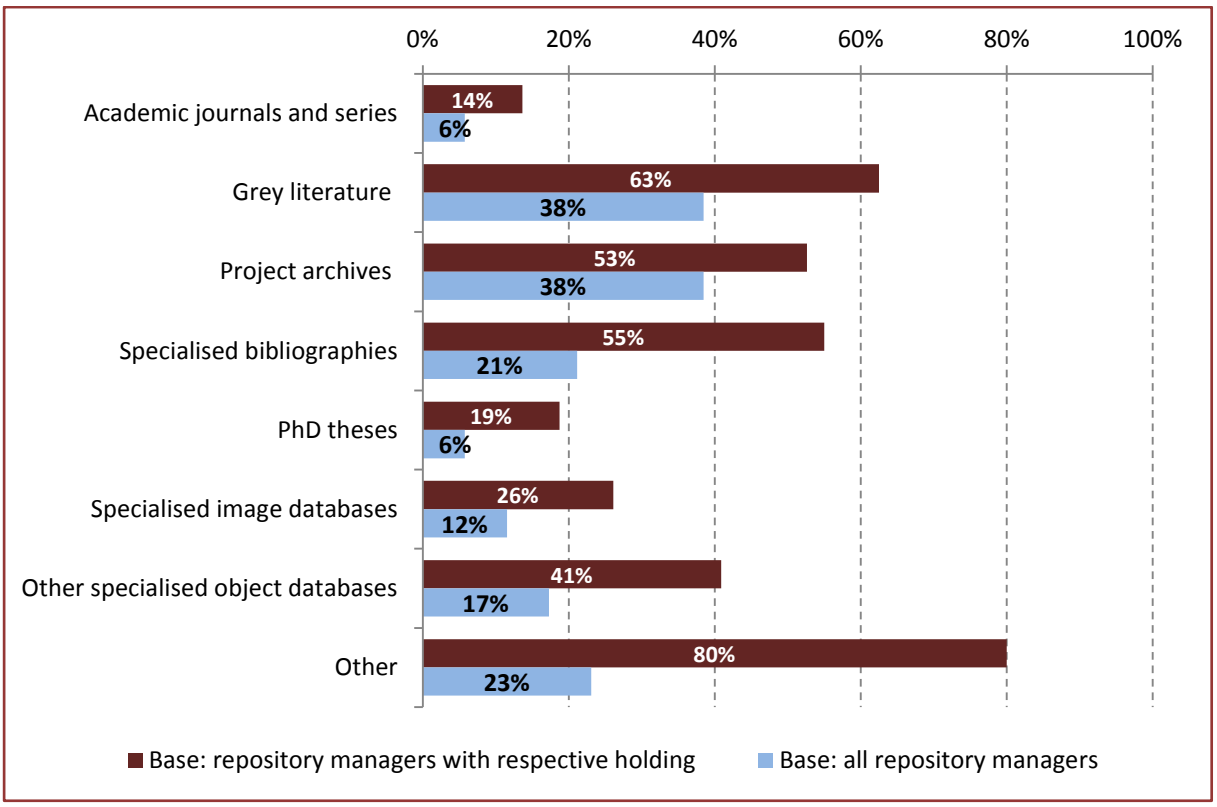
	B.1: Collections held	#	%	N
a	Academic journals and series	21	40%	52
b	Grey literature	32	62%	52
c	Project archives	37	71%	52
d	Specialised bibliographies	20	38%	52
e	PhD theses	16	31%	52
f	Specialised image databases	23	44%	52
g	Other specialised object databases	22	42%	52
h	Other	15	29%	52

Project archives and grey literature are not only the most common asset classes; they are also regarded as the most *important* ones (in terms of user demand) by the repository managers (see

Figure 6.3-4). Specialised bibliographies are also seen as an important asset by those that actually have them in their repository.

The comments provided by the respondents offer further insights on specific aspects and details that should be considered in this context. For instance, "holding" a certain type of collection can mean that the repository actually holds the originals, or that the repository "only" links to the place where the originals are available (*"The [repository] signposts to many resource collections. In some cases we hold the original material (…); in other cases we may just point to the resource (e.g. a reference to archive material held in a museum*)."

Figure 6.3-4: Question B.2 – "Which of these collections are the most important ones (e.g. in terms of downloads)?"



N = 52 (for base = all respository managers) /
N = 15-38 (for base = managers with respective holdings)

Comments from respondents on collections held:

- *"Within our research we are also resynchronizing data developed by others. We have not found a way to store this for future users. Example: I have retrieved e.g. Ernst Hollstein's data from his book (the mean value drawn CURVES not his tables!). It seems that the ITRDB are not interested!? When we are gone that big retrieval work might get lost."*
- *"Our repository contains raw datasets (images, databases, context form, GIS files, CAD files and so on) from professional and research archaeology and grey literature, especially archaeological reports."*
- *"some unpublished primary archives come to us first for recoding of important elements (i.e. photographs etc.) before being passed to the Historical archive at Surrey History Centre which is equipped for primary materials."*
- *"16,000 identifications of decorated sherds in thousands places in western Europe (30 years collective research)"*

- *"The HER signposts to many resource collections. In some cases we hold the original material (e.g. we have a large collection of grey literature); in other cases we may just point to the resource (e.g. a reference to archive material held in a museum)."*
- *"Information on protected historic environment and/or spatial datasets for download."*
- *"The repository is an implementation of tDAR and has the potential to hold all of the above. As a 2013 development project there are very few datasets in the repository at the moment."*

Comments from respondents on importance of various collections:

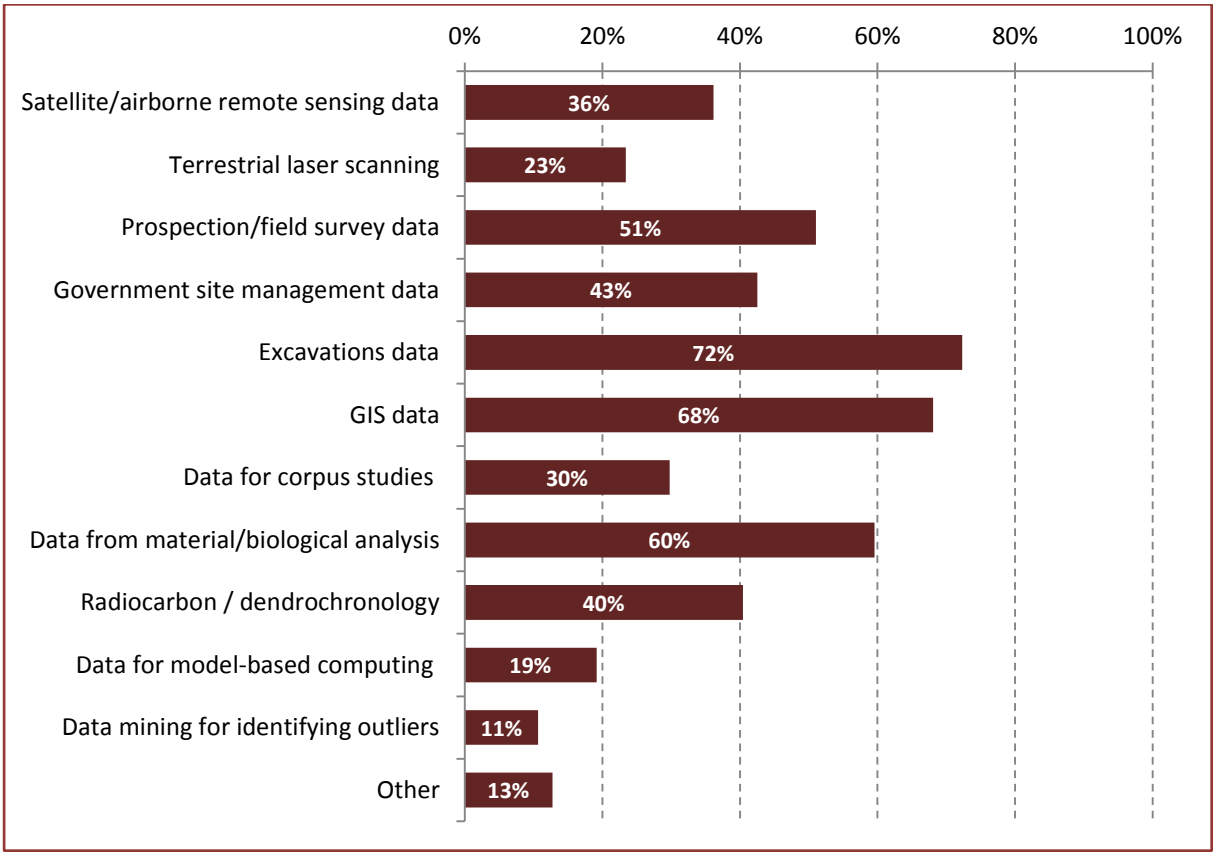
- *"They are all the constituent part of the database. I do not treat them separately. The database makes difference between monographs and longer studies on one side, and short papers on the other side. My personal search strategy is to use this distinction in the case of too many query results."*
- *"Data from professional archaeology are the most important: nowadays in Italy 90% of archaeological interventions are related to preventive or rescue archaeology carried on by professionals."*
- *"All our holdings are very important to us."*
- *"These data are not downloadable directly. We have no internet presence."*
- *"We do not monitor the access to the repository, so we haven't specific information."*

Types of data covered by collections

The next question focused on the specific types of data that were covered by the various collections held by the repository. The items proposed were aligned with those from the two questions posed to researchers (representing the main user community of data repositories) about the types of data they needed for their research and which they generated themselves (see "Importance of different types of data" in Section 6.2.2 and "Types of data generated" in Section 6.2.5).

The results mirror almost precisely the figures that had been obtained from the users. As documented in Section 6.2, for researchers, excavation data represented the single most important type of data during the search phase (75% of the respondents said that this data was "very important" for them) and was also the type of data most frequently generated. The other most important and most frequently produced types of data were GIS data, data stemming from material or biological analysis, and data from field surveys. In fact, these data types are also the ones most frequently held by repositories (see Figure 6.3-5). The results can be regarded, to some extent, as a validity check. It stands to reason that the data that is most frequently generated by users (researchers) will also be widely deposited and represented in digital repositories. Metaphorically speaking, the "goods" (data) that are most frequently produced (by researchers, in this case), are also most widely represented on the shelves in the "stores" (the digital repositories) and most often requested by the "customers". At least at this high level of aggregation, demand and supply seem to be well aligned.

Figure 6.3-5: Question B.3 – "Which of the following types of data are covered by holdings of your repository?"



N = 47

6.3.3 Issues and challenges for digital repositories

Key issues and challenges

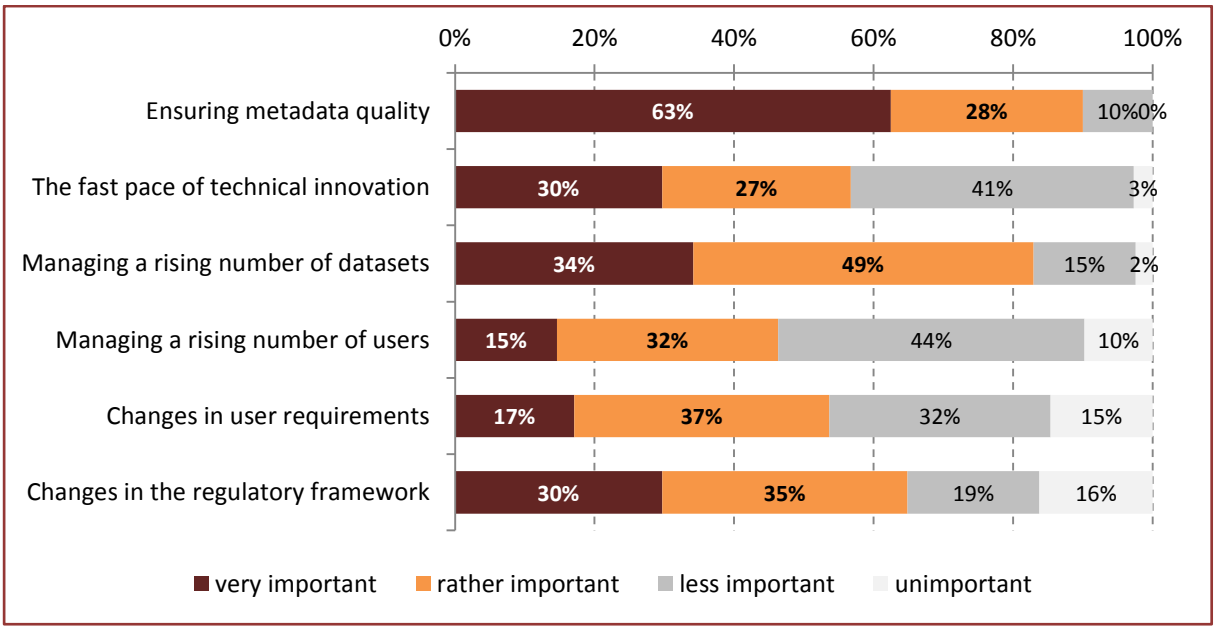
The second module of the questionnaire for repository managers addressed current issues and challenges which repositories are confronted with. The respondents were presented with a list of potential challenges and asked to assess how important they were for the repository they represent, and had the opportunity to make comments. **Ensuring metadata quality** came out as the by far most important issue. More than 90% of the respondents said that this was an important issue (with more than 60% saying it was "very important"). **Increasing scope** in the number of data sets to be managed was also seen as a critical issue (more than 80% said it was an important challenge).

The **fast pace of technical innovation** and **changes in the regulatory framework** seem to be an important challenge for some repositories but not for others – about 60% said that these issues were at least "rather important", 40% said it was not a relevant concern. Only a few repository managers reported user-driven challenges (such as "changes in user requirements" and "managing a rising number of users") as highly relevant (about 15%).

The comments point to another challenge which is probably critical for many initiatives in this field, in particular for those which are not firmly based on an institutional framework: managing the **costs for developing and/or operating the digital repository**. The comments indicate, for example, that several repository projects are "idealistic" ventures which are highly dependent on the "in kind" commitment of individuals, rather than on institutional budgets ("[our repository] *is an enthusiastic, zero-budget project, which is endangered in the moment of my retirement*"). If the organisational

framing of a repository is that of a project, there is uncertainty if the funding can be maintained after the lifetime of the project ("*Our main problem: Who will take care of our data when we cannot maintain it any longer?"*). However, even if the repository is established as an organisation, securing the budget can be difficult ("*An increasingly important issue within my sector is decreasing resources."*). In short, these remarks are evidence that many initiatives in the field seem to struggle to secure the required budgets and infrastructures for a longer-term maintenance. If cost had been included as an issue in the list of issues and challenges, it would possibly have been rated as another key challenge next to ensuring metadata quality.

Figure 6.3-6: Question C.1 – "What are the main issues and challenges your repository is currently confronted with in order to fulfil its mission in the best possible way?"



N = 37-40 (depending on number of respondents without answer)

		++	+	-	--	
	C.1	very important	rather important	less important	not important	N
a	Ensuring metadata quality	25	11	4	0	40
b	The fast pace of technical innovation	11	10	15	1	37
c	Managing a rising number of datasets	14	20	6	1	41
d	Managing a rising number of users	6	13	18	4	41
e	Changes in user requirements	7	15	13	6	41
f	Changes in the regulatory framework	11	13	7	6	37

Comments from respondents:

- *"the promotion of the database; the budgeting of the database (now it is an enthusiastic, zero-budget project, which is endangered in the moment of my retirement)."*
- *"Our main problem: Who will take care of our data when we cannot maintain it any longer?"*
- *"Data are produced with different file formats, especially with proprietary formats instead of open source formats, and with text formats such as .doc and .pdf that are not machine-readable."*
- *"The technology behind the ITRDB is quite dated now. Our major challenge is to locate the funding necessary to enable us to take advantage of the latest technological advances"*

- *"Main problem is to ensure that data producers and data managers from different agencies get the grip of the advantages of linked open data, and make them understand how to proceed in order to deal with the costly and inefficient silos that is the norm today, especially within the research sector."*
- *"An increasingly important issue within my sector is decreasing resources."*
- *"Coping with the vast diversity of data and data documentation methods in archaeology, addressing researcher needs (specifically, rewards and incentives for data sharing)."*

Most important technical developments

The repository managers were then asked to describe the most important technical developments which they considered as relevant for their repository. It was an open question (see list of answers received below), there were no standardised items proposed. The respondents mention a wide variety of specific issues which are difficult to summarise; several responses are related to data standards and protocols and to technical requirements for the provision of new services. The broad picture that emerges from the responses, however, is that technology as such (information and communication technology, software, database systems) may not be the most critical issue for data repositories. The answers do not indicate any fundamental shifts or changes (stemming from technical innovation) which affect the provision of the services ("*The repository infrastructure has remained rather static in recent years"; "Since (…) 2004, there have been no major changes*.").

Some respondents describe examples where technological issues are the result of changes in framework conditions other than the technological area. In other words, the change or innovation is not technologically induced, but technology has to respond – for instance to changes in legal requirements ("*Implementation of certain categories and identifiers due to changing legislation*") or to new research practices (for instance researchers publishing their data in Open Context).

Question C.2: Please describe briefly the most important recent technical developments which have been relevant for your repository or data centre.

Answers received:

- *"Major change has been a drop in staff. Some developments in the database design."*
- *"Transition from MS Access to SQL Server for database functions. (Near) complete digitisation of the Grey Literature collection to PDFA standard. Integration of archaeological data with the council's corporate GIS system to allow greater access to the records (currently working on smart phone compatibility to allow access in the field)"*
- *"Two years ago we moved from AutoCAD to ArcGIS and HBSMR database, which has enabled us to meet data standards. We have also recently added a new HER website (http://www.wiltshire.gov.uk/artsheritageandlibraries/museumhistoryheritage/wiltshireandswindonhistoricenvironmentrecord.htm) to enable searching of the data online."*
- *"Since the County Council bought Exegesis's HBSMR software in 2004, there have been no major changes."*
- *"The HER software is open source and is updated and developed on a regular basis by the HER managers. We are currently developing a development control module that will enable our DC Officers to manage and record their casework through the HER."*
- *"Additional data collection including x-ray and pxrf technologies"*
- *"we can mint DOI's directly"*
- *"The repository infrastructure has remained rather static in recent years. We are now at the start of a new round of improvements."*

- *"Changes to types of information required by new types of enquirers to our service. That is, new users have been generated by the NPPF that we currently have to accommodate and adapt to. We are having to create new systems to cope with the additional workloads and associated administrative effort, as well as increased staff time and a new knowledge base requirement."*
- *"We are developing and continuously publishing research support to the research community. We have recently started to put DOI on surveys..."*
- *"Provision of web feature services and creation of data as an open dataset freely available with no restrictions of use commercially/non-commercially. Currently going through a system upgrade to improve data delivery and upgrade of spatial dataset delivery."*
- *"Implementing a protocol for uniform data exchange for Dutch archaeologists"*
- *"Internet version of the 3D game and 3D books ready to be launched soon"*
- *"SOCH (K-Samsök) has been up and running since 2009. Most important is the development of the protocol which enables distribution of semantic data. This is an ongoing process to improve support for semantic web."*
- *"We created our repository on June 2012. Since then, we added a new functionality for the advanced search and we offer a new service as a recommended repository of the Journal of Open Archaeological Data."*
- *"Implementing protocol SIKB 0102, exchange format based on XML"*
- *"Provision of linked images/documents to web-sites, improved cross-linkages to other web-sites."*
- *"Implementation of certain categories and identifiers due to changing legislation"*
- *"Adding new identifiers and categories due to legislation change."*
- *"NPPF produced last year"*
- *"The new competing system of archaeological digs oblige me to sell less archaeology for each dig"*
- *"algorithms to facilitate searches of identifying patterns (2012)"*
- *"Support of "linked data" (using an RDF framework)"*
- *"We've realized a SPARQL endpoint that provides access to RDF metadata structured according to the CIDOC - Conceptual Reference Model in the implementation of Erlangen CRM/OWL."*
- *"WebGIS WebDatabase Web editing also for GIS geometries Web services OGC compliant Codifying standards for new archaeological field documentations and data archiving"*
- *"3D visualization, spatial queries, customizable interfaces, improvements in database complexity"*
- *"We've recently received funding for some thematic working groups, where researchers publish their data in Open Context and the work together to address research questions that draw on the large, integrated data sets. This is providing a lot of very useful information about data documentation methods and ways to facilitate data reuse."*
- *"Only substantial change, which was made since 2000, was adding the links to the database of the early mediaeval findspots of Eastern Alps - ZBIVA (http://zrcalo1.zrc-sazu.si/zbiva/frameset.php?lang=en)."*
- *"The museum until the end of 2015 will have a new museum collection database system that will permit to upload selected information directly in the museum website. Also, until the end of 2015, the museum will create various applications using the new media that will present in the internet online visit of the permanent exhibition, creation of virtual tours in past exhibitions, online educational programs etc."*

6.3.4 Access rules (for downloading, depositing, maintaining)

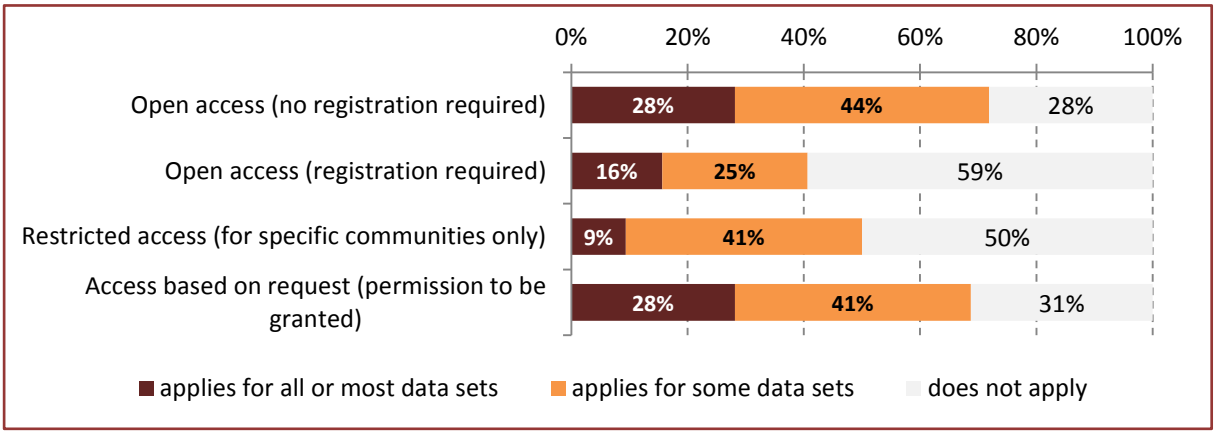
A specific module of the questionnaire aimed to explore the access rules which repositories apply, i.e. the requirements for being allowed to download or deposit data from/to the repository. Four basic principles were proposed, and respondents were asked if these applied either to all data sets of for some.

Access rules for downloading data: The two most prevalent approaches are the "extremes" in terms of being open or closed: about a third of the repositories covered by the survey apply an open access policy to all or most data sets without even asking for registration. Another third of the repositories provides access to their data sets only on request, i.e. permission needs to be granted (see Figure 6.3-7a).

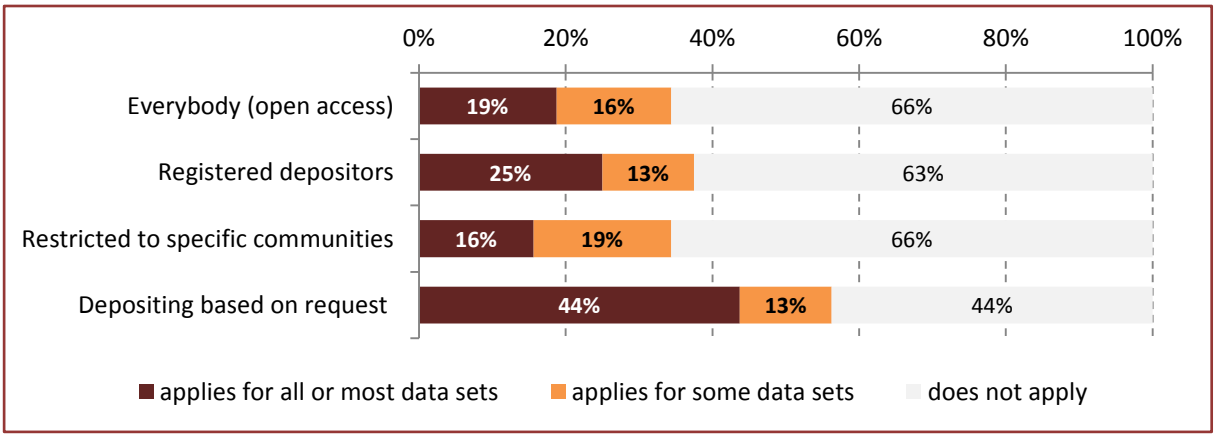
Access rules for depositing data: The access rules repositories apply to allow users to deposit data seem to be more "standardised" (i.e. applicable to all or most holdings) than the access rules (which vary depending on the data set). About half of the repositories require users to state a request for depositing data for all or most of the data sets covered. A quarter of the sample said they accepted new data from registered users, and 20% seem to be totally open, i.e. apparently anybody can deposit data without having to register or to file a request (see Figure 6.3-7b).

Figure 6.3-7: Question E.1 – "Who can download data from your repository? / Question E.2: Who can deposit data in your repository?"

(a) Downloading



(b) Depositing



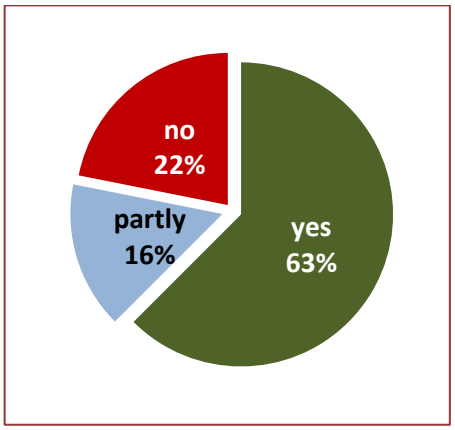
N = 32

	E.1 – Downloading data	for all datasets	for some datasets	does not apply	N
a	Open access (no registration required)	9	14	9	32
b	Open access (registration required)	5	8	19	32
c	Restricted access (for specific communities only)	3	13	16	32
d	Access based on request & permission	9	13	10	32
	E.2 – Depositing data				N
a	Everybody (open access)	6	5	21	32
b	Registered depositors	8	4	20	32
c	Restricted to specific communities	5	6	21	32
d	Depositing based on request	14	4	14	32

The repository managers where then asked if depositors could manage and update data which they have uploaded to the repository themselves. As to be expected, a broad majority of repositories (20 out of 32 that responded to this question) enable this function. Together with those that said that this was "partly" the case, close to 80% provide this functionality.

Figure 6.3-8: Question E.3 – "Can depositors manage and update data which they have uploaded to the repository themselves?"

	E.3	#	%
a	yes	20	63%
b	partly	5	16%
c	no	7	22%



N = 32

6.3.5 Perceived customer requirements and barriers

Perceived "customer" requirements

The repository managers were then asked to assess the needs of their customers, i.e. people and institutions who use the repository for downloading or depositing data. They were given a list of potential needs and asked to state how important they believed these needs to be, and how satisfied they believed their users to be with this factor.

Although the survey has only a small sample, the responses clearly show that good data quality and accessibility are "must have" factors for user satisfaction. Almost all of the curators argued that these were very (or at least rather) important (see Figure 6.3-9a). Data transparency and metadata quality are seen as almost equally critical. The only factor that is considered less important by a significant share of the respondents is the international dimension of the data sets. Clearly, not every repository has an international dimension. Many of the repositories covered by the survey focus on national collections and address, primarily, the national community – i.e. both demand and supply are mainly in a national use context. This does not undermine the goal and rationale of the ARIADNE e-infrastructure, though, as this e-infrastructure addresses explicitly international use cases. Even among the sample of this online survey, about 50% of the respondents said that the international dimension of the data was very or rather important for their users. These repositories represent a key community who could benefit from having the collections of their repository linked with other (international) database through the ARIADNE infrastructure and services.

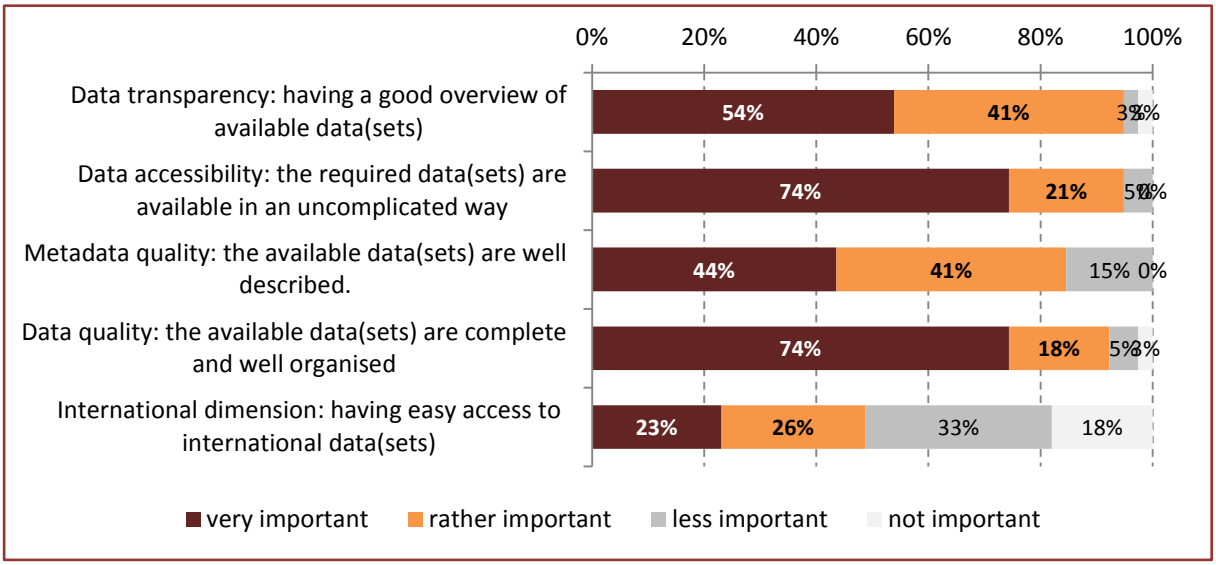
These results are mostly in line with the actual views of the researchers (as shown in Section 6.2, see Figure 6.2-24). Thus, the perspective of data users (researchers) and data providers (repository managers) are to a large extent consistent, even if researchers were asked for a 'global' assessment of the situation, while the curators were asked to assess the situation with regard to the specific repository they represented.

By contrast, the views of data managers and data users are somewhat different when it comes to assessing the satisfaction with regard to these aspects. Most of the curators, when asked to assess how satisfied their users would probably be with the current situation with regard to the various aspects (see Figure 6.3-9b), believed that the users were at least "rather satisfied" with the available data transparency and data quality. However, only 30-35% of the researchers said that they were actually satisfied with the current situation in these respects. As stated above, the views of researchers relate to the overall situation and not to a specific repository. Nonetheless, this gap in the data could be worth analysing in more detail. Is it possible that those who manage and operate data centres are not fully aware of the needs of the user community?

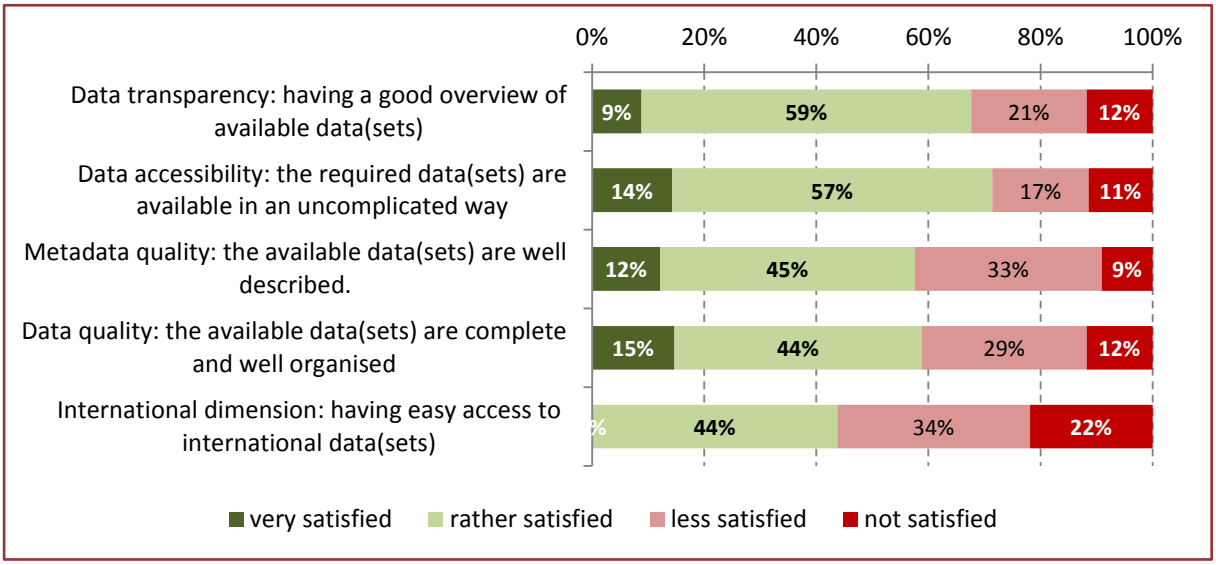
An interesting aspect in this context is to what extent the repositories themselves explore user needs and user behaviour. While some major repositories have launched surveys and initiatives in this field, many of the smaller data centres do not have an opportunity to systematically collect user feed-back. This was confirmed in a comment from one of the respondents ("*I do not know the opinion of the users. I'd never make such an inquiry. In personal contacts they are satisfied*.").

Figure 6.3-9: Questions D.1 and D.2 – "Please say how important the following aspects are for the users of your data repository in your view (D.1) and how satisfied they are with the current situation (D.2)."

(a) Importance



(b) Satisfaction



N = 32-39 per item (depending on number of respondents without answer)

Comments from respondents on importance of user requirements:

- *"It would be nice to have a common, indexed full-text repository guaranteed free of charge, without commercial tricks in the small text as they are by Academia.edu."*
- *"Access to data. Today most Swedish and German dendrochronology data are kept hidden. The originating institutions plainly refuse to leave out their data. A Swedish high court has even declared that it would be too much work for a university lab to leave out their data. These institutions are doing commercial dating and do not want their data used also by others. It is of course also a comfortable position not to have other researchers examine their work."*
- *"First of all they must be open."*

- *"Multi-language support for vocabularies and thesauri Shared and wider ontologies for archaeological domain Shared new Knowledge Experiences Bases"*
- *"The repository is primarily intended for internal use (Conservation in planning procedures) and not (yet) distributed widely. So the first goal is to complete the contents, organise and give the users a good and simple overview and use of this data when dealing with conservational issues in spatial planning procedure."*
- *"Workload constraints (due to under-resourcing) have an impact on my ability to keep the database up-to-date."*
- *"Clear citation, guidance on how to use data that others created"*

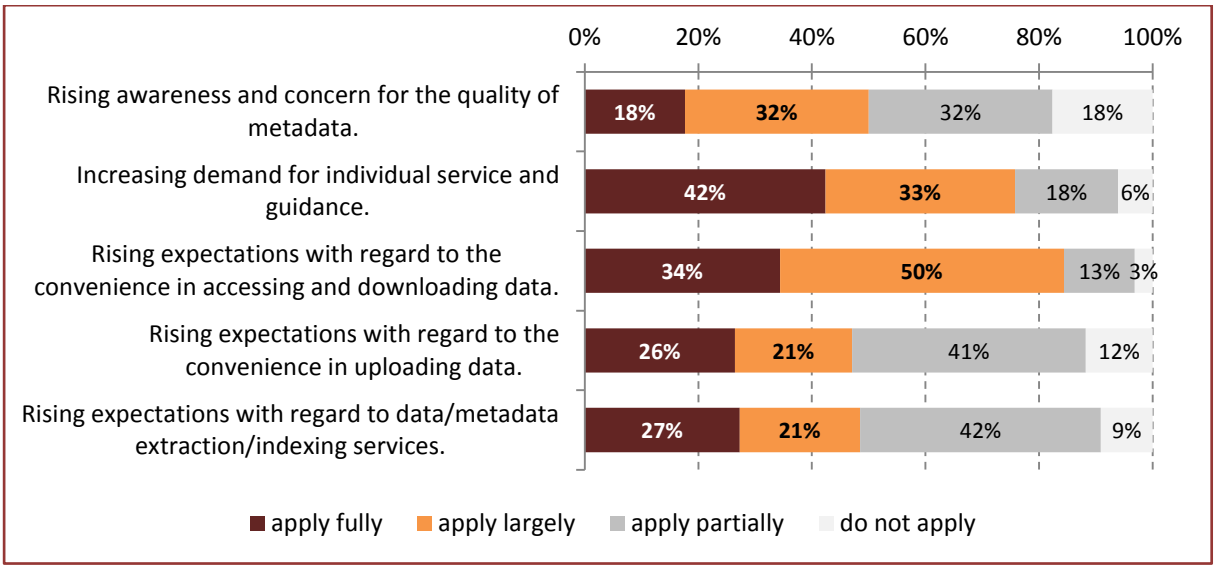
Comments from respondents on users' satisfaction:

- *"I do not know the opinion of the users. I'd never make such an inquiry. In personal contacts they are satisfied. If they are not able to find the searched topic, the reason is, they are not able to use the keywords."*
- *"We have issues with global coverage of our repository. Cultural differences around the globe with regards willingness to share data openly mean than certain areas (e.g. Asia and Africa) are not as well represented as others (e.g. North America)."*
- *"links to external DBs (i.e. libraries) to improve data and reduce errors in data entry improve the multilingual platform, searches, …"*
- *"The National Heritage Board has recently started a national project in order to deal with the lack of a national repository for field data. Recently a repository for field reports been implemented."*
- *"Not all datasets are available to users outside. User knowledge and computer skills are very different; the number of users is low. Software development (web based and implemented in GIS) aimed for end user is planned."*
- *"We are in a starting phase, so the above answer reflect the state where the center has not yet been functional"*
- *"Being cynical, I think users are in the main satisfied as they are not quite there yet in terms of buy-in to the whole process. This varies across disciplines eg Astronomers are fully engaged, as are geologists, other disciplines less so"*
- *"It is difficult to say since we are an organization that are working on spreading the knowledge about the organization..."*
- *"Users always want more!"*

Major changes observed in user behaviour

The interviewees were also asked to assess and describe changes in user needs and user behaviour. According to the repository managers, users have become more demanding with regard to convenience in the use of repositories, and there is an increasing demand for individual service and guidance (see figures below). Other potential changes that had been proposed in the survey seem to be less relevant, according to the curators' view (see Figure 6.3-10). Even if the sample is small, and the results should not be considered as precise statistics, it is interesting to see that these two trends are outstanding as compared to the other factors. For instance, curators do *not* see a similarly significant trend towards a rising awareness and concern for the quality of metadata (50% said that this applies only partially, if at all). This is something that should be further analysed in the update to this report.

Figure 6.3-10: Question D.3 – "Are there any major changes in the needs and behaviour of the users of digital repositories? Please say if the following statements apply."



N = 32-34 per item (depending on number of respondents without answer)

		++	+	-	--	
	D.3	apply fully	apply largely	apply partially	do not apply	N
a	Rising awareness and concern for the quality of metadata.	6	11	11	6	34
b	Increasing demand for individual service and guidance.	14	11	6	2	33
c	Rising expectations with regard to the convenience in accessing and downloading data.	11	16	4	1	32
d	Rising expectations with regard to the convenience in uploading data.	9	7	14	4	34
e	Rising expectations with regard to (meta)data extraction/indexing services.	9	7	14	3	33

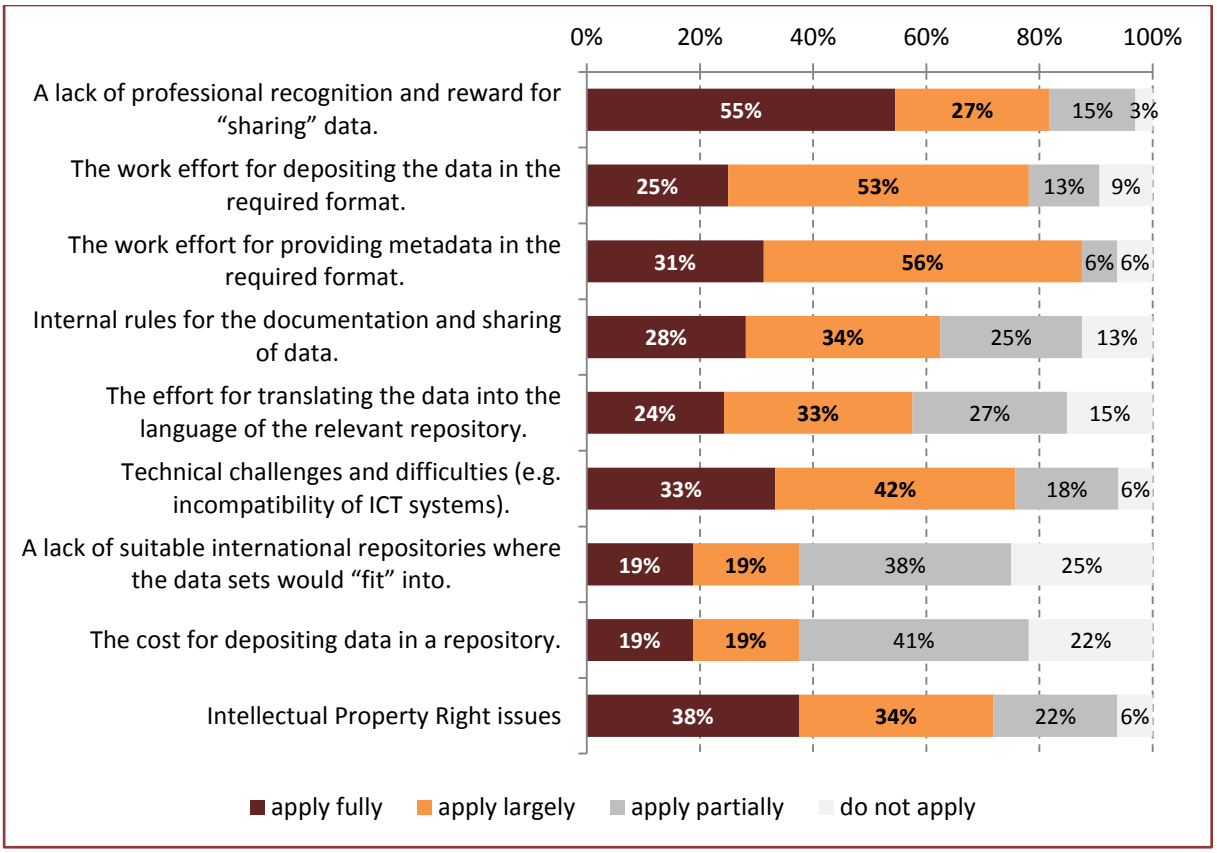
Barriers for depositing data

The repository managers were then asked what they considered to be the main barriers for researchers for depositing data in digital repositories (and thus sharing the data with colleagues). The same question was also directly asked to researchers (see Section 6.4.5, Question C.9), which allows for a comparison of the assessment from different perspectives. Repository managers see a lack of professional recognition as the most important barrier (more than 50% said that this was very important, see Figure 6.3-11); several other barriers are also seen as highly relevant, in particular the work effort (both for depositing the data and for providing metadata).

All in all, the assessment of the repository manager is close to the answers received from researchers. Two factors which are directly related to repositories are the exception: while the "customers" (researchers) also have some complaints about a lack of suitable international repositories and the cost for depositing data, the service providers (repository managers) tend not to regard these issues as relevant. Even if the figures are not fundamentally different and (in the case of the repositories) based on a small sample only, it indicates that the users' perception of the available offer may be somewhat different than that of the providers.

One of the comments on this question also pinpoints differences in the perspectives between researchers and repository managers. A respondent said: "*I think that a main problem is that too many researchers think this is MY data or my organization's data though its development has been mainly paid by tax money.*" This observation is clearly related with the different professional (and ultimately economic) incentives of these ARIADNE user communities. The project is well advised to carefully consider the motivation and incentives of different stakeholder communities when addressing them, for instance to develop plans for liaising with specific projects or communities.

Figure 6.3-11: Question D.4 – "The following table describes potential barriers (for researchers) for depositing research data in digital repositories. How important are these barriers based on your experience?"



N = 31/32 per item (depending on number of respondents without answer)

Comments from respondents:

- *"The humanities publications must be accessible free of charge or there will be no humanities any more. The owner of the repository must not be allowed to make the money with it."*
- *"I think that a main problem is that too many researchers think this is MY data or my organization's data though its development has been mainly paid by tax money. When we asked for some archaeological German data (referred from a paper in a journal) we got the answer: "Get out into your own forests and collect your own data!" Very nice!"*
- *"We do not specify a format for data to be deposited in. We are happy to receive raw data and reports, which we can index and deposit for future reference, if the researcher sees fit. We are also happy to keep research results confidential to protect intellectual copyright."*
- *"Multi-language support for uploading, inserting, maintaining and using deposited/collected data"*
- *"Property rights and institutional limitations"*

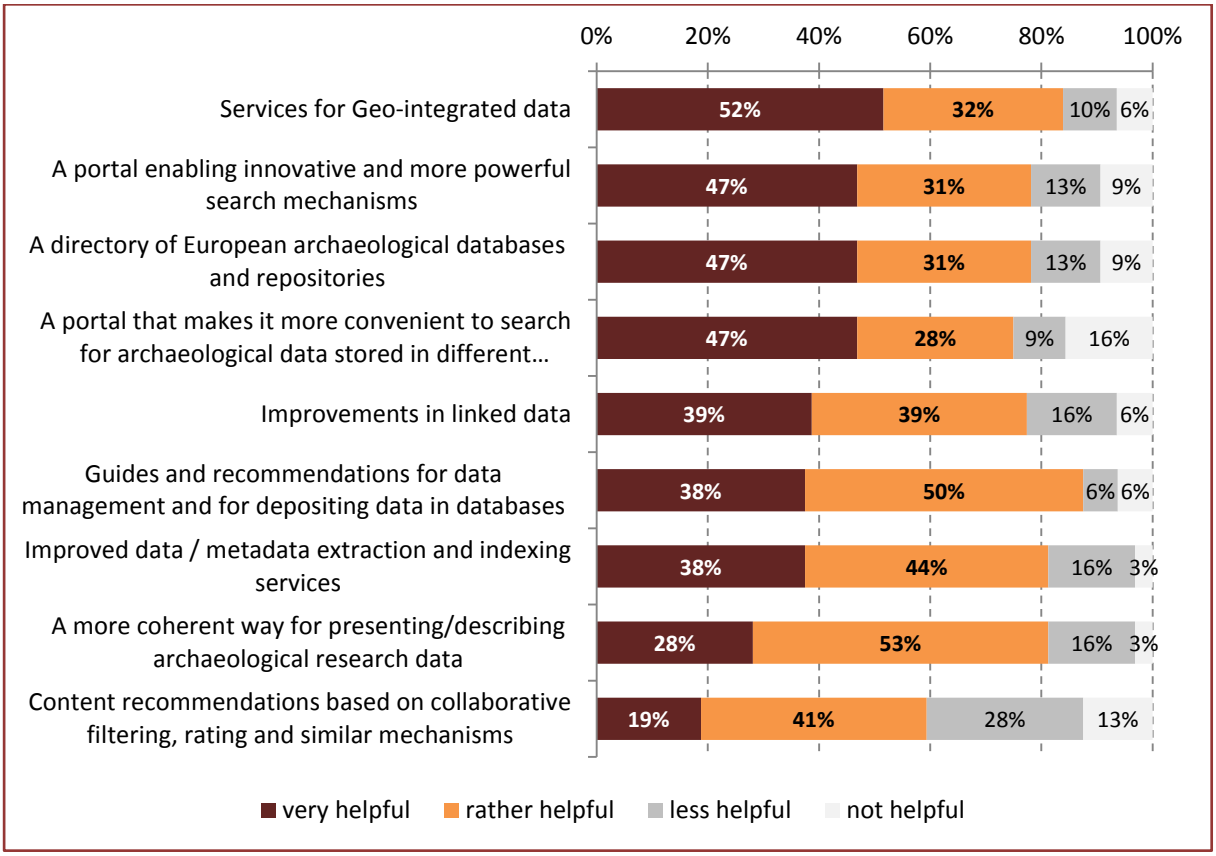
6.3.6 Usefulness of potential ARIADNE services for repositories

In the last module of the questionnaire, the repository managers were given a list of potential types of services which the ARIADNE project may consider for development, and asked to assess how useful each of these services would be for their repository. Researchers were also asked the same question (see Section 6.2.6), which enables a comparison between the two main user communities of the project: what do researchers expect from ARIADNE, what do data centres hope to get out of it? The overall picture of the repository managers' responses is not too different from that of the researchers. The results (see Figure 6.3-12) can be summarised as follows:

All of the potential services suggested are regarded as useful: all of the services are evaluated as "very useful" or "rather useful" by at least 60% of the respondents, most of the items even by more than 80%. There is no difference in that respect to the researchers' view. However, the results from researchers were more nuanced (in terms of how many respondents considered a particular service as *very* useful), allowing for a ranking of the various services. The figures from the researchers' survey are much more robust in this regard, however, due to the much large number of responses obtained. A specific difference between the two user communities is that researchers are even more focused on potential portal functionalities (for improving the transparency and convenience of data search). Clearly, this reflects their specific use case: for them, ARIADNE services are predominantly a search tool. For data repositories ARIADNE serves different purposes – such as "market extension" (addressing a wider user community) or improving the service to the existing customers.

Only a few respondents made use of the opportunity to provide feed-back on the proposed services. One of them is particularly important, though, as it points to a risk that too many portals create fragmentation rather than an integrated infrastructure. ("*There seems to be a lot of portals bringing together heritage datasets (e.g. Europeana, Heritage Gateway) and I don't think a new one is necessarily the answer. I think money and research should focus on improving existing portals (…).*"). This is certainly an issue worth further exploration and analysis, possibly also within the Special Interest Groups. The ARIADNE project is confronted with a challenge and possible dilemma in that respect. On the one hand, a broad majority of the targeted users suggest that portal functionalities would be very useful; on the other hand, the risk that another European portal per se, unless it provides really powerful search functionalities and access to rich data sources, will not effectively improve the underlying user requirements such as having a better transparency of available data sources.

Figure 6.3-12: Question D.5 – "To what extent could your repository benefit from the following services or improvements (as to be provided by ARIADNE)?"



N = 31-32 (depending on number of respondents without answer)

Comments from respondents:

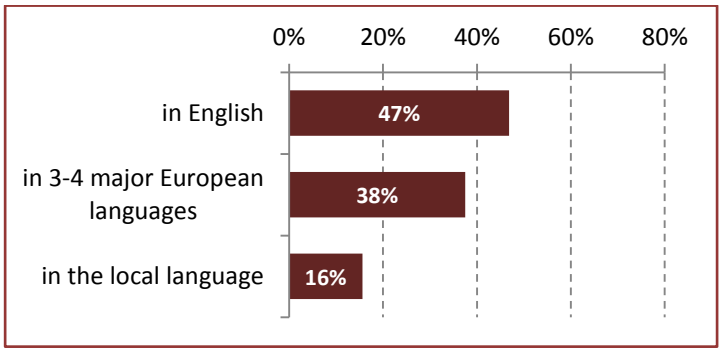
- *"ARIADNE's services would be desirable and relevant to a subset of our user community interested in European archaeological research"*
- *"There seems to be a lot of portals bringing together heritage datasets (e.g. Europeana, Heritage Gateway) and I don't think a new one is necessarily the answer. I think money and research should focus on improving existing portals, utilising the improvements and services above would be very worthwhile."*
- *"innovative methods to represent time and other values; semiautomatic extraction of data from texts"*

Language requirements

The repository managers were also asked about the language requirements for ARIADNE services for being able to integrate different data sets, in particular with regard to metadata. About half of the respondents believe that metadata in English would be sufficient, and about 40% said that metadata should be available in 3-4 major European languages. Only a few said that it was required in the local language. This assessment is similar to those of the researchers (who were asked the same question), the number who were confident that English would be sufficient was even higher.

Figure 6.3-13: Question D.6 – Language requirements for integrating data sets

	D.6	#	%
a	English	15	47%
b	3-4 major European languages	12	38%
c	local language	5	16%



N = 32

Further suggestions for ARIADNE

The last question of the survey invited the repository managers to make further suggestions or comments on any of the issues that were raised by the survey. About a third of the respondents made use of this opportunity. Their suggestions are listed below. Several comments highlighted the importance of ensuring wide access to research data, not only to the research community in the narrow sense, but possibly also to wider communities of practice. Two comments addressed the language issue, suggesting that Russian and Spanish would be particularly important languages.

Comments and suggestions received:

(F.3) Would you like to make any further comments on related issues (i.e. the access and management of archaeological research data)?

- *"Minor remark: the biggest European language is Russian. Europe ends at Ural and not at the border of EU. The centre of Europe is at Vilnius (Lithuania). The actually world lingua franca is at the moment English. But it will not be forever. So even recently strategy must be the equality of all languages (which are extraordinary important part of our cultural heritage), which could be facilitated by computer translating. Important remark: it is rather easy to find budgeting to start some database, but there is no system money to keep it alive after some more years. We depend on commercial offers which are intellectual cannibalism. I hate such system. If there is anything good in EU, there had to be found the solution of this problem."*
- *"It should not be difficult or time consuming to store or retrieve data from a database. My wife recently spent two days on retrieving a reasonably small set of data from the DCCD system. In contrast, dendrochronology data stored within the ITRDB can be easily handed over to the ITRDB and it can be easily retrieved (within minutes) by a user. All the requirements on metadata within the DCCD seem to have made the data scattered around so it has become a real time consuming work to retrieve it."*

- *"Archaeological data had to be released as open data (without access restrictions of any kind), they are produced with public money and they had to be public. Intellectual property had to be acknowledged to data producers with DOI and CC-BY license (at least a CC-BY-SA licence)."*
- *"If accessible in more than one European language, I would recommend Spanish, as this would enable access to many North and South American research institutions."*
- *"It will be a very valuable aspect if the archaeological research data will be available also for schools, citizens and other communities of interest, with appropriate personalisable access levels."*
- *"Interested in seeing the outcome of ARIADNE and how we might be able to apply systems like this to at least regions in North America"*
- *"Ariadne seems like a great idea, it is such a pity that so many repositories and data are not shared by wider communities, for developing new ideas, teaching, creating new collaborations, sharing the competence of each other."*
- *"Access and management of archaeological research data from different sources is not only based on databases and digital documents but for a large part also unique analogue information."*

7 Case studies: issues and challenges with regard to specific types of data

This section presents preliminary results from the work of three **ARIADNE Special Interest Groups** (SIGs). The SIGs are a mechanism of the ARIADNE project to explore specific issues related to user requirements in more detail (than it is possible with the general user survey). The SIGs have been established under WP 2 (as Task 2.2). Their assignment is to survey the state-of-the-art of a specific theme (typically related to specific types of data), focusing on user requirements and on the strengths and weaknesses of the available infrastructure and tools. On this basis, the SIGs assess the gaps (issues, challenges) and how the ARIADNE project could contribute to addressing the identified challenges. The work of the SIGs thus feed into the analysis of user requirements. By the end of the first project year (Dec. 2013), nine SIGs have been established. Each SIG has a coordinator, and typically 10-15 core members recruited from the consortium. Most of the SIGs have taken up their work during the second half of the year; they use different mechanisms for exploring their themes. Some SIGs have conducted or planned workshops (typically using a larger conference or annual meeting of an association as a platform), most of the SIGs have organised an online consultation among debate on specific issues (using the ARIADNE intranet as a platform).

In this section, we present the considerations and preliminary results of the following SIGs (the three "case studies" differ in terms of their scope and in the way how the SIG teams addressed the issues):

- the ARIADNE SIG on Excavation and Monuments Data (chaired by Elizabeth Fentress, AIAC);
- the ARIADNE SIG on Grey Literature (chaired by Julian Richards, ADS);
- the ARIADNE SIG on 3D Data and Visualisation (chaired by Roberto Scopigno, CNR-ISTI).

The SIGs will be an important mechanism for addressing specific issues with regard to user requirements particularly in the second project year. The results will be reported in the update to this document (Deliverable 2.2 – Second Report on User Requirements).

7.1 Case 1: Excavation and monuments data

7.1.1 Introduction

When the general public thinks of 'archaeology' the first image that springs to mind is that of an excavation (preferably on a sunlit field), while the second is a major monument like Pompeii or the Acropolis. The success of a European infrastructure will thus almost inevitably be judged on how well it answers the twin questions of access to information about both of these categories, as well as the related category of field-survey data.

All three of these are complex, with multiple data sets potentially attached to each single individual. For excavation these would be stratigraphy, description, images, finds, GIS data, geophysical data, environmental data, radiocarbon data; for monuments physical description, location data, images, access data, conservation data; for field surveys descriptions, images and finds data for individual sites, and GIS data, by period, material and site type for the survey as a whole.

Access to this type of data varies, although the respondents to the ARIADNE Stakeholder survey agree that they are generally abysmal. In the case of excavation data, a report on the site in .doc or .pdf format may be available, but the finds data is unlikely to be found in a database format, and is even less likely to be online. The three categories will now be discussed separately, introducing their problems and the current state of play in various European countries.

Excavations

Discovery: finding sites

At the start of any search for the purposes of research the primary query would be "what excavations have taken place that answer my search criteria?" (Roman villa, Bronze-Age sanctuary, medieval cemetery, sites Hertfordshire). In some European countries – UK, Bulgaria, Romania, Italy, and Holland through websites such as OASIS (ADS, UK), DANS EASY (NL) and Fasti Online (AIAC) – this information is at least partially available for sites excavated after c. 2000. OASIS (http://oasis.ac.uk) provides an online index to a large number of reports (in PDF/A format) and other material (images, spreadsheets, etc.) of archaeological fieldwork in England, Scotland and Wales.

The digital archiving system DANS EASY (https://easy.dans.knaw.nl) includes the e-depot for Dutch archaeology (EDNA), which provides comprehensive coverage of surveys and excavations in Holland. An initial search reveals a summary record: further data is available to registered users, generally in PDF format. In January 2014, 21,500 data objects were archived: 18,500 archaeological reports and 3,000 datasets which consist of photos, GIS, data-tables, drawings etc.

In Sweden, SND is a digital repository and works in a similar way as ADS and DANS. Data deposited at SND is made available online either directly or by ordering data (depends on the data). The data is presented in a catalogue with the metadata extracted from the data files or produced by SND who manages and documents the data to ensure availability and promote reuse. The documentation is compiled according to the international standard DDI (Data Documentation Initiative, xml-based and exportable). SND also provides Persistent Identifiers (DOI) to each dataset.

Fasti Online (http://www.fastionline.org) furnishes a short summary of each campaign on an excavation in one of the member countries (principally Italy, Bulgaria, Romania, Macedonia, and Albania). There are c. 3150 excavations registered to date, with around 15,000 records. However, the metadata available on the Fasti site means that the director or institution responsible can be contacted for further information. All three databases are accessible through search terms, keywords and a mapping interface, which is bound to be helpful to planning officials as well as those interested in archaeology.

A much newer site is that of INRAP, also with a map interface. It now contains records of several hundred rescue excavations carried out by INRAP in France (http://www.inrap.fr/archeologie-preventive/Sites-archeologiques). These are searchable by type and period, though not by keyword. The information available beyond a brief summary includes articles, videos and in some cases recordings: finds data and detailed reports are not included although here again could presumably be made available on application to the institution. This site is reserved exclusively for those rescue sites excavated by INRAP, and complements the site of Archéologie de la France Infos (http://www.revues-gallia.cnrs.fr/spip.php?rubrique19), run by the review Gallia, which records various excavations carried out by universities and local associations, and is searchable, though has no map interface. In Greece, the British School in Athens and the École Française d'Athènes collaborate on the production of Archaeology in Greece Online (http://www.chronique.efa.gr). Resembling Fasti in its summary reports (although these are very short indeed), the site so far contains a few dozen sites, searchable in various ways, including a map interface.

Outside of these projects there are a few very intensive projects which produce exhaustive coverage of all the excavations in their respective cities. Foremost is the Museum of London's LAARC archive (http://archive.museumoflondon.org.uk/laarc/catalogue) which provides an online resource for the study of all archaeological research on in the city, and may be the world's largest archaeological archive. However Rome's SITAr and Pisa's MAPPA also aim to provide comprehensive records of all excavations which have taken place in their respective towns. Both projects are linked to ARIADNE as associated partners, but they are not yet available to the public. For Vienna a government site maps all monuments and archaeological sites in the city (http://www.wien.gv.at/archaeologie), with date

of excavation, site codes and bibliography. Apart from these (and other projects which we have undoubtedly missed) a member of the public looking for excavation reports from a given country will have a difficult job knowing where to start.

Long-term storage and documentation of excavation data

Archiving and storage of excavation documentation is addressed by only two of the partners, ADS and DANS. This partly because it represents a far more serious undertaking, including questions of sustainability, access, copyright and data interoperability. The solution adopted by ADS requires the excavator to provide metadata on the components of the submission, which tend to include full reports of most aspects of the excavation. The finds reports are not always in a database format, making their use and linking much more problematic. DANS EASY adopts a similar policy, giving persistent identifiers to the documentation received. These are the only national archives of excavation data currently available: other excavation data is stored in the institutions that carried out the research, dispersed through university departments and are very seldom available online.

Two American projects, the Digital Archaeological Record (http://core.tdar.org) and Open Context (http://opencontext.org), guarantee long-term, sustainable storage of project data in return for a fee. Data may be presented as CSV files with metadata, which helps retrieval, as well as storage of images and project reports. tDAR is not easy to browse, and lacks map interfaces except at the level of the individual site. It is clear that one would have to know that a site of interest existed in order to access any data related to it, although at the level of an individual site the data may very complete. Open Context contains large amounts of data from a range of sites, some of them in Europe but most in the Middle East. It is fairly easy to search the site and to retrieve data from it.

A minority of sites serve their own excavation data (stratigraphy, finds catalogues and so on) on dedicated websites, usually housed on university servers. Examples are the Via Gabina villas project (http://viagabina.rice.edu) and the excavations at Cosa between 1990 and 1997 (http://www.press.umich.edu/special/cosa). The Via Gabina site is essentially the whole publication, including the catalogues in database format, while the latter complements a published volume, providing the stratigraphic and anthropological data for the volume, as well as tables of sherds. Both of these sites are in a very simple HTML format, and present few problems for sustainability, in spite of the general lack of metadata. A more dynamic solution is that of the Prescot street excavations in London (http://www.lparchaeology.com/prescot/) which serves live data via the ARK system.

Among the most elaborate examples are the excavation database of the Silchester Town Life Project (https://www.reading.ac.uk/silchester/i3/later/index.php) and the Çatalhöyük database, http://www.catalhoyuk.com/database/catal/Search.asp. This sort of solution seems best-applicable to large scale research excavations that are capable of convincing institutional servers to house them. However, universities have become increasingly wary of such projects, because of the apparent risks posed by interactive websites, and issues of long-term sustainability. Furthermore, although data may be available on such sites, it is difficult to link it to other such data, although the individual catalogues can generally be downloaded in CSV or Excel formats.

Although paper documentation of excavations has the disadvantages of inaccessibility in most cases, the vast majority of excavation records that were 'born digital' are at risk. The rapid changes in technology can be difficult for museums and universities, upgrades and migration to new solutions may be neglected, and the link between excavation and objects can thus easily be lost. Complex relational databases built for excavations a decade ago using Microsoft Access can now no longer be read without a computer still running Access 6: those using more obscure proprietary databases are entirely inaccessible. Thus some form of sustainable deposit is a major necessity.

Monuments

Sites and Monuments Records

National sites and monuments records are generally extensively catalogued by governmental institutions (for the UK the Ordinance Survey and English Heritage at the national level, and then at the county level). In Sweden, for instance, an online database "Fornsök" (fmis.raa.se in Swedish only) created by the Swedish National Heritage Board provides basic identification of sites, often with geo-references, sometimes with fuller descriptions and links to images, management information and other resources including GIS. Such resources are generally designed to support the management of archaeological sites, with different levels of protection (from recognition on world heritage registers, through national listing, local listing, to simple identification). They provide a primary source for desk-top research and other forms of archaeological investigation carried out in advance of development. They also provide a resource for example for identifying areas of archaeological potential, for landscape archaeology and for understanding the context and settings of archaeological sites.

Many of these databases are available online (an example is the Archaeological Survey of Ireland (http://www.archaeology.ie/ArchaeologicalSurveyofIreland). A more ambitious project is the CARARE project, coordinated by MDR, which focused on aggregating archaeological and architectural sites and monuments for the Europeana initiative. The remit of the CARARE project (http://www.carare.eu) concentrated on the infrastructure needed to support the aggregation of inventories (amongst other resources) created in different countries according to slightly different standards and in many languages. The project produced a prototype map tool (http://carare.eculturelab.eu/Carare50m/Map.html) which provides the locations of sites ranging from major monuments to excavations and artefacts on a GIS map base. The integration of several monument inventories and other resources (including data from DAI's ARACHNE database amongst others) the information is served on a single map where it can be browsed using keyword searches or by using route-planning tools.

A remarkable experiment in the combination of various forms of archaeological data on a single internet site is the Getty-financed MEGA Jordan (http://megajordan.org) which provides GIS documentation of sites ranging from major monuments to excavations to field-survey sites. More of a sites and monuments record than an excavation database, it is still interesting in its integration of this information and the fact that it serves it on a single map: we will return to this resource at the end.

Survey Archaeology

A major change in the archaeological focus in the last quarter of the last century was the growth of archaeological field survey, studying large tracts of land through intensive field-walking, combined with aerial photography and occasional geophysics. These projects, rather than providing in-depth information about a single site, created a more-or-less precise view of a landscape, with the degree of detail depending on the intensity of the survey. As time has passed many of these projects have been published on paper, some as volumes, others a journal articles. In the case of final publication there are also catalogues of materials. A favourite thesis in Italian universities, in Italy alone they number in the hundreds. Many projects were later retrofitted to GIS format: a case in point is the decades-long Tiber Valley Survey of the British School in Rome (an associate of ARIADNE). However, the prospects for on-line availability of these projects, where their aggregation would create a resource substantially greater than the sum of its parts, are still inexistent. Indeed, the Tiber Valley Survey plans long-term sustainability through deposit with ADS, with the consequent loss of GIS interface as it currently appears.

Since the 1990s surveys in Bulgaria are recorded on an online information system "Archaeological map of Bulgaria" (AIS AKB). All archaeologists are obliged to fill in information about newly registered sites and monuments as a result of their annual fieldwork. However, access to the database is restricted and protected by a regulation of the Ministry of Culture who is its actual owner. NIAM-BAS is responsible for its protection and maintenance. Despite its name, it is not GIS-based system, although the intention is to create a mapping interface.

A very few surveys are fully available online: an example is the Jerba Project, held on a university server (http://www.sas.upenn.edu/jerba/index). The map can be searched by site type and date, and various catalogues, such as architectural fragments, are available. A discovery tool for identifying 317 survey projects is represented by Mediterranean Archaeology GIS (MAGIS: http://cgma.depauw.edu/MAGIS). It is more or less limited to Anglo Saxon projects, never having reached the national universities and foreign schools that are the generators of many of these projects, and, unfunded, has not been updated for a few years. Survey archaeology is thus the least available on line of all major forms of archaeological research.

7.1.2 Issues and challenges to be addressed

The working group identified two main challenges with regard to the collection and preservation of data about excavations and monuments.

Discovery – improving the transparency of available data records

While the scope of ARIADNE's remit concentrates on the creation of an infrastructure that will make cross-platform searching possible, in the hierarchy of needs the aspect of discovery remains a major challenge. Digital storage of excavation records is certainly vital, but below that there are, of course, thousands of excavations whose records were born on paper and have never been digitized. These, however, are generally accessible in some form (excepting those which are thrown out with the death of the excavator). It is the existence of the excavation that needs to be recorded in some format easily available online, complete with metadata that will give some idea where and who the archives are housed. The very simplicity of the Fasti and INRAP records makes them quick and easy to fill in, though the legal obligation excavations to be reported in Italy and Romania makes their records far more comprehensive. In the same way discovery of survey projects is a major step forward towards their exploitation for research.

Sustainable storage and metadata

There is no doubt that the amount of work that may be involved in the deposit of digital excavation records represents a major obstacle, perhaps even more important than the cost of the deposit, although that may not be insignificant, and is rarely considered in excavation budgets. Without the mapping of the data no cross platform searching is possible, but getting excavators to budget this into their time and resources may represent a considerable challenge. The question of long-term sustainability is also a major challenge: even such rock-solid institutions as ADS are occasionally threated with de-funding. As one of the stakeholders commented, maintenance is not politically popular, nor is funding easily available for long-term projects (witness the collapse of MAGIS).

7.1.3 Implications for ARIADNE

The common denominator for excavations, monuments and survey projects is the existence of point data that can be mapped and viewed online. However, national monument inventories are mapped using national mapping systems, integration on a European or world map requires conversion of the national coordinates to an international spatial reference system such as WGS84. Providing tools,

which support and enable this conversion, will enable the development of a GIS-based European portal to sites, excavations, monuments and buildings.

A second issue is the creation of international standards for the documentation of excavations and monuments so as to render it transparent and comparable. Free access to tools, particularly for data mapping, to make it easy to comply with these standards will be important, as will be offering the means and guidance to archaeologists to deposit their digital records. The sustainability of digital datasets must also be high on the agenda.

Once these problems are resolved, however, it would be possible to create a complex map with a GIS of sites that could be enhanced by differentiation of the points (survey sites, excavation sites, listed monuments) and layers that could be turned on and off. Such a map would make discovery of broad categories of sites in a given landscape far easier, while allowing the user to understand at a glance what sort of further information might be available: MEGA Jordan is the model here for the geographic display of disparate types of site. Data points would then create links to the site from which they were drawn, where further research could be carried out. Once such a structure is in existence individual institutions could contribute datasets (this would be particularly true of field survey data) which could then be served on institutional computers. The result would be a multi-sourced cloud of data, from which a variety of institutional datasets could be discovered. The institutions would then become stakeholders in the care and maintenance of the collective site.

Of course, such a project has its utopian aspects, and there are parts of it that might best be built from the ground up. For some time, AIAC and the University of Southampton have been planning to create a site from which survey data could be served, and plan a survey of the contents of individual institutional projects in Italy in order to discover the scope of the problem. But at a larger scale the creation of such a resource would be a preliminary result that would allow both professionals and the general public to gain some idea of what is out there. Rather than a traditional portal, which always implies some kind of list of sites whose purpose is more or less opaque, a GIS-based European portal for sites, excavations and monuments would provide on an international scale the sort of resource that ADS and DANS offer nationally.

7.2 Case 2: Grey literature

7.2.1 Introduction

In archaeological terms "grey literature" means unpublished archaeological fieldwork reports. These are generally produced as a consequence of fieldwork undertaken as part of the mitigation of commercial or public property or infrastructure development. In most cases they are the only record of the results of such fieldwork, and they have become a substitute for conventional publication. However they are strictly speaking unpublished and generally lodged as a single hard copy in the local museum, archive or planning office, with additional copies provided to the client, and retained by those undertaking the fieldwork. Access presents a significant problem in most European countries, and there is a general view that much of this important new data is not being fed back into the research cycle. In the UK, for example, it has been estimated that academic teaching and research is 10 years out of date. Although most reports are generally generated in a digital format their online distribution is patchy and fragmented at best, with one or two notable exceptions (DANS and ADS).

In the UK, the value of free access to this important resource has been demonstrated recently in an impact assessment carried out at the ADS by a group of external economists. The findings of this assessment showed that not only does it provide significant economic benefit to the commercial sector; reliance on the resource has begun to change the how the sector operates. The ability to see the value of access to grey literature also motivates further deposition, and fuels a cycle of good

practice which is also becoming more apparent. Bringing grey literature together with the fruits of academic research in one place also sends a message that this work is important and valued by the sector. Digital dissemination has the potential to unlock these research resources and to make them far more accessible than the conventional journal literature, particularly to those working outside a university environment. With the introduction of DOIs the grey literature also becomes as citable as conventional publications. Grey literature was specifically mentioned in the ARIADNE proposal as an area of focus.

7.2.2 Issues and challenges to be addressed

The ARIADNE working group on grey literature has identified **four main challenges** in this domain which will be further analysed and assessed by the group in terms of their implications for the ARIADNE project:

- **Collection**: How should we go about ensuring that grey literature reports are collected within a research infrastructure? The experience of ADS and DANS may be useful here. As many organisations digitise their backlogs of reports we need to ensure that they follow best practice guidelines in terms of digitisation and plans for access.
- **Access**: How should the ARIADNE partners seek to promote access to grey literature? How far should grey literature feature within the ARIADNE cross-search portal?
- **Preservation**: How can it be ensured that the reports, and the data they contain, are preserved? There are specific aspects of digital preservation relating to reports. For example, the growing trend to embed various types of data e.g. plans, photos, tabular data, within a PDF/A file creates particular challenges for the long term sustainability of the reports. Protecting grey literature from obsolescence or loss is identified as one of the key contributions of the ARIADNE Integrated Infrastructure.
- **Indexing and interoperability**: Providing metadata for reports by hand is a time-consuming activity. Can techniques of Information Extraction be of assistance? (There is a crossover here to the work being undertaken within the NLP work package). Should ARIADNE seek to define a core metadata schema for grey literature? Members of the group noted that grey literature documents may contain useful photographs, drawings and other visual elements which are not addressed by available indexing techniques. Consequently useful information held on graphical elements is normally discarded from indexing. Another challenging task is to be able to extract and index tabular information held in such documents.

The Group agreed that it would be useful to define a core metadata standard for grey literature, as this would be useful to partners setting up grey literature databases, such as CSIC, which are now to contribute data. The core metadata schema would probably include, as a minimum, Dublin core fields such as contributor (organization), coverage, subject, report type and themes.

7.3 Case 3: 3D data and visualisation

7.3.1 Introduction

The working group on 3D and visualisation organised and held a workshop on "Ariadne infrastructure for multimedia data" on October 7-8, 2013, in Pisa. This workshop brought together data providers and technical partners with the goal of establishing a consensus on what we mean by the term "multimedia" (MM) or "visual data" and what kind of support the ARIADNE project can provide in

these domains. It was agreed that the concept of MM/visual data can be broadly defined as any type of visual representation of archaeological findings or assets, including standard 2D images (including high resolution and high dynamic range, HDR), advanced images (PTM, RTI, panoramic images), 3D models, and videos. The discussion on visual representation media started from the acknowledgement that the extent of the represented archaeological artworks is usually very wide; our focus should include the small findings (few centimetres) up to an entire archaeological site (tens or hundreds of meters).

The Pisa workshop was followed by a feedback and discussion phase which contributed to consolidating a more structured definition of the current needs and perspectives.

7.3.2 Issues and challenges to be addressed

The members of the working group have identified the following challenges in the management of visual media:[11]

- Access can be difficult, especially if making 3D models available to the general public on the Web. Some authorities are very protective of their assets;
- For scanning of high buildings and interiors, drones/UAVs may be required to capture inaccessible data to produce a complete model;
- Metadata is under development, CARARE2 schema yet to be tried and tested;
- Storage of large models, availability of information about these and accessibility is also an issue (no standards, many different formats and proprietary systems and technologies…);
- Having the means for dissemination of new services;
- Standards-related issues (widespread use of metadata forms developed by 3D-ICONS will go a long way, i.e. nothing else is needed after that);
- Knowledgeable public (peer-review community building, lots of good case studies; mostly covered in ARIADNE's WP4 by a planned expansion of Good practice guide for 3D in archaeology);
- Assessment on the relation resolution of data acquisition - quality of data (a heritage perspective);
- Adequate digital library with "intelligent" tools for interaction with content;
- Web access of high quality media content;
- Definition and creation of standard documentation (i.e. paradata) during the acquisition process of visual media;
- Interoperability and metadata standardization for an efficient storage/retrieval of visual media
- Technologies for online collaborative work;
- Technologies for visualization and management of different visual media in integrated (web?) environments;
- Support data acquisition and management with a standardized procedure;
- Enhance quality and performance of 3D models visualization;
- Enrich 3D visualization environment with tools for analyse and explore the content;
- According to ADS' perspective, we do not, at present, provide the visual media in its advanced/interactive state as our primary role is to preserve digital data and disseminate the

[11] For a detailed presentation of the discussion and the various issues and challenges that were raised, we refer to the summary document prepared by R. Scopigno, M. Dellepiane (CNR-ISTI) (internal ARIADNE working document, to be obtained on request).

material we are given rather than processing it or creating more advanced visualisations, though this is an area we are hoping to develop in the future;

- For more complex files there is a high need to get accurate metadata in order to rebuild and effectively re-use the data (this is very challenging);
- Viewing large pointcloud datasets over the web;
- More interactive tools for interrogating data;
- Embedding intelligence in 3D pointcloud data, assigning attributes to datasets with archaeological information;
- Enhance scientific collaboration on archaeological analysis (not only nice 3D visualization and reconstruction but enhance analytical tools);
- Best Practise guides for standardized data management;
- Also work out best practise for licensing and copyright issues of Multimedia items (might help for re-use and co-use of media in the future, collaboration with other projects: Wikimedia, (3D icons?);
- Use 3D not to create harmonic scenarios but to question single solutions to a research hypothesis (visualization of uncertainty);

7.3.3 Implications for ARIADNE

Which types of visual media should be considered and supported by Ariadne?

After the workshop, the participants were asked to rank a number of different types of media (presented in a list) in terms of their relative impact and importance, on a scale from 1 (very important) to 5 (less important). The results are shown in the table below. The attendees supported the choice of demanding the support of video and panoramic images to commercial tools (e.g. YouTube in the case of videos).

Priority	Content	MDR/ ADS	ZRC SAZU	Cyl	PIN	Disc Prog	DAI	Pref. *
	Images:							
1	High-Resolution images	1	1	1	1	1	1	6
2	Panoramic images	2	4	2	2	2	2	14
3	High Dynamic Range Images	4	2	4	3	4	4	21
4	Reflection Transformation Imaging	3	3	5	4	3	3	21
5	Image Pilebars	5	5	3	5	5	5	28
	3D – single object:							
1	3D on web	2	1	2	1	2	2	10
2	Tools creating digital 3D models	1	3	1	3	1	1	10
3	Enhanced Visualization tools	3	2	3	2	3	3	16
	3D – large scenes:							
1	Support via OSG4Web	1	1	1	1	1	1	6
2	Develop case studies (for different geographic regions)	2	2	2	2	2	2	12

* Legend to the "Preference" values: those values indicates the priority (1: best, 5: less important). Therefore the "Preference" sum in the rightmost column has to be interpreted as "lower the value, higher the preference".

How to link different initiatives with the ARIADNE project

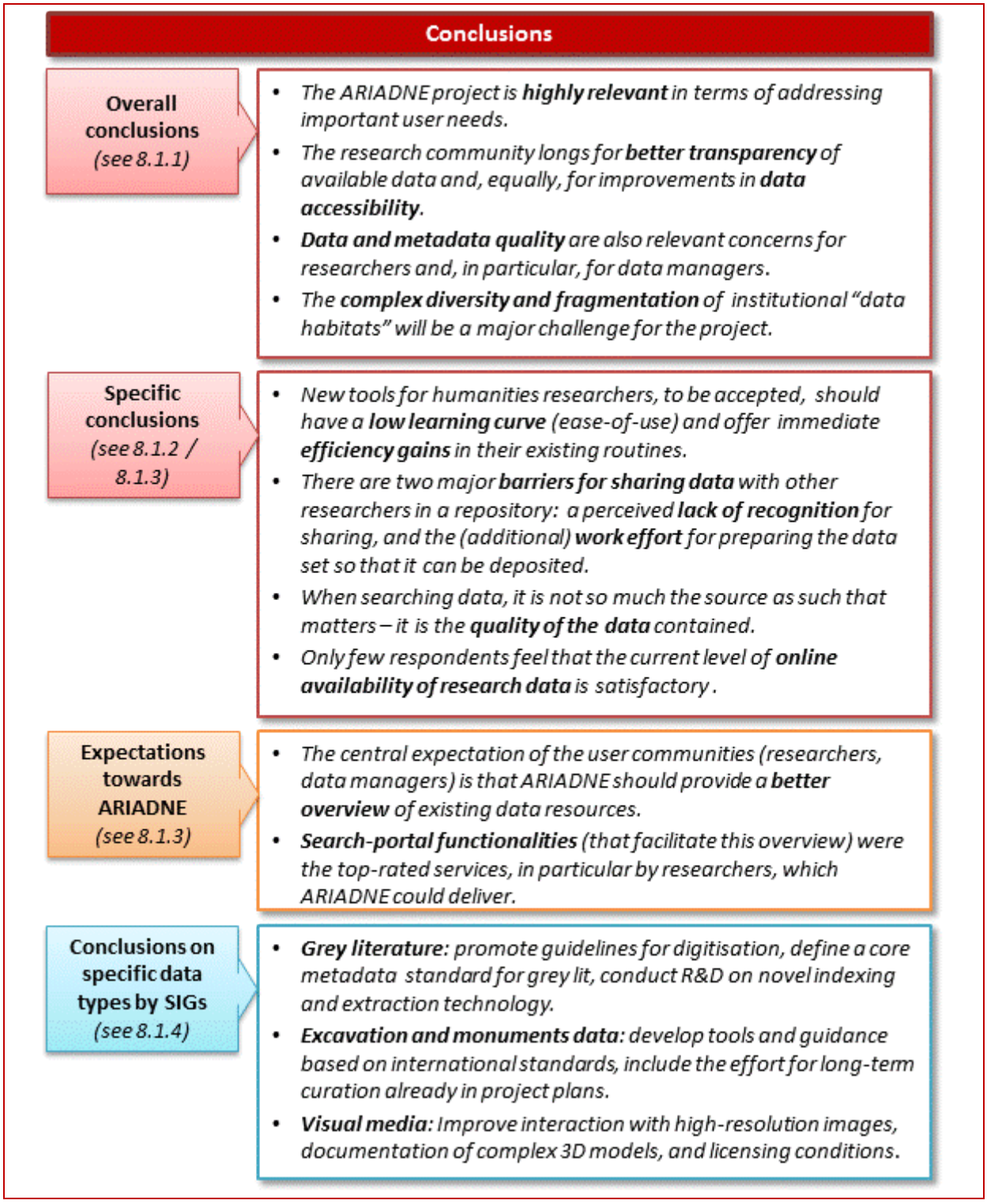
During the PISA workshop, the speakers who represented repositories or institutional databases were asked to assess if they saw an opportunity for ways the 3D- and multimedia-related initiatives of their repositories could be linked with the work of the ARIADNE project, and how these initiatives could possibly benefit from the work of ARIADNE (in this particular domain). Here is a summary of their suggestions and perspectives.

- **ADS**: The ADS is trying to develop and use multimedia data to bridge the gap with the museum sector. Here the focus is not just to provide repository services for data that museums do not have the skills or resources to deal with, but trying to integrate as curators of digital data within their collections. There is a relationship with Southampton Museum, but also with the Yorkshire Museum on the Star Carr exhibition where objects were shown alongside specially-made digital content evoking the sights and sounds of Mesolithic Yorkshire as well as footage of archaeological excavations. Briefly then, in closing the gap between museums and data repositories we can also bridge the physical/digital media gap using multi-media to go full circle from physical objects to 3D scan to 3D print creating a duplicate physical object. This work is not an attempt to replace museums but to enhance access and perhaps publicise to a greater extent what museums have to offer.
- **DAI**: Currently, the DAI databases contain several types of heterogeneous data, but their connection and availability is not so well defined. Actions to handle this are already ongoing, through the collaboration with other partners (external to ARIADNE). A more structured approach to 3D, GIS and CAD linking, retrieval and visualization would be needed. The metadata issue hasn't been faced in a structured way yet.
- **Discovery programme**: The participation in this and other projects (like 3D-ICONS) could help in having better structure and integration in the work of the organization.
- **STARC Cyprus Institute**: The ARIADNE infrastructure should be able to support the archiving of MM data by taking into account some important issues (that are especially critical for 3D data):
    - the need for a clear distinction between "observation/measurement" and "inference";
    - the need to know under which conditions data were obtained, acquired and processed;
    - a clear demarcation between a technical process and a cognitive process; the visualization of a "cognitive path" along which nodes of decisions are defined and their criteria well represented;
    - the possibility to visually investigate the 3D model (through cross-sections, slices of parts of 3D objects, measurements along any surface).
- **Virtual Gardener** on Blender: further effort in developing and extending of the Virtual Gardener, availability for the partner to test and give feedback (including consulting and training support).

8 Conclusions and recommendations

8.1 Conclusions

This section summarises the main findings of the empirical work. First, we draw some general conclusions (see Section 8.1.1); these confirm not only the high relevance of the ARIADNE project, but also its main challenge: the complex fragmentation of the research data to be integrated. Then, specific conclusions from the literature review (8.1.2), the interviews and the online survey (8.1.3), and from the work of the Special Interest Groups (8.1.4) are presented. These conclusions were the basis for our recommendations to the ARIADNE project (see 8.2).



8.1.1 General conclusions

The survey confirms the relevance of the ARIADNE project

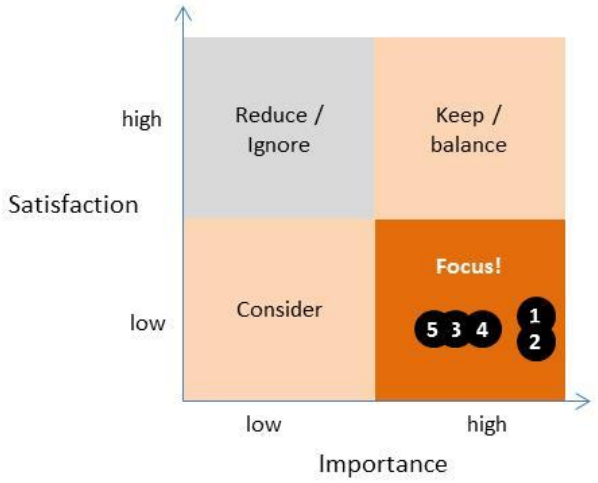
The interviews and survey with researchers clearly confirm the **high relevance of the ARIADNE project** rationale. The project directly addresses very important user needs which are not well catered for by existing services. More than 60% of the researchers interviewed in the survey said they were either "not satisfied" or "less satisfied" with the current situation.

The research community expresses, in particular, a need for an improved **transparency** of available research data (it is difficult to know which data actually exists, due to the enormous fragmentation of repositories and data sets in the field) and improvements in data **accessibility**. Major barriers to accessibility are costs (e.g. for obtaining licences to use pictures, for subscription fees) and the problem that relevant literature and data is often kept in places other than where it is supposed to be (e.g. in private collections). **Data and metadata quality** are further concerns; regarding metadata in particular from the perspective of data managers. Our evidence suggests that any improvements in these fields would be highly appreciated by the wider user community of researchers and data managers.

Essentially, this means that ARIADNE has a broad field of opportunity to create real value for users. While it is clear that the project cannot solve all these problems, ARIADNE has a **high impact potential** if its services can deliver improvement in any of the above mentioned areas (see matrix – all five domains of user requirements are in the segment which suggests focusing on them).

However, in order to take a **strategic decision** on priority areas, and to facilitate the choice and design of technical solutions, a further analysis of specific user requirements in the five domains is required. This will be the main goal for the updated version of this report in D2.2 (see *Section 2.3*), where requirements will be further broken down and explored with lead users among researchers and repository managers.

User requirements of archaeological researchers according to their importance and the satisfaction with the existing situation in a strategy matrix



User needs as shown in the matrix:

(1) = Data transparency

(2) = Data accessibility

(3) = Metadata quality

(4) = Data quality

(5) = International dimension of available data

Conclusions on the fragmentation of the research data landscape

The institutional diversity of "data habitats"

Based on the survey sample of repository managers, we perceive a complex diversity of "data habitats", comprising a variety of organisational and institutional mechanisms and regimes under which research data is collected, archived and maintained. This includes project-level repositories or databases (e.g. regional or city level, single site, digital corpus of artefacts, etc.), single institutes (research centres, museums and other), supra-institutional data centres, and heritage authorities and related services at county, province or national levels. Within this variety of organisational "habitats", some entities are important hubs – for instance authorities' services, data centres and national aggregators.

There are also different types of repositories, including authorities' records repositories and services (monuments and sites registers, archaeological surveys, physical archives), museum collection management systems (e.g. artefact collections, conservation and other archives), individual project archives and databases (typically with Web GIS frontend), data centre systems (e.g. OAIS based), national aggregators (metadata pool). Furthermore, the data habitat of a research institute typically comprises various "databases" from past and current projects, a catalogue of literature and reports, project archives, and special contents like aerial photographs.

Project-centred practices in research reinforce fragmentation

Major factors that lead to fragmentation are to be found (i) in common perspectives on research data and (ii) in the way how data is organised in the archaeological research sector (in particular in research institutes). The pilot interviews and the survey comments demonstrate that researchers have a predominantly project-centred rather than institutional perspective on data. As a result, the major formats for organising data are "project archives" (one per excavation site) and "database projects" of small research groups or even a single researcher. The format of "collection" is much less common. Unfortunately, this does not represent favourable conditions for linking and integrating. To link these project archives with a common repository requires a system, workflows, and dedicated staff which may not be available in many research institutions.

The case for leveraging "data tanks and pipelines"

Most archaeological data is not "big data", but rather "small" and "dark" data which is difficult to manage and prepare for aggregation and sharing. ARIADNE will have to focus on data sets that are already available in existing institutional repositories and national data centres and, on the other hand, promote the flow of currently "dark data" into the repositories or centres. However, in many countries national centres do not exist and also the state of institutional repositories may not be optimal (e.g. lack of Open Archive Initiative compliant systems). Hence, ARIADNE can play a significant role in leveraging the "data tanks and pipelines".

The case for an alignment with close neighbours

Archaeology has close scholarly neighbours like classics, medieval studies, epigraphy, iconology and others. A multitude of directories, catalogues, bibliographies, reference collections, text and image corpora, digital editions have been produced in those fields. These resources have not been fully considered yet by ARIADNE, but the neighbours are relevant both as providers and users of data and knowledge resources (e.g. vocabularies). 30% of the repository managers surveyed said that their holdings include data/content relevant for corpus studies.

8.1.2 Conclusions from the literature review: basic needs for tools and services

In the literature review we focused on studies which explored user requirements (with regard to digital tools and services) of humanities researchers in general as well as archaeologists in particular. In most reports humanities scholars are understood to present a special case, because the type of "data" they are working with (cultural content like texts and images) tends to be quite different than the data used in natural sciences. As a result, they have different expectations towards data tools; moreover, implicitly, the assumed average technical skills of the scholars are different. Archaeologists in general may have a higher affinity towards technology-supported work, however, because they produce most of their research data themselves.

In some cases, the requirements can be quite similar though, for instance when tasks like searching for archival content is concerned. Considerable differences exist where different research methods and types of data are used, as the research objects are quite different (e.g. an archaeological field survey vs. a scholarly edition of inscriptions). Research groups and collaborative tasks may play a greater role in archaeology. Below we summarise the characteristics of humanities research data, general requirements for tools and services, and specific requirements of archaeologists.

Data/content characteristics

The typical research data/content in humanities is characterised by the following features:

- Mostly small-volume and heterogeneous data (i.e. not "big data");
- Widely distributed, i.e. high demand for networking of resources;
- Difficult to bring together or to integrate into a single datasets, e.g. for advanced computing;
- Data is not particularly sensitive as compared to other social sciences or the health sector (i.e. there is normally no need for fine-grained authentication and authorization systems);
- High importance of semantics (cultural meaning, different languages, etc.);
- A large stock of relevant analogue material (e.g. older grey literature, various archival material).

Requirements for new tools and services

Basic requirements of humanities researchers with regard to data tools and services are:

- **Focus on efficiency gains:** tools/services must be practical in the sense that they increase the efficiency of existing research routines; they should in some way simplify or accelerate essential tasks, and the facilitation of existing practices must come before creating or suggesting new ones;
- **Provide clear valued added**: new tools must be clearly "better to use" than what is already in use (novelty *per se* is not a relevant criterion);
- **Low learning curve**: new tools should be easy to adopt and use (ease of use);
- **Low cost**: New tools must be affordable and sustainable.

The **most commonly stated needs** are:

- **Searching** across distributed resources;
- **Filtering** hit returns more effectively;
- *Not appreciated:* online collections pre-culled by others.

Archaeologists' "**dream tools**", mostly for the initial phase of data collection and for presenting project outcomes, are characterised as follows:

a) General criteria:

- Capable of accommodating existing practices and vocabularies;
- Highly flexible, ability to customise various things easily (e.g. data schemata, interfaces);
- Enabling provision of comprehensive project information.

b) Initial data capture/entry:

- Field tools and mobile applications for data recording;
- Efficiency gains, e.g. removing the "double-entry problem" (getting rid of paper templates).

c) Online publication:

- Linking project information with various other accessible data/content (e.g. grey literature, images, databases), also beyond the project level;
- Integration of information from many sites, especially map interfaces (e.g. what information is available for a region).

d) *Not* appreciated ("nightmares"):

- Adoption of unfamiliar (but predefined) data standards, schemata, vocabulary, user interfaces;
- Required markup of data to align it with more general Web or semantic standards (perceived as disconnected from immediate needs, and outside of practitioners' area of expertise).

Overall, is seems that archaeologists are rather reserved towards significant changes in existing research designs, workflows or vocabulary. Achieving data compatibility and interoperability are a concern, but the willingness to adopt unfamiliar standards is generally low.

These results are generally relevant for Task 2.1 and Task 12.2 with regard to the definition and specification of "most wanted" tools and services. Tools and services for facilitation of initial data capture and entry will not be directly relevant for ARIADNE. What has been identified as "nightmares" of archaeologists seems highly relevant for WP3, WP14 and WP15.

Section 3.6.8 includes suggestions on how to enhance archaeologists' data curation practices to enable re-usability of data.

8.1.3 Conclusions from the surveys and interviews

The researchers' perspective

Data sources used

Online publications with supplemental data are seen as particularly important, equal to printed publications. However, the feed-back indicates that it is not the source as such that matters – it is the quality of the data contained. There is no single most important source; researchers make use of all kinds of sources.

Types of data used and generated in research projects

Three out of four respondents said that excavation data was "very important" for them. Other types of data which is very important for a large group of researchers are GIS data, data stemming from material or biological analysis, and data from field surveys. These types of data are also the ones most frequently produced by researchers. However, other types of data are clearly relevant as well. Comments suggest that specific types of data are seen to have a significant potential for future research; however, they are yet difficult to collect (e.g. isotope, residue and DNA analyses).

Online accessibility of research data

Only few respondents feel that the online availability of research data is satisfactory. For most of the types of research data listed in the survey the online accessibility was rated more often as "fair" than

"good" or "very good" together. Exceptions were excavation data and GIS data; the figures were on a par with satellite/airborne remote sensing data, and quite close with data for corpus studies. Some respondents commented that they could not rate the availability of data they do not use, while others complained about lack of downloadable "raw data" for re-use.

Sharing data and depositing data in online repositories

Whether to share data with others or not is a very important issue for researchers. Survey results and literature on this topic indicate that research data may not only be scattered across different institutional databases, but a good deal of data might not even make it to the institutional database but remain on the computers of the individual researchers. Nearly half of the respondents store data (at least temporarily) on their own computer for all or most projects, and about a third in a shared project archive or institutional server. Within the survey it was not possible to explore to what exact extent this data will be organised and deposited in a shared database or repository after the completion of the project. However, all in all, the publishing and sharing of data in national data archives or international repositories is not yet common practice. Only about 15-25% of the researchers surveyed said that they would do so in many of their projects, while 50-60% do not make use of repositories at all. **Three main challenges** were identified for depositing data in an online repository:

- a lack of recognition for sharing the data;
- the work effort for preparing and documenting the data set so that data can be deposited in an institutional repository;
- (in some cases) a lack of opportunity.

This was also confirmed by the data managers who participated in the survey when asked to assess the barriers for sharing research data from their perspective.

This is a big issue for the ARIADNE project, as it presents a barrier that undermines the rationale of the project (see also recommendation on this issue).

Metadata creation at project level

Most researchers seem not to worry much about metadata (like repository managers do) and, consequently, often do not produce metadata for the various data (data sets) they generate in projects. Researchers are aware and concerned that producing adequate metadata is a considerable additional effort, and that they might lack required expertise. To allow for effective data sharing, these additional efforts (costs) will have to be covered somehow (e.g. by research grants).

Maintenance of data at the institutional level

About 40% of the respondents worked in a research organisation that operates an institutional repository which is managed by dedicated staff. The responsibility for maintaining data after the completion of a project is in many cases still part of the responsibility of the project manager (54%) or of a dedicated member of the research team (27%). In contract research, data is often maintained by the client to whom it has been delivered.

Important published digital data sources

Archaeological online databases, online publications, and printed publications with supplemental data (e.g. a monograph with data on a DVD or an accompanying website) were each seen as "very important" by about half of the respondents. The favourite were online publications with supplemental data (67% "very important"). However, the feedback indicates that it is not the source as such that matters – it is the quality of the data contained.

Awareness and use of online repositories

Except for the UK-focused ADS, the level of awareness and use for online repositories appears to be rather limited, in particular outside the respective country. This presents an opportunity and risk for ARIADNE at the same time (see recommendations).

Most important criteria for data sources

Most researchers perceived the following three criteria as "very important"

- Data transparency: having a good overview of available data(sets) – 77%;
- Data accessibility: the required data(sets) are available in an uncomplicated way – 73%;
- Data quality: the available data(sets) are complete and well organised – 64%;

But only 30-35% of the respondents were satisfied with regard to these criteria (see general conclusions, 8.1.1).

International dimension

Having access to international data(sets) was perceived as "very important" by 45% of the respondents, much lower than the other criteria but still substantial. According to the pilot interviews one major reason for seeking data beyond national boarders is comparative research for broad synthesis or meta-analysis.

Language requirements for data/metadata

There is mixed evidence concerning the question as to whether data and metadata need to be available in different languages (other than English). While a majority of the respondents did not regard this as a major issue, some respondents recommended that this should be seriously considered (in particular Spanish and Russian were seen as important languages for an effective outreach).

The data managers' perspective

Most present types of data

Most present and also perceived as most important by the data managers are project archives, grey literature, various databases, and specialised bibliographies. Academic journals and series and PhD theses are less common information items.

Metadata quality is the major challenge

The major challenge data managers see themselves confronted with in their daily work is ensuring metadata quality. This was by far the most important item out of a list of six challenges. Further relevant challenges include managing a rising number of data sets, and changes in the regulatory framework.

Technical innovation and user-driven challenges (changes in user requirements, rising number of users) are perceived as much less critical by many data managers.

According to several comments received on this question, the sustainability of project-based repositories and costs for operation and further development might be another key issue (it was not included in the list of main issues).

Technology is not the (major) issue

Most managers have a rather stable data management environment, and mainly carry out upgrading and refinement of services, sometimes in response to external demands (e.g. new regulations). The main themes with regard to important recent technical developments (open question) include: database functionality, data standards and exchange protocols, spatial data (GIS), Digital Object

Identifiers (DOIs). According to the pilot interviews with data centre managers, there is an increasing demand for DOIs because researchers need to link publications with deposited data.

Three worlds of access rules

About one third of the repositories have an "open access" policy (no registration required) for all, most or at least some data sets, while another third grant access only on request (and not necessarily to all or most data). Furthermore there is a "shadow world" of access restricted for specific communities only. With regard to depositing there is a clear preference by most repositories not to allow uncontrolled deposit by anybody, but rather by request.

Perceived customer requirements

There are two requirements the data managers perceive as most important for the users of their repository, namely:

- Data accessibility: the required data(sets) are available in an uncomplicated way;
- Data quality: the available data(sets) are complete and well organised;

The first requirement contradicts the practise of access only on request, but probably not access restricted to specific communities – if one is part of it.

Changes in user needs

Two needs were perceived as increasingly important:

- Convenience in accessing and downloading data;
- Individual service and guidance.

Expectations for ARIADNE

The main areas where researchers face problems are finding and accessing relevant data. Therefore, major expectations or hopes for the ARIADNE project are that the resulting services can improve the transparency of what is available, the search capability and, possibly, the conditions of access (e.g. promote open access repositories).

Improving transparency: A central expectation is that ARIADNE should provide a broad overview of existing data resources, beyond the partners' resources. The current approach is the ARIADNE Registry (based on the DCAT standard), which could be a stumbling block for many holders of small as well as large and varied collections interested in making their resources visible through ARIADNE. In order to create a broad overview another way to "register" many data resources may be required. In general, geo-spatial, GIS and map-based overviews and access may be perceived as particularly suitable (cf. the approach suggested by the Excavations and Monuments Data SIG in Section 8.1.4).

Capability of cross-searching data repositories: We assume that this is one of the main advantages users of the ARIADNE portal will expect.

Improved conditions of access: This is not a technical requirement but a research policy objective, which the ARIADNE project can support by promoting open access principles as well as leading by example ("walk the talk").

Filtering "useful" and re-useable resources: Concerning data that are accessible online, researchers mentioned that they are sometimes not as useful as they could be, because data is structured in different ways, not up to date, incomplete or lack important details (e.g. how it was collected or processed). Moreover, a lot of data are not re-usable but "canned content" (such as data tables in pdf documents) or not available under an adequate license. Therefore, we believe that users of the ARIADNE portal would benefit from innovative mechanisms which allow a quick scan of data

resources as a fast way of assessing the usefulness (e.g. previews, snapshots, "look inside" functionality). Also some pre-filtering or ranking services might be provided (e.g. only openly licensed data) - see recommendations in Section 8.2.

Portal service portfolio and specific user requirements: Respondents suggested that ARIADNE should establish a new portal for data search. If such a new portal (on top of existing data resources) is established, users will clearly expect an added-value – i.e. it must have other or better features, or provide access to more resources. While an improved overview, cross-searching and filtering of data resources would be quite some progress on the current situation, the specific requirements are not fully clear, however. There will be further requirements once users have discovered "useful" data(sets). For example, if such data has been found users might wish to also find related publications from the researchers who have shared it. Therefore, as next steps we suggest having some "lead users" evaluate a larger number of portals to identify where and how various needs and requirements are met (or not met), including examples of existing innovative approaches, and further ideas on specific services or features the ARIADNE user community might appreciate (see Section 2.3).

Services for repositories and other websites: Thinking beyond individual users, ARIADNE should also be seen as a service for data repositories, other websites and for specific communities of practitioners. ARIADNE might help enrich services of underlying repositories for instance by suggesting (and providing) links to similar or complementary collections or individual items held by other repositories. For other websites, ARIADNE could provide RSS feeds on available new data that are relevant for particular subjects or geographic regions. An Application Programming Interface (API) would be useful for application developers seeking to combine (mashup) ARIADNE data/metadata with other information resources.

8.1.4 Conclusions and suggestions from the SIGs

Three ARIADNE Special Interest Groups (SIGs) provided initial conclusions on issues and suggested approaches in their fields of interest.

Grey Literature SIG

In archaeological terms "grey literature" means unpublished archaeological fieldwork reports. Access to such reports presents a significant problem in most European countries, which impedes the availability of important data for further research. Initial conclusions of the SIG on approaches and means for tackling the issue are:

- *Promotion of guidelines for digitisation and access:* Ensuring that grey literature reports can be collected within a research infrastructure requires elaboration and promotion of best practice in terms of digitisation and plans for access.
- *Metadata:* Definition of a core metadata standard for grey literature would be useful for project partners and affiliated projects who intend setting up grey literature databases.
- *Indexing tools:* Providing metadata for reports by hand is a time-consuming activity. Techniques for Information Extraction could help. However, documents will often contain tabular information and useful visual elements, which would require novel indexing and extraction technology.
- *Long-term preservation:* Ensuring that the reports, and the data they contain, are preserved may pose specific challenges (e.g. new reports in PDF format contain various embedded data).

Excavation and Monuments Data SIG

Documentation of monuments, excavations and field surveys is complex, with multiple data sets potentially attached to each record. Access to the documentation varies, but in general it is far from satisfactory (as confirmed by the online survey). A major step ahead would be enabling an overview of where work has been conducted and where the records – "born on paper" or digital – are held. Initial conclusions of the SIG on approaches and means for addressing the issue are:

- *Documentation of monuments, surveys and interventions:* Tools and guidance for transparent and comparable documentation based on international standards are required.
- *Data/metadata preparation for long-term curation and access:* This requires considerable work on formatting, metadata production, mapping, etc. To leverage the accessibility, raising of awareness is necessary so that the effort is considered in project plans and budgets.
- *Mapping of data:* The common denominator of monuments, surveys and excavations is point data that can be mapped and viewed online. In order to develop a GIS-based European portal to available documentation, tools should be provided which support the required conversion of national mappings to an international spatial reference system.
- *Map-based discovery and access:* This would allow users to understand at a glance what sort of information is available for a region or area, and provide entry points to available documentation on institutional data servers (e.g. field survey data, excavation project archives).

3D Data and Visualisation SIG

The SIG addresses the production and use of various forms of visual representation of archaeological entities, small finds as well as structures or an entire site. The online survey did not address the wide range of products in this field. Therefore the initial report of the SIG provides an important review of perceived issues and needs of partners that might be supported by ARIADNE e-infrastructure and services. The SIG identified many challenges in the creation and management as well as access to advanced visual media which are listed in section 7.3.2. Among the challenges that appear to be of particular relevance to ARIADNE are:

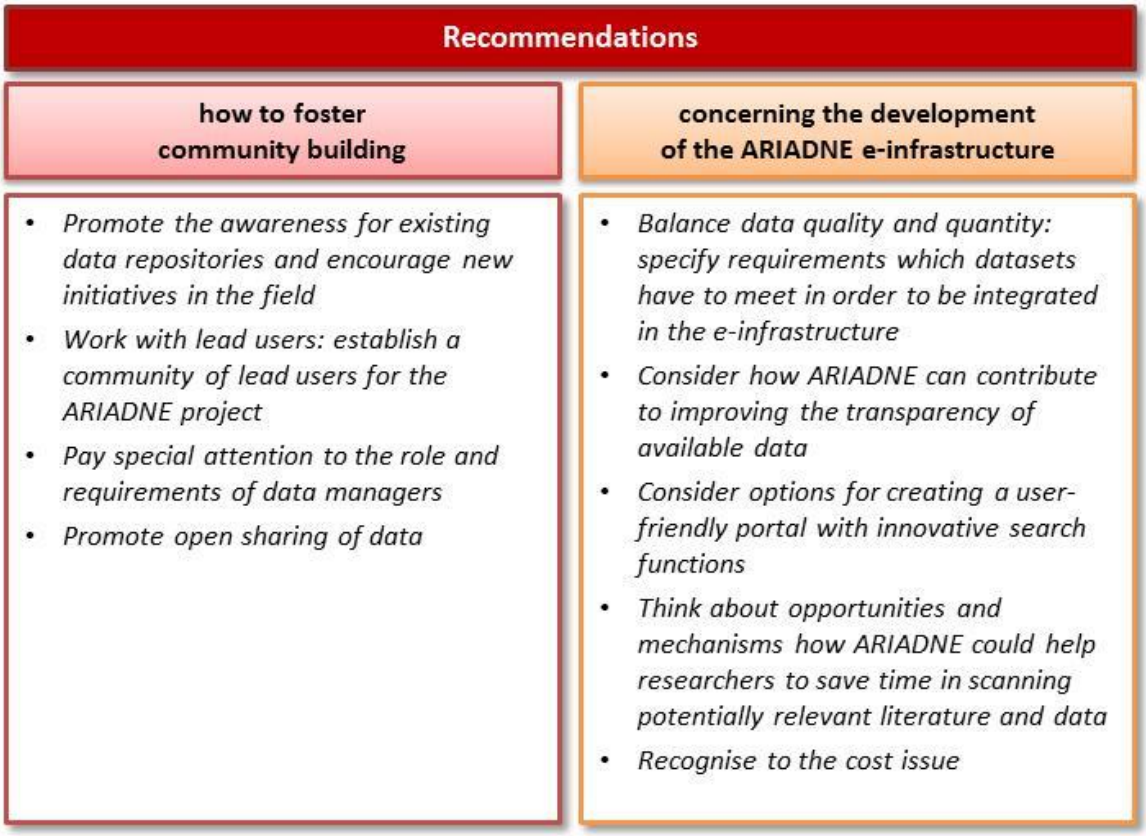
- *Standards and interoperability:* Many different formats are used and various metadata standards are under development (i.e. not tried and tested yet).
- *Web-based access and interaction:* High-quality visual media and models are still not easy to access and interact with for other purposes than just viewing and manipulating them in simple ways (e.g. rotation).
- *Re-use of data and models:* For complex files there is a need to have accurate documentation to enable effective re-use. Also promoting good practice with regard to matters of IPR and licensing for re-use is required.

SIG members conducted a ranking of different types of visual media in terms of their relative importance and impact:

- Image-based media: high-resolution images were ranked highest among five listed types of data.
- 3D: for single objects, improvement of web-based access was seen as a priority; for large scenes, support via OSG4Web was ranked as most important.

8.2 Recommendations for the development of ARIADNE services

The following recommendations are addressed, if not stated otherwise, to the ARIADNE project itself. This reflects the objective of this deliverable to facilitate evidence-based, informed decisions regarding the specification of the ARIADNE e-infrastructure and services. For each recommendation, we provide a short rationale (what is it about, what is the evidence on which it is based) and propose a set of specific, practical project activities how (and by whom) it could be addressed. We have grouped the recommendations into two categories which reflect the two main objectives of WP2: recommendations how to foster community building (Section 8.2.1), and recommendations regarding the design, offer and focus of the future ARIADNE e-infrastructure, derived from the analysis of user requirements (Section 8.2.2).



8.2.1 Recommendations with regard to community building

Promote the awareness of existing data repositories among the research community

The online survey results indicate that most of the existing digital repositories are not yet widely known among the research community, in particular outside the home country of the repository. The UK-focused ADS was the only internationally well-known repository (60% of the respondents said they had heard of it). Other repositories were known by up to 35% of the researchers interviewed. Even in the case of international repositories, awareness appeared to be limited to specific research domains or practices. This raises some issues for the development of ARIADNE services. On the one hand, this evidence can be framed as a case for an international integration of the (mostly) national initiatives. Integration could be exactly the added value that boosts the use of repositories for international research purposes and thus, ultimately, the awareness for holdings of other repositories which have previously not been on the radar of researchers. On the other hand, it also poses a certain risk for the acceptance of the ARIADNE e-infrastructure, namely whether aggregation *per se* will be sufficient to boost the acceptance and use of the services provided. We recommend

therefore that the ARIADNE project should also be used as a platform to raise awareness for digital repositories in general. In this context, the project could/should encourage as much as possible initiatives to establish new repositories, as these are an important foundation for the future integration of data.

Recommended activities:

- WPs 2 and 4: Develop an action plan how the ARIADNE project could be used as a platform to promote the use of digital data repositories among the international research community.
- WP2 (as part of the community building activities): Consider opportunities how associations, for instance AIAC could leverage their networks to promote digital repositories in Europe.
- Organise further workshops where digital repositories present themselves (the workshop organised in the context of the EAA Annual Meeting in Pilsen, September 2014, could serve as a model for organising a "promotional tour" with similar events at other events).
- WP 2 (liaison plan): Identify initiatives to set up further repositories, in particular in countries which do not yet have a major repository holding archaeological data, and offer cooperation and support – make them future contributors to the ARIADNE e-infrastructure early on.

Work with lead users: establish a panel or even a community of lead users for the ARIADNE project

"Lead users" (cf. van Hippel, 1986) are users of a product or service who experience specific needs months or years before the mass market will express the same needs, and who will benefit significantly from obtaining a solution to their needs. Translating this concept to the case of archaeological research and the use of data resources, lead users will be researchers who make intensive use of (cross-)searchable repositories in their daily work, and who have therefore specific needs and a genuine interest in developing solutions to these needs. Often, lead users are at the same time early adopters of new technologies and services in their field. We recommend therefore that the ARIADNE project should make every possible effort to identify such lead users and closely work together with them. Ideally, therefore, the project should try to establish a lead user community and a platform where ideas can be exchanged with this community and thus be "tested". This will help to identify detailed, specific user needs, indicate possible solutions to these needs, and help the technical project partners to develop solutions that effectively respond to these needs. Moreover, these lead users could serve as "ambassadors" for the project and the future e-infrastructure it develops. Lead users will play an important role, for instance, in the analysis of specific user requirements on the basis of the evidence shown in this report.

Recommended activities:

- WP2: Establish a focus group (or panel) of lead users which make intensive use of digital repositories and portals in preparation for the update to this report; conduct interviews or a workshop with such lead users to identify advanced user needs, gaps and possible suggestions for solutions to close the gaps. Coordinate this activity with WPs 12/13.
- WPs 12/13: Think about identifying and interviewing lead users of existing archaeological repositories and portals as part of conducting the technical user requirements analysis. Coordinate this activity with WP2 – consider synergies with the work of WP2.
- Research partners: Help WPs 2, 12 and 13 in identifying lead users and establishing contacts for them (there will be some lead users in organisations represented in the project consortium).
- Project management: Think about a mechanism for how lead users could get systematically involved in the (technical) development of ARIADNE services.

Pay special attention to the role and requirements of data managers

A crucial success factor for developing research e-infrastructures is to establish "common ground" among the key stakeholders involved; i.e. it requires close cooperation between researchers, data managers and technology experts. These parties need to achieve a common, shared understanding of objectives, design options, implementation, future use and overall operation. Literature suggests that the criteria for the fit for use (usefulness) should predominantly come from the research community ("demand pull") and not from a technological perspective ("technology push") – see 3.2.2). Within this demand-driven framework, special attention should be given to the role of data managers in e-infrastructures and service development. Their role should be emphasised, and appropriate recognition and career paths offered. As for ARIADNE, this means that data managers represent a highly important group of stakeholders who should be specifically addressed both with regard to exploring user requirements and as part of the dissemination and awareness raising activities. They can be important "ambassadors" for the ARIADNE project who promote it to the users of "their" data.

Recommended activities:

- WP2, Task 2.1: Conduct in-depth interviews with data managers for the preparation of D2.2 – explore their specific needs (as users) and recommendations towards research e-infrastructures.
- WPs 12/13: Consult with data managers when specifying the technical user requirements of the ARIADNE e-infrastructure.
- WP4 & Task 2.2: Think about mechanisms to liaise with data managers (Task 2.2) and how to promote the project specifically to them (WP 4), as data managers will be important ambassadors for the ARIADNE project.
- ARIADNE project as a whole: consider how e-infrastructures like ARIADNE could help data managers in their profession and possibly provide new career opportunities.

Promote open sharing of data

Initiatives in e-infrastructure and services must address data sharing practices heads on, because most researchers are reluctant to share their data, at least not "open access" and in re-useable form. This often goes hand in hand with data management practices with high risk of data loss. Research funders increasingly demand data management plans and open sharing of research products, though the impact will take quite some time and could be lower than expected, e.g. if the mandates are tooth-less concerning open data formats. These considerations are confirmed by literature (see Section 3.2.4) and by many comments and much feed-back received from respondents to the online survey, when asked about barriers for data sharing (see Section 6.2.5). The current system of incentives in academic research, as well as institutional requirements, are major barriers to data sharing. Researchers and institutes tend to regard "their" data (at least raw data) as an asset that loses value if shared with others.

As an Integrating Activity, ARIADNE is well placed to promote open sharing of data in the archaeology sector in Europe. Going beyond advocacy, this requires advice and support in matters of data management (targeted at sharing), effective metadata generation, licensing, data citation standards (e.g. DataCite) and, overall, trust building and recognition of data sharers.

Recommended activities:

- Project management: Think about a general strategy for how ARIADNE could/should position itself in this field, and how to promote data sharing – e.g. how can the project provide advice in relevant areas (such as data management targeted at sharing). According to the strategy, the project might then participate in selected European and international initiatives.
- WP4: Collect good & innovative practices in data sharing and disseminate them.
- WP5: Think about means and opportunities for how to promote data sharing in the execution of the transnational access and training activities.
- Partners managing an institutional repository (several) or national repository (ADS, DANS, SND): The activities of WP5 are targeted at individual (younger) researchers. Partners with expertise in running a repository could provide advice to other institutions on how to promote "open archives" (both in terms of policy as well as technical matters, e.g. Open Archive Initiative compliance).

8.2.2 Recommendations with regard to the development of the ARIADNE integrated e-infrastructure

Balance data quality and quantity: specify the requirements datasets have to meet in order to be integrated in the e-infrastructure (in particular with regard to "legacy data")

The ARIADNE project will have to carefully consider and specify the quality requirements for specific collections or data sets to be integrated in the e-infrastructure, so that the users regard the resulting services as valuable. In other words, the project needs to think about where and how "to draw the line". These criteria may be different for various types of data. In particular, ARIADNE will have to discuss how to deal with "legacy data". The project needs to specify practically applicable definitions and criteria regarding the data requirements. This recommendation is not a direct result of the survey or interviews, but is a logical consequence of the high importance attributed to data and metadata quality. When asked about needs with regard to data services, about 90% of the survey respondents said it was important that available data is "complete and well organised" (= data quality) and about 80% said it was important that the available datasets are well described (= metadata quality) – see Section 6.2.6.

Recommended activities:

- Project Management: Identify the WP or task which is best positioned to deal with the overall technicalities of the issue and can coordinate the development of general guidelines and requirements
- WP2, Task 2.2: Some of the Special Interest Groups should put this issue on their agenda and develop recommendations for specific types of data (depending on the focus of the SIG)
- Start an online consultation on this issue among the consortium members
- Optional: organise a dedicated workshop or meeting on this issue in the context of a forthcoming project meeting (e.g. in Istanbul in September)

ARIADNE as a means for improving the transparency of available data

Almost 90% of the researchers who participated in the survey confirmed (by agreeing fully or partially) that they often did not know what data is actually available, because research data are stored in so many different places (see Section 6.2.4). Moreover, when given a list of five basic needs with regard to research data, data transparency (having a good overview of available data) was the one which most researchers regarded as very important – but expressed rather low satisfaction with the current situation (see Section 6.2.6). In short, the researchers expressed a strong need for having a better overview of available data and thus confirmed the validity of the rationale for the ARIADNE project in an impressive way.

The question is how this lack of transparency (due to the enormous fragmentation of the data landscape) can be most effectively addressed. We suggest that ARIADNE should consider how to improve the overview of available digital data in Europe. This does not necessarily mean that the data will always have to be accessible directly through the integrated infrastructure, but it will already be an improvement over the current situation if relevant data sources are identified and described in a systematic way, with information on how data can be accessed. This might be easier to achieve than the full integration, but still create significant benefits for users.

Recommended activities:

- WPs 12/13: Consider options (data requirements, tools) for how ARIADNE could easily create an overview of digital data repositories and their resources, beyond the partners' resources (which are to be registered in the ARIADNE Registry based on the DCAT (Data Catalog Vocabulary) standard).
- WPs 12/13: Consider whether there could be different levels of "integration" in the e-infrastructure, which would make it easier to quickly integrate further data resources at least on basic levels and thus contribute to establishing a better market overview.

Acknowledge the importance of the "front end": consider approaches for creating a portal that improves considerably on users' current search options

Researchers would greatly appreciate (and probably expect from ARIADNE) services that remove their current difficulty when searching for data resources, i.e. using different repository websites with different interfaces but maybe little relevant material. Ideally, ARIADNE would not only integrate many different data resources, but improve considerably on current search options, experiences and outcomes. When asked about the usefulness of potential services from ARIADNE, the top-rated were portal services that make it *more convenient* to search for archaeological data stored in different databases (79% of the researchers) and enabling *innovative and more powerful* search (63%). While this needs to be explored further with regard to the details (both from a demand and supply perspective), it implies that users think about such a project mainly in terms of the front-end. We recommend that the ARIADNE project should devote special attention to the services and interfaces for the targeted users. What would be perceived as convenient and powerful when cross-searching data resources, in terms of search paradigm, interactivity, selectivity, navigation, for instance?

Recommended activities:

- WP2, Task 2.1: Explore user needs in this field in more detail, e.g. what important needs and expectations are currently not met by relevant portals, specific services that would be appreciated, etc.
- Technical WPs: Think of the "front end" not as something to be added when the technical implementation has been done, but put user-centred design and improvement on current services and interfaces high on the technical development agenda.

Think about opportunities and mechanisms for how ARIADNE could help researchers to save time in scanning potentially relevant literature and data

Researchers mentioned the lack of time to scan relevant literature and data as a major problem. Even if potentially relevant literature or data has been identified and would be accessible, they just do not have the time to review it. While this is a fundamental challenge which ARIADNE cannot solve, it might also present an opportunity for the project: any innovative mechanism that allows a quick scan of literature or data sets would probably be highly welcome and present a real asset for the integrated infrastructure.

Recommended activities:

- WP2: Explore user needs in this field in more detail and develop creative suggestions how innovative solutions could help to address the needs in a better way than existing ones (even if these "solutions" are technically hard to realise).
- Technical WPs: Think about mechanisms for how the ARIADNE infrastructure could provide an opportunity not just to access metadata and data, but offer a quick "first glance" of what the data actually looks like (previews, snapshots, "look inside" functionality, …).

Recognise the cost issue

The interviews and the survey confirmed that costs are a major barrier for data access, in particular with regard to digital resources. Nearly three quarters of the respondents agreed (fully or partially) that cost was "a major problem for access to online resources" (see Section 6.2.4 for details), for instance because single articles are often not available without a full subscription, or because prices for images or other items are too high. Of course, ARIADNE as a meta-infrastructure project does not have a direct impact on pricing schemes of the underlying institutions. However, the fact that costs are already a concern for many researchers should be carefully considered when developing a business model for offering ARIADNE services. Any fees charged to users could severely affect the acceptance and use of these services.

Recommended activities:

- When developing a business model for the ARIADNE e-infrastructure, be very careful in demanding fees for using the services (if necessary at all). Cf. Task 2.6 – Long-term sustainability plans.

Annex I: Roll-out plan for the ARIADNE Online User Survey

Overview of invitation mailings

Consortium members were requested to support the roll-out of the ARIADNE Online User Survey by sending out invitations to participate in the survey to their own networks and communities. The following overview describes the ARIADNE communities that were contacted by partners.

Consortium member	Promotion support provided (for communicating the Online Survey)
AIAC	AIAC (International Association for Classical Archaeologist) promoted the ARIADNE Online Survey through its mailing list to more than 2,350 contacts.
Institute of Archaeology ZRC SAZU, Slovenia	Has organised a public call through the "Rosa" mailing list followed by personal appeals. The estimate is that about a third of all archaeologists in Slovenia has been reached in this way.
Hungarian National Museum National Heritage Protection Centre (HNM NHPC)	Has announced the survey within the 1) Hungarian National Museum National Heritage Protection Centre (HNM NHPC) 2) Association of Hungarian Archaeologists (AHA). AHA includes field archaeologists, researchers, heritage managers etc. from several different archaeological institutions (museums, Institute of Archaeology of the Hungarian Academy of Sciences) and universities in Hungary.
Incipit CSIC	Has posted a note on the "Information Technologies and Cultural Heritage" LinkedIn group, asking members to participate in the survey.
University of York, ADS	UoY ADS has promoted the survey via its website, through social media and through its eNewsletter.
MDR	Announced the survey via the CARARE network, which includes national heritage agencies from more than 20 countries, asking the members of the network to distribute the survey to their network of archaeological contacts. Has also announced the survey via the LoCloud network which includes cultural heritage agencies (including some CARARE partners) and asked them to distribute the survey to their own networks of archaeological contacts.
DANS	DANS supported the roll-out by sending the invitation for the stakeholder survey by mail to • Dutch archaeologists/students registered in the electronic archiving system (EASY) • Teachers/professors Archaeology at Dutch Universities who are registered in the National Academic Research and Collaborations Information System (Narcis) • To the mailing list of contacts at Dutch archaeological organisations (commercial and academic) • The international community of dendrochronologists registered in the Digital Collaboratory for Cultural Dendrochronology (DCCD) The survey was also promoted on the DANS and EDNA websites.
SND	SND has sent out information about this survey to contacts among archaeologists at museums, universities and County Administrative Boards, the National Heritage Board and its different sub-organizations. SND also spread information about the survey via the CAA-se LinkedIn group and the CAA-se conference in November.
OEAW	Has sent the survey link to the ÖGUF mailing list, one of the most important archaeology mailing lists in Austria.
SRFG	Has sent out information about the survey to several archaeology newsgroups and listservs, such as newsgroups as listed on the following websites: JISCmail archaeology lists: http://www.jiscmail.ac.uk/mailinglists/category/V6.html http://www.stonepages.com/megalinks/mailing_lists/

	ArchNet: http://archnet.asu.edu/resources/selected_resources/newsgroups%20&%20listservs/topic.php
CulturaItalia	CulturaItalia, an Italian platform for archaeology and other cultural studies disciplines, e-mailed information about the survey to 2000 registered users, published an article about the survey on the website and informed about it via their RSS Feed and other social network channels. (http://www.culturaitalia.it/opencms/it/contenuti/focus/Ariadne__sondaggio_online_sull_accesso_ai_dati_archeologici.html).

Specific mailing lists used (details)

ID: 01	
ROSA (Slovenia)	
List is owned and maintained by	Slovensko arheološko društvo (Society of Slovenian archaeologists)
Person authorised to use the list	Members (to become a member another member's endorsement is required and a member needs to be actively involved in archaeology)
Number of people	200
Geographic focus	Slovenia
Thematic/domain focus	Important events pertaining to Slovenian archaeology
Other information	Since Slovenian is a small archaeological community the reach of the mailing list is about 90% of all archaeologists and is very effective vay of communication.

ID: 02	
Internal mailing list of the HNM NHPC	
List is owned and maintained by	HNM NHPC
Person authorised to use the list	anyone within the institution
Number of people	about 100 people
Geographic focus	all Hungary
Thematic/domain focus	- researchers - archaeometry specialists - artefact data specialists - field archaeologists - site survey specialists - heritage management specialists
Other information	The majority of archaeologists are joined to AHA, therefore distributing the survey within the AHA and HNM NHPC may result in some cross-posting. Nevertheless we can reach the most people in Hungary by asking AHA to distribute the survey. The survey doesn't need to be translated into Hungarian, it should be fine in English.

ID: 03	
Internal mailing list of the Association of Hungarian Archaeologists (AHA)	
List is owned and maintained by	AHA
Person authorised to use the list	AHA designated person (to be identified)
Number of people	about 600 people
Geographic focus	all Hungary
Thematic/domain focus	researchers, archaeometry specialists, artefact data specialists
Other information	The majority of archaeologists are joined to AHA, therefore distributing the survey within the AHA and HNM NHPC may result in some cross-posting. Nevertheless we can reach the most people in Hungary by asking AHA to distribute the survey.

ID: 04	
"Information Technologies and Cultural Heritage" LinkedIn group	
List is owned and maintained by	Cesar Gonzalez-Perez, Incipit CSIC
Person authorised to use the list	Cesar Gonzalez-Perez, Incipit CSIC
Number of people	2633 individual members as of 29 October 2013; grows at about 10 members per week
Geographic focus	Global
Thematic/domain focus	Information technologies and cultural heritage

ID: 05	
ADS Website	
List is owned and maintained by	UoY ADS
Person authorised to use the list	Webmaster of ADS Website
Number of people	Homepage has over 3,500 unique page views per month
Geographic focus	Worldwide, but primarily UK
Thematic/domain focus	Archaeology

ID: 06	
CARARE Mailing List	
List is owned and maintained by	MDR Partners
Person authorised to use the list	Kate Fernie, Sheena Bassett
Number of people	70+ people from 27 organisations
Geographic focus	Denmark, UK, Italy, Cyprus, Malta, Belgium, Brussels region, Netherlands, Poland, Germany, Estonia, Slovakia, Greece, Slovenia, Iceland, Lithuania, Bulgaria, Spain, Romania, Czech Republic
Thematic/domain focus	Archaeology and architectural heritage

ID: 07	
LoCloud Mailing List	
List is owned and maintained by	MDR Partners
Person authorised to use the list	Kate Fernie, Sheena Bassett

Number of people	70+ people from 30 organisations
Geographic focus	Norway, Austria, Belgium (Flanders), Bulgaria, Croatia, Cyprus, Czech Republic, Demmark, France, Germany, Greece, Iceland, Ireland, Italy, Lithuania, Netherlands, Poland, Portugal, Spain, Romania, Sweden, Turkey and UK
Thematic/domain focus	Local heritage

ID: 08	
ARIADNE Stakeholder Survey database by DANS (mailing list established specifically for this purpose through selection from the DANS contact database)	
List is owned and maintained by	DANS
Person authorised to use the list	Lucas Pasteuning : Lucas.pasteuning@dans.knaw.nl
Number of people	about 1000 archaeologists, dendrochronologists, commercial/non-commercial, professionals, researchers, students, universities
Geographic focus	The Netherlands and mainly Europe for dendrochronology
Thematic/domain focus	Archaeology, Dendrochronology

ID: 09	
SND internal mailing list	
List is owned and maintained by	SND – Swedish National Data Service
Person authorised to use the list	Liaison person for mailing: Ulf Jakobsson
Number of people	unknown
Geographic focus	Sweden
Thematic/domain focus	Archaeologists at museums, universities and County Administrative Boards, the National Heritage Board and its different sub-organizations.

ID: 10	
CAA-se (LinkedIn Group and Conference)	
List is owned and maintained by	n/a (LinkedIn)
Person authorised to use the list	Liaison person for mailing: Ulf Jakobsson
Number of people	
Geographic focus	Sweden
Thematic/domain focus	Archaeology
Other information	Information to be posted at LinkedIn Group and to be announced at CAA-se Conference in November 2013

ID: 11	
ÖGUF - Austrian Society for Pre- and Protohistory	
List is owned and maintained by	Austrian Society for Pre- and Protohistory
Person authorised to use the list	Jakob Maurer: jakob.m@gmx.at
Number of people	about 280
Geographic focus	mainly Austria
Thematic/domain focus	prehistoric archaeologists and related sciences, also laymen (about 60-70% are archaeologists)

ID: 12	
AIAC (International Association for Classical Archaeologist) Mailing list	
List is owned and maintained by	General Mailing List; Rome Mailing List
Person authorised to use the list	AIAC, Helga Di Giuseppe - Project Manager for Fasti Online
Number of people	about 2,350 people, covering hundreds of institutions
Geographic focus	Italy, USA, Great Britain, France, Switzerland, Spain, Belgium, Austria, The Netherlands, Germany, Albania, Romania, Bulgaria, Morocco, Japan
Thematic/domain focus	Archaeology, History, Philology, Geology
Other information	ur mailing list includes Institutions and Organization linked with archaeologist research, such us University, Soprintendenze for Culture Heritage and relative Ministery, Culture Associations, cooperatives, societies. People in the mailing list are Professors, scholars, independent researches, archaeologists representing the Soprintendenze, archaeologists in general

ID: 13	
EAA Mailing list	
List is owned and maintained by	EAA – European Association of Archaeologists
Person authorised to use the list	Manager of mailing list: Sylvie Květinová: kvetinova@arup.cas.cz Contact to EAA via DAI: Friedrich Lüth, President of EAA (from DAI)
Number of people	1100
Geographic focus	European/world-wide (over 40 countries)
Thematic/domain focus	The EAA currently has over 1100 members on its database from 41 countries world-wide working in prehistory, classical, medieval and later archaeology. They include academics, aerial archaeologists, environmental archaeologists, field archaeologists, heritage managers, historians, museum curators, researchers, scientists, teachers, conservators, underwater archaeologists and students of archaeology.

ID: 14	
CulturaItalia – mailing list and various social networking channels	
List is owned and maintained by	N.N.
Person authorised to use the list	N.N. (contact point was Sara Di Giorgio)
Number of people	Tweeted it to our 1450 followers Posted it on Facebook to our 9.842 fans Posted it on Linked-in to our group with 2693 members Posted on Google+ to our 240 followers Emailed it to our 2.000 registered users
Geographic focus	Italy
Thematic/domain focus	Archaeology and related sciences
Other information	Link to article on website: http://www.culturaitalia.it/opencms/it/contenuti/focus/Ariadne_sondaggio_online_sull_accesso_ai_dati_archeologici.html

Annex II: Questionnaire for researchers and directors of institutes

(see separate document)

Annex III: Questionnaire for repository managers

(see separate document)

Annex IV: Pilot interviews

This annex comprises the detailed summarization of 26 interviews with different stakeholders and user groups. The first two sections cover 21 interviews and a more general summarization of the results is included in Chapter 5. The processing of 5 interviews took too long for them to be included in the analysis. These interviews are documented in the final section.

A.IV.1 Directors of research institutes, researchers and project data managers

1.1 List of participants

The following persons were interviewed or provided written input in the interview template which is included in Annex IV. All input has come from members of project partners, except of the joint input of two researchers of BIAX Consult (which has been counted as one unit of analysis).

- Edeltraud Aspöck, researcher, ÖAW - Institute for Oriental and European Archaeology (OREA), Austria;
- Michael Ann Bevivino, research assistant, Late Iron Age and Roman Ireland (LIARI) Project, The Discovery Programme, Ireland;
- Edel Bhreathnach, CEO, The Discovery Programme, Ireland;
- Emanuel Demetrescu, fellowship researcher, CNR-ITABC, Italy;
- Ger Dowling, assistant project director, The Discovery Programme, Ireland;
- Elizabeth Fentress, independent scholar, AIAC, Italy;
- Kirsti Hänninen and Caroline Vermeeren, researchers, BIAX Consult, Netherlands;
- Marc Haendel, database manager and excavation technician, ÖAW - Institute for Oriental and European Archaeology (OREA), Austria;
- Sorin Hermon, assistant professor, The Cyprus Institute, Cyprus;
- Michaela Lochner, head of the research group Urnfield Culture, ÖAW - Institute for Oriental and European Archaeology (OREA), Austria;
- Susan Lyons, PhD candidate, Department of Archaeology, University College Cork, Cork, UK;
- Anja Masur, researcher, University of Innsbruck, Austria;
- Simona Simionescu, researcher, ARHEO VEST, Romania;
- Máté Stibranyi, researcher, Hungarian National Museum, National Heritage Protection Centre, Hungary;
- Ingelise Stuijts, company environmentalist, The Discovery Programme, Ireland;
- Benjamin Štular, researcher, ZRC-SAZU, Institute of Archaeology, Slovenia;
- Despoina Tsiafakis, researcher, ATHENA, Greece;
- Jacqueline Cahill Wilson, project director, The Discovery Programme, Ireland;

In order to keep the referencing with the interview protocols, in the sections below statements are referenced with [number]. These numbers do not correspond to the alphabetical listing of the interviewees above.

1.2 Search for relevant data, main data sources used

Questions: When working on research projects: Where do you and/or your research group typically search for relevant data? What are the main data sources you are using?

The researchers mentioned very different data sources which they consider as the main or most important sources for projects in their fields of research. Some researchers focused on their specific research, while most addressed the availability of data sources for their research group or the situation in general.

Most often mentioned were academic/scientific publications "for traditional literature review" (scholarly monographs, books, conference proceedings, print and electronic journals, "university libraries", "sometimes online library-search", "bibliographic references") and unpublished reports (excavation, field surveys etc.). [1, 4, 6 7, 8, 9, 11, 12, 14, 15, 18]

As described by one researcher, when looking for available literature in the initial stage of a project online search is very important, however, "in later stages, when analysis of specific national data and findings is conducted, I need to spend more time physically in the library" [14]. Another researcher mentioned: "We also use the internet, but in the field of archaeology, the informatical publications are not that widespread" (one might find relevant titles but not the actual article or report) [13].

Also another researcher noted: "However, the relevant literature is hardly available (directly) in public libraries (of universities etc.), but rather in specific digital documents which can be stored in very different places." The researcher also noted the importance of data made available from team members internally which would not be available otherwise [15].

Notably, publications were also considered as "data", i.e. "for qualitative analysis" [7] or "information and comparanda on data we are looking for" [13]. Consequently the "own archive and that of others" [12], "unpublished archives of individual practitioners and archaeological companies" [10], information of "private archaeological consultancies" [4], and also the "memory of colleagues" [1] are important sources for relevant information.

In a situation where a lot of "data" is contained in unpublished reports, personal contacts are of course very important. The same applies to "raw data" (i.e. data not summarised in tables, charts, etc.) As one researcher noted, "if we are lucky enough, or if I know the excavator, I can get to see them (…) the personal contact matters very much" [13].

However, major sources are monuments authorities, e.g. the Federal Monuments Office (Bundesdenkmalamt), Austria [14, 15]; the "site monument records maintained by the counties" in the UK (which are provided on request) [14]; the National Monuments Service of the Ministry/Department of Environment, Heritage & Local Government, Ireland [10].

One researcher thought that the research for relevant data "would depend on the category of materials searched for. Amphorae people search for online amphorae databases, coins for numismatic ones, etc." [3]. However, most researchers typically would look for "data from other excavations, data from results of other archaeological research projects (as documented in publications and research reports), including GIS data so that we can explore distribution patterns (e.g. for pottery), field survey data, and laboratory analysis (as much as we can get, this becomes more widespread)" [13].

In some cases researchers would also consider museum collections, "to look at physical anthropology, faunal remains, lithic raw material" [7]. However, with regard to museum collections one researcher noted: "sometimes we have no idea where they come from but they are there, so we use them" [13]. Also mentioned were "deposits" held by regional museums, departments of antiquities [9, 11], National Museum of Ireland Archives [10].

Websites, Internet sources (such as online museum catalogues), and "online databases (where available)" were considered as one important source of information among others [4, 6]. Concerning online databases, two researchers in Austria thought that there are "no open access databases" published in their fields of research which could be used for quantitative analysis. For quantitative analysis the researchers "rely on the data they produce (current excavation and previous excavations)" [7], cf. [8]. This also applies to the field of 3D reconstructions, where "the main data sources are 3D survey raw data (point clouds), images and texts as references for 3d reconstruction at detail (e.g. statue), architectural (e.g. temple) and territorial level (e.g. a city and its surroundings)" [2].

Particular databases where mentioned by researchers in Ireland, namely the National Roads Authority Archaeological Database and WODAN: Archaeological Wood & Charcoal Database". The Late Iron Age and Roman Ireland (LIARI) project of The Discovery Programme involved setting up a database of information collected from "a variety of public and private institutions that held datasets relating to the immense number of archaeological excavations that had taken place through development and infrastructure programmes around Ireland over the past fifteen years" [5]. The database contains relevant details of excavations, finds, new radiocarbon dates, etc. and has been combined with details of burials and available scientific data such as dates and isotope analysis, e.g. from the Mapping Death project.

1.3 Importance of digital repositories

Questions: How important are digital repositories as a source of data? For which kind of data in particular do you use digital repositories in your search?

Several researchers considered digital repositories to be "invaluable sources", "very important" and similar characterizations [3, 4, 6, 9, 10]. Mostly this related to data of excavations, for example: "excavation results, dating databases, etc." [4], "especially those with information on unpublished excavations" [6], "excavations that publish online are invaluable sources of data; if there were an Italian version of OASIS we would certainly use that" [3]. One researcher also considered that digital repositories "are paramount for centralising metadata and allowing for a consistent and standardised model to be used with entering or accessing data" [10].

In some cases, databases are available, for example: "For macro-remains we use RADAR [Relational Archaeobotanical Database, NL], for wood there is no system yet, it would be very important" [1]. "Some colleagues in the project have access to specialised database, e.g. faunal remains" [7]; "There is a pollen database with literature in Ireland that I use, www.ipol.ie; this gives info on Irish locations for pollen and publications. Eventually I would look for raw data" [12].

Yet, overall the situation was considered to be unsatisfactory, for example:

- "there are no digital data repositories in the field (…) "the only thing accessible online are electronic journals"[7];
- "so far there are very few and dispersed repositories for such data (e.g. describing the process of reasoning when publishing a conclusion of a research result – primary data for example)" [9];
- "repositories would be important, but wood/charcoal data are not available in a general repository yet" [12];
- "one large project (Thunau) has a research database which is accessed by various researchers who worked on parts of the project. However, the database is only accessible by the participating researchers" [8].

One researcher summarised why digital repositories are not so much used at the moment: "There are not many repositories; if there are, they do not contain much information and relevant data; access problems: access is either limited to certain people or only specific data/information can be accessed." Therefore the researcher emphasised: "If digital repositories are used, it is often in combination with personal contacts: first, the repository is accessed to search for data; if potentially relevant datasets are found, the holder of the data is contacted to obtain further information about the possibilities to obtain the data. Usually this works." [13]

While common, domain or subject-based digital repositories seem to be often missing in the relevant fields of research, research groups and projects of course often develop datasets or collections of digital items (e.g. "we are just making a digital repository mostly for asset reuse purposes (specifically 3D models)" [2]).

As one project director explains, datasets "[These] are essential and form the basis of comparative analysis for our research and enable us to use our specialised software and skills of our technical staff to plot and map highly detailed distributional analysis against topographical and ordnance survey data, to gain a much clearer picture of both the local, regional and national occurrence of archaeological sites and finds." [5]

1.4 Types of data searched for and/or generated in projects

Questions: What type of data do you and your research group typically search for in preparing and carrying out your projects? And what type of data do you and your research group typically generate with your projects?

Type of data searched for	Importance to find/use, own production, and availability
Excavations:	***General assessment*** Generally considered as "very important", "essential" or "crucial"; for example: "synthesis of all relevant information gleaned from excavations" [4]; "context, sample and stratigraphy information essential for understanding and interpreting datasets" [10]. ***Produced also by the researcher (or research group)?*** Eight researchers (or research groups) did not produce excavation data [1, 2, 6, 8, 9, 15, 17, 18]. For others it was the main or an important part of the work of their research group [3, 4, 5, 7, 10, 11, 12, 13, 14, 16]. ***Comments on availability:*** "in most cases not available" [2]; "there are very few such data published entirely" [9, cf. 11].
Field surveys:	***General assessment*** Considered variously as "crucial" [2], "essential" (5] "very important" [7, 8], "important" [4, 6], though not for archaeobotanists: "not important" [1], "mostly no ecofacts retrieved" [12]. One researchers considered the relevance of field survey data as "moderate – can be helpful in identifying site types and extant of sites for preparing potential sampling strategies" [10]. ***Produced also by the researcher / research group?*** For eight interviewees field surveys were an important part of their work [2, 3, 5, 6, 7, 11, 16, 17], seven did not mention field survey data [4, 8, 9, 14, 15, 10, 18], and for three [1, 10, 12] such data was not particularly relevant.

	Comments on availability: "Important - but lack of absolute dates in surveys of this type reduces its significance for the project" [4]; "there are very few complete data on the topic" [9, cf. 11], "we manage the data collection by ourselves" [2], "only when publishing a survey" [3].
Laboratory measurements & analysis:	***General assessment*** Mostly considered as "very important" [4, 7, 8, 9, 10, 12], or "important" [1, 3, 5, 6]. For example, "basis for comparison results" [1], "particularly for thin-sections" [3], "synthesis of all relevant absolute dates" [4], "the basis of the interpretations" [12], "particularly in relation to absolute dating methods" [4], "esp. for isotopic analysis" [6], "these are additional analysis that supplement the traditional archaeological approaches" [5]. ***Produced also by the researcher / research group?*** For most such data was an important part of the work of their research group or individually [1, 4, 5, 7, 9, 10, 11, 12, 15, 16, 17]. Six researchers / research groups did not produce such data [2, 3, 18; 6: "use sometimes", 8: "pays someone to do analysis (e.g. human anthropology, animal bone)"; 11: "we do not have the means for large scale use of such data". Examples included: "Weight, measurements, pictures, drawings, recording sheets, excel sheets with names" [12]. "There is a paleo-environmentalist on the team who produced analysis of elements from excavations such as bones, pollens, charcoals and wood" [16]. ***Comments on availability:*** "these are usually published in scientific articles" [9], "few data is published entirely and coherently [11], "we have a database which is only accessibly to members of the institute" [15].
Data for model-based computing, simulation, etc. (i.e. "in silico" research):	***General assessment*** Mixed results: Several researchers saw such data as not important or not used often [3, 4, 6, 8, 10]. Others commented: "still in its infancy" [11], "sometimes models" [1], "may be used for visually attractive illustrations" [12] or "yet to be defined how to present it" [9]. But quite some researchers considered it as important [2, 5, 7, 9, 11, 16], for example "important – data on dispersal of modern humans, there is much from the Neolithic for comparison" [7]. ***Produced also by the researcher / research group?*** According to the general assessment such data was not produced and only seldom used many researchers / research groups [1, 3, 4, 6, 8, 10, 12, 17, 18]. Two exceptions were: "essential and produced alongside our colleagues who are specialists in LiDAR, photogrammetry, and GIS modelling" [5]; "we work with GIS data analysis for landscape reconstruction (…) raster and vector data from GIS analysis" [2]. ***Comments on availability:*** No comments, however we can assume that not many databases are openly available that would allow for model-based computing and simulation.

Results of data mining for identifying patterns or interesting outliers:	***General assessment*** Mixed results: Either not relevant or not used [1, 7, 8], or "important" or "very important". Some examples. "In landscape analysis we look for settlement patterns" [2]; "Essential and this allows us to create a much more detailed distributional analysis using comparative studies but building clearer models in-house" [5]; "Use very often; important" [6]; "Very important - to identify trends and patterns in the data as well as merge with other existing datasets is essential for interpreting regional or landscape projects" [10]. ***Produced also by the researcher / research group?*** Most researchers / research groups did not produce such data. Two researchers understood it as "the result of a process – it is important to understand the process itself and what is the initial data used" [9, cf.11], whereas one saw as "essential for highlighting trends/patterns and unknown unknowns" [10]. ***Comments on availability:*** Some researchers considered it as "very important", however, "if such results were easily accessible" [4]; "practically inexistent so far online" [9, cf.11], "but this aspect has not been extensively used. Usually through exchange with colleagues or literature research, or comparing own data" [12].
Any other type of data:	***General assessment*** Other types of data mentioned were: "Any data of potential relevance is important" [4]; "Data describing reasoning processes in archaeology" [9]; "Very important are collections for comparison" [7]; "artefact studies" [18]; "Geomorphology, geology and vegetation data" [12]; "A very important area is the philological 3D reconstruction to obtain 'validated' models" [2]. ***Produced also by the researcher / research group?*** We assume that some of the data mentioned above are also produced by the researchers / research groups. Explicitly mentioned were: "Mostly 3D models from 3D survey and 3D reconstruction" [2]; "Data from 14C dating and carbon and nitrogen isotope analysis is useful for reconstruction of past diets in populations. This is an area of research which we are building with the LIARI project" [5]; "Bibliographical research, historical data, maps, etc." [11]; "Discovery Programme research projects normally incorporate large amounts of data in design-specific databases. These databases include images, maps and texts." [16] ***Comments on availability:*** We assume that some of data is also made available online; though one researcher mentioned "most data online as of today regards administrative matters or is incomplete" [11].

1.5 Gaps and problems when searching and accessing data

Question: What are the main gaps and problems you are experiencing in your research work with regard to searching or accessing data?

On this question some exemplary challenges were mentioned: Restricted access to data (limited access, costs of access, legal barriers), lack of data quality (e.g. missing metadata), or language problems. Very often personal contacts rather than formally established mechanisms enable access to required data.

Not accessible or only in a very restrictive form:

Most comments of the researchers related to the accessibility of data/databases. Some exemplary statements were:

- "Biggest issue: to find relevant data – we do not really know what is actually available" [13].
- "Restricted access to data, particularly accessing the results of archaeological excavations that remain unpublished and/or are inaccessible (e.g. in the hands of private archaeological contractors)". [4]
- "Required documents cannot be accessed, because they are deposited in a private collection of a researcher; often documents and data only become accessible when a researcher dies. Often, access is based on informal matters: you need to have the right contacts, you need to be established." [14]
- A registry of about 70,000 sites (descriptions with references, no reports or data) is available at the State authority [Hungary]; but this not openly accessible (only a few people, including the interviewee can access it); the people at the authority are "scared" about providing information about sites in an open manner, there is a strong resistance to this. [17]
- "Metadata of reports may be available, for example about reports scholars since the 19th century used to send to museums in Hungary, mainly the national museum, though older reports are often missing. (…) There is a huge amount of grey literature that is not accessible online, this needs to be digitised, we should work on that." [17]
- Two major databases of sites exist in Slovenia: one with information about some 10,000 sites, but one has to apply to get access; the other covers medieval sites which is also not open access. [18]
- "Sometimes it is difficult to get access to material held by museums for artefact studies. A lot of information is held privately by excavators that are retired or already died, and it is often difficult to find records" – "this should be solved on the national level, a repository of reports and data should be established" [18].
- "There are databases, but they are not accessible online or via a formal way" (e.g. rarely for site formation). [7]
- "There are no online sources. Info about databases is missing – where is which database with which content. (…) So, there is restricted access to data. If there are databases, they are not online. Access only possible via informal ways, e.g. ask personally." [8]
- "The archaeological available open data from the Italian public administration is the major issue in our research work" (e.g. relevant GIS is not easily available to researchers; lack of a repository to manage 3D models or related metadata). [2]
- "Indeed restricted access to data is a serious issue. Partly because of on-going research that may stretch many years, the length of publication time before material becomes available, retrieval of permissions from licensed excavators/companies, lack of access to literature including theses if not linked to a university." [12]

It was also noted that "there are some individuals who put parts of their research data online" [7]. Furthermore, one research director mentioned "that colleagues in both Ireland and the UK and Europe generally are happy to share data for comparative purposes"; this applied to both public and private institutions "if intellectual property rights are respected". The director also noted: "In the absence of full publication by individual authors we have to ensure that we do not pre-empt that opportunity for individual researchers" [5]. One researcher mentioned, "at the institution the policy is to help researchers interested in some data". [18]

Lack of data quality and usefulness

Where data/datasets are accessible online the researchers perceived a lack of usefulness because the data are incomplete or lack important details. Some exemplary statements are:

- "Really the lack of useful data online. For, say, comparative villas in Italy there is the Fasti Online, but for anything significant one still needs recourse to books." [3]
- "Incomplete datasets; online databases that aren't kept up to date (this is a big problem)" [6].
- "The main problem is the variability and inhomogeneity of data content and structure." [7]
- Lack of standardisation in "methods and practices which creates difficulties in comparing like with like and can lead to long delays in the re-formatting or re-processing of data to ensure comparability". [5]
- "Lack of data quality, metadata, incomplete information, amount of data in digital format compared to printed (data in digital format represents perhaps 5% of all data produced during an archaeological process)." [11]
- "In archaeological publication often only summaries of our research area are given, therefore we miss important information" [1, archaeobotanist].
- "Majority of analysis is confined to commercial works, so many aspects of analysis was not carried out due to budget/time constraints." [10]
- "Lack of details on how data was collected – it is difficult to assess the quality of data published online – for example how assessment were made when assigning an object to a particular typological category, or what were the methods of excavation and recording of data." [9]

Availability and quality of metadata

A few researchers specifically mentioned lack of or insufficient quality of metadata, e.g. "Missing metadata esp. from archaeological sites is an issue especially in the past" [12]; "lack of metadata in datasets" [13]; "lack of standardisation in meta-data" [5], "metadata is scarce" [18].

Language

Only few researchers mentioned that language is an issue for accessing and using data/datasets. For example, "Some languages are difficult (French, Spanish, Danish), but the basic data/tables are mostly understandable" [1]; "language is an issue in many cases (photographs help here)" [13]; "Language: in general not such an issue, maybe with languages such as Russian, Bulgarian etc. but mostly not immediately relevant. If raw data are available they can overcome languages such as measurements and Latin names" [12]; "in Hungary all information is in Hungarian, reports, etc. nothing in English" [17].

Costs

Also costs were addressed only by a few researchers. For example, "Restricted access to articles, being a private company not related to a university" [1]; "costs of access (especially to reference sources, like JSTOR)" [6]; "Often, data or other resources would be available online, but not for free, and often a subscription would be needed. However, I do not have the means for subscribing to a journal, in particular if I need only one specific article." [14]

1.6 Storage and management of data in collaborative projects

Question: In case of a collaborative project: where and how is this data stored and managed?

The researchers were asked to describe the typical practice, for example, if the data are

- (1) stored locally on computers of individual researchers and then shared through sending files directly to colleagues?
- (2) on servers with restricted access or on a repository of your local research group, unit or institute?
- (3) or stored in an open access institutional or subject-based repository or data centre?

The typical practice appeared to be storage and management on computers of individual researchers and on restricted access servers. Some statements were:

- "Mostly the second option, sometimes the first option" [1]
- "Normally second option. Never third option." [7]
- "First two options, for collaborative projects there are shared netdrives for everyone who works on the project." [8]
- "Local storage, personal computer of researchers; in collaborative work data is shared on USB sticks, servers have been used seldom. We know that this is an improper way of data management." [18]
- "As for my PhD: on my computer, and in back-ups. As for institutional research: we store data on the institutional server and specific members have access. Access restrictions depend on the project." [14]
- "Mainly individually" [15]
- "Data are stored on servers of the institute with restricted access (specific researchers or groups, internal only)." [13]
- "Mainly kept on our computers and server, e.g. field survey data, shape files, geophysical results and a few LIDAR data (…) only PDFs are shared, but we should also make raw data available." [17]
- "The data is stored on the organisation's server and the administration of the data is maintained by a central technology unit." [16]
- "Our data sets are stored locally on in-house servers on shared access drives. It is hoped that we will be able to make our data-sets available for on-line access for researchers over the next two years" (similar model as used for the "Mapping Death" on-line database and the WODAN open access database)." [5]
- "The data is stored on a cloud service (a linux based, virtualized, RAID6, 5Tb, server of our institute with restricted access we manage)" (related to 3D data, virtual reconstructions)." [2]
- "Data for my most recent projects (Villa Magna and Utica) is stored on a GIS database, ARK (LP archaeology), available on-line but password protected. On publication the ARK database will become available, although the publisher is not yet certain: there is a major reluctance to publish on-line catalogues in any format but HTML." [3]
- "Data generated by the collaborative research is typically stored locally on the computers of individual researcher and then shared later." [4]
- "Usually (1), files are shared through internet when working on a publication on a site-by-site basis." [12]
- "Data rarely/only in special cases stored locally on computers and then sent – this is only if someone works on a special case and develops a new approach (e.g. a new find category that does not fit into current database)." [7]

1.7 Responsibility for storing, archiving and maintaining data

Question: Who is responsible for storing, archiving and maintaining the data?

Most researchers considered that the storage and archiving is the responsibility of the research group or team, i.e. each of the individual researchers who work on a project.

- "Data is stored on the project database, the maintenance of which is the responsibility of all team members". [4]
- "Supervisors, finds specialists, myself. After each day in the field at least two hours are spent registering data from site, including contexts, drawings, finds etc." [3]
- "At the moment individual team members who are updating and inputting information on a rolling basis as this becomes available. The nature of the data set is such that it will always require maintenance and inputting to keep it up to date with all the most recent research." [5]
- "All members of the project team who are working with the data are responsible for storage and archiving, but the IT department of the Discovery Programme looks after the technical side of things." [6]
- "Each researcher that produces data will fill the necessary metadata and will archive it in our repository." [9]
- "Storage and archive is the responsibility of each individual user." [10]
- "The institute does not request data management plans, I have included it in grant applications with the objective of open access datasets, but no one is produce as yet." [18].

Though, in many cases the overall responsibility for the data, particularly concerning the maintenance is with one researcher or research director. For example,

- "The scientist in charge of the project is also responsible for storing such data." [11]
- "Analysts produce tables and store them in our central computer (second option), project leader stores them per project, makes report and sends this (sometimes with data) to client." [1]
- "Data generated by fieldwork (especially geophysical data) is the responsibility of the principle surveyor, who initially stores it on the organisations fieldwork computer before transferring it to a second computer for processing and later analysis." [4]
- "In our group two persons are involved in data management: me and another researcher. When a project is finished, all the data are moved to the long lasting FTP server storage." [2]

In an excavation project the data is collected by the research team but the excavation director supervises the input of the data and final reports in the project database [13]. More specifically, in the case of excavations conducted by researchers of the ÖAW the process is: "1. Input of field data during excavation; 2. Distribution of data to the responsible team members for further processing; 3. Team members can enter data in the data record present, they add information, correct; they have read and write options, but cannot create or delete data records; 4. M.H. controls data; able to create/delete; MSQL system on server is maintained by the IT services of the OEAW; they are responsible for the backups." [7]

Similar complex data handling can be found in the case of collaborative databases, for example, WODAN [Archaeological Wood & Charcoal Database], where the responsibility is with the individual researchers: "Data are written in paper format/tables, calculated and after the manual work entered into a computer file, usually excel. Reports are made. Currently after the report stage data can be entered manually into WODAN. This step is still time consuming. When in WODAN, the data can be managed by individual researchers. Maintaining WODAN is responsibility of the Discovery Programme at the moment. The raw (paper) data are archived in the DP." [12]

1.8 Forms and extent of data publication

Question: To what extent is data which your research group is producing being published, for instance in research reports, supplementary material to research papers or otherwise?

Data is published in various forms, however, mostly in traditional formats such as research reports, journal articles, proceedings, book chapters, monographs (sometimes with supplementary material on carrier media), and online webpages.

The description of publications practices are:

- "Data always as research report (biaxiaal, available on website and royal library), often as supplements in publication by clients=archaeologist , sometimes as cooperative publication or book, sometimes own research papers." [1]
- "At the moment just hardcopy or online traditional publication: no links to the long lasting repo." [2]
- "I have only one past excavation unpublished in book form, although interim reports have been published." [3]
- "Publication via a number of media (monograph, journal articles and online webpages)." [4]
- "Part of the data is published in form of catalogues which are part of books. Analysis published in form of research articles. Data often published in form of supplementary material (CD)." [8]
- "We aim to publish all our data in scientific publications – our association [ARHEO VEST] is publishing periodic reports, online and hard copies of our research results." [11]
- "Mostly, but not always, are data published in (internal) reports, articles or section of books. This depends on the results and whether they are important for the archaeological research of the Discovery Programme, or more specialistic of interest to the wood/charcoal research community." [12]
- "Data are published to a great extent, mainly in traditional ways (papers for conferences, articles in journals, volumes and monographs, contributions to specific publications, PhD theses). Annual Conference where new findings are presented." [13]
- Only "publications" in the narrow sense (abstracts, articles) are published. The underlying institutional data sets are not being published or made available publicly. They are an asset for the research organisation and are therefore not being shared with external communities. [15]
- "The primary outcome of a research project is the publication of a major peer-reviewed monograph. Lesser projects are published in a dedicated journal series or in article in peer-reviewed journals." [16]

The publication of actual research data/datasets presents a different case, again with different variants:

- The institute has one large dataset of over 1000 excavations, which is not available online, metadata for the related content is lacking (e.g. reports, GIS data, shape files, etc.). [17]
- In Slovenia an online open access archive of digitised maps of the 18th and 19th century (30-40,000 maps) with metadata is available. This was produced in a special project and is quite useful. [18]
- "There are some examples of MS Access databases that have been produced, but at the end of the project they were not prepared for web-based access. (…) We lack means to share data" (the researcher mentioned legacy technology, software and data and that new servers, costly software licenses, etc. would be required at the institute)." [18]
- "Cosa V was published as a volume, with the stratigraphym pottery tables and skeleton data on the web. Villa Magna will publish the full catalogues and stratigraphy on line." [3]

- "We have yet to publish our data, but hope to do so in a publication that will be out next year. We hope that eventually our full databases will be openly accessible online." [6]
- "CD with tables as part of monographs. Part of the excavation data has to be sent to BDA [Bundesdenkmalamt], according to Austrian BDA law. One part was sent to the NÖ Landesarchiv [Country of Lower Austria, County Archive] (it may at some point in the future be made accessible there – but no solid plans to do so at the moment)." [7]
- Concerning excavation data in Austria, researchers have to submit standardised reports to the Bundesdenkmalamt [BDA]. Several forms have to be filled for each excavation ("it is like writing a report"), however the information cannot be openly accessed by other researchers. [14], [15]
- "Online Publications: 1. Late Bronze Age Sites database; 2. Cemetery database: most of the archaeological evidence and most of ceramics is published; science data (e.g. animal bone and physical anthropology data) only partly published online." [8]
- "We tend to publish or make accessible freely online all our data." [9]
- "Basic project data is accessible to researchers on request and the Discovery Programme's databases are increasingly accessible online without charge." [16] "More work is needed but in keeping with the academic and education remit of the Discovery Programme we have made our resources available to students and scholars who have requested access." [5]

Concerning the integration of datasets one researcher mentioned: "I have to admit that if I had not got involved in the ARIADNE project (…) the idea of sharing and bringing together institutional data sets or even whole repositories would not have occurred to me as an option." [14]

1.9 Percentage of data deposited in (open access) digital repositories

Questions: According to representative surveys across many disciplines, it is assumed that researchers make available about 6-8% of their data in repositories which are also accessible to researchers not involved in the project. Can you estimate the percentage of the data produced by your research group that is deposited in a digital repository?

- *What kind of data is this, and in which format(s) is it shared?*
- *In which repositories or other ways do you make the data available?*

In general the accessibility of data for researchers not involved in a project is rather limited. Though there are projects that make some data openly available or aim to do so in the near future.

- "For people not involved in a project it's zero" [17], "fairly close to zero." [18]
- "No data from the LIARI [Late Iron Age and Roman Ireland] project has yet been deposited in an openly accessible digital repository." [6]
- "None is made available at the moment. Summary data via NÖ Landesarchiv." [7]
- "They are not published in open repositories; they are stored in our database." [13]
- "None, although there are plans to store Volubilis with the ADS. Plans, photos, reports, strat reports. ADS, if I can handle putting in all the metadata (perhaps in another life)." [3]
- "At present, no data generated by the project is accessible to researchers outside the project, but it is intended to make a significant percentage of the collated data available via an online, digital repository in the coming year. The aim is to produce a digital repository that is both searchable and open to updating by outside parties." [4]
- "As this is a 'live' project and the data sets are in use on a day to day basis, we have not made it available yet to other researchers. Our aim is to develop new software platforms that enable us to create an open access on-line searchable resource for future research in this area." [5]

Examples where some data has been made available included:

- "For macroremains in RADAR [Relational Archaeobotanical Database, NL] 100%, for wood and pollen no digital repository, only the publications." [1]
- "We are involved in the 3D-ICONS European Project where 10% of our data (high-resolution 3d survey models and metadata) will be shared (100% metadata and 1% 3D model's since we share just low-poly version according to the Italian Superintendency legal restrictions)." [2]
- "Above online publications [Late Bronze Age Sites database; Cemetery database] have been made available via the institutional homepage, the ÖAW server, with a stable identifier. The publication represents about 90% of the data. Formats are pdfs, the user interface allows searching individual graves. No download of tables possible." [8]
- "Our data is stored in our repository online accessible – it regards mainly 3d models available online as x3d or 3dpdf. Also jpeg and videos, all with related metadata." [9]
- "All our archaeological survey data is published online through maps (location of site and general description). Publishing ALL data is a very difficult task, since much of collected data is recorded manually and yet to be transferred in a digital format." [11]
- "WODAN [Archaeological Wood & Charcoal Database], is intended to be the digital repository for charcoal, maybe in the future also wood, especially for Ireland but extendable to Europe. At this time WODAN has only been used on a trial basis with less than 1% of data entered." [12]

1.10 Technical or other changed conditions that would ease data search and access

Question: Which technical advancement or other changes in framework conditions would you like to see in order to facilitate your research work (with regard to data search and access)?

Most of the researchers suggested implementation of "open access" principles and technical improvements that would allow easier searching and access. Metadata were mentioned only by one researcher.

- "Would be very important to have a tool to link metadata in an easy-way during the production phase (e.g., file system level manual metadata linking to the assets)." [2]
- "More of it. Also in summary form, so that it is possible to contact the project directors (ex. Fasti)." [3]
- "Online access to any data generated by archaeological and other relevant disciplines (excavation reports, geophysical surveys, palaeo-environment studies, etc.)." [4]
- "In conjunction with colleagues who are experts in the areas of software and hardware development we aim to create a platform which is fit for purpose now but that will be viable for a generation of users. Developing this will take time and the expertise of software developers." [5]
- "Open access to online academic articles and papers would be immensely helpful; as would access to excavation reports and excavation archives." [6]
- "Make transfer of data easier, allow larger quantities to share." [8]
- "A transparent and detailed description on how data was acquired, in order to assess data quality." [9]
- "A coherent method of data publication, accessibility (intelligent interfaces) and transparency of the process data has been created." [11]

- "WODAN with better working conditions" [1]; "Further developments of WODAN, with European applications." [12]
- "Large open databases, easily accessed, well managed – that would help to find relevant information (would be very useful in particular for classical archaeology). Would also make it easier for me to enter my material or parts of my material there." [13]

1.11 Technical or other changed conditions that would ease deposit of data in a digital repository

Question: Which technical advancement or other changes in framework conditions would make it easier for you or your group to deposit data in a digital repository?

Responses to this question included availability of a relevant repository, ease of data provision as well more specific technical requirements. As non-technical conditions mainly incentives and funding were mentioned.

- "Online access to a central repository." [4]
- "There is no archaeological data repository in Austria." [8]
- "Not having to spend hours and hours on metadata a process I barely understand." [3]
- "An online database template that is easy to use (and allows you to deposit data in a simple, straightforward way) would be really great." [6]
- "Would like to collect useful information, i.e. which is useful for me (and not information which is only needed by the excavator), and which can be accessed with a simple technology that I am familiar with and do not have to study before in order to use it." [13]
- "A major issue is that the institutional database is highly 'personalised' (e.g. it contains notes and comments from researchers that cannot be published), therefore access cannot be simply opened to external researchers. But would not object to contribute information about the results to open databases (all the basic information of the type which I need from other research projects)." [13]
- "This is part of our discussions with colleagues and hopefully through collaborations with colleagues who have successfully made such resources available to users elsewhere we can develop a framework to allow access to our own master data sets." [5]
- "A translation programme from excel to the programme used in WODAN." [1, cf. 12]
- "Would be important to be able to export an own cloud instance with the metadata included into a digital repository (with a metadata conversion tool like MINT)." [2]
- "No technical problem. Incentive has to come to share data, e.g. from funding bodies." [7]
- "Funding for digitization." [9]
- "Availability of funds." [11]

1.12 Expected benefits from ARIADNE infrastructure and services

Questions: The ARIADNE project is currently aiming to bring together and integrate existing archaeological research data infrastructures, so that researchers have better access to various distributed datasets: Do you think that your own research would benefit from such an integrated infrastructure? How exactly? What would be the single most important service you would expect from such a project?

Most responses related to common approaches (e.g. metadata, data structures), open access, ease of search and access, better access to data in relevant areas of research – particularly to leverage comparative research, and fostering of collaboration on the national and international levels.

- "The e-Infrastructure could improve the transparency of what is actually available (facilitate a better knowledge of existing datasets, improved access conditions). The most important achievement would be the development of a common approach for the presentation of archaeological research data (data structure, …)." [13]
- "Online repositories of archaeological data with good metadata would help a lot." [18]
- "Cross-searching data repositories." [3]
- "The most important service would be a metadata ingestion and management tool (1. metadata schema maker, 2. *ad hoc* form editing) that can be easily installed in our server to organize the workflow during the production phases." [2]
- "I would like to have simple-to-use tools for documenting and sharing data which can be used without having to be an expert in this domain." [15]
- "Make finds/sites/evidence comparable, easy search. User-friendly queries/interface." [8]
- The project "could be a big step to remove barriers to open access and move forward" [for ex. in Hungary], "make it open and international"; technically "ARIADNE should have a strong GIS support." [17]
- "I would recommend that ARIADNE should take a thematic or methodological approach, rather than just combining data in an arbitrary way. For instance, it would be very useful to get up-to-date state-of-the-art descriptions of specific methods or approaches, in a 'Wikipedia' style." [14]
- "More data means more accurate analysis. Raising awareness on the need to share high quality data." [9]
- "Having access to data always is of help, in understanding the 'big picture', in comparing results, in fine-tuning research, etc. There are two equally important services – one is opening existing data to free access and the other is providing funds for digitizing our content to be integrated in an Ariadne infrastructure." [11]
- "Depends on who else publishes their data (of the same field). More potential for quantitative comparisons. At the moment comparison is only possible for some material categories (lithics, fauna – these lend themselves for comparison, e.g. lithic artefact morphology), no comparison on a site level." [7]
- "The LIARI [Late Iron Age and Roman Ireland] project would definitely benefit from an integrated infrastructure; in the past, we have relied on the kindness of various organisations and good personal relationships to access data (especially excavation reports). It would be much more helpful if there was a streamlined process to access archaeological data within Ireland, or a service like the Archaeological Data Service in the UK." [6]
- "Easier access to relevant archaeological data, and comparing to other datasets not only for own ecofacts i.e. wood and charcoal, but also comparing to other environmental matters such as seeds and macro-remains, beetles etc. Most importantly, speaking from our experiences with the WODAN project is the process of standardization. If metadata sheets are produced in a similar way, language barriers can be overcome then, too." [12]
- "Access to the following data would be useful: Data of Late Bronze Age; cremation graves in general." [8]
- "The most important service would be an integrated European WODAN wood and charcoal database with easy entering process, easy access, easy downloadable results and other output." [12]

Also considerable benefits are expected with regard to fostering the collaboration between researchers and institutes on the national and international levels:

- "The research between different institutes (with different specialisations and their own databases) would benefit greatly." [1]
- "Our involvement in the Ariadne project is essential so that we can access both the latest technological developments and also methods and practices of other users in bringing comparative data sets into a digital format. We believe that it is in the interests of research in Ireland but in the UK and Europe more widely that comparative data sets be made available for furthering and developing advanced research in archaeology." [5]
- "The ARIADNE project will be used by the Discovery Programme in collaboration with the Heritage Council as a vehicle to organise cultural heritage data sets in major institutions in Ireland more efficiently and to seek to support the closer integration of these datasets." [16]
- "Access to a wider geographical datasets will in time help facilitate cross collaboration and enhance funding opportunities. Wider collaboration for the sustainability and future prospects for research in my discipline so that students and researchers with similar background can help it grow and develop into the future." [10]

A.IV.2 Directors or managers of data centres

Three interviews have been conducted with directors or managers of data centres:

[1] Julian Richards, director, Archaeology Data Service (ADS), United Kingdom,

[2] Hella Hollander, Project Manager, Data Archiving and Networked Services (DANS), Netherlands,

[3] Ulf Jakobsson, Data manager, Swedish National Data Service (SND), Sweden.

The data centres have similar missions:

Archaeology Data Service (ADS): "ADS's mission is to support research, learning and teaching with high quality and dependable digital resources. Fundamental to us fulfilling this mission is an ongoing programme of research into all areas of digital preservation, resource discovery and data sharing. Consequently we are actively engaged with research projects working with partners in all sectors of UK archaeology; academic, government, commercial and local."

Data Archiving and Networked Services (DANS): "DANS promotes sustained access to digital research data. For this purpose, DANS encourages researchers to archive and reuse data in a sustained manner, e.g. through the online archiving system EASY. DANS also provides access, via NARCIS.nl, to thousands of scientific datasets, e-publications and other research information in the Netherlands. In addition, the institute provides training and advice, and performs research into sustained access to digital information. Driven by data, DANS ensures that access to digital research data keeps improving, through its services and by taking part in national & international projects and networks."

Swedish National Data Service (SND): "The Swedish Research Council has appointed SND as a national resource for the coordination of existing and newly established databases within the social sciences, humanities and health sciences. SND offers support to Swedish research by facilitating researchers access to data within and outside of Sweden as well as offer support for research during the whole research process. SND presents Swedish research outside of Sweden."

2.1 Community of users served

Question: Please describe the user community of the data centre or repository your organisation is managing (key numbers, structure of users in terms of research domains, geographic distribution).

[1] Archaeology Data Service (ADS)

ADS has "users" and "depositors" from 3 main categories: About 40% academic/educational (in particular researchers who have their activities funded by bodies, such as the AHRC, who specify that digital outputs are deposited with the ADS; about 30% commercial users (contract research); various others like governmental agencies and public institutes.

The primary target group for ADS consists of scientific researchers. Students who are being prepared for doing research and researchers in training are also part of this group. In addition, the general interested public is also welcome to use the services. The majority of users/depositors is from the UK, but there are visitors who access the database from all countries.

[2] Data Archiving and Networked Services (DANS)

In general, the primary target group for DANS consists of scientific researchers, including research students and researchers in training. All Dutch archaeologists archive their data at DANS. They come from different organisations (in total about 40-50 organisations that regularly bring in their data), including commercial institutes, universities, and governmental institutions. There is an agreement among these organisations that they deposit their data at DANS within two years after an excavation.

The DANS collection comprises more than 20,000 archaeological datasets, comprising in total 1,561,838 data files (stored in the EASY database): about 17,000 reports (pdf) and some 3,000 large datasets (multiple files like pictures, maps, tables, report etc.) In 2012 around 20,000 downloads of a dataset (could be a download of the same dataset as well) and 94,564 downloads of files. Of all the datasets in EASY, about 42% has been looked at from the deposit time on.

[3] Swedish National Data Service (SND)

SND curates data of research domains of the humanities, social science and health science. Data depositors are Swedish researchers at universities and research centres in Sweden even though a few projects concern material outside Sweden (e.g. Finland and Italy). The users are mainly researchers and students in Sweden, though in 2012 about 25% of all orders for data in came from 16 countries outside Sweden.

In 2013, archaeological data comprised more than 360 archaeological surveys (388 datasets; shape-files, reports, Access databases, >40,000 files, >7.3 GB); the long-time storage of parts of the Swedish Rock Art Archive material (some 120,000 images, 5.8 TB), and a few thematic databases. Access to the openly available data (currently mainly GIS data) is increasing.

2.2 Main critical issues and challenges in fulfilling the centre's mission

Question: What are the main critical issues and challenges your organisation is currently confronted with in order to fulfil its mission in the best possible way?

[1] Archaeology Data Service (ADS)

The greatest initial challenge (when ADS was set up) was to establish credibility among the target users ("Why should we give ADS our data?"). In the meantime, this has been achieved.

Current challenges are:

- Greatest challenge: getting adequate metadata for the data that are to be deposited.
- "Navigating our position in the landscape", in particular our relationship with institutional bodies. What is the specific role of ADS? How to communicate this to target groups? What is our added value to libraries etc.
- Automating services as much as possible (to minimise costs) while still being able to provide a personalised service (which people value).
- General problem: the economic downturn (negative implications on funding).

[2] Data Archiving and Networked Services (DANS)

DANS has set four strategic priorities for the next few years:

- strengthen its services by serving more users more efficiently;
- develop into a discipline-independent data organisation;
- conduct research to support and improve its services;
- be an important building block in data provision in Europe.

The critical issues and challenges are:

- "We have a mass of information to offer and we only do this because we want to get it used." The data quality is therefore crucial.
- Quality of the metadata – we put a lot of effort into this, because data without metadata is meaningless.
- Also try to explain to researchers why it is important that they should invest the effort to provide good metadata together with the data they deposit in DANS. "This is not really a challenge, but something we find very important."
- Certification is an important issue and goal in this context.

DANS supports the Open Access principle, while being aware of the fact that not all data can be available freely and without limitations at all times. Therefore, DANS applies the principle 'Open if possible, protected if necessary'.

[3] Swedish National Data Service (SND)

At the moment the main challenges are:

- SND is a rapidly growing organization. We have increased the personal from 12 in 2009, 19 in 2011 and 34 in 2013. The plan is to further employ staff so that we in 2015 will be approximately 50. This means that there are technical, organizational and administrative challenges that we have to deal with continuously.
- The reason for the rapid growth was the extended tasks the Swedish Research council gave SND, from being an archive for social science data only to also include humanities and health science data.
- Legal issues: The Swedish legal system means in some cases concerns about availability of research data that contains personal information about the respondents.

2.3 Most demanding technical issues

Question: What are the most demanding technical issues your organisation has currently to deal with? (for example challenges such as managing already large and increasing volumes of data, heterogeneous data coming from many different projects, the integrating of data, …)

[1] Archaeology Data Service (ADS)

- "Managing depositors is the greatest challenge – getting them to provide metadata for comprehensive datasets."
- Technology (e.g. "big data") itself is not so much the challenge.

[2] Data Archiving and Networked Services (DANS)

- Presenting linked data in an archive: you have to think in new ways how to organise your archive.
- Getting the ISO standard (ISO 16363:2012 Audit and Certification of Trustworthy Digital Repositories). Otherwise, the technology itself is not so much the challenge.

[3] Swedish National Data Service (SND)

- Requirement for secure systems for handling registered data, especially for health science and social survey data, but also in other domains such as archaeology.
- Archiving archaeological data started in 2011, but this is a rapidly growing area of work; a tenfold increase in data is expected within the next few years.
- SND's Digital Archive is based on the Open Archival Information System (OAIS) model, all data is handled according to its workflow model. However, archaeological data has some specifics, the greater importance of GIS data, mapping services, etc.; also quick acquisition of know-how in the metadata, documentation, structure, etc. for archaeological data has been necessary.

2.4 Trends in user needs

Questions: What are the main trends you are observing with regard to user needs (whether of data depositors or consumers)? For example: Are there any changes in the user behaviour? Are there any new needs or requirements, or specific needs which you expect to increase in importance?

[1] Archaeology Data Service (ADS)

- Linked Data ("but put a question mark over this"): integration of linked data into the datasets;
- Increasing pressure of underpinning publications with data – growing user need to work with publishers for providing a repository with data;
- GIS data as a growth area;
- Mobile apps (access to database in the field through mobile applications).

[2] Data Archiving and Networked Services (DANS)

- DANS has become much better known and is increasingly "embedded in research practice";
- The main trend is a considerable growth in the amount of data that is deposited; currently about 400 data sets per year and growing;
- Researchers would like to see that data are cross-searchable across borders which requires improved links between data and data-sets;
- Demands made by researchers with regard to data services are continuously changing. To explore changes in user requirements, DANS plans to conduct a research programme on the subject of the life cycle of digital data.

[3] Swedish National Data Service (SND)

- In general the main needs are an easy way to store data, to find data no matter the original language, and to refer to the data consistently;
- More requests for DOI/PID, e.g. for linking publications and data;
- Demand for advanced search systems (currently under development);
- Increasing expectation of open access by research funders, the research communities, and the interested public.

2.5 Measures implemented in response to emerging needs

Question: Has your organisation implemented any measures in response to these emerging needs? (For instance: new services offered, organisational changes or technical innovations)

[1] Archaeology Data Service (ADS)

- Enhancement of our business model and tools in order to make it easier and cheaper to deposit small to medium sized archaeological archives (ADS-easy), depositors of commercial archaeology companies, for instance.
- Staying ahead of the development of institutional repositories, for example by providing enhanced and more comprehensive services, e.g. allowing universities "outsource" the long term access and preservation for the archaeological research data produced by their academics, providing an institutional view/index of all their archived results, etc.

[2] Data Archiving and Networked Services (DANS)

- A user panel has been established (both with depositors and consumers), they provide feed-back on specific issues;
- A review system is currently developed so that users can rate the quality of the data/information they have obtained;
- DANS wants to further strengthen the bond with university researchers by creating data windows at universities;
- DANS-EASY already meets the standards of the Data Seal of Approval (DSA), a next step is certification of the online archiving system EASY according to the international standard ISO 16363 (currently in progress);
- A special focus is also on the technical development of the database, aimed at increased efficiency, improved functionality, and cost control of storage and access to data;
- Another objective is further integration of current research information (e.g. about funded research projects, institutions and researchers involved, etc.).

[3] Swedish National Data Service (SND)

- Organisational change is on-going to fulfil the considerably extended mandate of SND (e.g. staffing, training, formation of service teams, etc.);
- New user services are under development (e.g. enhanced search services);
- Required new technical tools and services for emerging user needs in data archiving and access are investigated. What we have understood from some of our users is that researchers wish to be able to select what information they want to extract/download from available datasets. It would also be good to be able to run calculations in datasets and to do that online.

2.6 Relevance of ARIADNE for own service development

Question: In what ways could the results of the ARIADNE project (the envisaged e infrastructure and integrated services) have an impact on the services your organisation is providing?

[1] Archaeology Data Service (ADS)

- Access to expertise in ICT: ARIADNE could be an opportunity for ADS to work with key people in the information sciences;
- Opportunity to make ADS better known internationally;
- Opportunity to enhance the ADS collections.

[2] Data Archiving and Networked Services (DANS)

- ARIANE will allow DANS to enhance or develop new services which the organisation might not be able to finance from other funds (e.g. integrating XML schema, specific controlled vocabulary, etc. into the archiving system);
- Specifically the project allows increasing the capacity of the organisation to focus on the domain of archaeological research which has some specific needs and requires trialling of effective solutions.

[3] Swedish National Data Service (SND)

- A well-developed international infrastructure with common standards and interoperability will allow researchers here in Sweden to make their data visible and also make it easier for them to find material outside the Swedish border;
- Common standards to relate to will make it easier for us to promote the usefulness of documenting data in accordance with specific "rules", because that will make the data more visible and easier to find;
- Also, integrated services and common standards etc. will help SND to increase our knowledge about data-types we don't currently have and it will give us tools to further develop services on a national level.

2.7 Most important ARIADNE service for the data centre

Question: What would be the single, most important ARIADNE service you can imagine to make the services of your organisation even more attractive for its clients (depositors, consumers of data)?

[1] Archaeology Data Service (ADS)

- Improved integration of data mining in the ADS services, a field were ADS wants to do a lot more (ARIADNE has work package dealing with this aspect).

[2] Data Archiving and Networked Services (DANS)

- Use of vocabulary, thesauri, and open data approaches are of particular interest.

[3] Swedish National Data Service (SND)

- To find data wherever it is, whatever language (almost) would probably be the most beneficial service. This includes enough metadata to understand what the material is about and also finding the data the researcher produced himself/herself.

2.8 Relations to other e-infrastructure and services projects

Question: Which forms of cooperation, interoperability or integration with which other e-infrastructures and services of other providers will be paramount for the success of ARIADNE?

[1] Archaeology Data Service (ADS)

- DARIAH (http://www.dariah.eu), but also engage with international infrastructures outside Europe.

[2] Data Archiving and Networked Services (DANS)

- DARIAH (http://www.dariah.eu) and Europeana (http://www.europeana.eu)

[3] Swedish National Data Service (SND)

- There are a few major EU-projects that might be useful to connect to, both when it comes to standards and tools but also making data visible, e.g. CESSDA, DARIAH, E-Cloud and others.

A.IV.3. Further interviews with managers of data repositories and services

The German Archaeological Institute (DAI) has conducted 10 interviews, five with researchers and five with data services and repository managers. The five interviews with researchers have been included in the analysis documented in the Annex Section A.IV.1 above and the summarization in Section 5.1 of the main document). The processing of the five other interviews took too long for them to be included in this work. Therefore we document these five interviews separately below. They include three with managers of archaeological data services (two private companies and one state cultural heritage department) and two with data repository managers. The summaries of these two different groups are presented in the sections below. The results have been taken account of in the overall conclusions drawn from the pilot interviews.

3.1 Institutional and commercial data services

Three participants of this group of stakeholders could be interviewed: One interviewee works for a state cultural heritage department – LVR - Cultural Heritage Department Rhineland, Germany. The two other interviewees work at archaeological service companies – Denkmal 3D (Germany) and Wissenschaftliche Baugrund Archäologie (Germany). The companies generate and deliver data on behalf of purchasers of different background (e.g. building owners, development companies, public bodies like governmental departments, museums etc.).

Typical types of data

Asked about the type of data they typically produce, the attendees were requested to follow a given structure as follows:

- **data from** excavations**?**
    - The official documentation deliverables for the cultural heritage departments requires to be in hardcopy. Digital data is delivered as .doc (97-2003) (reports), xcl (2004)(lists and tables), dwg/dxf (CAD-data), jpg (photos), ASCII (measurements)
        - Moreover in our company we capture the whole documentation additionally also as pdf (texts and tables), ASCII (measurements), xml (databases).
        - important to us is a long term storage of all our data in xml format
    - Images, 3D data Tables, cad, doc
    - Databases, images, cad plans, raw data from DEM

- **data from laboratory measurements and analysis**?
  - Not upraised by any attendee
- **data from field surveys?**
  - DGM, 3D-pdf (Scans) are the most common formats we produce, but this is dependent on customer's needs, ASCII point clouds
  - Dwg/dxf (CAD plans)
- **data for model-based computing, simulation, etc. (i.e. "in silico" research)?**
  - Same as above
- **results of data mining for identifying patterns or interesting outliers**
  - Not upraised by any attendee
- **any other type of data?**
  - Museum archives
  - Images, 3D data, tables

Management of new research data

The interviewees were asked to inform about how new research data usually is managed. For example, does each research group manage data on its own with access only for that group (e.g. on a restricted access server of the institute/centre); or does the institute/centre operate an institutional content/data repository, managed by dedicated staff?

The following answers were given:

- "All projects of our company data are stored centrally on a local server and additionally on a hired online hard drive."
- "As we are service providers for industry and cultural heritage we transfer our reports and data to our customers and original data and reports always go to the heritage department in charge for the project."
- "In our department data is managed centrally in our database of archaeological sites. We encourage archaeologists use central tools and databases, some of these tools have been developed for their special needs, e.g. the stratigraphic excavation tool "Stratify", others are purchased from companies e.g. FAUST 3.0 plus museum archive."

Is the metadata shared in a content/data federation (e.g. through providing an OAI-PMH target, or otherwise)? One participant answered that an OAI target is provided.

Responsibility for data archiving and maintenance

The data responsibility is handled differently in the three organisations. While one institution has an IT-department at its disposal, at another there is one office manager in charge of the data storage of different offices in an office sharing building (overhead management for 3 small companies). In the third institution essential data is handed over to the responsible cultural heritage department and not generally stored at the office.

Documentation of research results

Typically documentation of research results after the project.

The participants were asked about the typical workflow of the final data maintenance and storage when a project is completed.

One referred to the issue that as a service provider the company is not the official owner of the data. Nevertheless they store and maintain the data and as creators of the data they negotiate in all their

projects rights to use the data for research purposes. Currently they deliver data to other researchers on request (e.g. for master theses etc.).

Another participant provides a fixed structure and concept for data recording. In all their projects (field surveys, excavations) data is entered in a standardised database by the researchers, partly already in the field. Afterwards the data is complemented with additional information, quality checked, and stored on a central server with decentralised backups.

The third interviewee described the following process: data collected during fieldwork is entered database templates stored on a local computer with hard drive backups. From the database report data and raw data are extracted for delivery to customers. A copy of all data is kept on the company server with external backup.

Availability of data

The attendees were confronted with the statement, that according to representative surveys across many disciplines, it is assumed that researchers make available about 6-8% of their data in repositories, which were also accessible to researchers not involved in the project. They were asked to estimate the percentage of the data produced by their organisations that was deposited in an accessible digital repository. In detail they were asked of what nature this data is and which formats are shared. Moreover it was asked if that data was contained in an institutional repository or elsewhere, e.g. in a national data centre or an international subject based repository.

One participant said that the data is available to researchers on request. The range of delivered material depends the research topic and on the availability of data. Data that is still under work by the researchers is not fully available. The type of the delivered data is datasets, photos, drawings, measurements, maps and shapefiles.

One participant remarked that they generally do not share any public digital data themselves, but relegate researchers to the heritage department in charge.

In one case there was not an answer given to this question.

Conditions of storage

In case they store data in a repository, interviewees were asked under which conditions (e.g. licences) the institute/centre would make it accessible to other researchers who have not been involved in the research projects.

Two participants stated that they have secured rights to use data produced for clients for research purposes.

Recent important technical developments

The participants were asked about the most important recent technical developments of their repositories, workflows, etc. (for example regarding the implementation of new services).

The participants described their work as rather "standardised", with no current need for major technical developments due to special needs of clients.

Incentives from ARIADNE projects

The participants were asked about their knowledge of the EU project "ARIADNE" and if they could think of ways their organisations might benefit from such an integrated infrastructure. Moreover they were asked what would be the single most important service they would hope to get from such a project.

One participant stated that they wished to deliver their data to the research community and also have the publishing rights to the data they produce. They would appreciate an external service to spread the knowledge about the data they offer. The most desirable service would be to make data visible to other researchers. A platform, where they could search for connected data in certain topics that is really used by a majority of researchers and institutes would be a real improvement.

Another participant mentioned that they would appreciate to have a low threshold service to upload their data, because as a small institution they would never be able to offer public data services and facilities.

The third interviewee imagined that research activities would be improved by additional services: Researchers would appreciate a registry to provide information about the data they hold. Input to such a registry needs to be automated for repositories with many data sets.

3.2 Managers of data repositories

This interview scheme used specifically addressed directors and managers of data centres, domain/subject-based repositories, portals or other services. Two interviewees from the following institutions have taken part: the manager of ARCHNE at the German Archaeological Institute (Germany) and one manager of the archaeological data repository/service Open Context, which is managed by the Alexandria Archive Institute (USA).

Main critical issues and challenges of organisations

Asked about the main critical issues and challenges the individual organisations were currently confronted with in order to fulfil its mission in the best possible way, one of the participants addressed the securing of the quality of data and the intelligibility of the data to the research community:

"The research community increasingly expects access to high-quality data. Open Context specializes in the review, documentation, and publication of research data contributed by scholars. Archaeologists today build complex databases to fully document and describe objects and deliver contextual information. Without sharing and preserving this data the information are lost. Yet, archaeologists find it difficult to share and preserve this irreplaceable information that is key to understanding ancient societies. The challenge is to enhance conventional publications through comprehensive dissemination and preservation of rich digital data and media. Important to us is moreover data quality. Our approach to handle this issue is our peer reviewing system.

Another problem we face is the intelligibility of data. There are important questions about how we make data easy to understand to enable reuse to be credible. We challenge these problems by our use of very simple Web architecture approaches. These also let us work together with the California digital Library, the main data archive of the University of California. It is important that the data works well with the WWW and can easily be networked with other qualified data published by museums, other scientific domains etc. It also means anyone can use the data provided by us in their own apps (we require CC licenses that remove standard copyright restrictions to make reuse legally possible)."

The other manager stressed the availability of a sufficient amount of authoritative data in order to address real research queries and to overcome language problems in their own system. As a third challenge, he considered the democratisation of data for the research community and beyond:

"For us availability of archaeological data for research and public use is one of the most striking issues in the next years. In order to use the stored information properly a comprehensive search function is the basic challenge in ARACHNE. To achieve this issue we map ARACHNE datasets to several protocols

and interfaces. A further challenge to our own system is to overcome its monolingualism. Here concordant archaeological vocabularies and thesauri need to be developed. Currently, what we will do now is including the term lists (thesauri, etc.) of our system into the vocabulary and a heavy Arab column, there is financing for both until the end of 2014 right now. The primary use is a word net for further text mining and multilingual information systems.

A long term challenge is the democratisation of the whole data of DAI, because in order to research and evaluate archaeological data properly the whole picture is condition sine qua non for archaeological science."

Current technical issues

This question was about the most demanding *technical* issues the organisation has currently to deal with, for or example:

- Challenges such as managing already large and increasing volumes of data (the issue of "big data" in the archaeological sector?);
- Managing heterogeneous data coming from many different projects;
- Managing the integrating of data (e.g. linking publications and data/datasets, also from external sources).

One interviewee referred to his previous answers. The other mentioned that data sharing and preservation was a huge issue across the humanities. *"Most efforts focus on library like 'digital archives' which are absolutely necessary. We think that while archives are needed, they aren't always sufficient, and that's why we think of ourselves as a "publisher" (through free and open access). There are important questions about how we understand and promote quality of information and how we make data easy to understand to enable reuse to be credible. These are hard problems, and there are some other efforts that are also exploring ways to make higher quality, more intelligible data via similar models."*

User community

The interviewees were asked to describe the user community of the data centre or repository their organisations are managing (key numbers, structure of users in terms of research domains, geographic distribution).

One participant distinguished between users in the sense of data providers and users who retrieve the data for further research: "*Our peers are various kinds of researchers in archaeology and associated science. Currently 45 archaeological projects share their data in Open Context. The persons authoring the data mainly come from North American Institutions, but there are some more Nations involved, coming from Europe, Africa and as well. The archaeological records are worldwide. We've now got a rapidly growing and dynamic community of researchers active in producing open data."*

He also mentioned that the datasets are mainly from research in the Mediterranean region and used by archaeologists in classics and ancient history. The service website has about 140.000 hits a year.

Main trends

The interviewees were also asked about the main trends they observe with regard to user needs (whether of data depositors or consumers). For example: Are there any changes in the user behaviour? Are there any new needs or requirements, or specific needs, which you expect to increase in importance?

One participant did not make a comment on this. The other one answered the first part of the question with the following comment: "Our customers all recognize the need to better share and

preserve data. They also appreciate the idea of having editorial and peer-review systems to help promote greater quality of data." New needs and requirements were seen in tools for visualisation of complex research queries and theories, in particularly in order to be used for teaching. Comment:

"There is more and more recognition for the need for computational systems to support research and teaching. Here the challenge will be to derive services from real research questions. People easily can combine data from different places for mashups, new visualization, and new research queries, if you only follow some simple principles of Web architecture. Moreover data and services should be networked with other data published by museums, other scientific domains, etc."

As an obstacle he mentioned the problem of IPR issues: *"But still, the issue of restricted licenses is the main obstacle to create real networks. PR initiatives to overcome these restrictions in the archaeological world will be a challenge to face."*

Current technical innovations

Has your organisation implemented any measures in response to these emerging needs? For instance new services offered, or organisational changes or technical innovations?

One interviewee referred to answers given before. The other respondent addressed different plans as follows:

"A new (but desperately needed) challenge is the utilisation of data for analysis on a wide range in collaboration with other institution. In Germany our institute together with other data providers in archaeology currently develops a structure for data centre classical and ancient studies (IANUS). At our institute we are currently working on an elaborate excavation database that organises and structures the complete data management of our archaeological projects – Open Infra. In order to complete the data life circle we also invent a digital journal-server in order to publish reports and interpretation of our data. All current and future digital projects are dedicated to achieve an information balance in long term perspective and to invent a knowledge eco system. We want to encourage researchers to do big data analyses in future in which and we want our data to be part of it."

Your organisation and the ARIADNE project

This battery of questions was only to be addressed to members of the ARIADNE consortium; so only one of the respondents (ARACHNE manager) responded to the questions:

Technical advancement or other changes in framework conditions for more attractive services

This question was about technical advancement or other changes in framework conditions that would make it easier for the organisation to provide even more attractive services for its clients.

Advancement to the service was seen in the ability of using combined data from different services. For this ARIADNE needs to develop ways to map the existing vocabularies as they are and build ontologies.

Impact of ARIADNE on own service

Asked in what ways the results of the ARIADNE project (the envisaged e-infrastructure and integrated services) could have an impact on the services the organisation is providing the following answer was given:

"A main issue of archaeological research is the access to different kinds of data on a certain subject, on order to get the best range of information. It there was such a service (described above), all data provided by DAI (also others than ARACHNE) could easily be exploited by researchers all over the world."

Opinion on most important ARIADNE service

What would be the single, most important ARIADNE service to make the services of the organisation even more attractive for the clients:

(a) for depositors of data?

Here the answer was "Advice and help to create and implement metadata".

(b) for consumers of data ?

The reply to this question was "Integrating platform for research data".

Technical advancement or other changes to create more attractive services

Asked for technical advancements or other changes in framework conditions that would make it easier for the organisation to provide even more attractive services for its clients, the answer was:

There should be a way to provide all archaeological data produced by public means under Creative Commons licensing. A second demand was to ensure quality data: "If people become aware of the advantages and the needs of digital delivered data, they will create better data in consequence."

Impact of ARIADNE

The response to this question was that any effort in encouraging, supporting and advising researchers to digitally share their quality data to the community for free use would lead to improvement for the whole community.

References

ACLS - American Council of Learned Societies / Commission on Cyberinfrastructure for the Humanities & Social Sciences (2006): Our Cultural Commonwealth. New York: ACLS, http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf

Adamick J. & Reznik-Zellen R. (2010a): Representation and Recognition of Subject Repositories. D-Lib Magazine, Vol.16, Nr. 9/10 (September/October 2010), http://www.dlib.org/dlib/september10/adamick/09adamick.html

Adamick J. & Reznik-Zellen R. (2010b): Trends in Large-Scale Subject Repositories. D-Lib Magazine, Vol. 16, Nr. 11/12, http://www.dlib.org/dlib/november10/adamick/11adamick.html

AHDS - Arts and Humanities Data Service (2006): E-Science Scoping Study, documentation, http://www.ahds.ac.uk/e-science/e-science-scoping-study.htm

Aitchison, Kenneth (2009): Discovering the Archaeologists of Europe: Transnational Report. Institute for Archaeologists, January 2009, http://www.discovering-archaeologists.eu/DISCO_Transnational_Report.pdf

Anderson B. & Carlson S. (2006): Entangled Data: Knowledge & Community making in E (Social) Science. University of Essex, Chimera Working Paper Number: 2006-15, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.916&rep=rep1&type=pdf

Anderson S., Blake T. & Dunn S. (2010): Methodological commons: arts and humanities e-Science fundamentals, in: Watson, P. *et al*. (eds.): e-Science: past, present and future, I. Phil. Trans. R. Soc. A, 368: 3779-96, http://rsta.royalsocietypublishing.org/content/368/1925/3779.full.pdf

Anderson, Sheila (2007): The Arts and Humanities and e-Science: Scoping Survey Report and Findings. Commissioned by the UK Arts and Humanities e-Science Initiative. King's College London, Department Centre for e-Research, December 2007, http://rae2007.cch.kcl.ac.uk/scoping-sa/e-science_report_final.pdf

Anichini F., Fabiani F., Gattiglia G. & Gualandi M. (2012): MAPPA. Methodologies applied to archaeological potential predictivity. Rome, https://www.academia.edu/2156935/MAPPA._Methodologies_Applied_to_Archaeological_Potential_Predictivity

ARACHNE, http://arachne.uni-koeln.de

Archaeology Data Service (ADS), http://archaeologydataservice.ac.uk

ARIADNE - Advanced Research Infrastructure for Archaeological Dataset Networking in Europe, http://ariadne-infrastructure.eu

Armbruster C. & Romary L. (2010): Comparing Repository Types. International Journal of Digital Library Systems, 1(4): 61-73, manuscript, http://arxiv.org/ftp/arxiv/papers/1003/1003.4187.pdf

Association of Research Libraries (ARL), Contemporary User Studies, http://www.arl.org/rtl/plan/userstudies2009.shtml

Association of Research Libraries (ARL), E-Science Working Group, http://www.arl.org/rtl/escience/

Austin T. & Mitcham J. (2007): Preservation and Management Strategies for Exceptionally Large Data Formats: 'Big Data'. English Heritage / Archaeology Data Service, 28 September 2007, http://ads.ahds.ac.uk/project/bigdata/final_report/bigdata_final_report_1.3.pdf

AVROSS (2007): Accelerating Transition to Virtual Research Organisation in Social Science. Final Report, http://web.fhnw.ch/plattformen/avross/papers-and-prensentations/final-report/Final%20report,%20full%20draft,%20v2-2.pdf

Babeu, Alison (2011): Rome Wasn't Digitized in a Day: Building a Cyberinfrastructure for Digital Classicists. Washington, DC: Council on Library and Information Resources, http://www.clir.org/pubs/reports/pub150/pub150.pdf

Barjak F., Lane J. *et al*. (2009): Case Studies of e-Infrastructure Adoption. In: Social Science Computer Review. Vol. 27, Nr. 4, November 2009, pp. 583-600, http://ssc.sagepub.com/content/27/4/583.full.pdf

Benardou A., Constantopoulos P., Dallas C. & Gavrilis D. (2010) Understanding the Information Requirements of Arts and Humanities Scholarship. In: The International Journal of Digital Curation, Vol. 5, Issue 1: 18-33, http://www.ijdc.net/index.php/ijdc/article/viewFile/144/206

Benardou, Agiatis (2011a): Classical studies facing digital research infrastructures: from practice to requirements. The Digital Classicist, http://www.digitalclassicist.org/wip/wip2011-05ab.html

Benardou, Agiatis (2011b): Classical Studies facing Digital Research Infrastructures: From Practice to Requirements. Digital Classicist & Institute of Classical Studies Seminar, London, July 1st, 2011, http://www.digitalclassicist.org/wip/wip2011-05ab.pdf

Berry, David (2011): The computational turn: Thinking about the digital humanities. In: Culture Machine, Vol.12, http://www.culturemachine.net/index.php/cm/article/viewArticle/440

Blackwell C. & Crane G. (2009): Conclusion: Cyberinfrastructure, the Scaife Digital Library and Classics in a Digital age. Digital Humanities Quarterly 3.1, Winter 2009, http://www.digitalhumanities.org/dhq/vol/3/1/000035/000035.html

Blanke T., Bryant M., Hedges M., Aschenbrenner A. & Priddy M. (2011): Preparing DARIAH. Seventh IEEE International Conference on eScience, Stockholm, https://de.dariah.eu/fileadmin/publications_presentations/Blanke_Bryant_Hedges_Aschenbrenner_Priddy_Preparing_DARIAH.pdf

Bodard G. & Mahony S. (eds., 2008): 'Though much is taken, much abides': Recovering antiquity through innovative digital methodologies. Digital Medievalist 4 (special issue), http://www.digitalmedievalist.org/journal/4/DCintro/

Bodard G. & Mahony S. (eds., 2010): Digital Research in the Study of Classical Antiquity. Ashgate, http://www.ashgate.com/isbn/9780754677734

BoneCommons, http://alexandriaarchive.org/bonecommons/

Borgman C.L (2007) Scholarship in the Digital Age: Information, infrastructure and the Internet. Cambridge, MA: The MIT Press

Borgman, C.L. (2009): The Digital Future is Now: A Call to Action for the Humanities. Digital Humanities Quarterly, 3(4), Fall 2009, http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Borgman, C.L. (2010): Research Data: Who will share what, with whom, when, and why? Fifth China – North America Library Conference 2010, Beijing, 8-12 September 2010, http://works.bepress.com/borgman/238

Bos N., Zimmerman A., Olson J. *et al*. (2007): From shared databases to communities of practice: A taxonomy of collaboratories. Journal of Computer- Mediated Communication, 12(2), http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00343.x/full

Bradley, John (2012): No job for techies: Technical contributions to research in digital humanities, pp. 11-26, in: Deegan M. & McCarty W. (eds.): Collaborative Research in the Digital Humanities. Farnham: Ashgate (DH2009 conference script online: https://www.yumpu.com/en/document/view/5530294/script-no-job-for-techies-kings-college-london)

Brockman W.S., Neumann L., Palmer C.L. & Tidline T. J. (2001): Scholarly work in the humanities and the evolving information environment. Washington, DC: Council on Library and Information Resources, http://www.clir.org/pubs/reports/pub104/pub104.pdf

Brown S., Ross R., Gerrard D. *et al*. (2006): RePAH - A User Requirements Analysis for Portals in the Arts and Humanities. Final Report, http://repah.dmu.ac.uk/report/pdfs/RePAHReport-Complete.pdf

Burdick A., Drucker J., Lunenfeld P. *et al*. (2012): Digital Humanities. MIT Press, 2012

Burns C.S., Lana A. & Budd J.M. (2013): Institutional Repositories: Exploration of Costs and Value. D-Lib Magazine, Vol. 19(1/2), http://www.dlib.org/dlib/january13/burns/01burns.html

CARARE - Connecting Archaeology and Architecture in Europeana, http://www.carare.eu

Carlson S. & Anderson B. (2007): What are data? The many kinds of data and their implications for data re-use. Journal of Computer-Mediated Communication, 12(2): 635-651, http://onlinelibrary.wiley.com/enhanced/doi/10.1111/j.1083-6101.2007.00342.x/

Carusi A. & Reimer T. (2010): Virtual Research Environment - Collaborative Landscape Study. A JISC funded project. Oxford e-Research Centre and Centre for e-Research, King's College London, http://www.jisc.ac.uk/media/documents/publications/vrelandscapereport.pdf

CESSDA - Council of European Social Science Data Archives, http://www.cessda.org

CHARISMA - Cultural Heritage Advanced Research Infrastructures: Synergy for a Multidisciplinary Approach to Conservation/Restoration, http://www.charismaproject.eu

Chavan V. & Penev L. (2011): The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 2011, 12(Suppl 15):S2, http://www.biomedcentral.com/1471-2105/12/S15/S2

CIDOC - Conceptual Reference Model, http://www.cidoc-crm.org (ISO 21127:2006: Information and documentation - A reference ontology for the interchange of cultural heritage information, http://www.iso.org/iso/catalogue_detail?csnumber=34424)

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

CLIR - Council on Library and Information Resources (2009): Working Together or Apart: Promoting the Next Generation of Digital Scholarship. Washington, DC: CLIR, http://www.clir.org/pubs/reports/pub145/pub145.pdf

COAR - Confederation of Open Access Repositories (2013): Incentives, Integration, and Mediation: Sustainable Practices for Populating Repositories, June 2013, http://www.coar-repositories.org/files/Sustainable-best-practices_final.pdf

Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age (2009): Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. Washington, D.C.: National Academy of Sciences, http://www.nap.edu/catalog/12615.html

Constantopoulos P., Dallas C., Androutsopoulos I. *et al.* (2009): DCC&U: An Extended Digital Curation Lifecycle Model. In: The International Journal of Digital Curation, Volume 4, Issue 1, pp. 34-45, http://www.ijdc.net/index.php/ijdc/article/viewFile/100/75

Cragin M.H., Palmer C.L., Kogan M. *et al*. (2009): Data Sharing, Small Science, and Institutional Repositories. UK e-Science AHM 2009: Past, Present and Future, Oxford, December 7-9, 2009, http://www.allhands.org.uk/2009/09/Cragin_UK_All-Hands_2009_final2.pdf

Crane G., Babeu A. & Bamman D. (2007): eScience and the humanities. International Journal on Digital Libraries, 7(1): 117–122, http://www.perseus.tufts.edu/~amahoney/jdlesciencecomm.pdf

Curtis+Cartwright Consulting (2010): Using the Cloud for Research. Final Report. Guildford, UK, http://www.jisc.ac.uk/whatwedo/programmes/researchinfrastructure/usingcloudcomp.aspx

DANS - Data Archiving and Networked Services, http://www.dans.knaw.nl/en

DARIAH - Digital Research Infrastructure for the Arts and Humanities, http://www.dariah.eu

Data Curation Profiles (2011) website, http://wiki.lib.purdue.edu/display/dcp/Data+Curation+Profiles

Deegan M. & McCarty W. (eds., 2012): Collaborative Research in the Digital Humanities. Ashgate

Digital Antiquity (see also tDAR), http://www.digitalantiquity.org

Digital Collaboratory for Cultural Dendrochronology (DCCD), http://dendro.dans.knaw.nl

Dutton W.H. & Jeffreys P. (eds.): World Wide Research: Reshaping the Sciences and Humanities. Cambridge: MIT Press

EC & ESF - European Commission and European Science Foundation (2007): Trends in European Research Infrastructures. Analysis of data from the 2006/07 survey. July 2007, http://ec.europa.eu/research/infrastructures/pdf/survey-report-july-2007_en.pdf

e-depot for Dutch archaeology (EDNA), http://www.edna.nl

Edwards P.N. *et al*. (2007): Understanding Infrastructure: Dynamics, Tensions, and Design. Workshop report, NSF Office of Cyberinfrastructure, http://cohesion.rice.edu/Conferences/Hewlett/emplibrary/UI_Final_Report.pdf

EGI-InSPIRE (2011): Integration of clouds and virtualization into the European production infrastructure. Project deliverable 2.6, March 2011, https://documents.egi.eu/document/258/

e-IRG - e-Infrastructures Reflection Group (2013): White Paper 2013. The Hague, The Netherlands, http://www.e-irg.eu/images/stories/dissemination/white-paper_2013.pdf

e-IRG - e-Infrastructures Reflection Group (2005): e-Infrastructures Roadmap. Prepared by M. Leenaars, M. Heikkurinen, P. Louridas & F. Karayannis. Published by e-IRGSP, The Hague, December 2005, http://isoc.nl/michiel/eirg-roadmap.pdf

e-IRG - e-Infrastructures Reflection Group (2012): Cloud Computing for research and science: a holistic overview, policy, and recommendations. e-IRG taskforce, http://www.e-irg.eu/images/stories/dissemination/e-irg_cloud_computing_paper_v.final.pdf

eIUS (2009a), University of Oxford & NCeSS: eIUS – e-Infrastructure Use Cases and Service Usage Models, website, http://projects.oucs.ox.ac.uk/eius/

eIUS (2009b): Archaeology Experience Report, July 2009, http://projects.oucs.ox.ac.uk/eius/outputs/Archaeology_ExpRep-pub.pdf

eIUS (2009c): Archaeology : eIUS e-Infrastructure Case Study, video, http://www.youtube.com/watch?v=LxZcI0ikKV0

eIUS (2009d): Use Case 11 – Archaeology, http://projects.oucs.ox.ac.uk/eius/outputs/Archaeology_FinalUseCase-1.pdf

ELIXIR - European Life Sciences Infrastructure for Biological Information (2011) Business Case. Cambridge/UK, http://www.elixir-europe.org/sites/elixir-europe.org/files/documents/businesscase_2011_high_res.pdf

Ellrott K., Zmasek C.M., Weekes D. *et al*. (2010): TOPSAN: a dynamic web database for structural genomics. Nucleic Acids Research, Vol. 39, Database issue, D494–D496, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3013775/

Engage - Engaging research with e-Infrastructure project, http://www.engage.ac.uk

EPOCH – Excellence in Processing Open Cultural Heritage, http://www.epoch-net.org

eResearch2020 (2010): The Role of e-Infrastructures in the Creation of Global Virtual Research. Study for the European Commission, DG Information Society and Media by Empirica GmbH *et al*. Full Report. Bonn, http://www.eresearch2020.eu

e-SciDR (2008): Towards a European e-Infrastructure for e-Science Digital Repositories. A report for the European Commission. Prepared by The Digital Archiving Consultancy Ltd, UK, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/e-scidr.pdf

ESF - European Science Foundation (2011): Research Infrastructures in the Digital Humanities. Science Policy Briefing 42. Strasbourg: ESF, http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf

ESFRI - European Strategy Forum on Research Infrastructures (2006): European Roadmap for Research Infrastructures. Report 2006. Luxembourg: Publications Office of the European Union, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap

ESFRI - European Strategy Forum on Research Infrastructures (2011): Strategy Report on Research Infrastructures. Roadmap 2010. Luxembourg: Publications Office of the European Union, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap

ESFRI - European Strategy Forum on Research Infrastructures, http://ec.europa.eu/research/esfri/

EUDAT – European Data Infrastructure, http://www.eudat.eu

e-Uptake - Enabling Uptake of e-Infrastructure Services project, http://www.engage.ac.uk/e-uptake

euroCRIS - The European Organisation for International Research Information, http://www.eurocris.org

European Desktop Grid Initiative (EDGI), http://edgi-project.eu

European E-Infrastructure Forum (2010): ESFRI project requirements for Pan-European e-infrastructure resources and facilities, Geneva, 28. April 2010, http://www.osiris-online.eu/RI%20Reports/eef-report.pdf

European E-Infrastructure Forum (EEF), http://www.einfrastructure-forum.eu

European Grid Infrastructure (EGI), https://www.egi.eu

European Middleware Initiative (EMI), http://www.eu-emi.eu

FAIMS - Federated Archaeological Information Management System (Australia), https://www.fedarch.org

Fasti Online, http://www.fastionline.org

Gaimster D. & Majewski T. (eds., 2009): International handbook of historical archaeology. New York: Springer, http://mayrasandy.files.wordpress.com/2012/08/36192861-hist-or-archaeology.pdf

Geertz, Clifford (1973): Thick description: toward an interpretive theory of culture, in: The interpretation of cultures: selected essays. New-York: Basic Books, 1973, pp. 3-30, available online at: http://hypergeertz.jku.at/GeertzTexts/Thick_Description.htm

German Archaeological Institute (DAI), http://www.dainst.org

Gray J. & Szalay A. (2004): Where the Rubber Meets the Sky: Bridging the Gap between Databases and Science. Technical Report, MSR-TR-2004-110, http://arxiv.org/ftp/cs/papers/0502/0502011.pdf

Gray, Jim (2007): Jim Gray on eScience. A Transformed Scientific Method, pp. xvii-xxxi, in: The Fourth Paradigm. Data-Intensive Scientific Discovery. Edited by T. Hey, S. Tansley and K. Tolle. Redmond, WA: Microsoft Research 2009, http://research.microsoft.com/en-us/collaboration/fourthparadigm/

GRDI2020 - Towards a 10-year Vision for Global Research Data Infrastructures, http://www.grdi2020.eu

GRDI2020 (2012): Global Research Data Infrastructures: The Big Data Challenges. Final Roadmap Report, February 2012, http://www.grdi2020.eu/Repository/FileScaricati/e2b03611-e58f-4242-946a-5b21f17d2947.pdf

Green D. & Roy M.D. (2008): Things to Do While Waiting for the Future to Happen: Building Cyberinfrastructure for the Liberal Arts. EDUCAUSE Review, vol. 43, no. 4 (July/August 2008), http://www.educause.edu/ero/article/things-do-while-waiting-future-happen-building-cyberinfrastructure-liberal-arts

Gregory I.N. & Ell P.S. (2008): Historical GIS: Technologies, Methodologies, and Scholarship. Cambridge University Press (Cambridge Studies in Historical Geography)

Hall M. & Silliman, S.W. (2006): Introduction: Archaeology of the Modern World. In: Historical Archaeology. Oxford: Blackwell, http://www.faculty.umb.edu/stephen_silliman/Articles/Introduction%20%28Hall%20and%20Silliman%29.pdf

Harley, Diane (2007): Use and Users of Digital Resources. A survey explored scholars' attitudes about educational technology environments in the humanities. Educause Quarterly, Vol. 30, No. 4, https://net.educause.edu/ir/library/pdf/eqm0742.pdf

Harley, Diane *et al*. (2010a) Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines. – Executive Summary and Full Report. University of California Berkeley, Center for Studies in Higher Education, eScholarship, http://escholarship.org/uc/cshe_fsc

Harley, Diane *et al*. (2010b): Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines. – Archaeology Case Study. University of California Berkeley, http://escholarship.org/uc/item/15x7385g#page-37

Hedges, Mark (2009): Grid-enabling Humanities Datasets. Digital Humanities Quarterly, 3(4), Fall 2009, http://digitalhumanities.org/dhq/vol/3/4/000078/000078.html

Heidorn, P. Bryan (2008): Shedding Light on the Dark Data in the Long Tail of Science. Library Trends 57(2): Institutional Repositories: Current State and Future. Edited by S. Sheeves & M. Cragin, http://hdl.handle.net/2142/9127

HERA-NET (2006): The HERA Survey on Infrastructural Research Facilities and Practices for the Humanities in Europe. S. Kaur-Pedersen and G.M. Kladakis, Danish Agency for Science, Technology and Innovation, September 2006, http://www.ucm.es/info/eurohum/docs/heraonhumanities.pdf

Hey T. & Hey J. (2006): E-science and its implications for the library community. In: Library Hi Tech 24 (4): 515–528; manuscript, http://eprints.rclis.org/9202/1/heyhey_final_web.pdf

Hey T., Tansley S. & Tolle K. (eds., 2009) The Fourth Paradigm. Data-Intensive Scientific Discovery. Redmond, WA: Microsoft Research, http://research.microsoft.com/en-us/collaboration/fourthparadigm/

Huggett, Jeremy (2012): Core or Periphery? Digital Humanities from an Archaeological Perspective. In: Thaller M. (ed.): The Cologne Dialogue on Digital Humanities 2012 – Controversies around the Digital Humanities (Special Issue, Historical Social Research, Vol. 37, No. 3), http://www.cceh.uni-koeln.de/files/Huggett_final.pdf

IANUS - Research Data Centre for Archaeology and Ancient Studies (Germany), http://www.ianus-fdz.de

IANUS - Schäfer F., Heinrich M. & Jahn S. (2014:) Stakeholderanalyse 2013 zu Forschungsdaten in den Altertumswissenschaften. Teil 1: Ergebnisse [Version 1.0] Hrsg. IANUS, http://www.ianus-fdz.de/projects/ergebnisse/wiki

Imperial College (2010): Research information management: Developing tools to inform the management of research and translating existing good practice. Imperial College London & Elsevier, sponsored by JISC, http://www.jisc.ac.uk/media/documents/programmes/RIM/RIMTNT_FinalReport.pdf

International Association for Classical Archaeology (AIAC), http://www.aiac.org

Internet Archaeology (journal, incl. data papers), http://intarch.ac.uk/authors/data-papers.html

IWGDD (2009): Harnessing the Power of Digital Data for Science and Society. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. Washington/DC, January 2009, http://www.nitrd.gov/About/Harnessing_Power_Web.pdf

Johnston, Lisa (2010): User-needs assessment of the research cyberinfrastructure for the 21st century (June 21, 2010). International Association of Scientific and Technological University Libraries, 31st Annual Conference, http://docs.lib.purdue.edu/iatul2010/conf/day1/5

Jones, Elisabeth *et al*. (2008): E-science talking points for ARL Deans and Directors. Association of Research Libraries, October 24, 2008, http://www.arl.org/bm~doc/e-science-talking-points.pdf

Journal of Intercultural and Interdisciplinary Archaeology, http://www.jiia.it

Journal of Open Archaeology Data, http://openarchaeologydata.metajnl.com

Kansa E.C., Kansa S.W. & 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 S. & Kansa E. (2009): Yes, it is all about you: User needs, archaeology and digital data. CSA Newsletter, Vol. XXII, No. 1 (April 2009), http://csanet.org/newsletter/spring09/nls0902.html

Kansa S. & Kansa E. (2011): Enhancing Humanities Research Productivity in a Collaborative Data Sharing Environment. White Paper to the NEH Division of Preservation and Access Advancing Knowledge. Alexandria Archive Institute, 27 June 2011, http://ux.opencontext.org/blog/wp-content/uploads/2011/06/white_paper_PK_50072.pdf

Kilbride, William (2006): Grand challenges: Grand opportunities? Archaeology, the historic environment sector and the E-science programme. 24 July 2006, http://www.ahds.ac.uk/e-science/documents/Archaeology-grand-challenges.pdf

Kintigh K.W. & Altschul J.H. (2010): Sustaining the Digital Archaeological Record. Heritage Management, Vol. 3, Issue 2: 264–274, http://www.digitalantiquity.org/wp-uploads/2011/01/20110127-Kintigh-Altschul-Forum-on-Sustaining-the-Digital-Archaeological-Record-from-Heritage-Management.pdf

Lee C.P., Bietz M.J., Thayer A. (2010): Research-Driven Stakeholders in Cyberinfrastructure Use and Development. Proceedings of the International Symposium on Collaborative Technologies and Systems, Chicago, 17-21 May 2010, pp. 163–172, https://depts.washington.edu/csclab/wordpress/wp-content/uploads/Lee-Stakeholders-CTS-2010.pdf

Lee C.P., Dourish P. & Mark G. (2006): The Human Infrastructure of Cyberinfrastructure. CSCW 2006 conference, Banff, Canady, 4-8.11.2005, http://www.ics.uci.edu/~gmark/CSCW06.pdf

Leydesdorff L., Hammarfelt B. & Salah A. (2011): The structure of the Arts & Humanities Citation Index: A mapping on the basis of aggregated citations among 1,157 journals. Journal of the

American Society for Information Science and Technology, 62(12), 2414-2426, http://arxiv.org/ftp/arxiv/papers/1102/1102.1934.pdf

MacColl J. & Jubb M. (2011): Supporting Research: Environments, Administration and Libraries. Dublin, OH: OCLC Research. http://www.oclc.org/research/publications/library/2011/2011-10.pdf

MAPPA Open Data (University of Pisa, Italy), http://mappaproject.arch.unipi.it/?lang=en

Marcial L.H. & Hemminger B.M. (2010): Scientific Data Repositories on the Web: An Initial Survey. Journal of the American Society for Information Science and Technology, 61(10):2029–2048, http://onlinelibrary.wiley.com/doi/10.1002/asi.21339/full

Meier zu Verl C. & Horstmann W. (eds., 2011): Studies on subject-specific requirements for open access infrastructure. Bielefeld: Universitätsbibliothek Bielefeld 2011, https://pub.uni-bielefeld.de/luur/download?func=downloadFile&recordOId=2445229&fileOId=2479864

Microsoft Research (2006) Towards 2020 Science, http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/

Nawrotzki K. & Dougherty J. (2013): Writing History in the Digital Age. Ann Arbor, MI: University of Michigan Press, 2013, http://dx.doi.org/10.3998/dh.12230987.0001.001

Nelson, Bryn (2009): Empty archives. Most researchers agree that open access to data is the scientific ideal, so what is stopping it happening? 10 September 2009. In: Nature, Vol. 461, 10 September 2009, http://www.nature.com/news/2009/090909/full/461160a.html

Niccolucci F. & Richards J.D. (2013): ARIADNE: Advanced Research Infrastructures for Archaeological Dataset Networking in Europe. The European Archaeologist, Issue 39, Summer 2013, http://e-a-a.org/TEA/archive/TEA_39_SUMMER_2013/rep1_39.pdf

Nicholas D., Rowlands I., Watkinson A. *et al*. (2012): Digital repositories ten years on: what do scientific researchers think of them and how do they use them? Learned Publishing 25: 195-206, http://ciber-research.eu/download/20120620-Digital_repositories_ten_years_on.pdf

NSF - National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure (2003): Revolutionizing Science and Engineering Through Cyberinfrastructure, https://arizona.openrepository.com/arizona/bitstream/10150/106224/1/report.pdf

NSF - National Science Foundation, Cyberinfrastructure Council (2007): Cyberinfrastructure Vision for 21st Century Discovery. Arlington, VA., USA, March 2007, http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

NSF – National Science Foundation, National Science Board (2005): Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. Washington/DC: NSF, http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf

ODE - Opportunities for Data Exchange (2011): Report on integration of data and publications. ODE project, 17 October 2011, http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf

OECD (2004): Declaration on Access to Research Data from Public Funding. Paris: OECD, http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157&Lang=en&Book=False

OECD (2007): Principles and Guidelines for Access to Research Data from Public Funding. Paris: OECD, http://www.oecd.org/dataoecd/9/61/38500813.pdf

Open Context (2009): Sharing Archaeological Data: Part 3 – What is your Dream Tool? Heritage Bytes, 29 April 2009, http://ux.opencontext.org/blog/2009/04/29/sharing-archaeological-data-part-3-%E2%80%93-what-is-your-dream-tool/

Open Context (Alexandria Archive Institute, USA), http://opencontext.org

Open Context, bibliography, http://opencontext.org/about/bibliography

OpenAIRE – Open Access Infrastructure for Research in Europe, http://www.openaire.eu

OpenDOAR - Directory of Open Access Repositories, http://www.opendoar.org

Palmer C.L., Teffeau L.C. & Pirmann C.M. (2009): Scholarly information practices in the online environment. Dublin, OH: OCLC, http://www.oclc.org/research/activities/past/workagenda/researchinfo/support/practices.htm

PARSE.Insight (2009): Insight into digital preservation of research output in Europe. Project deliverable D3.4: Survey Report, 9 December 2009, http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf

Pritchard S.M., Carver L. & Anand S. (2005): Collaboration for knowledge management and campus informatics. EDUCAUSE – Centre for Applied Research. Research Bulletin, Vol. 2005, Issue 2, January 18, 2005, http://net.educause.edu/ir/library/pdf/ERB0502.pdf

Procter R., Voss A., Asgari-Targhi M. (2013): Fostering the human infrastructure of e-research. In: Information, Communication and Society, Vol. 16, Nr. 10, December 2013, 1668-1691.

Procter R., Williams R., Stewart J. *et al*. (2010): If you build it, will they come? How researchers perceive and use web 2.0. London: Research Network Information, http://wrap.warwick.ac.uk/56246/

Procter R., Williams R., Stewart S. *et al*. (2010): Adoption and use of Web 2.0 in scholarly communications. Phil. Trans. R. Soc. A 2010 368, pp. 4039-4056, http://rsta.royalsocietypublishing.org/content/368/1926/4039.full.html#

Pryor, Graham (2009): Multi-scale Data Sharing in the Life Sciences: Some Lessons for Policy Makers. International Journal of Digital Curation, Vol. 4, No 3, http://ijdc.net/index.php/ijdc/article/view/135/178

Rains, Michael (2011): Creating a virtual research environment for archaeology, pp. 159-169, in: Kansa E. *et al*. (eds.): Archaeology 2.0. Cotsen Institute of Archaeology, UC Los Angeles, http://escholarship.org/uc/item/1r6137tb

Reichman O.J, Jones M.B & Schildauer M.P (2011): Challenges and opportunities to open data in ecology. Science 331: 703-705, http://www.nefmc.org/research/cte_mtg_docs/110316/Challenges%20and%20Opportunities%20of%20Open%20Data%20in%20Ecology.pdf

Research Data Alliance (RDA) – 'Research Data Sharing without barriers', https://rd-alliance.org

Richards, Julian (2009): From anarchy to good practice: The evolution of standards in archaeological computing. Archeologia e Calcolatori, 20: 27-35, http://eprints.whiterose.ac.uk/10707/

Richards, Julian (2012): Digital Infrastructures for Archaeological Research: A European Perspective. The CSA Newsletter, Vol. XXV, No. 2, September 2012, http://csanet.org/newsletter/fall12/nlf1202.html

Riding the Wave (2010): How Europe can gain from the rising tide of scientific data. Final report of the European High-level Expert Group on Scientific Data. A submission to the European Commission, http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

RIN - Research Information Network & British Library (2010): Patterns of information use and exchange: case studies of researchers in the life sciences, http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/patterns-information-use-and-exchange-case-studie

RIN - Research Information Network & Key Perspectives (2008): To Share or not to Share: Publication and Quality Assurance of Research Data Outputs. Main Report and Annex. RIN: London,

http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs

RIN - Research Information Network (2010): If you build it, will they come? How researchers perceive and use web 2.0. London: RIN, http://www.rin.ac.uk/system/files/attachments/web_2.0_screen.pdf

RIN - Research Information Network (2011): Reinventing Research? Information Practices in the Humanities. London: RIN, http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities

RIN & NESTA - National Endowment for Science, Technology, and the Arts (2010): Open to All? Case Studies of Openness in Research. Report prepared by the Digital Curation Centre. London: RIN, http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies

ROAR - Registry of Open Access Repositories, http://roar.eprints.org

Robey, David (2013): Infrastructure needs for the Digital Humanities. e-Research South website, http://www.eresearchsouth.ac.uk/news/infrastructure-needs-for-the-digital-humanities

Ross S., Sobotkova A., Ballsun-Stanton B, & Crook P. (2013): Creating eResearch Tools for Archaeologists: The Federated Archaeological Information Management Systems project. Australian Archaeology, No. 77, December 2013, https://www.academia.edu/5690498/Creating_eResearch_Tools_for_Archaeologists_The_Federated_Archaeological_Information_Management_Systems_project

Science magazine (2011): Science Staff introduction to the Special Issue 'Dealing with Data'. Science, Vol. 331 no. 6018, 11 February 2011: 692-693, http://www.sciencemag.org/content/331/6018/692.short

Shotton, David (2011): Why researchers don't publish data. Opencitations, 4 August 2011, http://opencitations.wordpress.com/2011/08/04/why-researchers-don%E2%80%99t-publish-data/

Sinnott R.O. *et al*. (2005): Large-scale data sharing in the life sciences: Data standards, incentives, barriers and funding models (The Joint Data Standards Study). August 2005, e-Science Technical Report Series, http://www.nesc.ac.uk/technical_papers/UKeS-2006-02.pdf

Sparks, Sue (2005): JISC Disciplinary Differences Report. Rightscom Ltd, London, http://www.jisc.ac.uk/uploaded_documents/Disciplinary%20Differences%20and%20Needs.doc

Strasser, Charly (2013): Closed Data… Excuses, Excuses. Data Pub, 24 April 2013, http://datapub.cdlib.org/closed-data-excuses-excuses/

Strategic Environmental Archaeology Database (SEAD), Umeå University, Sweden, http://www.sead.se

SURF Foundation (2009): The European Repository Landscape 2008. Inventory of Digital Repositories for Research Output. M. van der Graaf and M. Vernooy-Gerritsen. Amsterdam University Press, http://dare.uva.nl/document/150724

Sustainable Archaeology (Canada), http://sustainablearchaeology.org

Svensson, Patrik (2010): From Optical Fiber to Conceptual Cyberinfrastructure. Digital Humanities Quarterly 5(1), Winter 2010, http://digitalhumanities.org/dhq/vol/5/1/000090/000090.html

Swedish National Data Service (SND), http://snd.gu.se/en/start

tDAR - The Digital Archaeological Record (Digital Antiquity consortium, USA), http://www.tdar.org

Tenopir C., Allard S., Douglass K. *et al*. (2011): Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6(6): e21101, http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101

Terras M., Nyhan J. & Vanhoutte E. (2013): Defining Digital Humanities. A Reader. Ashgate, http://www.ashgate.com/isbn/9781409469636

Thessen A.E. & Patterson D.J. (2011): Data issues in the life sciences. ZooKeys 150: 15–51, http://www.pensoft.net/journals/zookeys/article/1766/data-issues-in-the-life-sciences

UCL & Emerald – University College London & Emerald Group Publishing Ltd (2010) Social media and research workflow (14 December 2010), http://ciber-research.eu/download/20101111-social-media-report.pdf

Unsworth, John M. (2000): Scholarly Primitives: What methods do humanities researchers have in common, and how might our tools reflect this? King's College, London, http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html

von Hippel, Eric (1986): Lead Users: A Source of Novel Product Concepts. Management Science, 32(7): 791–806, https://faculty.fuqua.duke.edu/~moorman/GeneralMills/Section3/Section3Documents/Source%20of%20Novel%20Product%20Concepts.pdf

Warwick C., Terras M., Galina I. *et al*. (2008): Library and information resources and users of digital resources in the humanities. Program: electronic library and information systems, Vol. 42 Issue 1: 5-27 (manuscript online: http://eprints.ucl.ac.uk/13807/1/13807.pdf)

Warwick C., Terras M., Huntington P., Pappa N. & Galina I. (2006): The LAIRAH Project: Log analysis of digital resources in the arts and humanities. Final report to the Arts and Humanities Research Council, http://www.ucl.ac.uk/slais/claire-warwick/publications/LAIRAHreport.pdf

Weller, Toni (ed., 2013): History in the Digital Age. New York: Routeledge

Williams, Caroline (2006): Intute: The New Best of the Web. Ariadne, Issue 48, July 2006, http://www.ariadne.ac.uk/issue48/williams

Witt M., Carlson J., Brandt D.S & Cragin M.H. (2009): Constructing Data Curation Profiles. In: International Journal of Digital Curation, Vol. 4, No. 3, 2009, http://www.ijdc.net/index.php/ijdc/article/view/137

Wouters P. & Beaulieu A. (2006): Imagining e-science beyond computation, 48-70, in: Christine, H. (ed.), New infrastructures for knowledge production: Understanding e-science. London: Information Science Publishing, http://virtualknowledgestudio.nl/documents/_annebeaulieu/new-infrastructures-for-knowledge-production.pdf

Wouters, Paul (2007): Conservative culture or diverging identities? Studying resistance to technology in academia. E-Social Science conference 2007, Ann Arbor, Michigan, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.8644&rep=rep1&type=pdf

Wright S.J., Kozlowski W.A., Dietrich D., Khan H.J., McIntosh L. & Steinhart G.A. (2013): Using Data Curation Profiles to Design the Datastar Dataset Registry. D-Lib Magazine, Volume 19, Number 7/8, July/August 2013, http://www.dlib.org/dlib/july13/wright/07wright.html

Zenodo repository (CERN, related to OpenAIRE), http://www.zenodo.org