D2.1 Initial Report on Community Needs

31 October 2019

Grant Agreement number:	823914
Project acronym:	ARIADNEplus
Project title:	Advanced Research Infrastructure for Archaeological Dataset Networking in Europe - plus
Funding Scheme:	H2020-INFRAIA-2018-1
Project co-ordinator:	Prof. Franco Niccolucci, PIN Scrl - Polo Universitario "Città di Prato"
Tel:	+39 0574 602578
E-mail:	franco.niccolucci@gmail.com
Project website address:	www.ariadne-infrastructure.eu

ARIADNEplus is a project funded by the European Commission under the Horizon 2020 Programme, contract no. H2020-INFRAIA-2018-1-823914. 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 - Advanced Research Infrastructure for Archaeological Dataset Networking in Europe – Plus project. The report presents the Initial Report on Community Needs (D2.1) that has been produced under WP2 “Extending and Supporting the ARIADNE Community”. The research leading to these results has received funding from the European Community’s Horizon 2020 Programme (H2020-INFRAIA-2018-1) under grant agreement n° 823914. But the content of this document cannot be considered to reflect the views of the European Commission.

Partner in charge of the deliverable:	Salzburg Research Forschungsgesellschaft mbH (SRFG)
Authors:	Guntram Geser (SRFG)
Contributing partners:	Julian Richards (UoY-ADS) contributed to the elaboration of parts of the survey questionnaire. Andres Dobat (AU), Attila Kreiter (HNM), Daniel Löwenborg (Univ. Uppsala) and Benjamin Štular (ZRC-SAZU) tested and provided suggestions on the survey questionnaire. All partners contributed to the survey.
Quality review:	H. Wright, UoY-ADS
Version	Final Version, v1.0 (public)

Table of contents

1 Introduction and Overview

1.1 ARIADNE/plus objectives

The overall objective of the ARIADNE/plus initiative is to help the archaeological research and data management communities in Europe (and beyond) to more effectively share and reuse data resources which are dispersed and often difficult to discover and access. For this purpose the initiative has developed a digital infrastructure and services that enable registration, aggregation, integration, and search and retrieval of data records which describe and link to the available data(sets).

The initial ARIADNE project has already implemented e-infrastructure and services for this, and ARIADNEplus aims to provide several new or enhanced services which can be used in virtual research environments on a Cloud-based data management and access platform. The existing datasets in the ARIADNE Catalogue will be updated and the pool of data records extended geographically, temporally and thematically by incorporating additional datasets. The records will be integrated using a Linked Data approach that enables novel ways to search and browse data based on detected relations between them.

Thus ARIADNEplus aims to take the next steps in enabling data sharing and collaborative (re)use for archaeological research across institutional and national as well as disciplinary boundaries. Fostering a culture of data sharing and reuse and joint capacity building will be crucial for successful development and takeup of the ARIADNEplus open/FAIR data resources, common e-infrastructure and services.

1.2 Task brief for this report

This report is the first deliverable of Task 2.2 “Reviewing the Community Needs and the Market” of WP2 “Extending and Supporting the ARIADNE Community”.

According to the task brief this initial report

- provides information on recent EU research and e-infrastructure strategies, including Open Science, FAIR data, and the European Open Science Cloud;
- presents the results of the ARIADNEplus user needs survey including, where possible, comparison to those of the ARIADNE 2013 survey;
- also describes how the results match with the planned new technical and other services, and gives suggestions on activities likely to enable an optimal match.

A special focus of the user needs survey was the intended portfolio of the ARIADNEplus user services for data search & access and several other new or enhanced services for researchers and data managers. The task brief mentions that the analysis and suggestions regarding these services will take account of the development of related technological markets.

In this report new technological concepts and developments are addressed regarding the relation of ARIADNEplus to the European Open Science Cloud (EOSC), digital Open Science, and FAIR data. Studies on particular technological markets relevant for the portfolio of ARIADNEplus service were not undertaken, as finding out which services the ARIADNEplus user community appreciates more or less was only a first step towards the intended Virtual Research Environments (VREs) which will integrate relevant user services and tools.

In-depth studies on particular technology markets, open source as well as commercial, will be conducted alongside the design and development of the VREs. This will also require evaluation of the

intended Cloud-based VREs, service bundles and particular tools from the perspective of their user groups, involving group members, who can also propose VREs for their research. Therefore the technology market studies will be carried out in future when the envisaged VREs are defined.

1.3 Report structure

The report comprises three main parts:

Summary of main results and suggestion

Development in policies and technologies

- The European Open Science Cloud (EOSC) initiative
- The Open Science policy
- The FAIR data principles
- The current status of open research data policies

The ARIADNEplus user needs survey

- The survey approach and demographics
- Research interests
- Data publication/sharing
- Barriers to data deposition and sharing
- Readiness to share data
- Reuse of data, including purposes and types of data & sources
- ARIADNEplus data search & access services
- ARIADNEplus special services for researchers and data managers
- Related training needs

The final chapter gives a brief outlook for the next phase of the user needs survey.

2 Main results and suggestions for ARIADNEplus

2.1 Development in policies and technologies

Important recent EU research and e-infrastructure strategies addressed are the European Open Science Cloud initiative, Open Science and FAIR data. Recent revision of related policies are noted.

European Open Science Cloud (EOSC)

European Open Science Cloud (EOSC) initiative, launched in April 2016, aims to remove the current lack of integration of research e-infrastructures of different disciplines and support Open Science practices which include sharing and reuse of FAIR data resources. From the perspective of end-users the EOSC will provide a common virtual platform for researchers to store, (re)use and analyse FAIR data for research, innovation and educational purposes.

The EOSC initiative is well on its way, as on 23 November 2018 the initial EOSC portal was launched, enabling access to the first sets of services and resources. But the realisation of the EOSC vision, particularly wide adoption and use by different stakeholders, will require much more than technical services. In June 2019 the EOSC Executive Board published the strategic implementation plan for the EOSC that gives a comprehensive overview of the planned activities which will contribute to the realisation of the EOSC goals for the period 2019-2020.

Alignment and integration with the EOSC is part of the innovation objective and approach of ARIADNEplus. The project follows the EOSC vision of providing a wide range of services for research in Cloud-based Virtual Research Environments (VRE), indeed is ahead of the EOSC in this regard. ARIADNEplus services, such as the data aggregation and data portal services, have already been implemented as VREs on the fully operational D4Science platform.

Cloud-based VREs will allow researchers to use online tools for different tasks and types of data. Providing research tools online in Cloud-based environments avoids researchers investing effort to acquire, implement, maintain and upgrade them. The approach allows cost-savings for the research community while at the same time opportunities for research groups to jointly address research questions.

While this will be enabled by ARIADNEplus VREs on the D4Science platform, the core relation of ARIADNEplus, and other domain research e-infrastructures, with the EOSC will be established based on catalogues of service and data resources. The optimal way to exchange catalogue information with the EOSC “marketplace” of service and data resources will be investigated. For example, ARIADNEplus services will be registered in the D4Science catalogue of service resources and could be shared from there with the common EOSC catalogue.

What does it mean for the ARIADNEplus user community?

In brief:

- ARIADNEplus will align and integrate via D4Science as far possible with the EOSC, contribute services to the EOSC as well as use relevant services from the EOSC.
- Archaeological research and data management organisations (e.g. repositories) will not need to describe resources they wish to share in the EOSC catalogue, their information about available datasets and services can be aggregated by ARIADNEplus and provided as a collective contribution to the EOSC.

- Research groups within the archaeological domain can be provided dedicated VREs on the D4Science platform, which allow them to use data and online tools in collaborative research, while avoiding needing to implement and maintain such tools themselves.

Open Science

Open Science as a priority and guiding principle of research policy at the European level was introduced in 2015 by Carlos Moedas, the Commissioner for Research, Science and Innovation (2014-2019). The Commissioner adopted the concept of "open innovation" and pushed for more openness of research in the European Research Area (ERA).

In European Commission research policy information and documents the term Open Science is now being used regularly when creating open access to shared research resources, and expected impact of such access, are being addressed. The expected impact of openness is a transformation of science leading to advances in knowledge, innovation and societal benefits. Innovative digital, ICT-enabled research is understood to play a key role in this transformation. Open Science using digital methods and tools will enhance research collaboration, involvement of citizens, and transparency and relevance of better accessible research outcomes.

The concept of Open Science is of course highly relevant also for archaeological research. The need and challenges of open research practices and resources is already being discussed in the archaeological community. An open data imperative is particularly strong in this field: excavation of sites destroys the primary archaeological evidence, the work on archaeological heritage is done in the public interest, and there is little commercial relevance of archaeological data.

Therefore openness should become embedded in archaeological research practices as *"the default modus operandi"* so that the advantages of accessible and reusable data gain priority over the interest of the individual researcher.

What does it mean for the ARIADNEplus user community?

- Open Science using digital methods and tools, supported by ARIADNEplus, is expected to enable more accessible and reuseable research results.
- Many archaeologists are not yet well prepared or equipped for the Open Science research paradigm. As the matter is complex, strong leadership with regard to policies/mandates, supportive institutional measures (e.g. capacity building, training of researchers), and use of state-of-the-art digital infrastructure are required.

FAIR data

Over the last few years the FAIR data principles, published in April 2016, have been adopted by different stakeholders for the sharing and reuse of research data. FAIR, Findable, Accessible, Interoperable and Reusable (described with 15 principles), replaces the general concept of Open Data by one that is more specific and measurable.

FAIR principles do not imply that the data is "open" or "free", as with Open Data, but rather only that data licenses must be available that formally state what users are allowed to do with the data. The different approach of FAIR regarding openness and costs allows participation of data holders that otherwise could not be involved in the FAIR data initiative.

While there is a FAIR "boom" in the international research data management community no wide awareness, lest knowledge, of the principles among researchers can be assumed. However, the effort to realise and manage large volumes of FAIR data as envisaged for the EOSC will be substantial.

It is expected that at some point research funders will start making data-related costs eligible for FAIR data only. This will require a lot of investment in training of FAIR data managers of research organisations, repositories and other research infrastructure.

What does it mean for the ARIADNEplus user community?

- The ARIADNEplus project is committed to develop and share expertise in the application of the FAIR data principles.
- In the ARIADNEplus survey FAIR data ranked at the top of suggested training offers.
- The project work plan includes tasks dedicated to policies and good practices for FAIR data management for both researchers (e.g. Data Management Plans) and data repositories.

Recent revisions of research data policies

According to recent reports, since 2016 more governments have put open research data on their agenda. Previously the focus was mostly on information and data produced by public sector agencies, and often reused by researchers. Increased investment in digital research infrastructures, both national and common European ones, the Open Research Data Pilot for Horizon 2020, and other factors shifted the focus to research data of funded research organisations and projects. But still there is no consensus position across EU Member States or even within countries.

Recent revision of two related policies for research data should be recognised:

The European Commission's Recommendation C(2012)4890 on Access to and Preservation of Scientific Information in Europe is the core common strategy regarding open access to research publications and data. In April 2018 it was replaced by the Recommendation EU 2018/790, which takes account of recent developments in research practices relating to Open Science and the European Open Science Cloud initiative.

In April 2019, the EU Directive on the Re-use of Public Sector Information (Directive 2003/98/EC and amendments) was replaced by the Directive (EU) 2019/1024 on Open Data and the Re-use of Public Sector Information, to be implemented by Member States in the national law by 16 July 2021. The Directive, among other points, includes that Member States have to establish policies and actions *"aiming at making publicly funded research data openly available ('open access policies') following the principle of 'open by default' and compatible with FAIR principles"* (Article 10).

What does it mean for the ARIADNEplus user community?

- The new Directive on Open Data and the Re-use of Public Sector Information is particularly relevant for public sector research institutions among the ARIADNEplus stakeholder community.
- The new Recommendation on Access to and Preservation of Scientific Information in Europe should be recognised as it integrates recent research policies in a core EU strategy document.

2.2 Results of the ARIADNEplus user needs survey

2.2.1 Data publication

Summary of main results

In the ARIADNE/plus surveys we used the concept of "data publication", mainly to emphasise the common understanding that publication means that the data indeed is publicly available. Researchers often share data directly with colleagues but do not make them publicly available (e.g. in a repository). This means that valuable data remains within small circles of peers and is not available to other researchers and the wider public. Moving more data from closed-circle or not sharing to "open data" requires overcoming strong barriers (as addressed in the next section).

About ten years ago surveys across many disciplines showed that the data practices of researchers run against what advocates of proper data management and open data sharing would advise. The surveys found that after the completion of research projects most data remains locked away, resides on PCs, storage devices, and restricted access servers, out of reach of other researchers, and in danger of loss. Over the last few years the situation seems to have improved, arguably mainly due to the expectation of research funders that data from funded projects is being deposited in appropriate repositories for long-term preservation and access.

More sharing of data through accessible repositories

In the ARIADNE 2013 survey around 50% of respondents shared from some to all of their project data through an accessible repository, while in 2019 around 65% ("not at all" shared respondents in this way 50% in 2013 and 35% in 2019). A comparison of the 2013 and 2019 results suggests that from 2013 to 2019 in the ARIADNE/plus communities of respondents the sharing of data through accessible repositories increased significanty by 10-15%.

Results of other surveys point to a general increase in repository-based data sharing, e.g. PARSE Insight (2009) compared to Tenopir et al. (2015), +10%. The Figshare surveys (2016, 2017, 2018) found a year-on-year increase in researchers' willingness to make their data openly available in various ways. In 2017, 29% said they made data available in a specific data repository while it was 33% in 2019, +4%.

ARIADNE/plus surveys found 30% more repository-based data sharing than others:

- In the Tenopir et al. 2013/14 survey (reported 2015) 30% said that they store from some to all of their data in a repository, in the ARIADNE 2013 survey 20% and in the 2019 survey 30%.
- In the Figshare surveys 2017 and 2018 sharing of data through a specific data repository was reported by 29% and 33% respondents, respectively. In the ARIADNEplus survey 2019 by over 30% more.

Differences between ARIADNE/plus 2013 and 2019 respondents

Considerable differences between the samples of respondents 2013 and 2019 should be noted:

- (presumably) more responses from ARIADNE partners,
- considerably fewer from countries with a mandated data repository (Netherlands, UK),
- more responses from Eastern and Southeastern Europe, and
- overall more established, permanently employed researchers and data managers (i.e. significantly less with a fixed-term contract and Ph.D. students).

In 2019, there were fewer responses from the Netherlands and the UK, where archaeologists are obliged to deposit data from publicly funded projects in an accessible repository. Despite this, the responses of all respondents amounted to 10-15% more data publication through different types of

accessible repositories than 2013. Very likely more data publication was reported by other ARIADNEplus partners.

Particularly interesting are the figures reported for supplementary material:

- PARSE.Insight (2009) 15%; Tenopir et al. (2015) 19.4%; Figshare (2017) 34%, (2018) 35%; *Springer Nature* (2018) 42%.
- In the 2013 and 2019 ARIADNE/plus surveys, many more said that they make supplementary material available; 2013: 82% of 520 respondents, 2019: 81% of 449 respondents. In the 2019, 13% in all or most, 25% in many, and 42% at least in a few projects (the percentages 2013 are roughly the same).
- The explanation for the difference to the other surveys could be that many of the ARIADNE/plus survey respondents are obliged to provide fieldwork reports to a national heritage authority, and do this with supplementary material added. Compared to the Figshare figure for supplementary material of around 35% (which is quite high), it appears that 45% of the ARIADNE/plus respondents do so.
- For the comparison between the ARIADNE and ARIADNEplus survey participants, the almost identical figures for supplementary material could mean that overall the differences are not as considerable as other survey results suggest.

Suggestions for ARIADNEplus

While the ARIADNEplus survey shows good results for sharing reports and data through institutional repositories (e.g. repositories of heritage authorities or research centres), many archaeologists in European and other countries do not have available yet a state of the art digital repository for archiving and sharing their data.

This issue is being addressed by the COST Action SEADDA, the Saving European Archaeology from the Digital Dark Ages network. SEADDA and ARIADNEplus share the goal of making archaeological data FAIR (Findable, Accessible, Interoperable and Reusable), especially by supporting knowledge exchange and collaboration on data repositories and e-infrastructure.

The core requirement for moving research data into accessible repositories is decisive open data mandates by research funders, coupled with funding of the basic costs of domain repositories and the researchers' data deposition costs (e.g. as part of research grants).

Suggestions to increase further the sharing of archaeological data through appropriate repositories are:

- Continue the good collaboration between ARIADNEplus and SEADDA on capacity building for new repositories and use of the ARIADNEplus digital infrastructure
- Support strict open data policies of funding bodies and institutions – data repositories and infrastructure should give full support to such mandates.

2.2.2 Barriers to data deposition & sharing

Summary of main results

Core functions of the ARIADNEplus digital infrastructure are to aggregate data from archaeological repositories and provide search and access services. Therefore the ARIADNEplus initiative depends on repositories richly filled with accessible data shared by researchers. It cannot ignore obstacles which hinder researchers in sharing their data in an open manner. Rather the initiative must support researchers in data sharing and help ensure that they receive appropriate credit for doing so.

ARIADNE/plus surveys 2013 and 2019

In the 2013 ARIADNE and 2019 ARIADNEplus surveys the participants were given a list of potential barriers for researchers to deposit their data in digital repositories and share it with others. The respondents were asked how important the different barriers are in their view. The question was answered by around 500 respondents in 2013 and 400 in 2019. The barriers which respondents perceived as most critical were the same, albeit with some differences regarding the percentages of "very" or "rather" important combined:

- *A lack of professional recognition and reward:* was considered as most critical by 75.5% of respondents in 2019, while 72% in 2013.
- *The work effort for providing the data and metadata in the required formats:* was an important barrier for 74% of respondents in 2019, while in 2013 more respondents worried about the work effort for metadata (80%) and data (80%).
- *Intellectual property rights issues:* was a concern for 75% of respondents in 2019, while significantly less in 2013 with 65%.

Two barriers were perceived as somewhat less important with about the same percentages: Lack of appropriate repositories with 67% in 2019, while 66% in 2013; the cost for depositing data in a repository with 59% in both years.

In 2019 many respondents noted also lack of awareness and interest, political, legal and institutional obstacles, and lack of skills and of curatorial and technical support. Respondents for example said, "*Researchers do not fully understand the benefits of data sharing", "Unwillingness to share data until all possible internal use has been extracted*", "*Lack of time and/or staff available to complete the data sets".*

The main barriers to data sharing in archaeology are the same as for researchers in other disciplines. One specific concern is disclosing information about the location of archaeological sites which looters could use to identify them; in some cases also indigenous communities have a stake in the protection of sites and artifacts of cultural or religious value.

Results of other surveys

It is worth noting some results of other surveys: Fecher et al. (2015) in a large survey with respondents from different disciplines found that *"if I were cited in publications using my data"* would motivate 79.3% of 1,420 respondents to make data available (9.5% said it would not, and 11.2% were undecided); *"if I had enough time beforehand, to publish on the basis of my data"* was the second strongest enabler of data sharing, 77.5% agreed to it. Obviously, researchers must be granted the time to exploit their data appropriately, i.e. sufficient time until data of funded research has to be archived or include embargos on deposited data. In the latest Figshare *The State of Open Data* survey (2018) the majority of respondents felt that they did not get sufficient credit for data sharing, 58%, compared to 9% who felt they do; 33% were not sure.

Suggestions for ARIADNEplus

Understanding obstacles to data sharing and helping to remove them is essential for infrastructures for research data such as ARIADNEplus as well as the underlying digital repositories. Advocates of open data argue that such data will often be (re)used and cited, bringing recognition and rewards to data publishers (incl. data repositories). The scenario is that data citations indicate and acknowledge providers of valuable data, promote further data sharing and (re)use, and enable the impact of open data to be tracked and measured. Most importantly, it would drive the emergence of an academic credit system that appropriately rewards open data sharing.

- Research infrastructure components, protocols and metrics for data citations are in development. ARIADNEplus should investigate how services of the research infrastructure could help identify and track (re)use of data based on data citations (e.g. article-data links) and other indicaters.
- As a general requirement for identifying data (re)use, the project could promote and support standardisation of data citation in the archaeological sector, i.e. how data should be cited in publications to ease the identification and tracking of data (re)use.

2.2.3 Increase of readiness to share data

The survey participants were asked if they agreed with the statement: *"In the last 5 years the readiness of archaeologists to share data through publicly accessible repositories or databases increased"?*

83.2% of 376 respondents agreed. This result suggests that the prospects for open data sharing in archaeology look positive. However, several respondents perceived a higher awareness among archaeologists that data should be made available, but little increase in readiness to do so. Others felt that the increase is taking place only slowly. More has to be done to foster data sharing.

2.2.4 Reuse of data

Summary of main results

Why reuse is important

Sharing data is important but without (re)use the benefits associated with open data sharing would not materialise. There are many good arguments for making data available, for instance, that reported research results can be scrutinized and duplicative data collection prevented. Particularly strong however is the argument that reuse of data, for example to investigate new research questions, allows exploitation of previous investment. Preserved data that is being reused gains in value, otherwise it might be perceived only as a cost factor.

"Return on investment" expected by research funders explains much of the increasing pressure on researchers to share their data from publicly funded research for reuse. It is also very important for repositories to document not only downloads but actual reuse. Metadata with rich context information is essential for reusing data, as is a license that clearly states what users are allowed to do with the data.

Results for reuse

Results of the ARIADNEplus survey confirm that archaeological researchers often (re)use available data and allow some insights about what and how. The survey participants were asked, *"Did you / your research group in the last 2 years use any data which other researchers made available through a publicly accessible digital repository or databases?".* An astonishing number of 220 respondents said they did and also briefly described the data types and/or the sources.

In comments 34 respondents also gave reasons why they did not (re)use other's data. Most said relevant data was not available or posed some problems, difficulty to access or use, missing licensing information, lack of support, among others. Some also said that they did not need data from other researchers.

Main purposes of data reuse

The 220 respondents were asked "*What was the main purpose of the data reuse?"*, and three predefined purposes and the option "Other - please specify" offered. Building a database for the research community was a purpose for 31%, comparison to own results for 55%, and use together with own research data for 63% (multiple answers were possible). Few mentioned other purposes for the data (re)use, for example, to use it as test data (e.g. "test algorithms and approaches") or as a conceptual resource (e.g. "typology terms", "data structure").

Patterns

188 respondents provided descriptions of data and/or sources used, often only one or two data types (e.g. *"14C data"*, *"genetic data"*, *"find data"*, *"coins and ceramics"*), one source, e.g. *"AMČR"* [Archeological Map of Czech Republic] or "Archaeological Map of Bulgaria", or general statements such as *"several sources from the ADS (U York)".*

An analysis of descriptions including two or more data types or sources suggests three patterns of (re)use of available with own data which could be quite frequent in archaeology. Most often reused together were

- Information on sites (maps, distribution) and fieldwork reports (e.g. *"Site and monuments database, online archaeological reports"*),
- Cartographic, GIS and LiDAR data (e.g. *"National LiDAR and cartography"*; *"GIS, LiDAR, descriptions of archaeological monuments"*),
- Databases/catalogs of different artefacts (e.g. *"Coin and ceramic catalogs"*; *"Databases of inscriptions and coins"*).

The report chapter on data reuse includes an overview of groups of data types or sources; a word cloud of over 50 types of data, data sources and specific research objects; and a list of 96 named sources from which respondents sourced data.

Suggestions for ARIADNEplus

- ARIADNEplus should promote and support data reuse so that the investment in the collection of archaeological data can be exploited further for research, education and other purposes. Ways to enable with the ARIADNEplus infrastructure and services easy and effective reuse of data should be investigated.
- Different purposes and forms of data reuse should be considered to understand better actual practices of data reuse in archaeology so that these can be supported effectively.

2.2.5 ARIADNEplus data discovery & access services

Summary of main results

This section of the report presents and discusses the results for three closely related survey questions on data search and access: online availability of the different types of archaeological data that is part of the focus of ARIADNEplus, and how helpful it would be to discover and access it via the ARIADNEplus portal at both the collection level and item level.

ARIADNEplus will incorporate data from a wider range of archaeological research domains than ARIADNE, including environmental archaeology, maritime and underwater archaeology, biological and inorganic materials studies, radiocarbon, dendrochronology and other dating data methodologies, among others. Furthermore, the project aims to integrate more datasets at item-level to provide advanced semantic data search to find data items based on semantically defined relations.

Online accessibility of data types

Survey respondents rated the current availability of the different types of data ARIADNEplus aims to mobilise and integrate into the dataset catalogue and portal. The analysis of the results showed:

- good availability: Archaeological sites and monuments data (usually provided by heritage authorities), National GIS data and maps (from mapping agencies), and Satellite or airborne remote sensing data (in Europe offered freely by the European Space Agency);
- less good availability: Data and documentation from fieldwork (excavation, field survey/prospection, fieldwork reports), and Databases and catalogs of various artefacts (e.g. museum collections);
- poor availability: dating data (e.g. dendrochronology, radiocarbon) and scientific data/analysis of biological and inorganic remains). Also the availability of environmental archaeology and maritime & underwater archaeology data was perceived as poor.

ARIADNEplus portal for data discovery & access

One particularly important objective of the survey was to identify if respondents perceive support by the ARIADNEplus portal to discover and access more helpful for some data types than for others. Data types for which help is more appreciated could then be prioritised regarding mobilisation and integration in the ARIADNEplus data catalogue and portal.

A comparison of the online accessibility rating of the data types and helpfulness of portal support for discovering and accessing datasets or collections of such data showed:

- surprisingly, the appreciation of support was lower for data types for which the accessibility was evaluated as insufficient;
- respondents were most appreciative of portal support for discovery and access of Sites and monuments databases or inventories, national GIS data & maps, and satellite or airborne remote sensing data (e.g. LiDAR, although the online accessibility of these data types was rated much better than that of other data types;
- The rationale for ARIADNEplus should not be to prioritise support for data types which are already much more accessible than others. The fact that the more accessible types are being provided by national mapping and heritage authorities indicated ARIADNEplus should prioritise other data types.

Suggestions for ARIADNEplus

The survey results tentatively suggest the following priorisation of data types for mobilisation and integration in the ARIADNEplus portal:

- Data types with high or medium appreciation of portal support, and currently medium or low online accessibility. These types are ranked according to the appreciation of support and level of accessibility:
  - Excavation data (e.g. excavation archive)
  - Artefact/finds databases or image collections
  - Radiocarbon, dendrochronology and other dating data
  - Environmental archaeology datasets
  - Unpublished fieldwork reports
  - Field survey/prospection data
- Subject-based data types of the following domains:
  - Maritime and underwater archaeology data
  - Scientific data/analysis of inorganic remains
  - Scientific data/analysis of biological remains
  - Inscriptions, coins or other special databases

  Some of these data types may also require more FAIRness to enable their integration in the ARIADNEplus dataset catalogue and portal.

Item-level access

The survey participants also rated which data types they would find helpful for their research if able to search items within datasets integrated from multiple sources. The results do not add much to the evaluation above, except that artefact/finds databases or image collections were ranked highest.

2.2.6 ARIADNEplus special services for researchers and data managers

Summary of main results

The survey results for the ARIADNEplus services for researchers and data managers can be summarised as follows:

Services which ARIADNE/plus already provides: A very encouraging survey result is that respondents appreciated such services most (and may have already used them). These services are:

- Register a dataset in a portal that allows searching data from many providers
- Discover & access archaeological data stored in repositories in different European & other countries
- Spatially and/or chronologically defined search options

"Register a dataset in a portal that allows searching data from many providers" is the second highest ranked of all services listed (behind the not yet available to *"Search and visualize geo-spatial/GIS datasets"*). This signals a high interest of survey respondents to make datasets available for search and access via the ARIADNEplus data portal.

Multi-lingual search: This service is also already available on the data portal but was appreciated significantly less than the ones above. This does not speak against support of multi-linguality when portal users search for particular subjects because it ranks much higher than other services on the list.

Top on the list of new services – Search and visualize geo-spatial/GIS datasets: These services are the highest ranked among the new services, and are part of the plan of services ARIADNEplus will implement. It is worth noting that in the ARIADNE 2013 survey "A portal that makes it more convenient to search for archaeological data stored in different databases" was seen as "very helpful" for their research by 79% of respondents (very or rather helpful 96%, N=481), while "Services for Geo-integrated data" by 52% (very or helpful 81%, N=471). With the portal in place, services for geo-spatial/GIS data now are on top of appreciated new services.

Lowest on the list – Map a database (schema) to the CIDOC-CRM extended for archaeological research data: This result does not come as a surprise because the service is specifically for data managers (databases, repositories) and these make up only 20% of the survey respondents.

Also low – Visualize in 3D the layers of an excavation and the related documentation: Survey respondents also evaluated this service as less helpful for their research or data management work than others. 53% of respondents were archaeological researchers (field work) but also many of these may have seen the service as beyond their expertise or what they might use in practice.

Services in the middle range: All other services were in the middle range of appreciation, judged as "very helpful" or "helpful" by between 78.3–83% of the responses per service. Listed according to the percentages (more to less):

- Use Linked Data to interlink own and other datasets
- Annotate images (e.g. artefact or laboratory images) and link them with other content,
- Post a picture of an object and get suggestions for similar ones
- Display and manipulate visual data objects (e.g. RTI images, 3D models, LiDAR data)
- Link and present together visual media (e.g. a 3D model) and related documentation
- Process many documents (using Natural Language Processing) to find those on certain topics or specific information they may contain
- Identify & extract information from textual sources (e.g. a document repository) to produce metadata
- Annotate texts (e.g. fieldwork or laboratory reports) and link them with other content
- Align own vocabulary terms with international thesauri (e.g. Getty Arts & Architecture Thesaurus)

"Future" services: Two services were suggested that are technically very demanding and not included in the work plan of the project, hence potential future service. These are text analysis and recommendation provided online on top of large document repositories *("Process many documents (using NLP) to find those on certain topics or specific information they may contain")* and image recognition, comparison and retrieval *("Post a picture of an object and get suggestions for similar ones").* The latter service was seen a bit more interesting.

General evaluation

The two lowest ranked services still were considered as very helpful or helpful by close to 75% (74.3%, 74.5%) of respondents, those in the middle range by 78.3–83%, and the ones on top by 93.5–96.3%. However, of the latter only *"Search and visualize geo-spatial / GIS datasets"* (93.5%) is not yet available on the data portal.

It is also worth noting that in the middle range, after the already available multi-lingual search (83%), first comes *"Use Linked Data to interlink own and other datasets"* (82.9%), followed by four services for working with visual content (e.g. high-resolution images, RTI images, 3D models, LiDAR), at 81.2–82%. This signals a high interest in visual content services, at least considerable more than for textual content.

Suggestions for ARIADNEplus

The main suggestions that can be derived from the survey results are:

- Devote special attention to the new services for search and visualisation of geo-spatial/GIS datasets.
- Prioritise also the use of Linked Data for interlinking datasets, particularly at item-level.
- Continue to enhance the existing and develop new visual content services of interest.
- Evaluate further which services for textual content are of interest to users, including services not yet considered.
- Promote further the use of CIDOC-CRM by making clear its capability to integrate research data conceptually, especially regarding the ontology extensions developed in the ARIADNE project for archaeology (e.g. excavations, standing structures, epigraphy).
- In the testing and evaluation of online tools with end-users investigate if there are any reservations against using them them as services in a Cloud-based virtual research environment instead of a stand-alone desktop or online tool.

2.2.7 Training needs

Summary of main results

Survey respondents who answered the question (around 330) thought that training on all of the listed activities would be very helpful or helpful between 86.6–94.5%. The percentages for "very helpful" ranged between 48.5–67.3%.

Regarding "very helpful" significantly less appreciation was expressed for training in how to create and implement a data management plan (DMP), manage a digital repository, produce metadata and use domain vocabularies to describe archaeological datasets. Data science skills, managing datasets of a large archaeological project, depositing project datasets in a digital repository and, above all, apply open/FAIR data principles in archaeology were scored higher.

The open/FAIR data principles are generally relevant for all participants and training on how to apply them ranked on top. Among the survey participants a higher than "average" awareness of the principles can be assumed.

That researchers are the largest group in the survey sample certainly had a considerable impact on the results. Researchers worry about additional data-related work, which explains why training regarding DMPs, metadata and vocabularies is appreciated less.

When researchers need to deposit data in a repository, the question of metadata comes up. All studies on data sharing through digital repositories, including the ARIADNE/plus surveys, found that researchers consider the effort to provide the required metadata as a barrier to open data sharing. While data repositories and users would benefit from high-quality metadata, data creators face the burden and usually prefer not to invest much effort on providing metadata.

Inconsistently, in the ARIADNEplus survey, training on data deposition appeared to be welcome, despite the (not recognised) fact that this would require dealing with metadata and vocabularies. Awareness of an increasing expectation that data from funded research projects should be deposited may have contributed to this result.

To define and implement a data management plan seems to be unappealing. But training on how to manage datasets of a large archaeological project was appreciated. Such management is necessary and the task to do it can be taken on by, or delegated to, one or two team members who are trained to do it.

Training to develop data science skills, i.e. use advanced data processing and analysis methods, promises to provide more value from the effort invested in the collection of data. It was considered as very helpful by more respondents than other five items on the list, only data deposition and the open/FAIR data principles were appreciated more.

Training to manage a digital repository of archaeological data was appreciated less. But managing a digital repository is a professional activity of a smaller group and distinct from being a researcher; only 13% of the survey respondents were managers of an institutional repository.

Suggestions for ARIADNEplus

The following suggestions take account of the ARIADNEplus plans for training, the survey results on training needs, and the background on data management plans, research data management, and data science. Furthermore, the fact that the closely related SEADDA project focuses on data respositories for archaeology is considered.

FAIR data principles

The survey found that training for the application of open/FAIR data principles in archaeology would be appreciated most, both by researchers and data managers, and ARIADNEplus is committed to support these principles within the archaeological sector.

The project has a work package comprised of six tasks dedicated to policies and good practices for FAIR data management. The tasks include evaluating implications of the implementation of the principles in the sector; provide policy support tools such as a flexible Data Management Plan template and supporting wizards; guidelines and support on repository creation, management and quality control; guidance on how to realise FAIRness of data taking account different regulations in participating countries, IPR-related and other issues; and offering practical training material and workshops.

Significant contributions to capacity building and take-up of the FAIR principles by these activities can be expected. The background given for training needs suggests that:

- the contributions should be as practical as possible, distinct from the broad wave of general information on the FAIR principles by ever more on the FAIR data “bandwagon”;
- project partners involved in the tasks mentioned should consider what falls, at least at a general level, within the remit of other organisations, e.g. support for DMPs and research data management in general by university libraries and repositories;
- ARIADNEplus training activities on FAIR data should focus on what matters for archaeological researchers and data managers specifically.

Data Management Plans (DMP)

Survey respondents ranked data management planning much below other training opportunities. To define and implement a DMP and related activities (metadata, vocabularies) adds work, but researchers are unsure they will benefit from this additional work. The background section on DMPs

notes that requirements defined by research funders for such plans are varied and may be inconsistent. Critiques argue that DMPs generate a lot of work while there is little evidence for positive effects.

Suggestions for countering these negative perceptions of DMPs are

- provide case studies on data management planning of archaeological projects making clear the benefits for projects and researchers involved;
- consider a DMP template that covers the minimum standard requirements while focusing more on the practicalities of different archaeological projects large and small;
- in general, rather than dry information about creating DMPs consider how to help the archaeological research community step up practical training of PhD students and early-career researchers in tried and proven data management practices.

Data managers of projects and repositories

Survey respondents appreciated training on how to manage datasets of a large archaeological project, while less so on how to manage a digital repository. The latter is a professional role and only a minority of respondents were managers of an institutional repository.

Data managers of projects

Training for data managers of archaeological projects will be provided in the ARIADNEplus trans-national access (TNA) programme, specifically under the TNA themes Data Stewardship and Implementing Interoperability. These researchers and data managers will take a keen interest in developing metadata for archaeological datasets with domain vocabularies in order to manage, use and share FAIR datasets.

Managers of repositories

In matters pertaining to archaeological repositories ARIADNEplus will benefit from coordinating activities with the SEADDA project, in which many consortium partners participate. SEADDA aims to foster the development of archaeological data repositories in countries where the research community lacks an appropriate repository, while ARIADNEplus supports finding and accessing data that is being shared through existing repositories.

Therefore, ARIADNEplus could

- help developers of repository initiatives plan participation of the repositories in its research infrastructure at an early stage,
- repository projects which are more advanced might benefit from available services, for example, by using data description and mapping services for representative initial datasets.

Research Data Management (RDM)

Training provided in the TNA framework as well as tutorials and workshops partners will organise on ARIADNEplus services and tools do not scale. In order to reach a higher number of researchers and data managers with information and guidance on RDM, ARIADNEplus can

- continue to make them aware of available guides to good practice, e.g. the guides offered online by the Archaeology Data Service/Digital Antiquity;
- provide a series of webinars on FAIR archaeological research data with contributions by experts from the ARIADNEplus partnership;

- organise “train the trainer” workshops at conferences or research institutes so that research data managers can serve as disseminators of good practices, including in questions of IPR and copyrights and sensitive archaeological data.

Data science skills

Survey participants appreciated training to develop data science skills, i.e. use of advanced data processing and analysis methods, more than the other five items on the list. Such training promises to get more from the effort invested in the collection of data.

ARIADNEplus has limited capacity to raise the level of data science skills of archaeological researchers. What the project can offer is:

- raise awareness of Open Science practices related to the sharing and (re)use of FAIR data;
- support the documentation and integration of archaeological datasets based on metadata standards and domain vocabularies;
- use of ARIADNEplus Cloud-based Virtual Research Environments (VREs) for data science tasks;
- a specific activity could also be to organise a group of high-potential young researchers for a collaborative synthetic project on human migration related to the initiative of the Coalition for Archaeological Synthesis (CfAS), to explore the potential of ARIADNEplus VREs, services and datasets for comparative analyses and synthetic research.

3 Development in policies and technologies

This chapter presents important recent developments in European Union policies regarding research and research infrastructures. Regarding research, the key policies are Open Science and FAIR Data. On the research infrastructure front, the key development is the European Open Science Cloud, as a policy for the "market" of research infrastructures including the general technological approach.

The developments are related because the European Open Science Cloud (EOSC) is meant to support Open Science practices, especially sharing of FAIR data, as well as other research resources. As Carlos Moedas, Commissioner for Research, Science and Innovation (2014-2019) of the European Commission put it, Open Science *"is a move towards better science, to get more value out of our investment in science and to make research more reproducible and transparent. (…) And that implies that research data are findable and accessible and that they are interoperable and re-usable. In essence, this is what the Open Science Cloud is about: an open and trusted environment where research data can be safely stored and made openly available"* (Moeda 2016: 4).

3.1 European Open Science Cloud (EOSC)

The European Open Science Cloud (EOSC) is intended to provide a common virtual platform for researchers to store, (re)use and analyse FAIR data for research, innovation and educational purposes. This summarises what the EOSC should help researchers to do. From the perspective of the European Commission the EOSC is a policy for research e-infrastructures and data. This section briefly describes the research infrastructure policy while FAIR research data are addressed in *Section 3.3*.

3.1.1 Lack of integration between research e-infrastructures

Research e-infrastructures and services are important pillars and drivers of collaborative and data-intensive research. They provide researchers access to distributed, but shared digital resources (data, services, tools), and can support advanced and innovative research across institutional and disciplinary boundaries.

Considerable investment has already been made at the European and national levels to implement research e-infrastructures for different disciplines. But the e-Infrastructure Reflection Group (e-IRG), the main advisory body for European e-infrastructures, regulary noted fragmentation, lack of coordination and interoperability between the existing e-infrastructures. The e-IRG asked all stakeholders to increase coherence and resource sharing (e.g. e-IRG 2013). Interoperability and resource sharing are required to enable synergies, cost-effectiveness and sustainability in supporting innovative ICT-enhanced research across disciplines.

Interoperability between e-infrastructures and sharing of resources has not been sufficiently realised. The European Open Science Cloud (EOSC) is now attempting to provide a solid basis for tighter integration of research e-infrastructures, research data and other resources.

3.1.2 Development of the EOSC

The EOSC has two essential dimensions: it is a policy for the "market" of research infrastructure, and a general technological approach to such infrastructure.

The European Cloud Initiative and the EOSC

In the discussion about the EOSC the overall European Cloud Initiative has often been overlooked. The initiative supports the strategy of the European Commission to help create a Digital Single Market in Europe. In this strategy the European Cloud Initiative should strengthen Europe's position in data-driven innovation of industries to improve economic competitiveness[1].

The initiative aims to provide Europe-wide data infrastructure and networking, high-speed connectivity for data transfer between big data centres, and advanced high performance computing (HPC), for both science and industry. In this setup the data infrastructure should provide storage, preservation, access and management of big data, including centres for high-capacity cloud solutions and HPC. The public and private investment needed to implement the European Cloud Initiative was estimated at €6.7 billion, of which the European Commission estimated that €2 billion would be spent from Horizon 2020 funds.

The European Open Science Cloud (EOSC) was proposed in April 2016 in the European Commission Communication "*European Cloud Initiative - Building a competitive data and knowledge economy in Europe"* (European Commission 2016a). The EOSC should offer Europe's 1.7 million researchers and 70 million science and technology professionals a virtual environment to effectively exploit large volumes of data and technologies for (big) data processing and analysis. The European Cloud Initiative highlighted the scientific community as the initial user base to be expanded to industries and businesses as well as public sector organisations. The EOSC, based on the Europe-wide data infrastructure, should provide a trusted environment for the scientific community to store, share and re-use scientific data and results across borders and disciplines. Technically it should allow integration of major existing but fragmented digital infrastructures, and provide a basis for emerging new ones.

In March 2018, two years after the European Commission had issued the Communication on the European Cloud Initiative, the Commission presented an *Implementation Roadmap* for the EOSC (European Commission 2018a). Before that, EOSC-focused project consortia were invited and selected for funding, for example, eInfraCentral, EOSCpilot and EOSC-Hub. Dedicated funding of OpenAIRE (advance) and the FREYA project (e-infrastructure for persistent identifiers) also belong to the initial wave of EOSC projects.

The EOSC-Hub, led by the EGI Foundation and funded until December 2020, plays a core role in the EOSC implementation. The project should create the integration and management system of the EOSC. The system is to deliver an integrated catalogue of services, software and data from the EGI Federation, EUDAT-CDI, INDIGO-DataCloud and several other pan-European research e-infrastructures. On 23 November 2018 the initial EOSC portal[2] was launched, enabling access to the first sets of services and resources. The portal has been built in a collective effort by OpenAIRE Advance, EOSC-hub, eInfraCentral and EOSCpilot projects.

Around that time €600 million in H2020 funding was allocated for setting up the EOSC by 2020 (European Commission 2018e). A full-cost study for running the EOSC will be conducted by the EOSC Governance to support defining an appropriate mix of funding (European Commission, Member States, and others).

[1] European Commission: Digital Single Market - The European Cloud Initiative, https://ec.europa.eu/digital-single-market/en/%20european-cloud-initiative

[2] EOSC portal, https://www.eosc-portal.eu

EOSC Model action lines

A Commission Staff Working Document (2018) gives an overview of the EOSC implementation roadmap. It describes an "EOSC model" by distinguishing six lines of required implementation actions. These action lines are for *"a pan-European federation of data infrastructures built around a federating core and providing access to a wide range of publicly funded services supplied at national, regional and institutional levels, and to complementary commercial services"* (European Commission 2018a: 9).

The working paper summarises the action lines in a figure that is extremely useful to understand the general setup of the EOSC (Figure 1). The "Architecture" section makes clear what is highlighted above: the EOSC, implemented as a federation of infrastructurs, should provide *"the solution to the current fragmentation in research data infrastructures which are insufficiently interoperable"*.

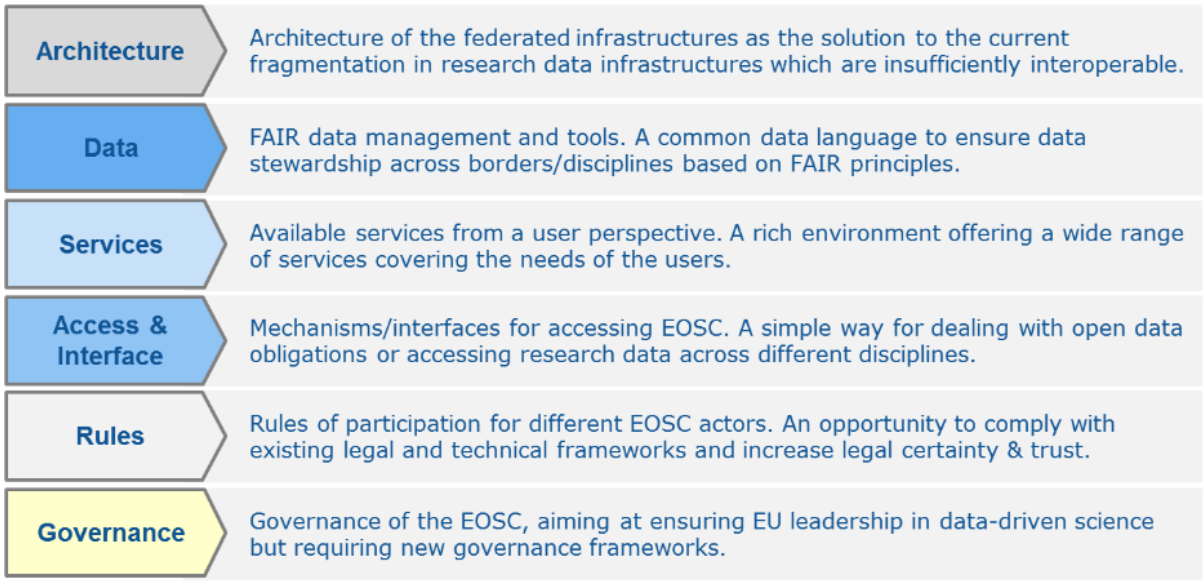

***Figure 1**: The six lines of action of the EOSC model. Source: European Commission 2018: 9.*

Regarding "Data" the core role of FAIR data principles is highlighted. Such principles should provide the basis of a *"common data language to ensure data stewardship across borders/disciplines"*. The FAIR data principles to make data Findable, Accessible, Interoperable and Re-usable are addressed in *Section 3.3*.

The overall governance of the EOSC is with the EOSC Governance Board of representatives selected by each EU member state and associate country. The role of this board is to supervise the implementation of the EOSC, taking account of the requirements and supporting coordination among the EU member states. The EOSC Executive Board is made up of representatives from the research and e-infrastructure communities. The board members are appointed by the European Commission to provide advice and support on the strategy, implementation, monitoring and reporting on the progress of the implementation of the EOSC. The EOSC Secretariat supports the two governance groups and provides the information platform[3].

In June 2019 the EOSC Executive Board published the strategic implementation plan for the EOSC that gives a comprehensive overview of the planned activities which will contribute to the implementation for the period 2019-2020. It is primarily intended for use by stakeholders engaged in building the EOSC, and includes the list of related Horizon 2020 projects (EOSC Executive Board 2019). Preliminary

[3] EOSC Secretariat, https://www.eoscsecretariat.eu

connection of most (major) data infrastructures and services to the EOSC is scheduled to be reached in the second quarter of 2020 (European Commission 2018b).

3.1.3 ARIADNEplus and the EOSC

The EOSC is variously described as *"a pan-European federation of data infrastructures"* that provides access to a wide range of services (European Commission 2018a: 9), *"Europe's virtual environment for all researchers to store, manage, analyse and re-use data for research, innovation and educational purposes"* (EOSC Portal)[4], or *"a one-stop-shop"* for users *"to find, access, and use research data and services from multiple disciplines and platforms"* (e-IRG 2018).

How does ARIADNEplus relate to this multi-facetted, difficult to grasp entity to which the project aims to align its research e-infrastructure and services?

As elaborated by Niccolucci & Richards (2019), alignment and integration with the EOSC is part of the innovation objective and approach of ARIADNEplus, which is "*based on the provision of innovative and advanced web services in a cloud environment, coherent with the vision, and integrated in the implementation of the EOSC. ARIADNEplus will progressively set up an ecosystem for digital archaeological research which incorporates data and services and enables the use of cloud-based Virtual Research Environments (VRE)"*. Further, referring to envisaged impact, *"The overall strategy with regard to improved use of resources, economies of scale and cost-savings is Cloud-based virtualisation and integration in the EOSC"* (Niccolucci & Richards 2019: 9 and 23).

Thus, ARIADNEplus follows the EOSC vision of providing a wide range of services for research in Cloud-based Virtual Research Environments (VRE), indeed it is ahead of the EOSC in this regard. ARIADNE services such as the data aggregation and data portal services have already been implemented as VREs on the fully operational D4Science platform[5], supported by project partner Institute of Information Science and Technologies (ISTI) of the National Research Council of Italy.

Cloud-based VREs for e-archaeology will allow researchers to use online tools for different tasks and types of data. Providing research tools online in Cloud-based environments avoids researchers investing effort to acquire, implement, maintain and upgrade them. The approach allows cost-savings for the research community while at the same time opportunities for research groups to jointly address research questions.

While this will be enabled by ARIADNEplus VREs on the D4Science platform, the core relation of ARIADNEplus, and other domain research e-infrastructures, with the EOSC will be established based on catalogues of services and data resources. Regarding datasets, ARIADNE was a front-runner in establishing a standards-based catalogue of datasets. The ARIADNE Data Catalogue Model has inspired the model of the humanities e-infrastructure cluster project PARTHENOS, which developed a model with additional entities (Frosini et al. 2018), which are also included now in the ARIADNEplus model. Thus a harmonisation of catalogues for humanities research resources has been achieved. The optimal way to exchange catalogue information with the EOSC "marketplace" of service and data resources will be investigated. For example, ARIADNEplus services will be registered in the D4Science catalogue of service resources and could be shared from there with the common EOSC catalogue.

[4] EOSC Portal: For providers, https://eosc-portal.eu/for-providers

[5] See the overview of the envisaged ARIADNEplus service portfolio in *Section 4.10*.

What does this mean for the ARIADNEplus user community?

- ARIADNEplus will align and integrate via D4Science as far possible with the EOSC, contribute services to the EOSC, as well as use relevant services from the EOSC .
- Archaeological research and data management organisations (e.g. repositories) will not need to describe resources they wish to share in the EOSC catalogue, their information about available datasets and services can be aggregated by ARIADNEplus and provided as a collective contribution to the EOSC.
- Research groups of archaeological domains can be provided dedicated VREs on the D4Science platform, which allow them using data and online tools in collaborative research, while avoiding to implement and maintain such tools themselves.

3.2 Open Science

3.2.1 Open Science as a research policy priority

The current debate about the system of scholarly research and communication is all about "open", including, among others, open access (publications), open data, open science and, of course, open research infrastructures. As many proponents suggest, "openness" provides much potential for novel forms of research collaboration, including participation of citizens, and innovative generation and publication of new knowledge.

Open Science as a priority and guiding principle of research policy at the European level has been introduced in 2015 by Carlos Moedas, the Commissioner for Research, Science and Innovation (2014-2019). The Commissioner adopted the concept of "open innovation" and pushed for more openness of research in the European Research Area (ERA). The title of his speech at the ERA 2015 Conference, *"Open Innovation, Open Science, Open to the World"* (Moedas 2015) has also been used for a commissioned study (European Commission 2016b) which describes the background and requirements for a research policy set to openness in all respects.

In European Commission research policy information and documents the term Open Science is now being used regularly when open access to shared research resources, and expected impact of such access, are being addressed. The expected impact of openness is a transformation of science leading to advances in knowledge, innovation and societal benefits. Innovative digital, ICT-enabled research is understood to play a key role in this transformation. For instance, the webpage on Open Science of the European Commission's Digital Single Market section states, *"Open Science aims at transforming science through ICT tools, networks and media, to make research more open, global, collaborative, creative and closer to society"*.[6]

3.2.2 Open digital science

A paper on *Digital science in Horizon 2020* by a unit of DG Connect, the Directorate-General for Communications Networks, Content and Technology of the European Commission provides a useful overview of the scope of open digital science (European Commission 2013).[7] The paper was issued before the public consultation in 2014 on "'Science 2.0': Science in Transition" (European Commission 2014).

[6] European Commission: Digital Single Market policy: Open Science, http://ec.europa.eu/digital-agenda/en/open-science

[7] See also: http://ec.europa.eu/digital-agenda/en/open-science

The paper summarised in a figure the guiding idea that take-up of digital (open) science will benefit both the research community and society at large. Innovative use of digital methods and tools by researchers would enhance research collaboration, involvement of citizens, and transparency and relevance of better accessible research outcomes.

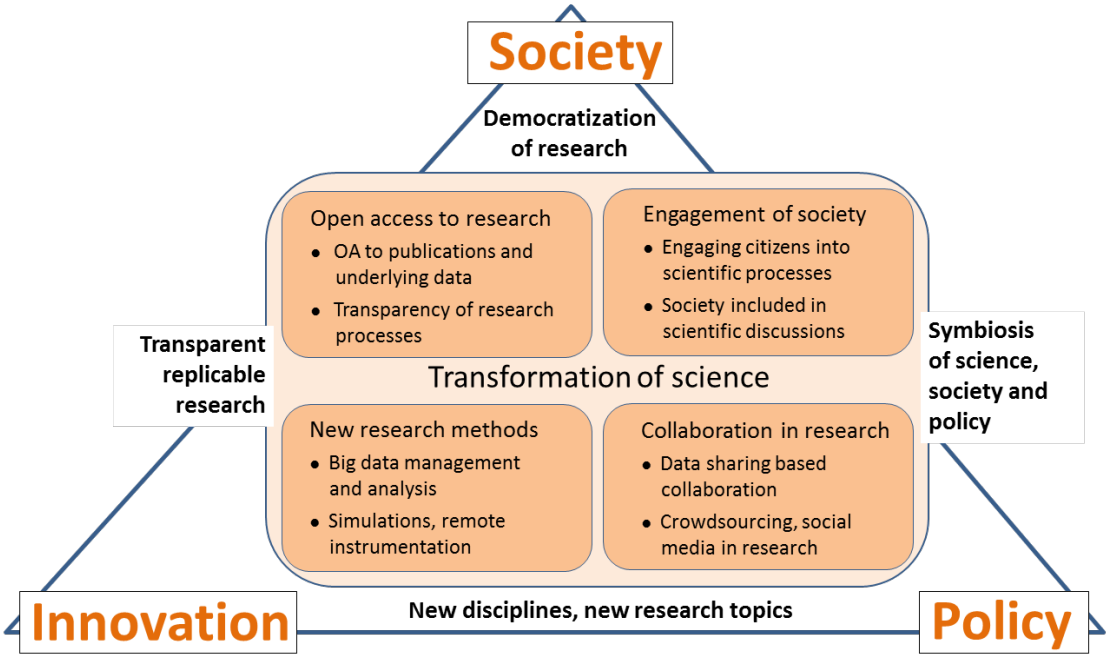

***Figure 2**: Digital Science Triangle. Source: European Commission 2013: 2 (adapted for better readability)*

Some of the points in this figure may appear as somewhat outside the research environment and data resources of most archaeologists, but the overall directions will also fit for archaeological institutes and researchers, if read in more familiar terms like building larger database for computational archaeology ("big data"), public and community archaeology ("engaging citizens"), while "crowdsourcing" may not strike a chord. Such sourcing means to employ a web application for collecting many inputs (ideas, data) from researchers and/or non-experts to a project.

3.2.3 Different views on openness

The overall vision of Open Science is making the research process and its results as transparent and accessible as possible in order to increase societal benefits which can be derived from scientific and technological knowledge. In practice, proponents have different perspectives on Open Science and emphasise some aspects more than others.

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

The core but difficult to grasp element of Open Science is "openness". A report of the European ERA-net project e-InfraNet provides a useful discussion of the concept of openness in various contexts (e.g. research, content/data, software, infrastructure, standards, innovation). The report suggests openness as the *"default modus operandi"* for all publicly funded research and educational resources,

with "open" as the preferable approach *"not as an end in itself or as an ideology"* (e-InfraNet 2013: 10).

The highly influential *Science as an Open Enterprise* report of The Royal Society (2012) urged researchers to strive for *"intelligent openness"* for which data should be *"accessible, useable, assessable and intelligible"* (The Royal Society 2012: 12 and 14). The report emphasises that research data *"must provide an account of the results of scientific work that is intelligible to those wishing to understand or scrutinise them".* Recipients should be able to understand and judge what is communicated, particularly the nature of the claims that are made, and the reliability of the source and evidence provided. The report also considers that (re)users of open data could range from the highly expert in the field to the non-specialist.

3.2.4 Open science and data in archaeology

The concept of Open Science is of course highly relevant for archaeological research. It includes open sharing of research resources (data, tools), novel forms of research collaboration, including a deeper citizen involvement, and can greatly extend the societal relevance and reach of archaeological knowledge.

The need and challenges of open research practices and resources have also been discussed in the archaeological community (e.g. Beck & Neylon 2012; Beck 2013; Costa et al. 2012; Kansa 2012; Lake 2012; Wilson & Edwards 2015). A recent publication on *Open Science in Archaeology* by Marwick et al. (2017); a large group of recognised archaeological researchers, has greatly added to the awareness of open research and data in the field.

An open data imperative is particularly strong in this field: excavation of sites destroys the primary archaeological evidence, the work on archaeological heritage is done in the public interest, and there is little commercial relevance of archaeological data. Therefore openness should become embedded in archaeological research practices as *"the default modus operandi"* (e-InfraNet 2013) so that the advantages of accessible and reusable data gain priority over the interest of the individual researcher.

However, many archaeologists are not yet well prepared or equipped for open data sharing. As the matter is complex, strong leadership with regard to policies/mandates, supportive institutional measures (e.g. capacity building, training of researchers), and state-of-the-art digital archives are required.

3.3 FAIR data

An important development over the last few years has been the strong interest in the FAIR data principles by different domains in the sharing and reuse of open data through repositories and other research infrastructures. Published only two and half year ago (April 2016), the FAIR data principles seem already have replaced the previously more familiar terms associated with Open Data. It appears that the stakeholder communities were looking for a fresh concept to promote the agenda of data sharing and reuse, particularly a broader and more technical one than Open Data. Meanwhile FAIR is being used for different things, including data(sets), digital repositories and other e-infrastructure.

3.3.1 Open versus/and FAIR Data

In the ARIADNEplus survey "FAIR" was only mentioned in one question on training needs. Here "open/FAIR data" has been used to also relate to the more familiar term of Open Data. Elsewhere FAIR or open has been avoided and instead data accessible in or shared through an "accessible repository" mentioned. Throughout this report "open sharing of data" is being used to make clear a very important

point: researchers often share data directly with colleagues but tend not to make it openly available, e.g. through an accessible repository, for others to (re)use.

Open Data

The term "Open Data" has been used for many years by the research data management community, as well as other user communities. Widely referenced is the "Open Definition" of the Open Knowledge Foundation which defines "open" briefly as: *"Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)"*.[8] Details are then given on criteria which should be fulfilled so that data, content or knowledge can be considered as open, especially that it should be shared under an open license or in the public domain.

Most widely used are Creative Commons licenses which allow different levels of openness, explained in simple as well as formalised legal terms, can be embedded in any digital content, and are machine-readable. Grabus & Greenberg (2019) present an extensive overview of existing and new initiatives for rights and licensing options for data sharing, and new initiatives

Licensing is also a key principle in FAIR, as "R1.1. (meta)data are released with a clear and accessible data usage license" (see section below). The main difference to Open Data is that the FAIR principles do not imply that the data is "open" or "free" in the sense of uncontrolled and free of charge access and (re)use.

As Mons et al. (2017) explain, *"None of these principles necessitate data being 'open' or 'free'. They do, however, require clarity and transparency around the conditions governing access and reuse. As such, while FAIR data does not need to be open, in order to comply with the condition of reusability, FAIR data are required to have a clear, preferably machine readable, license."* They also highlight that the different approach of FAIR in this regard allows participation of data holders that otherwise could not be involved: *"The transparent but controlled accessibility of data and services, as opposed to the ambiguous blanket-concept of 'open', allows the participation of a broad range of sectors – public and private – as well as genuine equal partnership with stakeholders in all societies around the world"* (Mons et al. 2017).

FAIR Data

The development of the guiding principles for FAIR data has been initiated in 2014 by participants of a research data workshop in 2014 ("Jointly Designing a Data Fairport" in Leiden, Netherlands) who discussed how obstacles to data discovery and reuse could be overcome. The meeting concluded with a draft formulation of a set of principles that were subsequently elaborated upon in greater detail *"namely, that all research objects should be Findable, Accessible, Interoperable and Reusable (FAIR) both for machines and for people"* (Wilkinson et al. 2016). The 15 principles are listed below, but no commentary is intended. This can be found in the paper and related publications (e.g. Expert Group on FAIR Data 2018; Mons et al. 2017).

[8] Open Knowledge Foundation: The Open Definition, https://opendefinition.org

The FAIR guiding principles

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1. the protocol is free, open and universally implementable

A1.2. the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

I2. (meta)data uses vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

To be reusable:

R1. (meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with data provenance

R1.3. (meta)data meet domain relevant community standards

Source: Wilkinson et al. 2016.

The listing of the FAIR guiding principles illustrates how detailed and largely technical these principles are. Consequently also the knowledge and effort that is required for implementing them is substantial.

The European Commission has brought together an Expert Group on FAIR Data to analyse what is needed for "turning FAIR into reality" and suggest concrete actions for all stakeholders for how to do so (Expert Group on FAIR Data 2018). A strong promoter of making research data FAIR is the GO FAIR initiative[9], launched in late 2017 by the Ministries of Science of France, Germany and the Netherlands, with support and coordination offices in these countries, and a growing number of implementation networks.

While there is a FAIR "boom" in the international research data management community no wide awareness, lest knowledge, of the principles among researchers can be assumed.

Ivanović et al. (2019) surveyed staff of 32 repositories on how FAIR their repository is. They found misconceptions of some of the FAIR principles and their implementation. Particularly highlighted is that the I2 FAIR principle [I2. (meta)data uses vocabularies that follow FAIR principles] is often missed and appears difficult to implement. Therefore it is suggested a service that could aid in this implementation would be particularly helpful.

[9] GO FAIR, https://www.go-fair.org

The Figshare 2018 *The State of Open Data* survey included the question *"How familiar are you with the FAIR principles in relation to open data?".* Of the 1,239 respondents who answered the question, 60% said they *had never heard* of the FAIR principles before, 25% had heard of the principles but were not familiar with them, while 15% claimed being familiar with the principles. Researchers from different disciplines participated in the survey, e.g. biology 19.3%, medicine 14.2%, social sciences 14.1%, earth & environmental sciences 10%, engineering 6.8%, humanities 3.7%. The survey directors state, *"This lack of awareness is concerning as the FAIR principles are being rapidly adopted by publishers, funders and institutions worldwide but there is a crucial gap in educating researchers on what is expected of them"* (Figshare 2018: 11; Figures for familiarity with FAIR: extracted from the available dataset, for disciplinary composition from the interactive visualisation based on the dataset).

3.3.2 Costs of (not) FAIR

A study of PricewaterhouseCoopers EU Services (2018a) for the European Commission, Directorate-General for Research and Innovation, estimates *"that the annual cost of not having FAIR research data costs the European economy at least €10.2bn [billion] every year".*

This figure is mainly based on estimated costs of time wasted by researchers due to not having access to FAIR data/content (44%) and storage costs for redundant copies of data/content (52%), which would be avoided if available in the EOSC. Another significant cost factor is licenses for not open data/content (3.5%). In addition, the study assumes that costs of a number of not quantified consequences such as lack of machine readability of data, shortcomings of research quality, etc. could account for another €16 billion.

PricewaterhouseCoopers EU Services (2018c) also made available a guidebook and calculation tool for research-performing organisations and data infrastructures (e.g. repositories) to produce a FAIR data cost-benefit analysis.

A study by Oxford Research and Højbjerre Brauer Schultz for the Danish Agency for Science and Higher Education (2018) provided a preliminary analysis of the potential for implementing FAIR data in Denmark. It includes an analysis of costs and benefits of implementing FAIR and an examination of barriers and opportunities in implementing FAIR data at the Danish research institutions. The analysis generally indicated a potential positive socio-economic value from introducing FAIR data in Denmark in the long run. It also suggests that the size of the benefit will largely depend on the extent of successful implementation of FAIR data and the actual reduction in researchers' effort spent on data-related work.

It is expected that at some point research funders will start making data-related costs eligible for FAIR data only. This will require a lot of investment in training of FAIR data managers of research organisations, repositories and other research infrastructure. Barend Mons, chairman of the first High Level Expert Group on the EOSC, estimates that over the next decade 500,000 data managers would need to be trained to make research data FAIR for the European Open Science Cloud to be successful, one data expert per 20 researchers (Mons 2016).

3.4 Current status of open research data policies

SPARC Europe & DCC (2019) give an overview of the current status of open research data policies in Europe. The study found that since 2016 more governments have put open research data on their agenda. Previously the focus was mostly on information and data produced by public sector agencies, and often reused by researchers.

Increased investment in digital research infrastructures, both national and common European ones, the Open Research Data Pilot for Horizon 2020, and other factors shifted the focus to research data of funded research organisations and projects. But the study notes that still there is no consensus position across EU Member States or even within countries.

The latest progress report on the implementation of the European Commission's Recommendation C(2012)4890 on Access to and Preservation of Scientific Information in Europe comes to the same conclusion (Tsoukala et al. 2018). The recommendation addressed the EU Member States and called for improved policies and practices on open access to scientific publications and research data, as well as on the preservation and re-use of scientific information. The Member States and participating associated countries regularly report on their measures and development towards agreed goals.

The implementation of this policy actually is core common strategy regarding open access to research publications and data. In 2018 the Recommendation was updated to reflect recent developments in research practices relating to Open Science and the European Open Science Cloud initiative (Commission Recommendation (EU) 2018/790, 25.4.2018; European Commission 2018b/c).

The mentioned priority of public sector information/data was due to required complicance with the EU Directive on the Re-use of Public Sector Information (Directive 2003/98/EC and amendments). In 2019 it was replaced by the Directive (EU) 2019/1024 on Open Data and the Re-use of Public Sector Information, in force since 16 July 2019. Member States have to transpose it into national law until 16 July 2021. The Directive, among other points, includes that Member States have to implement policies and actions *"aiming at making publicly funded research data openly available ('open access policies') following the principle of 'open by default' and compatible with FAIR principles"* (Article 10).

4 ARIADNEplus user needs survey

This chapter presents the results of the ARIADNEplus online survey that has been carried out as part of project Task 2.2 - Reviewing the Community Needs and the Market.

In line with the task brief a broad survey on the needs of the ARIADNEplus user community regarding data sharing, access and (re)use, new services (as developed by the project), and related training needs has been carried out. Core groups addressed by the survey have been archaeological researchers and data managers/providers.

Results of the ARIADNEplus survey are compared, where possible, to those of the ARIADNE 2013 survey and, particularly, to planned new technical and other services. The analysis of the results focuses on the match between the perceived user needs and planned ARIADNEplus service portfolio, and suggestions are provided on activities likely to enable an optimal match.

4.1 Surveying ARIADNEplus community needs

Objectives

The objectives for the online survey according to the Description of Work (DoW) of the project were to collect information on needs of the ARIADNEplus user community regarding data sharing, access and (re)use, new services (as developed by the project), and related training needs. Results of the ARIADNEplus survey were to be compared, where possible, to those of the ARIADNE 2013 survey and, particularly, to planned new technical and other services. Furthermore, the analysis of the results had to focus on the match between the perceived user needs and planned ARIADNEplus services, and suggestions to be provided on activities likely to enable an optimal match.

Target participant groups

The core target group addressed by the international survey was the archaeological research community, particularly archaeological researchers and data managers and providers; this group also includes curators of museum collections and databases. Other potential users of the ARIADNEplus digital infrastructure and services such as heritage management officers and professionals, or citizen scientists interested in archaeological research, were not excluded, but not specifically addressed.

Questionnaire development & testing

The survey questionnaire has been developed and implemented by Salzburg Research on the online survey platform Lime Survey[10].

The WP2 lead, Archaeology Data Service, in particular Prof Julian Richards, contributed to the design of the questionnaire; some questions have been discussed extensively.

Four members of the ARIADNEplus partnership tested the questionnaire and gave valuable suggestions for additions and improvements: Andres Dobat (AU), Attila Kreiter (HNM), Daniel Löwenborg (Univ.Uppsala) and Benjamin Štular (ZRC-SAZU).

Structure of the questionnaire

According to the task brief the online survey had to focus on needs of the ARIADNEplus user community regarding data sharing/publication, access and (re)use, new services (as developed by the project), and related training needs.

[10] https://www.limesurvey.org

The survey questionnaire comprised of the following modules:

Module: Survey demographics

- Country (location of the organisation or where professionally based)
- Organisation member of the ARIADNEplus project consortium (y/n)
- Type of organisation
- Current position in the organisation
- Main professional activity
- Main research interests (3-5 keywords):
- Gender

Module: Data publication and access

- Methods of publishing research data
- Barriers for depositing data in digital repositories
- Accessibility of different types of data

Module: Data re-use

- Use of data made available by other researchers in the last 2 years
- Kind of data and from which repository or database
- Increase of readiness of archaeologists to share data in the last 5 years (y/n)

Module: ARIADNEplus data services

- Interest in a range of archaeological data types
- Interest in item-level searching across datasets
- Dedicated services for researchers and data managers

Module: Training needs

- Interest in a range of data-related training offers for researchers and data managers

Module: Survey completion

- Final comments & suggestions
- Thanks, contact information & link to the project website

For most of the questions the option "Other – please specify" and free-text boxes for comments were included. Several of the survey questions were matrix table questions presenting different options and allowing to rate them on a 4-point Likert scale. All questions were non-mandatory.

Survey 2013 and/vs. 2019

The topics of the ARIADNE survey in 2013 were organised according to steps in the data-related workflow of researchers, e.g. searching information and data sources, accessing them (if found and accessible), depositing own data and making it available to others. The focus was on difficulties in the this workflow, such as barriers to sharing data through an accessible repository, for instance. In addition, respondents were asked about their expectations towards the (initial) services of the ARIADNE data portal.

The questions of the 2019 survey do not follow a workflow model but still centre on the critical questions of finding and accessing data sources and researchers' own sharing of data. A new topic in the questionnaire is data reuse, which has gained a lot of interest in recent years. A special focus of the 2019 survey is of course the perceived usefulness of the intended new services and tools.

Questions kept for comparison concern data sharing through accessible repositories and barriers to such sharing/publication of data. Results of the 2013 survey are also referenced in the interpretation of some of the outcomes of the new survey.

The greatest difference between the two surveys is that in 2013 ARIADNE began to develop data services for the archaeological research and data management community, while in 2019 the data portal is established and ARIADNEplus is working to provide new or enhanced services. Furthermore, the project will incorporate data from a wider range of archaeological research fields, and aims to integrate more data at a deeper level to provide advanced semantic data search and item-level access.

Different from the ARIADNE 2013 survey no separate survey has been implemented for managers of institutional and other repositories. However, some survey questions (or parts thereof) are also particularly relevant to data managers, including researchers taking care of project databases and managers of repositories.

For the 2019 survey it was decided to avoid a separate questionnaire for repository managers because the ARIADNEplus "sister project" SEADDA - Saving European Archaeology from the Digital Dark Age[11] focuses specifically on archaeological data repositories. SEADDA will carry out studies on the current situation of such repositories and support capacity building of repository managers. SEADDA is a COST Action involving over 100 members representing 31 countries, including archaeological partners of ARIADNEplus.

4.2 Survey dissemination and input

Survey duration

The survey was open for responses from May 25th to September 4th, 2019. The survey ran smoothly throughout, no technical difficulties were perceived by the survey team or reported by respondents.

Dissemination

The dissemination of the survey was carried out by the members of the ARIADNEplus consortium, through contacting members of their national networks and communities as well as some networks that span several countries. The survey was announced on the ARIADNEplus and partners' institutional websites, and invitations distributed through own and community mailing lists as well as social media channels such as twitter. Partners sent out invitations to participate at least twice. Also related projects

[11] SEADDA, https://www.seadda.eu

disseminated the survey, SEADDA and the humanities research infrastructures cluster project PARTHENOS[12], for instance.

Suvey input

The survey received 701 questionnaires of which 484 could be included in the analysis. As usual in online surveys, many started the questionnaire and began answering questions, but then decided not to continue and maybe return later. Also typically, a larger number of respondents went through all survey pages (70%), while others quit during the survey (30%).

The survey sample comprises all questionnaires in which at least a sufficient number of questions have been answered. At the minimum the demographic questions and the questions on one topic (e.g. data publication) had to be answered.

Because respondents quit during the survey the number of responses per question decreases from 480-400 to the first questions to around 330 to those at the end of the survey. The number of responses per question, or items in a table of questions, varies because the respondents were not forced to answer every single question before they could move on to the next page of the survey.

Response rate

The survey has been distributed widely by the ARIADNEplus partners and likely further disseminated by some of the recipients. It was conducted as an open inquiry, anybody who received the URL of the survey webpage was able to participate.

We estimate that the various invitations to participate reached 10,000 researchers in archaeology and cultural heritage generally and perhaps 4,000 more interested in the stated focus of the survey. The focus was described as “community needs regarding data sharing and access, new services and tools, and related training needs”, with reference to the ARIADNEplus digital infrastructure.

If the assumed distribution is roughly correct, the qualified response, which are the 484 questionnaires finally included in the analysis, would be nearly 5% of the wider group reached, or 12% of the assumed more interested recipients.

It is worth noting that some distribution channels were highly targeted, i.e. reached large or small communities of achaeologists, while others reached a large audience but not necessarily many archaeological researchers. For example, project partner MiBACT-ICCU through the e-mail list of the Cultura-Italia portal reaches some 2000 registered users, but only a small part of these are working archaeologists. In contrast, the Institute of Archaeology, ZRC-SAZU (Slovenia) reaches about 250 archaeologists directly through the “Rosa” mailing list of the Society of Slovenian Archaeologists.

Representativeness

The survey input is of course not representative for all archaeologists working across Europe. The Discovering the Archaeologists of Europe (DISCO) project estimated that the number of archaeologists active in various occuptions could be approximately 33,000 (DISCO 2014: 6). Figures for archaeologists that are engaged in research projects and interested in digital research tools and data management allow to put the survey input in perspective. The annual meetings of the European Association of Archaeologists (EAA) can have over 2,000 delegates and 150+ sessions in which attendants present and discuss current research and other activities[13]. The core conference for archaeology and ICT is the annual Computer Applications and Quantitative Methods in Archaeology (CAA) conference that

[12] PARTHENOS, http://www.parthenos-project.eu

[13] European Association of Archaeologists (EAA), https://www.e-a-a.org (conferences).

attracts over 300 participants every year[14]. The respondents of the ARIADNE/ plus surveys are likely to share some characteristics with members of these different groups (see the survey results on research interests in *Section 4.4*), and are perhaps more positive towards some of the survey topics than typical researchers in archaeology and related disciplines.

4.3 Survey demographics

Responses per country

Table 1 presents the list of countries where the respondents are professionally based (e.g. location of their organisation). For 449 (93%) of the respondents this is a European country, while 35 (7%) of the responses are from non-European countries. Among the countries are all of the ARIADNEplus partners, while a few others are also present (* indicates these countries).

***Table 1**: Survey respondents per country (N = 484, all respondents)*

European countries		**European countries (cont.)**	
Austria	10	Norway	9
Belgium	8	Poland*	1
Bosnia and Herzegovina*	3	Portugal	11
Bulgaria	34	Romania	34
Croatia	22	Slovakia*	1
Cyprus	10	Slovenia	23
Czechia	59	Spain	11
Denmark	10	Sweden	7
Estonia*	2	Switzerland*	1
Finland	5	United Kingdom	11
France	77		**449**
Germany	17	**Non-European countries**	
Greece	7	Argentina	11
Hungary	19	United States	8
Iceland	6	Japan	7
Ireland	5	Israel	5
Italy	30	Turkey*	1
Lithuania*	5	Canada*	1
Malta*	3	Australia*	1
Netherlands	7	Philippines*	1
North Macedonia*	1		**35**

Among the surprises regarding the distribution of the responses are the relatively few responses from the Nordic countries, while Bulgaria, Croatia, Czech Republic, Romania and Slovenia stand out among the countries in Eastern and Southeastern Europe.

[14] CAA International, http://caa-international.org/about/history/

The latter may be explained by the particularly strong enthusiasm of leading researchers from these countries about the ARIADNE/plus initiative. This is also shown in the case of Croatia that is a new country in the initiative. In France a larger number of responses are very likely from researchers of the over 40 regional preventive archaeology centres of project partner INRAP.

Regional distribution in Europe

In a regional view the largest group of the European respondents (449) is located in Eastern Europe (33%), followed by Western Europe and Southern Europe (with 27% each). Respondents from Northern Europe are much less present in the survey sample (13%). The United Nations geoscheme for Europe[15], on which the grouping of countries in Table 2 is based, Cyprus does not belong to Europe, but is of course included in the overview.

***Table 2**: Regional distribution of respondents from Europe (N = 484)*

Northern Europe	**60 (13%)**		**Eastern Europe**	**148 (33%)**
Denmark	10		Bulgaria	34
Estonia *	2		Czech Republic	59
Finland	5		Hungary	19
Iceland	6		Poland *	1
Ireland	5		Romania	34
Lithuania*	5		Slovakia *	1
Norway	9		**Southern Europe**	**121 (27%)**
Sweden	7		Bosnia & Herzegovina *	3
United Kingdom	11		Croatia	22
Western Europe	**120 (27%)**		Cyprus	10
Austria	10		Greece	7
Belgium	8		Italy	30
France	77		Malta *	3
Germany	17		North Macedonia *	1
Netherlands	7		Portugal	11
Switzerland *	1		Slovenia	23
			Spain	11

Participation by project members

Among the countries with responses in the survey sample are all of the ARIADNEplus partners in Europe as well as in non-European countries. The respondents were asked if the organisation they work for is a member of the ARIADNEplus consortium: 218 (46%) said “yes”, 258 (54%) “no” (eight respondents did not answer the question).

Participation of ARIADNEplus partners has been encouraged because their researchers and data managers will belong to the core users of the planned new services and data resources. Therefore it was important to learn from the survey what a larger group of them expect from these services and resources.

[15] https://unstats.un.org/unsd/methodology/m49/

Organisational background

***Table 3**: What type of organisation are you working for? (N = 482)*		
A university or public research organisation (such as an academy of sciences, a foundation or similar)	257	53%
A museum	92	19%
A governmental institution	71	15%
A private company or research institute	40	8%
I am not affiliated with an organisation (self-employed, free-lancing or similar)	8	2%
Other	14	3%

The majority of the respondents work at a university or public research organisation. The percentage of these respondent (53%) is about the same as in the 2013 survey (54% of 640 who answered the question). Also the percentages of respondents working for a governmental institution are nearly the same, 15% in the 2019 and 16% in the 2013 survey. 2019 we included the category "museum", as suggested by one of the colleagues who reviewed the survey template. Respondents from museums (19%) turned out as the second largest group in this survey. Other groups are less present in 2019: for a private company or research institute work 8% while 12% in 2013; not affiliated with an organisation are 2% compared to 13% in 2013. Of the 14 respondents who selected "Other" no one used the option to specify the type of organisation or other context of work.

Current position

***Table 4**: What describes best your current position in the organisation? (N = 483)*		
I am a permanent employee	373	77%
I have a post-doc / project-related research contract	45	9%
I am a Ph.D. student	27	6%
Other	38	8%

The majority of respondents (77%) said that they are permanent employees of their organisation. This is a much higher percentage than in the 2013 survey where 54% said so. Also other percentages differ: 9% have temporary work contracts (e.g. post-doc/project-related) while 14% in the 2013 survey, 6% are Ph.D. students while 15% in 2013. Further, 8% chose "Other" in 2019 while 17% did so in 2013.

Compared to the 2013 survey among the respondents in 2019 many more are established, permanently employed archaeological researchers, data managers, and other professional categories, while fewer with fixed-term contract or Ph.D. students. Some of this difference may be due to relatively more respondents from the project consortium (41 organisations in ARIADNEplus while 23 in ARIADNE).

Of the 38 respondents who 2019 said "Other" position 36 provided some information. Most mentioned a position in academic or professional terms (e.g. Adjunct professor, Lecturer in ancient history, Owner and principal, Director, Site manager, Project manager, Associate postdoc researcher,

Research Affiliate, Research assistant). Ten declared that they work on a non-permanent, fixed-term contractual basis, perhaps tied to a particular project (e.g. Non permanent employee, Contract employee, Research Scientist on temporal contract, Fixed-term researcher, Technician with a contract for three years).

Professional activity

***Table 5**: Which of the following describes best your main professional activity? (N = 484, all respondents)*		
Director of an archaeological institute or research centre/laboratory	33	7%
Archaeological researcher / field work	258	53%
Manager of an institutional repository or other service that provides access to archaeological data(sets)	61	13%
Laboratory-based researcher	43	9%
Manager of project databases	32	7%
Other	57	12%

In 2013 the ARIADNE project ran two surveys in parallel; one survey on archaeological researchers and directors (or heads of dept.) of research institutes and laboratories, the other on directors and managers of data repositories and other data access services. For the 2019 survey it was decided to avoid a separate questionnaire for the latter group and also not include specific questions only for data repository managers in the survey. One reason for this decision was that the closely related SEADDA project will carry out surveys on the situation of archaeological data management in Europe and beyond.

The respondents in 2019 were asked if their main professional activity is data management however, as manager of an institutional repository (or another data access service) or manager of project databases. Table 5 also distinguishes between archaeological researchers (field work) and laboratory-based researchers, a distinction that was not made in the 2013 survey.

Due to the different approach in 2019 a comparison to the results of the two parallel surveys in 2013 was not undertaken. However, it is noteworthy that the 2019 survey attracted more managers of a repository or other data access service (61 respondents) than the special survey on this group in 2013 (52). These respondents are the second largest group (13%) after the majority of archaeological researchers with a focus on field work (53%).

The percentage of respondents who chose “Other” (12%) is quite high. But 53 of these 57 respondents specified their main professional activity, inviting a closer look into their descriptions. These represent four groups of about the same size (12-14 respondents). One group is university professors and lecturers, and heads of departments (archaeology dept., laboratory, IT) of research organisations and governmental institutions. Another group is comprised of researchers (archaeology, history, heritage, architecture), scientists (e.g. geology, geoarchaeology), and researchers and technicians active in field or laboratory environments, or both. The third group is respondents with a focus on data acquisition (e.g. digitisation, photography/3D, geomatics), project databases, knowledge management, and digital/computational archaeology. Finally, there is a group of heritage and records managers, museum curators, librarians, and general categories such as project managers and consultants.

Gender distribution

With a share of 45% female respondents are well represented in the ARIADNEplus survey sample. In the ARIADNE 2013 survey the percentage was 43%.

***Table 6**: Gender distribution (N = 480)*		
Male	266	55%
Female	214	45%

These figures correspond to the results of an empirical study conducted by the Discovering the Archaeologists of Europe project in 2007/2008 (Aitchison 2009). The study found that the distribution of female and male archaeologists was 46% and 54%, respectively. These percentages are based on a total of 9,109 archaeologists from twelve European countries.

4.4 Research interests

4.4.1 Keywords and word clouds

The respondents were asked to describe their main research interests briefly with 3-5 keywords. The idea was to produce word clouds of research fields and topics of the survey participants. 452 respondents provided keywords, most 3-4 terms, some only one or two, or brief description of their research activity.

Many respondents described their research interests with a combination of archaeological and method- or technology-related keywords (e.g. *"landscape archaeology, GIS, spatial analysis"*; *"pre-history, artefacts, 3D data, use-wear analysis"*; *"GIS, landscape, middle ages, medieval fortifications"*).

Other large groups only used archaeological terms (e.g. *"Iron Age, metallurgy, settlements"*; *"Roman, funerary, ritual, skeletons"*; *"medieval, christianization, hill forts, rural"*), or only method-/ technology-related terms (e.g. *"data management, spatial field recording (2D/3D), GIS analysis"*; *"digital archaeology, 3D scientific visualisation, knowledge representation"*; *"GIS, 3D, Open Data"*).

In order to produce word clouds, the keywords were extracted, separated and harmonised (e.g. spelling). Some keywords were subsumed under more general terms, e.g. "rescue archeology" under "preventive archaeology". In cases of descriptive text the content had to be summarised in one or more terms. For example, *"manager of project databases"* became "database management" or *"Public and Community Archaeology – ways and possibilities of sharing archaeological heritage with public"* simply "Public Archaeology". Moreover, many terms had to be hyphenated to prevent separation of term words by the word cloud generation tool WordItOut[16], e.g. Bronze-Age instead of Bronze Age, or cases such as Open-Data or spatial-analysis.

The result of the procedure was nearly 1200 occurences of around 350 terms that were used from only once up to over forty times. Not included were the numerous occurences of archaeology and archaeological, except in cases such as digital-archaeology, bio-archaeology or zoo-archaeology. Among the often used terms are periods such as Prehistory (32), Neolithic (25), Iron-Age (21), Bronze-Age (15), Roman (34), Medieval (30) or Middle-Ages (11), in addition also in variants such as Late-Bronze-Age, Roman-Period, Late-Roman, Early-Middle-Ages, Post-Medieval, among others. Other often used terms are settlements (43), pottery (25) and ceramics (13), landscape (33) and, among the technical terms, GIS (35), 3D (24), databases (20), spatial-analysis (18).

[16] WordItOut, https://worditout.com

Most terms were used by one or two respondents, 128 and 74, respectively. Some examples of these among the geographic or regional terms are Alps, Balkans, Crete, Cyprus, Magna-Graecia. Among the terms for research topics and objects, for example, households, fishing, foraging, shells, stone-mortars, weapons, jewellery, sculptures. Some examples among the terms for research fields and methods are petrography, geochemistry, palynology, use-wear-analysis, trace-element, isotopes, bio-markers, ancient-DNA. Examples of data-related terms used by only one or two respondents are legacy-data, digital-recording, Big-Data, data-mining.

The sections that follow present word clouds that provide different lenses on the research interests of the survey respondents. The three facets presented are Cultural periods, locations & relations, Research fields, objects & methods, and Data, digital systems & methods.

4.4.2 Cultural periods, locations & relations

Figure 3 presents a word cloud of 115 terms for cultural periods, locations/regions, and relations and processes derived from the keywords of the survey respondents.

***Figure 3**: Word cloud of 115 terms for cultural periods, locations, relations & processes.*

Cultural periods

Due to the many variants only the most often used cultural period terms (e.g. Prehistory, Neolithic, Bronze-Age, Iron-Age, and others) and some special ones (e.g. Hallstatt-period, migration-period, Viking-Age) are included in the word cloud.

Locations

Respondents mostly mentioned regions (e.g. Latium and Magna-Graecia in Italy, Bohemia, Carpathian-Basin, Peloponnese, North-Africa) and countries (e.g. Bosnia, Bulgaria, Cyprus, Norway), but few used particular place names, which may indicate a much wider geographic research interest.

Under the category of locations, also a number of terms for important places of human activity are included (e.g. city, town, monastaries, castles, hillforts). The most often used term here is settlements, places of human occupation which may have existed only for a short time, or persisted and changed during one or more cultural periods.

It is worth noting that the general term "sites" appeared only five times in the keywords, e.g. where the main term was a technical one as in *"3D digitization of archaeological objects and sites"* or *"geometric documentation of archaeological sites".* Therefore it was not included in the word cloud. Included however are technologies which allow researchers to detect, map and analyse the distribution of places of archaeological interest, e.g. aerial-photography, remote-sensing, GIS and spatial-analysis.

Relations and processes

In addition to "when" and "where" terms, the word cloud includes terms of relations and processes such as long-distant-contact, contact-zones, route, interactions, trade, networks, Romanization, Christianization, colonialism.

4.4.3 Research fields, objects & methods

***Figure 4**: Word cloud of 100 terms for research fields, objects and methods.*

The word cloud of 100 terms for research fields, objects and methods derived from the keywords of the survey respondents illustrates the wide range of research and other activities of the survey participants. The keywords for their research interests include large fields and topics of research as well as specific objects and methods. The word cloud intends to illustrate this variety of interests.

Research domains

The research fields presented in the word cloud comprise domains within the humanities (e.g. epigraphy, numismatics, iconography, art history, religion) as well as natural sciences applied in archaeology, "ArchaeoSciences" as one respondents called these fields of research. Indeed, respondents mentioned many more of these than fields in the humanities, for example, geo-physics, sedimentology, petrography, mineralogy, palynology, geo-chemistry, bio-archaeology, archaeo-botany, zoo-archeology, anthropology, human-osteology, human-ecology, human-evolution.

The more general terms for research and other activities of respondents of course include prospection, field surveys, excavations, preventive-archaeology as well as public-archaeology, heritage-management, preservation, conservation, museology, collections, exhibitions.

Research topics, objects and methods

The research topics and objects include large topics such as environment, agriculture, technology and economy as well as more specific terms of objects and activities of material-culture and subsistence; too many to address in greater detail. Pottery/ceramics, lithics, metals, buildings, burials and funerary practices are of course among the few terms that stand out. Among the terms for methods are the general terms quantitative-methods, archaeometry, materials-characterization, and some specific ones such as use-wear-analysis, trace-element, isotopes, nanoparticles, micro-morphology, bio-markers, ancient-DNA. The term quantitative-methods is also included in the next word cloud together with other more specific terms for computational methods.

4.4.4 Data, digital systems & methods

***Figure 5**: Word cloud of 85 terms for data-related and other digital systems and methods.*

Figure 5 presents a word cloud of 85 terms for data-related practices, digital systems and methods derived from the keywords of the survey respondents.

Digital archaeology & archaeo-informatics

In this data-focused view, the main research interests are expressed in the terms Digital-Humanities, Digital-Archaeology and archaeo-informatics. Digital-Archaeology refers to new research approaches enabled by digital methods, whereas archaeo-informatics supports such digital methods, and is a field of research in its own right.

The main areas where they converge are GIS, 3D technologies, and databases. GIS relates to geo-referenced data and digital research methods such as spatial-analysis and network-analysis. 3D technologies provide novel methods for digital-recording/documentation of sites and objects. Big-Data, data-mining, predictive-modelling, artificial-intelligence, fuzzy-analysis, are among the single mentions of advanced computational approaches.

“Databases” is present as a general term of data organisation and relates to new methods of semantic knowledge-organisation and data interoperability. The latter are organised by mapping data to ontologies (e.g. CIDOC-CRM) and thesauri based on Linked-Data and semantic-web standards. Linked-Data is among the most often used technical research interests.

Also some general forms of content organisation are mentioned, for example, information-systems, CMS (content management systems), multimedia-systems, digital-libraries, digital-editions (e.g. in epigraphy).

Data acquisition, management and sharing

The word cloud contains some terms for specific methods of data acquisition (e.g. photography, laser-scanning, CT-scan, LiDAR, remote-sensing), while the term digitisation typically refers to content collections, archival material, legacy-data and grey-literature that are made available online. The main group of terms in this view of research interests concerns data-management and sharing, e.g. digital-archiving, digital-/data-preservation, digital-/data-curation, data-stewardship, data-publication, data-dissemination, data-sharing. In this context Open-Access and Open-Data were mentioned several times, while DMP (data management plan) and FAIR data were each mentioned only once.

4.5 Data publication

4.5.1 Background

General development

About ten years ago surveys across many disciplines showed that the data practices of researchers run against what advocates of proper data management and open data sharing would advise. The surveys found that after the completion of research projects most data remains locked away, resides on PCs, storage devices, and restricted access servers, out of reach of other researchers, and in danger of loss (e.g. PARSE.Insight 2009; Science 2011).

All studies confirm that data management by individual researchers or research groups does not ensure long-term access. Long-term access requires professional curation of data which goes beyond the core interest and expertise of most researchers, which centres on carrying out research work. Many factors such as changes in research groups (e.g. retirement or move of staff elsewhere), shifts in research focus, and others make proper management of the data of completed projects for future reuse very unlikely. Rather, funding for new projects needs to be secured, new avenues of research explored, fresh results published, etc. Therefore researchers *"will tend to regard data curation as a set of optional activities to complete once the pressure is off... and it never is!"* (Rusbridge 2007).

Data typically loses its value to researchers when the project results have been published, the data becomes obsolete and remains on PCs, carrier media, restricted access servers, and is eventually discarded or lost otherwise. Archaeologists, like researchers in other disciplines, will often keep data in ways that involve a high risk of loss, e.g. if a server crashes, carrier media become unreadable, or data are perceived as obsolete and deleted.

As noted by INCREMENTAL, a project that aimed to improve the data management capacity within the Universities of Cambridge and Glasgow, *"We found that many researchers: (i) organise their data in an ad hoc fashion, posing difficulties with retrieval and re-use; (ii) store their data on all kinds of media without always considering security and back-up; (iii) are positive about data sharing in principle though reluctant in practice; (iv) believe back-up is equivalent to preservation"* (Ward *et al*. 2011).

Over the last few years the situation seems to have improved, arguably mainly due to the expectation of research funders that data from funded projects is being deposited in appropriate repositories for long-term preservation and access. The impact of the increasing pressure from funders is already felt widely in the ecosystem of research. Because compliance requires efforts such as negotiation of open data mandates, implementation of appropriate digital repositories, solving intricate questions of IPR and licensing, and training of researchers (e.g. data management planning).

4.5.2 Data publication

Research data is generally understood to be data collected or generated to analyse and publish original research results. The difference between a research publication (e.g. paper) and the data that underpins the presented research results is generally clear. Not so clear, however, is what it means to publish research data. In the ARIADNE/plus surveys we used the concept of "data publication", mainly to emphasise the common understanding that publication means that the data indeed is publicly available. Researchers often share data directly with colleagues but do not make them publicly available (e.g. in a repository).

Published means public

Among researchers of data-related practices of scholars the meaning of "data publication" has been debated without a clear consensus (e.g. Callaghan 2019; Kratz & Strasser 2014; Lawrence et al. 2011; Parsons & Fox 2013). Kratz & Strasser (2015a) in an empirical study found that researchers' expectations of data publication *"center on availability, generally through an open database or repository"*.

Researchers often make data available to colleagues but not others. For example, in a survey of 1,564 academic researchers of different disciplines 58% said that they shared data with other researchers they know personally, 49% with colleagues at their institute, and 40% with scholars that work on similar topics. Only 13% shared data publicly (Fecher *et al*. 2015).

This means that valuable data remains within small circles of peers and is not available to other researchers or the public at large. Moving more data from closed-circle or not sharing to "open data" requires overcoming strong barriers (see *Section x.x.x* on these barriers).

Publications "count" – if peer-reviewed

The concept of data publication suggests acknowledging datasets which are deposited and publicly accessible in repositories as publications. It relates to the familiar notion of a product that "counts" in the academic system of recognition and reward. This has been emphasised by the Making Data Count project, but also that it requires peer review and citations of the published data (Kratz & Strasser 2014, 2015a/b).

An international study carried out in 2014 surveyed around 4,000 academic researchers on what constitutes trustworthiness of publications in the digital environment (Nicholas et al. 2015). Not surprisingly the study found that peer review was still the most trustworthy characteristic of all. Open access journals were perceived by many as dubious, as these might not have proper peer-review systems. The vast majority of survey respondents did not trust articles in social media

Regarding data, researchers generally agree that peer-reviewed datasets that are made publicly available should count as publications. For example, in a small survey in Australia of the Federated Archaeological Information Management System (FAIMS)[17] project included the question *"Would you agree that peer reviewed publication/sharing of data online should be given research credit or professional acknowledgement as a publication?"*. Of the 79 respondents 86% agreed while 14% did not like this idea (Sobotkova 2013).

It is anticipated that the open data policies of research funders and journals will bring about a wave of data in need of quality review, but who will do this is an open question.

Data review by journal reviewers

Some journals which require data deposit as part of the publication process expect reviewers to look also into the data record and conduct at least a cursory review of the data (e.g. compliance with data standards in the field). Moreover many journals now offer researchers the option to publish a peer-reviewed data paper which describes a publicly available dataset. Here a more detailed evaluation would be appropriate.

Carpenter (2017) conducted an analysis of publishers' policies regarding peer reviews of data and found that these policies do not match well with the expectations expressed by researchers in the Kratz & Strasser survey. For example, a review of methods was included in only 64% of policies in any form. Compliance with metadata standards was included in only 49% of policies. While the expectations of data peer review by researchers are high, Carpenter (2017) notes that the policies of

[17] FAIMS - Federated Archaeological Information Management System (Australia), https://www.fedarch.org

publishers for such reviews *"seem to be focused more on easily assessable qualities than those that match researcher expectations"*.

Data review by repository curators

Data curation experts defined numerous criteria for reviewing datasets when they are provided to a repository of the research community (e.g. Lawrence et al. 2011), but it is unlikely that these elaborated sets of criteria can be applied in many data reviews. Some repositories perform a systematic data review upon deposit (e.g. DRYAD, NASA Planetary Data System, Qualitative Data Repository), but most cannot afford such reviews so data curators mainly receive standard metadata and data formats appropriate for long-term preservation of the data.

Eric Kansa, the Program Director of the archaeological data publication platform Open Context[18], highlights the collaborative dimension of *"data sharing as publication"*. Similar to conventional forms of publications, proper publication of data is a collaborative undertaking between the researchers (authors), who know the data best, and professional data curators (editors), who know what is required to describe and make the data available online for others users (Kansa 2014). Data publication in Open Context involves much support by the data curators as all relevant data elements are published so they can be discovered, accessed and referenced individually (Kansa 2015). Archaeological data repositories such as ADS, tDAR or the archaeological section of DANS can invest less effort, mainly handling standard metadata and data formats.

Kratz & Strasser (2015a) surveyed reseachers regarding their expectations as to what published research data should include. The most frequently observed feature of a published dataset was open availability (68%), availability in a repository (54%), and the indication of links between the data and a paper (e.g. via a DOI). Rich metadata (39%), unique identifiers (39%), and formal metadata (25%) were less frequently cited. Only 28% of respondents felt that peer review was a necessary part of data publication. On the question of peer review, most respondents would welcome a review of the appropriateness of methods (90%), metadata that supports reproducibility (80%) and, ideally, a deep technical review (75%).

4.5.3 Types of repositories

Portals that provide information about repositories and researchers who studied the landscape of available research repositories distinguish between different kinds of repositories. Most widely used is the distinction Institutional versus Disciplinary or Subject-based research repositories. An Institutional repository is a repository of a single institution (university, research centre or other) that typically contains content only of affiliated researchers. A Disciplinary or Subject-based repository serves a whole discipline or a (sub-) domain of research.

The Directory of Open Access Repositories (OpenDOAR) distinguishes between Institutional (or departmental), Disciplinary (cross-institutional, subject based), and Governmental repositories. In addition, the category Aggregating is meant for systems which collect metadata and content from several repositories. Pampel et al. (2013) differentiate between Multidisciplinary, Disciplinary, Institutional, and Project-specific research data repositories. Armbruster & Romary (2010) use the classification Research, Subject-based, Institutional, and National. Here "research" is meant for repositories that in addition to content/data deposit and search also provide functionality for research tasks (e.g. annotation and linking). Adamick & Reznik-Zellen (2010a/b) distinguish in the category of Subject-based repositories between Single-subject, Multi-disciplinary and Inter-disciplinary. Here a Multi-disciplinary repository holds content from many different disciplines while an Inter-disciplinary

[18] Open Context (Alexandria Archive Institute, USA), https://opencontext.org

repository from different disciplines but related to common research topics and aimed to support integrative research.

Thus different typologies are being used to distinguish the existing variety of repositories. What complicates matters further is that the categories can contain repositories with a very different organisational setup. For example, Multi-disciplinary includes repositories such as Figshare[19] or ZENODO[20] as well as university repositories for scholars of all faculties. In the category Institutional most are multi-disciplinary repositories of universities. However, if the term Institutional is used for a research centre the repository would contain papers, reports and data only for research subjects present at the centre.

Categorisation of repositories

Figure 6 presents a categorisation of different types of repositories as background to the ARIADNEplus survey questions on the use of repositories and interpretation of the results:

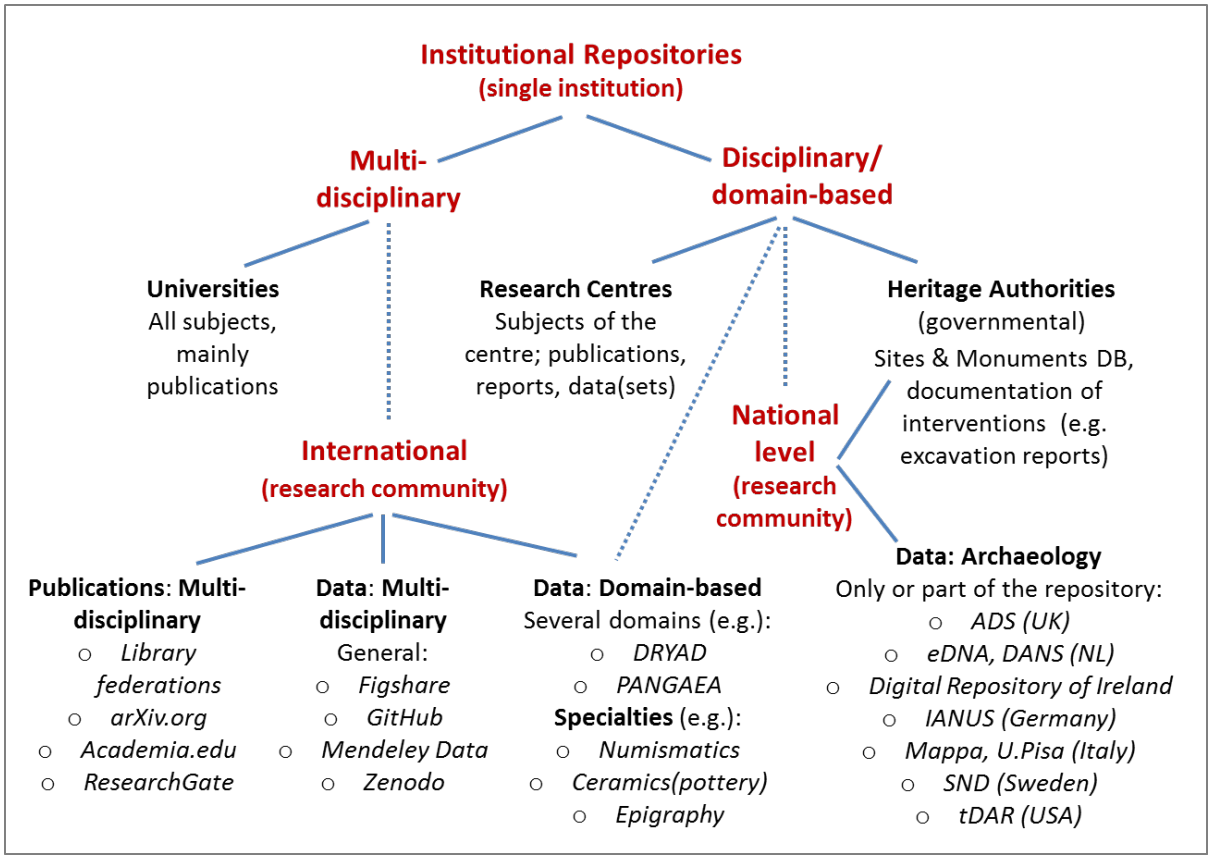

***Figure 6**: Categorisation of types of repositories (source: ARIADNEplus / Salzburg Research, 2019).*

Explanation of the categorisation:

- Institutional Repositories" vs. "National" and "International" repositories: The categorisation first distinguishes between "Institutional Repositories" (those of single institutions such as an university) and "National" and "International" repositories that are being used by researchers of a country or internationally.

[19] Figshare, http://figshare.com

[20] Zenodo, https://zenodo.org

- "Multidisciplinary" vs. "Disciplinary/domain-based": This distinction allows grouping of these further into those which hold content/data from many disciplines or mainly from one discipline or even a sub-domain of research.
- "Data repositories"

  Can be distinguished from repositories which only or mainly hold documents. Typically university repositories contain various documents (papers, theses, presentations, reading lists, etc.). *arXiv.org* provides access only to preprints of papers and other documents, while platforms such as *ResearchGate* to various content. These do not qualify as proper repositories although they have repository functionality and have been online for many years.

  "Data repositories" not only contain data but also various documents. Here the distinction between "Multidisciplinary" and "Disciplinary/domain-based" makes the difference:

  - Multidisciplinary repositories typically do not control what researchers deposit and require minimal metadata, with the effect the many user "dump" all sorts of content in them.
  - Disciplinary/domain-based repositories accept only relevant data (some also related documents) and set high standards regarding metadata. Such repositories can be for
    - data of several disciplines, e.g. DRYAD (bio-sciences) or PANGAEA (earth and environmental sciences),
    - data of specialties, e.g. ceramics (pottery) or numismatics studies, are typically databases of items held by physical repositories/collections.
  - Archaeological data repositories (under disciplinary/domain-based):
    - are typically at the national level, not in the sense of governmental status but because they are mainly used by archaeologists of a single country, e.g. the e-Depot for Dutch Archaeology (eDNA) in the archiving system of Data Archiving and Networked Services – DANS;
    - can be part of a repository that also holds data/content of other domains, typically but not only other humanities and social sciences (e.g. DANS, Digital Repository of Ireland, Swedish National Data Service);
    - or only accept archaeological data/content, e.g. Archaeology Data Service (UK), Mappa repository (University of Pisa, Italy), IANUS (Germany, currently not active).
  - Archaeological data repositories and/vs. repositories of governmental Heritage Authorities:
    - Repositories of Heritage Authorities typically hold documentation of archaeological interventions (e.g. excavation and other fieldwork reports) while the data of the fieldwork (e.g. survey, excavation) is stored at the archaeological research centre or deposited in national-level repository.

Importance of mandated data repositories

Archaeological data repositories or data collections in other repositories can be mandated (or not) by research funding agencies to receive deposits from funded research projects or interventions (e.g. preventive archaeological fieldwork).

From 2007, archaeologists in the Netherlands are formally obligated to deposit their data with the Data Archiving and Networked Services - DANS, according to the Quality Standard for Dutch Archaeology (Kwaliteitsnorm Archeologie). The DANS-EASY system includes the e-Depot Dutch

Archaeology (eDNA)[21]. Archaeological data deposited in eDNA is the largest part of DANS-EASY and over 80% of it is publicly accessible. In the UK, the Archaeology Data Service (ADS)[22] is the repository mandated for archaeological data by the Arts and Humanities Research Council and the Natural Environment Research Council; the repository is also recommended by the British Academy, Council for British Archaeology, English Heritage, and the Society of Antiquaries.

In other European countries archaeological repositories are not mandated, including tDAR in the United States. Archaeologists are not obligated to deposit data in DRI (Ireland), IANUS (Germany), Mappa (Italy) or SND (Sweden). Data has been deposited because these repositories exist and archaeologists felt it was a good thing to do. For example, a collaboration between the ARIADNE partner Discovery Programme, Transport Infrastructure Ireland (TII) and the Digital Repository of Ireland (DRI) enabled a first large collection of archaeological documentation in the DRI, more than 1,500 excavation reports commissioned by TII during Ireland's infrastructure building programme between 2001 and 2016, over 176 geophysical survey reports, and other content (TII 2017).

In countries where no mandated national-level archaeological data repository exists, the situation is very likely that revealed by a survey 2013 in Germany. The online survey was conducted by the IANUS initiative for a Research Data Centre for Archaeology and Ancient Studies[23] (IANUS 2014). The IANUS initiative was coordinated by the German Archaeological Institute and aimed to establish a national-level data centre and repository.

The survey was intended to support the evaluation of existing demands and expectations towards the services of the data centre/repository. The 243 respondents were from the field of "Altertumswissen-schaften" (Prehistory, Archaeology of the Ancient World, Classical Archaeology and other specific fields) and held positions at universities (45%), other research institutions (32%), museums (8%), cultural heritage departments (7%), and various others (8%).

Of the 243 respondents 43 (18%) said that they deposit data in a professional archive (11) or provide it to a data center (32). But only 21 (9%) allowed online download of their data from a portal. 19 provided only metadata to an openly accessible portal, and 29 to a restricted access portal. Most respondents kept data of completed projects at arm's length on carrier media (136), their computer (98) and/or local network/central server (82). The majority (161) declared that they would make data available to third parties only on individual request.

In the ARIADNEplus 2019 survey 16 of 17 respondents from Germany answered the question on data publication. The results for sharing data through an accessible repository were somewhat less encouraging than in the survey sample overall, especially regarding a national-level repository. While 34% of all respondents said they share data "not at all" through an accessible national data repository, 9 of the 16 German respondents, 60%, said this. In 2019 the "not at all" includes the IANUS data repository. The IANUS initiative was funded 2011-2017 by the German Research Foundation. After their data repository was implemented and data curators trained, in 2016/17 seven deposited data-collections were made available on the IANUS data portal[24]. Thereafter IANUS stopped accepting data deposits.

[21] Data Archiving and Networked Services (DANS): e-Depot Dutch Archaeology (eDNA), http://www.edna.nl

[22] Archaeology Data Service (ADS), http://archaeologydataservice.ac.uk

[23] IANUS - Research Data Centre for Archaeology and Ancient Studies (Germany), http://www.ianus-fdz.de

[24] IANUS: Data portal, http://datenportal.ianus-fdz.de

4.5.4 Survey results 2019

Question on data publication

The ARIADNE/plus surveys 2013 and 2019 asked participants in what way their research group/ organisation typically publishes data, i.e. make it available to others. The options given in the questionnaire were: selected data in a paper published in a journal or conference proceeding, in a research report, or as supplemental material to a paper or report. Typically data published in these ways is summary tables and/or illustrative images in papers, more detailed tabular data and/or visual documentation in reports or supplemental material.

A further three options were that the data is made available through an accessible institutional repository (managed by the organisation of the respondent), a national repository, or a subject-/domain-based repository. Implied in the repository option is that the published data is not contained as a summary table or selective extract in one of research products mentioned before, which are typically provided as PDF documents. Thus the data is be published separately and in a format appropriate for numeric, tabular, visual or other data. Furthermore, as published data it can be accessed and retrieved. Not stated or implied is that the data must comply with requirements defined for Open Data or FAIR data (see *Section 3.3* on these requirements).

While the ARIADNE initiative is mainly interested in data shared through accessible repositories, also the conventional ways of presenting data were included in the survey question. Respondents in 2019 commented that the notion of "data" in the survey is too vague (*"le mot Data est trop vague"*) or *"too broadly defined: a plot in a research report is definitely not 'data'"*. But also other surveys on data sharing include the conventional forms, and these are still the ways most respondents declare to publish data.

Survey results

The survey participants were asked how their research data is typically being published, i.e. made available to others beyond the research group and organisation. Respondents were given a number of options and asked whether this way of publishing research results is used in all/most projects, many, a few or none at all.

To what extent and in what way is data which your research group / organisation is producing typically being published (i.e. made available to a certain community beyond your own institute)? Please say which of the following methods apply for (1) all or most projects, (2) many projects, (3) a few projects, or (4) not at all:

	N	All or most projects	Many projects	A few projects	Not at all
Selected research data is published in journal papers and conference proceedings	483	**190** (35%)	**157** (33%)	**112** (23%)	**24** (5%)
Selected research data is published in research reports (for instance in tables and figures)	464	**168** (36%)	**168** (36%)	**97** (21%)	**31** (7%)
Data is published in supplemental material (such as annexes with large, detailed tables, or laboratory images)	449	**60** (13%)	**114** (25%)	**188** (42%)	**87** (19%)
Data is made available through an accessible institutional repository managed by our organisation	457	**98** (21%)	**96** (21%)	**128** (28%)	**135** (30%)
Data is made available through an accessible national repository	451	**71** (16%)	**93** (20%)	**135** (30%)	**152** (34%)
Data is made available through an accessible subject-/domain-based repository	437	**55** (13%)	**66** (15%)	**144** (33%)	**172** (39%)

***Table 7**: Publication of project research data (N = 437-483).*

Expectedly, the standard 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. Around 70% of the respondents said that they published selected data this way in many or all research projects they were involved in. Respondents also made available data in supplemental material, i.e. a document with large data tables or visual documentation of laboratory results, in many (25%) or at least a few projects (42%).

When it comes to making project data available through an accessible repository about one third of respondents said that data was shared "not at all" through such a repository; 30% did not deposit data in an institutional repository, 34% not in a national repository, and 39% not in a subject-/domain-based repository (39%). The percentages of publication of data from at least a few projects through such repositories were 28%, 30%, 33%, respectively.

"Other" methods and Comments

Respondents also had the options to state "Other" methods they typically use for publishing research data as well as give "Comments".

Under the 39 entries under "Other" there were many (18) no other, none, nothing, etc. Some entries confirm a broad understanding of data found by other surveys. For example, respondents mentioned *"Data is made available in exhibition catalogue"*, *"Research bulletin published by our organization"*,

"Monographs", *"Through publications of other institutions - museums and others"*. Also there were *"green open access"*, i.e. depositing a final manuscript or pre-print in a repository, and *"academia.org"*, a platform for posting publications.

The only data repository mentioned was *Zenodo*. One respondent wrote, *"Selected and classified results are made available through a subject-based webGIS application"*. Others mentioned *"project websites"* and *"private hosting platforms"*. Also there were data

- *"made available upon request (…) after the original report is made public"*,
- *"data kept in personal files (eg. excel spreadsheets, photos, articles etc.)"*,
- *"Kept on own PC"*,
- *"With very limited financial resources available, it is only possible to me to keep data in personal files, conduct pilot studies and publish selected results"*.

In such cases data is not available to others as well as there is a high risk of data loss.

Among the Comments (28) six respondents mentioned that no respository is available or that their organisation is developing a repository or database intended to be publicly accessible. For example one respondent mentioned, *"There is no repository in our organisation, just a server. Data are accessible but for the moment only on demand. If a repository would exist it would be managed by our organisation and subject/domain-based, not national (for political reasons)".*

Two respondents worried that due to lack of staff and technical means not all collected archaeological remains could be processed and the data made available.

Comments of other respondents make clear that different research data may be made available in different ways (or not), for example,

- *"Our lab is both commercial and research based, with different dissemination channels – all research project data in journals, all commercial work data in reports. (…) Which data are made available through domain repositories depends on the type of research material – e.g. insect results are deposited in a database, soil chemistry are not."*
- *"We publish selected data but make everything available in our database. Other departments (…) DO NOT make data available in the database. They publish the main results and selected data, the rest cannot be accessed."*
- *"Artefact catalogues are published through a national repository, excavation reports are published through an institutional repository."*
- *"I advocate print publication of synthetic reports on excavations, with full data online."*

4.5.5 Comparison 2013 / 2019

Table 8 compares results for the question on data publication of the surveys 2013 and 2019, specifically the results for making data available through an accessible repository. In 2013 around 520 and in 2019 around 450 respondents answered this part of the question.

	All or most projects		Many projects		A few projects		Not at all	
Type of data repository	**2013**	**2019**	**2013**	**2019**	**2013**	**2019**	**2013**	**2019**
Institutional repository	13%	21%	19%	21%	27%	28%	40%	30%
National repository	9%	16%	15%	20%	28%	30%	48%	34%
Subject-/domain-based	6%	13%	10%	15%	24%	33%	59%	39%

***Table 8**: Comparison of the results for data published through accessible repositories. Respondents 2019: N=457,451,437; respondents 2013: N=521,521,516 (ARIADNE 2014: 102, Figure 6.2.20).*

The results show a significant difference in the percentages of respondents who made data from projects "not at all" available through an accessible repository.

In the 2019 survey, respondents who answered the questions, 10% less said that they publish data "not at all" through an institutional repository, 14% less not through a national repository, and 20% less not through a subject-/domain-based repository. The percentages for sharing more or less data from projects through the different types of repositories are all higher than in 2013. Surprisingly, in the 2019 survey sample sharing data from "all or most projects" through an accessible repository was 7-8 % higher than 2013. Significantly more respondents also claimed that they share data from many projects through a national data repository (+ 5%).

The larger differences regarding a subject-/domain-based repository could in part be due to a change in the survey question: 2013 survey repondents were asked if they publish data in an "international" data repository of this type whereas 2019 international was omitted. In the 2013 survey it was used because there are few national-level repositories *only* for digital archaeological data, and hardly any such specialised repositories of universities.

This leaves among the institutional data repositories those of archaeological research centres (and perhaps some of heritage authorities) as the most likely candidates that hold archaeological subject-/domain-related data only. In addition, the category subject-/domain-based of course still includes contributing data to international repositories. This can be digital repositories or databases of specialities such as ceramics, numismatics or epigraphy, or of large fields such as bio-sciences (e.g. ancient DNA data in GenBank)[25] or earth & environmental sciences (e.g. pollen datasets, isotopes and other data in PANGAEA)[26].

Looking into the distribution of responses per country an interesting result regarding national repositories surfaces. In the 2013 survey many respondents were from the United Kingdom and the Netherlands where there are (mandated) national-level data repositories for depositing archaeological data (see *Section 4.5.3*). In the 2013 survey respondents from the UK (79) and the Netherlands (35) together were 24% of the 482 participants who indicated where they are professionally based. In the 2019 survey respondents from these countries (18) were only 4% of all 484 participants, all with known

[25] GenBank, http://www.ncbi.nlm.nih.gov/genbank/

[26] PANGAEA - Data Publisher for Earth & Environmental Science, http://www.pangaea.de

country of professional activity). But in the 2019 survey considerably fewer respondents said that they make data "not at all" available through a national data repository, 34% compared to 48% in 2013.

One explanation for this result could be that the 2019 survey sample includes relatively more archaeologists who are obliged to provide documentation of fieldwork (interventions) to a national heritage authority. While there were fewer researchers who are obliged to deposit project data in a dedicated national data repository for archaeology, more respondents may have considered providing documentation to a heritage authority as making it available through a national-level repository or database (for example, the database of the National Archaeological Information System - Archaeological Map of Bulgaria, AIS-AKB, of the Bulgarian Ministry of Culture).

4.5.6 Increase in data publication?

In the comparison of the 2019 and 2013 ARIADNE/plus survey results we found a significant difference regarding publication of data through accessible repositories of the three types distinguished in the surveys. In the 2019 survey between 60% and 70% of the respondents (N = 437-457) said that they make available data from all/most, many or at least a few projects in this way, while 2013 from below 50% up to maximal 60% said that they do so (N = 516-520).

This suggests that from 2013 to 2019 in the ARIADNE/plus communities of respondents, the sharing of data through accessible repositories increased significanty by 10% or even more. Before we compare this result to other recent surveys on the data sharing of researchers, considerable differences between the samples of respondents in 2013 and 2019 should be noted:

- *Presumably more responses from ARIADNE partners:* The 2019 sample very likely contains a larger share of responses from ARIADNE partners. These are 46% of the 484 responses. 2013 the respondents were not asked if their organisation is a member of the ARIADNE consortium, but the consortium at that time had 23 partners, while ARIADNEplus has 41 partners.
- *More permanently employed staff:* Another remarkable difference, possibly due to the stronger presence of responses from project partners is: 77% of the respondents 2019 said they are permanent employees, compared to 54% of the respondents in 2013. Thus the 2019 responses contain a larger share of more established, permanently employed archaeological researchers, data managers, and other professional categories, while significantly less have a fixed-term contract or are Ph.D. students.
- *Fewer responses from countries with a mandated data repository:* Regarding obligation to deposit data in a national-level data repository, in the 2019 sample of respondents there were fewer from the United Kindom and the Netherlands where publicly funded archaeologists are obliged to do so. In 2013, responses from these countries were 24% [114] of 482 participants who stated where they are based professionally, while 2019 only 4% [18] of the 484 respondents (all with known location). In 2019 far fewer respondents said they make data "not at all" available through a national repository (34% compared to 48% in 2013). Some of the difference may be explained by more respondents in 2019 who are obliged to provide documentation of fieldwork to a repository of a national heritage authority, understood as a national repository.
- *More responses from Eastern and Southeastern Europe:* The 2019 sample contains considerably more responses from these regions than the 2013 sample. Counting only countries with a larger number of responses, in 2019 Bulgaria (34), Croatia (22), Czech Republic (59), Romania (34) and Slovenia (23) together make up 35.5% of all responses (172 out of 484). 2013 the 53 responses from these countries were 11% of the 482 respondents for whom the location is known. Without Slovenia, with 38 reponses in 2019, 23 in 2013, the percentages would be 31% and 3%, respectively.

It is worth noting here also that 2013 responses from the UK were the largest "national" group in the sample with 79 responses (2019: 11), while 2019 respondents from France contributed most with 77 responses (2013: 53).

Among these points in 2019 very likely less responses from the Netherlands and the UK reduced while more from other ARIADNEplus partners increased the percentage of data publication through different types of accessible repositories, leading to at least 10% more such publication than 2013.

4.5.7 Comparison to other surveys

The results of our surveys suggest that between 2013 and 2019 in the ARIADNE/plus communities of respondents the sharing of data through different types of accessible repositories increased significanty by 10% or more. 2013 from below 50% to maximal 60% of respondents said that they publish data from all/most, many or at least a few projects in this way, while 2019 between 60% and 70%. How do these figures compare to other surveys?

Other surveys used for comparison

Except of the ARIADNE/plus surveys no other seems to have surveyed a larger number of archaeological researchers on questions related to data sharing and reuse. The surveys we are aware of have too few respondents to be considered for comparison (e.g. Austin & Mitcham 2007 [48 respondents]; Sands 2009 [37], Sobotkova 2013 [79]), but we add their results in *Section 4.5.9*).

From other available surveys we chose earlier and recent ones on researchers from many disciplines for comparison to the ARIADNE/plus results, with a particular focus on data sharing through repositories:

- PARSE.Insight 2009 [n=1387]: physical sciences 33%, social sciences 17%, technology 14%, life sciences 13%, humanities 7%, medicine 6%, behavioural sciences 5%, agriculture & nutrition 5% (no other disciplines). Countries: EU 44%, USA 33%, Canada 5%, Australia 4%, Japan 3%, other 11%.
- *Science* journal 2011 [n=ca.1700]: an international sample of peer-reviewers, multi-disciplinary but many from laboratory-based scientific research (survey results used for minor comparison).
- Tenopir et al. 2011 [n=1329] and 2015 [n=1015], in the combined dataset of the surveys: ecology 17.4%, environmental science 14.4%, biology 10.9%, social sciences 6.5%, (…), humanities 0.5%. Regions: North America 68.1%, Europe 14.6%, Asia 7.4%, South America 3.8%, Africa 1.8%, 1.8%.
- Fecher et al. 2015 [n=1,564]: natural sciences 33%, social sciences 31%, human sciences 12%, humanities 11%, engineering 8%, agriculture 5%, and others. Countries: Germany 88%, other 12%.
- Figshare 2018 survey [n=max.1800]: biology 19.3%, medicine 14.2%, social sciences 14.1%, earth & environmental sciences 10%, engineering 6.8%, chemistry 4.2%, humanities 3.7%, and others (percentages for the 1150 who stated their discipline at the end of the survey). Countries, e.g., USA 18.72%, India 7.59%, UK 6.58%, Germany 4.97%, Italy 4.13%, Spain 3.88%, France 1.77%, Canada 3.37%, Brazil 3.63%, Australia 3.12%, China 2.36%, Japan, 1.6%, Russia 0.93%.
- Schmidt et al. (Belmont Forum Open Data Survey) 2016 [n=1232]: earth & environmental sciences 68.7%, climate & atmospheric sciences 31.3%, biological sciences (20.9%), physical sciences 13.1%, (…), social sciences 5.4%, humanities none. Countries, e.g., Germany 16.4%, United States 14.7%, Italy 9.4%, UK 7.1%, France 5.4%, Australia 3.6%, Spain 3.4%, China 3.1%, Canada 2.6%, Japan 2.1%.

In most of these surveys the percentages of respondents from the social sciences are relatively high while from the humanities lower or none. This allows for good comparison of the ARIADNE/plus survey results to those of others with respondents from many disciplines. But archaeology is a multi-disciplinary field of research in which researchers have a background in different disciplines, natural

sciences, environmental & geosciences, social sciences, different domains within the humanities, for instance. They produce a wide range of data in surveys, excavations, laboratory analyses of physical and biological finds, etc. Furthermore, the array of data includes data from specialised laboratories which serve archaeologists among other clients (e.g. synchrotron facilities or sequencing labs with regard to ancient DNA), or airborne or satellite remote sensing and imaging data.

PARSE.Insight (2009) and Science (2011)

In the PARSE.Insight survey of 1,200 respondents from different disciplines and countries, 20% stated that they store data in a digital archive, of which 14% in an archive of the research organisation and 6% of the discipline. Most stored their their data on computer at work (81%), a computer at home (51%), a portable storage carrier (66%), or a server of the organisation (59%). 15% said that they submitted such data (material) to a journal. Further information, for example how openly available their data generally was, are given in the survey report (PARSE.Insight 2009: 32-34).

In 2011 the scientific journal *Science* polled their peer reviewers about the availability and use of data (Science 2011). They received about 1,700 responses from an international and multi-disciplinary sample 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.

Thus in these surveys few respondents said that they store data in a repository of the research community, 6% and 7.6%, respectively. In both it remained unclear if any of the data archived internally or managed on a server was accessible to users beyond the research group. On servers usually it is not. As a case in point, in the PARSE.Insight 59% of respondents said that they stored data on the server of the organisation and 58% that their data is available to their research group.

Tenopir et al., 2011 and 2015

Tenopir et al. (2011) and (2015) reported results of surveys on data sharing and reuse of researchers from different disciplines. There were 1,329 respondents in the first survey (Oct. 2009 to July 2010), and 1,015 respondents in the second survey (Oct. 2013 to March 2014). Most responses in both surveys came from North America (73% in 2011, 61% in 2015) and European countries (about 15%), while less from other world regions and countries. Researchers from all disciplines participated, with larger shares of the respondents in both surveys of Ecology (17.4%), Environmental Science (14.4%), Biology (10.9%), Engineering (6%).

In the first survey (responses 2009/10) about 400-450 (30+%) of the 1,329 respondents chose not to answer the question on data sharing. Of those who did (850-900), 46% said that they do not share their data with others. Among the remaining around 45% reported that they make at least some of their data available on a website (organisation, principal investigator, or own) or through a national, regional or global network. About 10% did not clarify how they share at least some of their data. As the study authors note, "*The high percentage of non-respondents to this question most likely indicates that data sharing is even lower than the numbers indicate"* (Tenopir et al. 2011: 9).

The follow-up survey of Tenopir et al. (2015, responses 2013/14) provides more detailed information on where the respondents put their data. Indeed, these are the most detailed and comparable figures of the surveys selected for comparison. The appendix to their survey provides figures for data stored in repositories, on internal or external servers (e.g. institution/department, principal investigator, Dropbox, Google, etc.) as well as personal means (e.g. PC, carrier media, on paper in the office). We are primarily interested in the use of accessible repositories for making data available to others beyond the researchers' institution. Not considered is placing data on a server so that project collaborators or others of the institution/ department can access it.

Table 9 compares the results of ARIADNE/plus (2013, 2019) and those of Tenopir et al. (2015: S1 Appendix, Table K) for relevant categories. In this comparison the ARIADNEplus percentages for publication of data through such repositories from "All/most" and "Many" projects have been combined, of the Tenopir et al. the percentages for "All" and "Most" data stored in repositories.

Tenopir et al. 2015: Where data is stored	***All or most***		***Some***		***None***	
My institution's repository	11.3%		21.5%		67.2%	
Discipline-based repository	9.5%		18.0%		72.5%	
Other data repository or archive	9.3%		22.6%		68.1%	
Average	*10%*		*20%*		*70%*	
ARIADNE/plus: Project data made available through an accessible…	***All or most/ many***		***A few***		***Not at all***	
	2013	2019	2013	2019	2013	2019
Institutional repository (own org.)	32%	42%	27%	28%	40%	30%
Subject-/domain-based repository	16%	18%	24%	33%	59%	39%
National repository	24%	36%	28%	30%	48%	34%
Average	*24%*	*32%*	*27%*	*30.3*	*49%*	*34.3*

Table 9*: Comparison of the use of repositories. Tenopir et al. 2015, N = max. 1000; ARIADNE 2013, N = 516-521; ARIADNEplus 2019, N = 437-457.*

Tenopir et al. 2011 versus 2015

The differences between the first Tenopir et al. survey (responses 2009/10) and the second survey (responses 2013/14) indicate a remarkable increase in stated data sharing in the quite similar populations of their survey samples. In the first survey over 30% did not answer the question on data sharing. Of those who did, 46% did not share data with others, while around 45% made at at least some of their data available somehow, on a website or through a network; about 10% remained unclear.

Due to a change of the sharing channels the results of the first and second survey are not directly comparable. However, asked in the second survey about the amount of data made available to others somehow around 9.5% said none, 43% some, 47.5% most or all (Tenopir et al. 2015: S1 Appendix, Table I). In this survey still 70% did not store any data in a digital repository, while 30% used one, 20% to store some and 10% most or all of their data. The 30% of respondents around 2013/14 who used a digital repository is 10% higher than the percentage reported by the PARSE.Insight survey 2009 (see above).

ARIADNE/plus (2013, 2019) versus Tenopir et al. 2015

No data made available through a repository: In the ARIADNE 2013 survey around 50% of respondents said that they publish data from projects "not at all" through an accessible repository, while 2019 around 35% said so. In the second Tenopir et al. survey (responses 2013/14) around 70% stored "none" of their data in a repository. Thus in the ARIADNE/plus surveys considerably fewer respondents did not share project data through a repository, about 20% in 2013, 35% in 2019.

From some to all data: In the ARIADNE 2013 survey around 50% of respondents shared from some to all of their project data through an accessible repository, while in 2019 it was around 65%. In the Tenopir et al. survey (responses 2013/14) around 30% said that they store from some to all of their data in a repository, presumably most accessible to others beyond the research organisation. Thus in the ARIADNE/plus surveys more respondents shared at least some project data through an accessible repository, 20% and 30%, respectively.

Figshare surveys, 2016-18

Since 2016 Figshare[27] has investigated the sharing of research data in their international "The State of Open Data" surveys (Figshare 2016, 2017, 2018). The Figshare repository is part of the portfolio of digital services and tools of Digital Science[28] that belong to the global media company Holtzbrinck. The Figshare surveys have a larger number of respondents, in 2018 about 1800 (2017: 2300, 2016: 2000). The respondents are mainly from universities and other research institutions, in 2018 77%, and from different disciplines, e.g. biology 19.9%, medicine 13.6%, social sciences 14.5%, earth & environmental sciences 10.3%, less from others such as chemistry (4.2%) or arts & humanities (3.8%).

The Figshare surveys found a year-on-year increase in researchers' willingness to make their data openly available. In the 2018 survey 64% of respondents said that they made data openly available to others (e.g. "posting in an open space"), up 4% on 2017 (60%) and 7% on 2016 (57%). However, this also means that 36% in 2018 and 43% in 2016 did not do so. In 2018 23.7% said that they never make their research data publicly available (2017: 21.4%). Among the respondents 2018 who made data available 24.2% said frequently, 28% sometimes, and 19% rarely (5.1% skipped the question).

Regarding different ways of sharing data, Figshare reported results only for 2017 and 2018. In 2018 more respondents said they made data available in a specific data repository, 33% in 2018 compared to 29% in 2017 (+4%). The percentage of those who provided supplementary material to a research article remained about the same, 34% in 2017, 35% in 2018. Slightly fewer said they made data available related to a data paper, 18% in 2018, 20% in 2017. As one expert noted, compared to the research community in general these figures are high (Baynes 2018: 17).

Datasets described in peer-reviewed data papers must be accessible online, i.e. in an online database or available in an accessible repository (Chavan & Penev 2011; Candela et al. 2015), supplementary material generally goes into the repository of the publisher or an archiving service and must be retrievable. However, as multiple responses were possible the percentages cannot be summated.

Regarding data papers it is worth noting that some publishers introduced them to offer researchers the opportunity for an additional paper instead of just providing supplementary material. In the first wave of data papers, this may have had the effect of turning the data description in supplementary material into a publishable data paper. For example, Elsevier's data journal *Data in Brief* [29], launched in 2014 for submissions from all research areas. Among the advantages of this approach, they state

[27] Figshare, https://figshare.com

[28] Digital Science, https://www.digital-science.com

[29] https://www.journals.elsevier.com/data-in-brief/

"Make your data, which is often buried in supplementary material, easier to find" and *"Thoroughly describe your data, facilitating reproducibility"*.

These goals are not easy to achieve. Kervin et al. (2013) provide an analysis of many errors in data papers identified in the quality review of a core ecological journal, Li et al. (2019) of shortcomings of papers describing biodiversity datasets in different data journals. These journals are more specialised than *Data in Brief* for such analyses could not be found.

ARIADNE/plus (2013, 2019) versus Figshare 2018

The Figshare surveys had a relatively high percentage of respondents who said that they made data openly available, 64% in 2018, up 7% on 2016 (57%), but how they did so is not specified.

In the 2019 ARIADNEplus survey, between 60-70% of the respondents said that they published data from all/most, many or at least a few projects *through a repository*, while in the 2013 ARIADNE survey between 50-60%. Thus in 2013 the percentage of repository-based data sharing was roughly the same and in 2019 higher than Figshare's figures for data sharing in general.

The Figshare figures for data made available through a specific data repository are 33% in 2018 and 29% in 2017. Compared to the ARIADNE/plus figures of repository-based data sharing these percentages are rather low. In the ARIADNE 2019 survey over 30% more said they made data available through an accessible repository.

The fact that Figshare respondents also made supplementary material, and perhaps some data related to data papers, available in a repository, does not make much of a difference here. As also many ARIADNEplus respondents make supplementary material available; in the 2019 survey 81%, of which 13% in all or most, 25% in many, and 42% in at least in a few projects.

4.5.8 Supplementary material

One major difference between results of the ARIADNE/plus and other surveys concerns supplementary material. The Tenopir et al. survey 2013/14 included the category *"Publisher or publisher-related repository"*, which typically means that supplementary material is made available for the research results reported in a publication. Of their respondents, 2.4% said that they stored all or most, 16.9% some, and 80.6% no such data (material) in a publisher/publisher-related repository. In the PARSE.Insight survey some years earlier 15% said that they submitted data (material) to a journal publisher (PARSE.Insight 2009: 32).

The ARIADNE/plus surveys include the category *"Data is published in supplemental material"*, mainly to distinguish this form of data publication from data summaries, charts and other overviews presented in published papers or research reports. When archaeologists make available supplementary material it is often provided to a heritage authority to document in greater detail results of fieldwork that has been carried out. Indeed, in the ARIADNE/plus surveys 2013 and 2019 the figures for respondents who did *not* make any supplementary material available are 18% and 19%, respectively, while in the Tenopir et al. survey it was 80.6%. In contrast, of the ARIADNE/plus respondents nearly the same percentage said that they make supplementary material available from all/most or at least a few projects (2013: 82%, 2019: 81%).

In the Figshare surveys 2017 and 2018 the percentages of respondents who provided supplementary material to a research article were 34% and 35%, respectively. A survey of *Springer Nature* in 2018 (around 7,700 respondents from 126 countries) specifically investigated the data sharing behaviour of researchers related to the publication of journal papers (Springer Nature 2018). When asked what they do with their data when submitting a journal paper, 22% said they submit supplementary material, 21% deposit the data in a repository, and 20% do both (37% neither). Thus, in this survey, 42% said

that they submit supplementary material to the publisher (41% deposit data in a repository). Also compared to these surveys the percentages of ARIADNE/plus respondents who made data available as supplementary material are much higher.

4.5.9 Results of small surveys on data sharing in archaeology

The ARIADNE/plus surveys seem to be the only larger surveys on data sharing and reuse in archaeology. Other surveys on data-related practices which included questions on this had too few respondents to be considered for comparison. However, we briefly summarise their results:

A questionnaire survey 2009 in Ireland included the question: *"Do you routinely use data created by others?"* (Sands 2009: 54-59). The question was answered by 37 respondents of which 30 said "yes". 29 respondents received data directly from colleagues (e-mailed, on carrier media, etc.), 25 extracted it from written reports, 17 downloaded it from dedicated websites (e.g. Excavations.ie or the National Roads Authority website), but only four used and contributed data to an online facility. The question *"When you reuse data how much restructuring is required?"* was answered by 29 respondents, of which, nine had to do extensive restructuring and 16 indicated light re-working, while four could use it directly.

An online survey undertaken in 2007 by the Archaeology Data Service on the creation, use and preservation of "big data" also provided evidence of data sharing and re-use (Austin & Mitcham 2007: 36). The survey addressed data from "big data" technologies such as airborne LiDAR (Light Detection and Ranging), terrestrial 3D laser scanning, maritime survey techniques and others. Responses were received from 48 respondents, not only archaeologists, but also from the earth sciences, for instance. 70% of respondents had somehow re-used data at least once a year (others "very infrequently"). Over 80% said that they had received large datasets from other researchers or organisations, and nearly 80% stated that they would allow access by others to their data. All respondents said that they consider using existing datasets for a new project, for example, to avoid duplication of costly data collection or conceive new surveys. The report notes, *"Clearly there is both a strong desire to, and sound reasoning for, reuse of data"*.

A survey conducted by the FAIMS project in Australia on the use of information technology by archaeologists included one question about sharing primary datasets (Sobotkova 2013). Of the 79 respondents 41% were academic and 37% consulting archaeologists; others such as government or museum employees were less present. 46% of the respondents were willing to share data after they had finished their own publication, 24% before publication but only with selected persons or groups, 20% without restriction (even before publication); 5% said that they are prohibited from sharing by their employer, and 5% were "not at all" willing to share their data.

4.5.10 Summary and suggestions

Summary of main results

In the ARIADNE/plus surveys we used the concept of "data publication", mainly to emphasise the common understanding that publication means that the data indeed is publicly available. Researchers often share data directly with colleagues but do not make them publicly available (e.g. in a repository). This means that valuable data remains within small circles of peers and is not available to other researchers and the wider public. Moving more data from closed-circle or not sharing to "open data" requires overcoming strong barriers (as addressed in the next section).

About ten years ago surveys across many disciplines showed that the data practices of researchers run against what advocates of proper data management and open data sharing would advise. The surveys

found that after the completion of research projects most data remains locked away, resides on PCs, storage devices, and restricted access servers, out of reach of other researchers, and in danger of loss. Over the last few years the situation seems to have improved, arguably mainly due to the expectation of research funders that data from funded projects is being deposited in appropriate repositories for long-term preservation and access.

More sharing of data through accessible repositories

In the ARIADNE 2013 survey around 50% of respondents shared from some to all of their project data through an accessible repository, while in 2019 around 65% ("not at all" shared respondents in this way 50% in 2013 and 35% in 2019). A comparison of the 2013 and 2019 results suggests that from 2013 to 2019 in the ARIADNE/plus communities of respondents the sharing of data through accessible repositories increased significanty by 10-15%.

Results of other surveys point to a general increase in repository-based data sharing, e.g. PARSE Insight (2009) compared to Tenopir et al. (2015), +10%. The Figshare surveys (2016, 2017, 2018) found a year-on-year increase in researchers' willingness to make their data openly available in various ways. In 2017, 29% said they made data available in a specific data repository while it was 33% in 2019, +4%.

ARIADNE/plus surveys found 30% more repository-based data sharing than others:

- In the Tenopir et al. 2013/14 survey (reported 2015) 30% said that they store from some to all of their data in a repository, in the ARIADNE 2013 survey 20% and in the 2019 survey 30%.
- In the Figshare surveys 2017 and 2018 sharing of data through a specific data repository was reported by 29% and 33% respondents, respectively. In the ARIADNEplus survey 2019 by over 30% more.

Differences between ARIADNE/plus 2013 and 2019 respondents

Considerable differences between the samples of respondents 2013 and 2019 should be noted:

- (presumably) more responses from ARIADNE partners,
- considerably fewer from countries with a mandated data repository (Netherlands, UK),
- more responses from Eastern and Southeastern Europe, and
- overall more established, permanently employed researchers and data managers (i.e. significantly less with a fixed-term contract and Ph.D. students).

In 2019, there were fewer responses from the Netherlands and the UK, where archaeologists are obliged to deposit data from publicly funded projects in an accessible repository. Despite this, the responses of all respondents amounted to 10-15% more data publication through different types of accessible repositories than 2013. Very likely more data publication was reported by other ARIADNEplus partners.

Particularly interesting are the figures reported for supplementary material:

- PARSE.Insight (2009) 15%; Tenopir et al. (2015) 19.4%; Figshare (2017) 34%, (2018) 35%; *Springer Nature* (2018) 42%.
- In the 2013 and 2019 ARIADNE/plus surveys, many more said that they make supplementary material available; 2013: 82% of 520 respondents, 2019: 81% of 449 respondents. In the 2019, 13% in all or most, 25% in many, and 42% at least in a few projects (the percentages 2013 are roughly the same).
- The explanation for the difference to the other surveys could be that many of the ARIADNE/plus survey respondents are obliged to provide fieldwork reports to a national heritage authority, and do this with supplementary material added. Compared to the Figshare

figure for supplementary material of around 35% (which is quite high), it appears that 45% of the ARIADNE/plus respondents do so.

- For the comparison between the ARIADNE and ARIADNEplus survey participants, the almost identical figures for supplementary material could mean that overall the differences are not as considerable as other survey results suggest.

Suggestions for ARIADNEplus

While the ARIADNEplus survey shows good results for sharing reports and data through institutional repositories (e.g. repositories of heritage authorities or research centres), many archaeologists in European and other countries do not have available yet a state of the art digital repository for archiving and sharing their data.

This issue is being addressed by the COST Action SEADDA, the Saving European Archaeology from the Digital Dark Ages network. SEADDA and ARIADNEplus share the goal of making archaeological data FAIR (Findable, Accessible, Interoperable and Reusable), especially by supporting knowledge exchange and collaboration on data repositories and e-infrastructure.

The core requirement for moving research data into accessible repositories is decisive open data mandates by research funders, coupled with funding of the basic costs of domain repositories and the researchers' data deposition costs (e.g. as part of research grants).

Suggestions to increase further the sharing of archaeological data through appropriate repositories are:

- Continue the good collaboration between ARIADNEplus and SEADDA on capacity building for new repositories and use of the ARIADNEplus digital infrastructure
- Support strict open data policies of funding bodies and institutions – data repositories and infrastructure should give full support to such mandates.

4.6 Barriers to data deposition & sharing

4.6.1 Background

About ten years ago, some research funders started or considered requesting deposition of data from publicly funded research in accessible repositories. Surveys showed that many researchers in different disciplines have a positive attitude to sharing of research data (with colleagues), however only a minority shares data in an open manner, such as depositing it in an accessible repository.

Studies based on interviews with researchers from different disciplines (e.g. Pryor 2009; RIN 2008; RIN & NESTA 2010) made clear that researchers would rather not make their data openly available to anybody. They perceived, and still perceive, more obstacles than incentives to do so (see also Bishop 2015; Borgman 2010; Costello 2009; LeClere 2010; Pearce & Smith 2011).

Among the obstacles or barriers: little academic recognition and reward for making data available; often unclear data ownership and rights of use, confidential and sensitive data; additional effort for providing shareable data (e.g. data preparation, metadata, licensing, etc.); concerns that data might be misused or misinterpreted, professional vulnerability if shortcomings of data are detected.

The main barriers to data sharing in archaeology are the same as for researchers in other disciplines. One specific concern is disclosing information about the location of archaeological sites which looters could use to identify them; in some cases also indigenous communities have a stake in the protection of sites and artifacts of cultural or religious value (Frank et al. 2015).

While there is an increasing pressure on researchers to make their data available the concerns did not disappear. The benefits of open data sharing are not apparent (e.g. lack of academic reward), rather the personal return on investment may be negative because of a competitive disadvantage. Consequently, attempts to make researchers more willing to share their data must demonstrate that it produces measurable personal benefits that outweigh the additional effort and potential risks.

Some convincing examples of researchers who benefitted will help (e.g. Popkin 2019), but the challenge is systemic and concerns the hierarchy of valuable contributions to academia. The core requirements for open data sharing are not technical but institutional, especially the need for appropriate academic recognition and reward (i.e. data citation, relevance for tenure and promotion). Other requirements must also be met, such as the existance of appropriate and trusted repositories, along with training and support in data and metadata preparation for sharing.

4.6.2 Survey results 2019

The survey participants were given a list of potential barriers to depositing their data in repositories and sharing it with other researchers. The respondents were asked how important the different barriers were for them.

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? (1) very important, (2) rather important, (3) rather unimportant, (4) not important:

	N	Very important	Rather important	Rather unimportant	Not important
A lack of professional recognition and reward for open data sharing	415	**173** (42%)	**139** (33.5%)	**72** (17%)	**31** (7.5%)
Intellectual property right issues (e.g. in collaborative projects)	413	**152** (37%)	**157** (38%)	**83** (20%)	**21** (5%)
Lack of a mandate or enforcement from research funding bodies	399	**131** (33%)	**146** (36.5%)	**92** (23%)	**30** (7.5%)
National heritage law does not require depositing digital data, not even with restricted access	399	**124** (31%)	**124** (31%)	**84** (21%)	**67** (17%)
Internal rules are in conflict with depositing data in open repositories	399	**51** (13%)	**117** (29%)	**116** (29%)	**115** (29%)
The work effort for providing the data and metadata in the required formats	406	**129** (32%)	**170** (42%)	**78** (19%)	**29** (7%)
The cost for depositing data in a repository for long-term preservation and access	404	**109** (27%)	**130** (32%)	**104** (26%)	**61** (15%)
Lack of an appropriate national or international repository where the data sets would “fit” into	398	**144** (36%)	**122** (31%)	**76** (19%)	**56** (14%)

***Table 10**: Barriers for data deposition and sharing (N = 398-415).*

Most of the respondents in 2019 who answered the question said that they perceive as “very” or “rather” important barriers for sharing their data through digital repositories. These barriers included: a lack of professional recognition and reward (75.5%), issues associated with intellectual property rights (75%), the required additional work (74%), and lack of an appropriate repository for their data (67%). Less important was the cost of depositing data in a repository (59%). The barrier with the highest percentage of “very important” was a lack of professional recognition and reward for open data sharing (42%).

In the 2019 survey, two new questions suggested by reviewers of the survey template were included. “Lack of a mandate or enforcement from research funding bodies” was perceived as a “very” or “rather” important barrier by 69.5% of respondents. At 62%, significantly fewer respondents thought that “National heritage law does not require depositing digital data, not even with restricted access” was a barrier. Only 42% thought “Internal rules are in conflict with depositing data in open repositories” was a “very” or “rather” important barrier.

"Other important barriers" and Comments

Respondents also had the option to state what they perceived to be "Other important barriers" as well as give "Comments". There were 55 entries under "Other" (12 "no" other or "none") and 16 comments. The respondents noted several barriers which impede the sharing of archaeological data. The many statements received are extremely useful because they reveal much information regarding the actual situation of data sharing, and feelings of respondents about it. Most of the responses under "Other" and "Comments" were combined and grouped together with those belonging to different categories of barriers.

Lack of awareness and interest

- *Awareness of the problem at a national/legislative level*
- *In all sectors in Greece (university, research, private, public) the conversation on primary data deposition has not been initiated*
- *Lack of awareness among non-specialists*
- *Lack of interest of institutions & researchers in sharing data*
- *People are unaware of the importance of depositing their data*
- *Some colleagues may not appreciate the importance of sharing data, especially of sharing it in an accessible way (consider e.g. the language barrier between Finland and other countries)*
- *Researchers do not fully understand the benefits of data sharing*

Political, legal and institutional situations

- *Politique d'établissement qui est encore loin de ces enjeux*
- *Lack of public commitment*
- *Absence de politique de diffusion*
- *Law for protection of buried heritage*
- *The main problem is national law, it should enforce data deposition even if access is restricted*
- *In my country (Bosnia and Herzegovina) we have the problem of two entities (…) Each entity has its own laws and rules. Unfortunately, there is no unique system nor collaboration between institutions.*
- *Across multiple countries provisions, awareness and requirements are varied*

Data ownership, IPR and copyright

- *The attitude of the Danish Museums who see the data they own as "theirs", which is simply not the case in the law*
- *People do not want to publish datasets in fear of someone "stealing" their future project/publications*
- *Desire to keep the data in order to be able to study it and publish it later*
- *Unwillingness to share data until all possible internal use has been extracted*
- *Competition for recognition*
- *Copyright on source data; Intellectual property right issues*
- *Intellectual property rights are brought forward as an excuse for not sharing (even after absurd amounts of time have passed since original data collection)*

- *We have seen a shift with our researchers from the fear of data being stolen when being in open access towards an acceptation of these conditions; being a public governmental service, data must be freely accessible*

Lack of skills and of curatorial and technical support

- *Skill level/training of colleagues; Lack of technical skills*
- *Teaching programs and on-line resources to demonstrate importance and challenges of sharing scientific data in archaeology*
- *Lack of experience with creation and reuse of digital data archives*
- *Lack of national authorities concerning the database structures. Data produced by individual researchers are unique and it is difficult to share them.*
- *Mapping to metadata standards is considered a major burden to many researchers, especially interdisciplinary research where disciplinary schemas do not cover the scope of the research. The mapping process also needs to be simple, transparent and fast. This is not so for CIDOC-CRM, for example.*
- *There is no digital repository for archaeologists and there is no internal rules for depositing.*
- *Lack of information about repositories*
- *Not all types of data can be meaningfully deposited in existing repositories*
- *Sometimes the existence of an appropriate repository is not well known, which also can be a problem*
- *Help of an IT service dedicated to SHS projects (on a long-term commitment, partnership)*
- *Lack of time and resources to publish*
- *Lack of time and/or staff available to complete the data sets*
- *The datasets get old*
- *New techniques allow bigger volume of data to be uploaded online, as a result some repositories that contain work of the last 5 years the data are so unappealing for the users*
- *Sustainability*

Financial

- *Lack of financial resources represent the main barrier to building digital repositories*
- *There is a serious lack of public funds for archaeological investigation in Portugal*
- *Lack of funds for investigation*
- *Lack of time and resources to publish*
- *Monetary costs even for researchers as individuals*

One respondent neatly summarised major issues as *"Time, money and lack of reward for effort for this kind of work are the issues as well as lack of training in how to do this".* Lack of funds was not mentioned often, perhaps because this is part of the work situation of many respondents. Several noted a lack of awareness of the importance and benefits of data archiving and access at all levels, national, institutional and individual researchers. In addition, existing regulations as well as institutional settings appear as barriers to making data openly available.

Arguably the strongest barrier is the perception of researchers who created/collected the data is that is their data, even if derived from publicly funded research, and that as much individual value as possible should be extracted from it. One respondent observed that intellectual property rights "*are brought forward as an excuse for not sharing".* The same might be said about the assumption that published data from a survey or excavation might be "stolen", e.g. used by other researchers for publications without appropriate credit or, where necessary, without contacting the data producers beforehand. Results of surveys regarding concerns about potential adverse use of openly shared data are summarised in *Section 4.6.5*.

Lack of experience with making data available through an accessible repository seems to contribute to researchers' reservations regarding open data practices. One respondent from a governmental service mentioned that they *"have seen a shift with our researchers from the fear of data being stolen when being in open access towards an acceptation of these conditions".*

Where researchers are interested in making their data available there is still much need for training and curatorial and technical support. Experienced researchers doubt that colleagues have the required knowledge skills for creating shareable project data and documentation (metadata) according to established standards. Training for the application of such standards, guidance on good practices as well as institutional support (data managers, IT services) are needed *"to complete the data sets"* for sharing.

There is also a worry that datasets might get outdated and repositories without much current data unattractive. More worrying, however, is a perceived lack of readily available information about appropriate repositories of archaeological data.

4.6.3 Comparison 2013 / 2019

In the 2019 and 2013 surveys the question on barriers to sharing research data through digital repositories was answered by about 400 and 500 respondents, respectively. The barriers which respondents perceived as most critical were the same, albeit with some differences regarding the percentages of "very" or "rather" important combined:

- *A lack of professional recognition and reward:* was considered as most critical by 75.5% of respondents in 2019, while 72% in 2013.
- *The work effort for providing the data and metadata in the required formats:* was an important barrier for 74% of respondents in 2019, while in 2013 more respondents worried about the work effort for metadata (80%) and data (80%).
- *Intellectual property rights issues:* was a concern for 75% of respondents in 2019, while significantly less in 2013 with 65%

Two barriers were perceived as somewhat less important with about the same percentages: Lack of appropriate repositories with 67% in 2019, while 66% in 2013; the cost for depositing data in a repository with 59% in both years.

4.6.4 Results of other surveys

Results of other surveys with respondents from many disciplines are hardly comparable to our surveys because of different sets of questions or missing information such as percentages for some results which could be compared. But some of their results are particularly worth noting.

Tenopir et al. 2011 and 2015

The Tenopir et al. 2011 and 2015 surveys asked respondents about reasons for not making data available electronically to other researchers. In the 2011 survey the most important reasons were "insufficient time" (54%) and "lack of funding" (40%). Less important were "do not have rights to make data public" (24%), "no place to put the data" (24%), "lack of standards" (20%), and ''sponsor does not require'' (17%). (Tenopir et al. 2011: 9).

In the Tenopir et al. 2015 survey, respondents who said that they do not share all of their data were asked why all or part of their data are not available to others, and then given a series of possible reasons. Where comparisons could be made to the 2011 survey, they found that perceptions of barriers had changed. There was less importance assigned to insufficient time and lack of funding, while a greater concern about not having the rights to make data public and that other people might not need them.

The top-ranked barrier in 2015, not included in 2011, was *"I need to publish first"*, 43.5% of respondents who said that they do not share all of their data. In the Fecher et al. survey in 2015 *"if I had enough time beforehand, to publish on the basis of my data"* was the second strongest enabler of data sharing (after *"if I were cited in publications using my data"*). On a 5-five point Likert scale 77.5% of 1,420 respondents agreed to it, 46.5% completely and 31% somewhat less (Fecher et al. 2015: 20). Obviously, researchers must be granted the time to exploit their data appropriately, i.e. sufficient time until data of funded research has to be archived or include embargos on deposited data.

Figshare 2018

In the Figshare 2018 survey the top six responses to *"What problems/concerns do you have with sharing datasets?"* (over 400 respondents) were: "Concerns about misuse of my data", "Unsure about copyright and licensing", "Not receiving appropriate credit or acknowledgement", "Unsure I have the rights to share", "Organising data in a presentable and useful way" and "Contains sensitive information" (Baynes 2018: 16).

The percentages for these concerns are not given in their report. Regarding credit for data sharing however, the majority of respondents in 2018 felt that they did not get sufficient credit, 58%, compared to 9% who felt they do; 33% were not sure. The importance of credit (i.e. data citations) is addressed in greater detail in the next section.

Particularly noteworthy among the findings of the Figshare 2018 survey is also that the percentage of respondents in support of national mandates for open data was higher at 63% than in 2017 (55%), but in 2016, 78% were in support. Further, the survey found a marked increase of uncertainty about where funds will come from to support making data open: 53% of respondents in 2018, while 36% in 2017 and 30% in 2016.

Springer Nature 2018

In the *Springer Nature* survey 2018 over 4,000 respondents answered the question *"What problems do you have in sharing datasets?"*. The most important reason for not sharing data was "Organizing data in a presentable and useful way", selected by 46% of respondents. Other common challenges were: "Unsure about copyright and licensing" – 37%; "Not knowing which repository to use" – 33%; "Lack of time to deposit data" – 26%; "Costs of sharing data" – 19% (Springer Nature 2018: 16).

In the 2019 ARIADNEplus survey, the results for these barriers to share data through digital repositories were roughly similar. IPR issues, the additional work required, and lack of an appropriate repository were perceived as important barriers, while less so, the costs of sharing data (e.g. deposit costs).

Size of datasets matters: A very interesting result of the *Springer Nature* survey is that the size of datasets has an impact on whether data are shared: respondents that generated the smallest data files had the highest proportion of data that were not made available through a repository or as supplementary information. This was the case for 42% of 2,036 respondents with the smallest data files of below 20 megabytes (MB). If they share such data, these respondents had a clear preference for making it available only as supplementary material. In contrast, among the 700 respondents with paper-related data greater than 50 gigabytes (GB), 70% make their data available, with a strong preference for sharing through repositories (59%).

4.6.5 Importance of data citation

Understanding the core role of professional recognition and reward for impeding or driving open data sharing is essential for infrastructures for research data such as ARIADNE as well as the underlying digital repositories, including recognition for their own work. Advocates of open data argue that such data will often be (re)used and cited, bringing recognition and rewards to data publishers (incl. data repositories). The scenario is that data citations indicate and acknowledge providers of valuable data, promote further data sharing and (re)use, and enable the impact of open data to be tracked and measured. Most importantly, it would drive the emergence of an academic credit system that appropriately rewards open data sharing.

A recent investigation of data citations in archaeology using the Crossref DOI citation dataset found *"that citation of datasets, although problematic to measure, appears to be almost nonexistent in archaeological literature"* (Marwick & Pilaar Birch 2018). An examination of the Thomson Reuters Data Citation Index found a few more citations for archaeological data, but inconsistencies that make them unreliable, and *"some signs that, when available, their insensitivity to context makes them of limited value as a means of assessing data reuse"* (Huggett 2017).

While researchers expect that shared data that has been (re)used by others is cited, there is little empirical evidence as yet of benefits derived from open data sharing. Researchers also have many concerns about adverse use of their data, such as data being scooped, misused or misinterpreted. They balance potential benefits of open data sharing against potential negative effects, and the outcome is not necessarily positive for sharing.

Researchers expect data citation

Survey results confirm that data citations, assumed to translate into benefits such as professional reputation and career advancement, could be a strong motivation for researchers to make open data available. In the Tenopir et al. 2011 survey of 1,291 respondents 91.7% agreed that *"It is important that my data are cited when used by other researchers"*, 1.6% disagreed and 6.7% were undecided. In the Fecher et al. survey (2015) 79.3% of 1,420 respondents said that *"if I were cited in publications using my data"* it would motivate them to make data available to others, 9.5% said it would not, and 11.2% were undecided. In the Belmont Forum's survey of around 850 respondents 69% agreed that *"dissemination and recognition of your work"* is a very important motivation for them to make their data openly available, only 5% said that it is not important, and 26% were undecided (Schmidt et al. 2016).

Thus in these surveys respondents from different disciplines considered it essential that researchers who make data available gain academic/professional recognition. Notably, other potential benefits such as co-authorship of papers or involvement in projects that build on the data appear as less important. For example, in the Tenopir et al. 2011 survey 59.7% of respondents considered co-authorship as a motivation for data sharing, in the Fecher et al. 2015 survey only 34% of respondents (Tenopir et al. 2011: 8; Fecher *et al*. 2015: 9).

However, there is a major problem regarding citations for open data sharing. Extensive analyses have shown a clear citation advantage for open access papers in journal and conference proceedings, in the recent Piwowar et al. (2018) study 18% more citations than average. SPARC Europe documented and evaluated citation advantage studies for many years and discontinued this work 2016 because they considered the advantage as far more common knowledge than in the early days of open access publications[30].

In comparison, little empirical evidence is available for a citation advantage of open research data. A briefing paper of SPARC Europe (2017), prepared by the Digital Curation Centre (UK), presents some of the scarce evidence that is available for a citation advantage of papers published with a link to underpinning data. Such studies are available for a few research fields such as clinical cancer trials, astronomy/astrophysics, and international relations. Evidence that the citing authors are actually influenced by data access is even more limited. For example, Piwowar & Vision (2013) analysed the citation counts of 10,555 papers on gene expression studies that created a microarray of data, and found that studies that made data available in a public repository received 9% more citations than others for which the data was not made available. But they also found that only a few citations more than on average came from papers that reused the data. Researchers have various reasons to reference available datasets, not necessarily because they (re)used data.

The most important caveat of the SPARC Europe paper is that data reuse patterns, i.e. if, how and to what degree available research data are being reused, can be assumed to be very much domain-specific, and so will citation or non-citation of the data by other researchers. Characteristics of research domains, such as the degree of collaboration, common procedures, typical data sizes, will have supporting or impeding effects.

Not receiving appropriate credit

In the ARIADNEplus 2019 survey, respondents perceived a lack of professional recognition and reward as the most important barrier for sharing their data through digital repositories. Of the respondents, 75.5% considered it as “very” or “rather” important, while 42% considered it “very important”. This was the highest percentage of “very important” for any one of the barriers suggested, e.g. more than IPR issues (37%), lack of a mandate or enforcement of open data by research funders (33%), or additional work effort for providing the data and metadata (32%).

In the Figshare 2018 survey “Not receiving appropriate credit or acknowledgement” was among the top concerns of respondents with sharing datasets, and 58% felt that they do not get sufficient credit for sharing data, while only 9% felt they do, and 33% were not sure. In this survey only 46% of respondents said citations would motivate them “very much” or “quite a lot” to make data openly available, although 7% more than in the 2017 survey. Thus the appreciation of citation was much less pronounced than in earlier surveys, i.e. 91.7% in Tenopir et al. (2011), 79.3% in Fecher et al. (2015), and 69% in Schmidt et al. (2016). These surveys also used scales of agreement and the percentages are also only for the first two points on 5-point Likert scales in Tenopir et al. and Fecher et al. and “very important” in the 3-point scale in the Schmidt et al. survey.

In the reference surveys, the question of whether data citations would motivate them to share data was answered by between 850 and 1420 respondents from different disciplinary backgrounds: Tenopir et al. (2011) most from ecology, environmental science and biology, Fecher et al. (2015) well-balanced across disciplines, Schmidt et al. (2016) most from earth & environmental, climate & atmospheric and biological sciences, and Figshare (2018) well-balanced across disciplines.

[30] SPARC Europe: The Open Access Citation Advantage Service, https://sparceurope.org/?page_id=978

If we consider only the results of Fecher et al. (2015) 79.3% and Figshare (2018) 46%, the difference in motivation is around 33% less in the Figshare survey. Notably, the Figshare 2018 survey also found a marked decrease in respondents who valued a data citation as much as an article citation, 55% in 2018, while 68% in 2016; those who valued it less were 30%, up from 20% in 2017.

It appears that over the years the enthusiasm for data citations for open data sharing declined because the researchers perceived that there has been little, if any, recognition and reward from the academic/professional credit system for data sharing. Correspondingly, concerns about adverse use of shared data may have increased.

Adverse use of shared research data

Adverse use of shared research data comprises any form of behaviour by data users that results in an undesirable outcome for the researcher/s who made the data available. These include data that might be scooped for competitive misuse, flawed interpretation by data users, and professional vulnerability if shortcomings of data are detected and published results falsified. Moreover, data sharers may be accused of not having dealt with ethical issues appropriately. Consequently, they may want to control who can access and use their data. They fear a loss of control if their data is accessible to others, and avoid it, bringing forward various excuses for not sharing data (Bishop 2015; Costello 2009; LeClere 2010; Pearce & Smith 2011; Rouder 2015; Strasser 2013).

Survey results on adverse use

In the Tenopir et al. 2011 survey a majority of respondents agreed strongly or somewhat strongly that across their research field data may be misinterpreted due to complexity of the data (75%) as well as due to poor quality of the data (71%). Furthermore three quarters (74%) believed that data may be used in ways other than intended. These results represent opinions and do not necessarily reflect actual practice, however, as the authors note, the level of agreement "*reveals many psychological barriers to good data sharing practice"* (Tenopir et al. 2011: 5 and 7-8). The follow-up survey in 2015 reports increases in scientists' concern over these issues based on a statistical analysis of the 2011 and 2015 answers, but the figures for the responses 2015 are not given (Tenopir et al. 2015: 18, S1 Appendix, Table Q).

Fecher et al. (2015) asked respondents about their concerns regarding unfavourable outcomes of making data available. A clear majority of 80% said that they would not share their data *"if other researchers could use my data to publish before me".* 46% said that data could be misinterpreted prevents them from sharing data. But only 12% were concerned that others could criticize or falsify their work. The majority of the respondents (72%) disagreed that criticism or falsification would prevent them from making data available.

In the Belmont Forum open data survey 41% of respondents thought that "loss of credit or recognition" is a major barrier to open data sharing, 38% a barrier, and 21% a minor barrier. Next came "misinterpretation or misuse", for 37% a major barrier, 37% a barrier, and 26% a minor barrier. For "loss of control over intellectual property the percentages were 34%, 38% and 28%, respectively (Schmidt et al. 2016)

Need to publish first

In the Tenopir et al. 2015 survey the top-ranked barrier, not included in 2011, was *"I need to publish first"*, while 43.5% of respondents said that they do not share all of their data. In the Belmont Forum survey 54% of respondents thought that *"desire to publish results before releasing data"* is a major barrier to open data sharing, 32% a barrier, and 14% a minor barrier (Schmidt et al. 2016).

The main barrier in the Fecher et al. survey 2015 was respondents' worry that others could use their data for publications before them (80%). But a strong motivator was *"if I had enough time beforehand,*

to publish on the basis of my data" (77.5%). The authors emphasise that "publish first" is the most important sharing condition, and that their results indicate that researchers see data publications as far less valuable for reputation building than research papers.

Most research funders that nowadays request that project data is deposited in an accessible repository require that this is done between 6 to 12 months after completion of the project. Thus at that time researchers worrying that others might "scoop" their data for publications would have to exploit the data as far as possible in own papers.

A period of even 12 months after project completion may seem way to short for some research communities. Detailed case studies based on interviews with members of different scientific research communities revealed a range of concerns regarding data sharing, including that many researchers wish to exploit their data and related intellectual capital to produce publications over an extended period (RIN 2009; RIN 2010). Pryor (2009) summarised results of the case studies focused on diverse research communities in the life sciences (RIN 2009) in provisios that should govern the sharing of data. One of these is that "*sufficient time must be given to allow the completion of their analysis of the data"* (Pryor 2009: 80). However, their study groups were unable to prescribe how much time would have to be granted to complete the analysis. One group based on a retrospective example thought four years, but also that a new method or tool could allow further exploitation of their data. Members of two other groups declared it impossible to predict when their data "could be deemed *finished with* and available for open sharing".

4.6.6 Summary and suggestions

Summary of main results

Core functions of the ARIADNEplus digital infrastructure are to aggregate data from archaeological repositories and provide search and access services. Therefore the ARIADNEplus initiative depends on repositories richly filled with accessible data shared by researchers. It cannot ignore obstacles which hinder researchers in sharing their data in an open manner. Rather the initiative must support researchers in data sharing and help ensure that they receive appropriate credit for doing so.

ARIADNE/plus surveys 2013 and 2019

In the 2013 ARIADNE and 2019 ARIADNEplus surveys the participants were given a list of potential barriers for researchers to deposit their data in digital repositories and share it with others. The respondents were asked how important the different barriers are in their view. The question was answered by around 500 respondents in 2013 and 400 in 2019. The barriers which respondents perceived as most critical were the same, albeit with some differences regarding the percentages of "very" or "rather" important combined:

- *A lack of professional recognition and reward:* was considered as most critical by 75.5% of respondents in 2019, while 72% in 2013.
- *The work effort for providing the data and metadata in the required formats:* was an important barrier for 74% of respondents in 2019, while in 2013 more respondents worried about the work effort for metadata (80%) and data (80%).
- *Intellectual property rights issues:* was a concern for 75% of respondents in 2019, while significantly less in 2013 with 65%.

Two barriers were perceived as somewhat less important with about the same percentages: Lack of appropriate repositories with 67% in 2019, while 66% in 2013; the cost for depositing data in a repository with 59% in both years.

In 2019 many respondents noted also lack of awareness and interest, political, legal and institutional obstacles, and lack of skills and of curatorial and technical support. Respondents for example said, "*Researchers do not fully understand the benefits of data sharing", "Unwillingness to share data until all possible internal use has been extracted"*, "*Lack of time and/or staff available to complete the data sets".*

The main barriers to data sharing in archaeology are the same as for researchers in other disciplines. One specific concern is disclosing information about the location of archaeological sites which looters could use to identify them; in some cases also indigenous communities have a stake in the protection of sites and artifacts of cultural or religious value.

Results of other surveys

It is worth noting some results of other surveys: Fecher et al. (2015) in a large survey with respondents from different disciplines found that *"if I were cited in publications using my data"* would motivate 79.3% of 1,420 respondents to make data available (9.5% said it would not, and 11.2% were undecided); *"if I had enough time beforehand, to publish on the basis of my data"* was the second strongest enabler of data sharing, 77.5% agreed to it. Obviously, researchers must be granted the time to exploit their data appropriately, i.e. sufficient time until data of funded research has to be archived or include embargos on deposited data. In the latest Figshare *The State of Open Data* survey (2018) the majority of respondents felt that they did not get sufficient credit for data sharing, 58%, compared to 9% who felt they do; 33% were not sure.

Suggestions for ARIADNEplus

Understanding obstacles to data sharing and helping to remove them is essential for infrastructures for research data such as ARIADNEplus as well as the underlying digital repositories. Advocates of open data argue that such data will often be (re)used and cited, bringing recognition and rewards to data publishers (incl. data repositories). The scenario is that data citations indicate and acknowledge providers of valuable data, promote further data sharing and (re)use, and enable the impact of open data to be tracked and measured. Most importantly, it would drive the emergence of an academic credit system that appropriately rewards open data sharing.

- Research infrastructure components, protocols and metrics for data citations are in development. ARIADNEplus should investigate how services of the research infrastructure could help identify and track (re)use of data based on data citations (e.g. article-data links) and other indicaters.
- As a general requirement for identifying data (re)use, the project could promote and support standardisation of data citation in the archaeological sector, i.e. how data should be cited in publications to ease the identification and tracking of data (re)use.

4.7 Reuse of data

4.7.1 Background

The 100th edition of ERCIM News, the quarterly magazine of the European Research Consortium for Informatics and Mathematics (over 7,500 subscribers), featured scientific data sharing and reuse as a special theme. The title of the article by Christine Borgman, a distinguished American scholar on this topic, is *"If Data Sharing is the Answer, What is the Question?"* (Borgman 2015). Borgman and colleagues at the UCLA Center for Knowledge Infrastructures[31] had investigated scientific data practices and infrastructure for some years. A new large project, with the question as project title, was carried out 2015-2018, funded by the Sloan Foundation.

In the article Borgman notes that data sharing had already become an unquestioned policy enforced by governments, funding agencies, journals, and other stakeholders, while arguments against data sharing were rarely expressed in public fora. She stresses that critical questions of data sharing should not be side-lined and lists many questions, especially concerning utility and required investments, that needed to be addressed in-depth, involving all stakeholders (see also Pasquetto et al. 2017). Our short answer to the question about the rationale for data sharing is: data reuse.

What does reuse of data mean

In the discussion about reuse of research data various definition were used. Van de Sandt et al. (2019) looked into a sample of 65 works related to reuse of research data, 20 provided a definition, 45 did not, although the term was used in the title. One definition which is often referenced in the literature is the *"the use of data collected for one purpose to study a new problem"* (Zimmerman 2008). But this definition limits it to use for new research, while data can reused for different purposes.

The more common notion of data reuse in research, covered by the term "secondary use", is that the data is being used for another purpose than the one(s) for which the producers collected or generated it. Also in the ARIADNEplus survey, no specific definition of data reuse was suggested. What counts as more or less valuable reuse of shared data, initiatives will have to sort out how to track reuse based on data citations (e.g. article-data links) and evaluate it with appropriate metrics, obviously this requires more than just counting citations. Research infrastructure components, protocols and metrics are in development (Burton et al. 2017; Cousijn et al. 2019; Fenner et al. 2018; Pierce et al. 2019).

Notably, the FAIR data principles define requirements for "reuseable", e.g. description with rich metadata, and release with a clear data usage licenses, but do not provide a basis for the evaluation of valuable reuse.

Why reuse is important

It is important that research data is being made available, but even more so, that it is being (re)used, otherwise the benefits associated with open data sharing will not materialise. There are many good arguments for making data available, for instance, reported research results can be scrutinized and duplicate data collection prevented.

Particularly strong however is the argument that reuse of data, for example to investigate new research questions, allows exploitation of previous investment. Preserved data that is being reused gains in value, otherwise it might be perceived only as a cost factor. "Return on investment" expected by research funders explains much of the increasing pressure on researchers to share their data from

[31] Center for Knowledge Infrastructures, UCLA Department of Information Studies, https://knowledgeinfrastructures.gseis.ucla.edu

publicly funded research for reuse. Furthermore, why for data respositories documentation of actual reuse is very important (although not easy).

The first things to consider and study regarding open research data is whether researchers in particular domains (sub-discipline level) actually do reuse data of others, or would do so, if more reusable data becomes available. If reuse is unlikely it would make little sense to build digital repositories and ask researchers to invest the effort to prepare their data and metadata for this purpose. "Just in case" certainly is not a good argument. The Royal Society's report *Science as an Open Enterprise* notes, *"Sharing research data can be complex and costly and needs to be tempered by realistic estimates of demand for those data"* (The Royal Society 2012a: 60). Doubts about "build it and they will come" not only apply to sharing but also to reusing data (Wallis et al. 2013).

High demand typically exists for data from core archives, but not necessarily for "small data" repositories. If substantial reuse of data from data repositories can be assumed, we may start asking questions about how to enable easier and more effective reuse of data once disconnected from their producers. Metadata with rich context information is one important requirement, regarding archaeologists see Faniel et al. (2013). But there are also others, for instance, related to different purposes and forms of reuse. Huggett (2018) discusses different forms of repurposing data in archaeology.

Overall, there is a need to better understand data reuse (and lack of reuse) practices in different fields of archaeology and related disciplines, so that data reuse can be fostered and supported effectively. The results of the ARIADNEplus survey confirm that archaeological researchers often reuse data and provide some insights about what and how.

4.7.2 Survey results: Data reuse in the last 2 years

The questions for reuse of data in the survey were as follows: First the respondents were asked *"Did you / your research group in the last 2 years use any data which other researchers made available through a publicly accessible digital repository or databases?"* Thus "use" not "reuse" was used in this question. The reason behind this was that "reuse" would have raised the question for what purpose which we did not want at this point.

When the answer was "yes", a free-text field was served and the respondents requested, *"Please briefly describe what kind of data and from which repository or database?".* Therefore we wanted to get information about the type of data the respondents used, from which sources, and possibly also learn about what their notion of reuse is (e.g. from description of what they did with the data).

Next came the question, "*What was the main purpose of the data reuse?"* and a table with only three predefined purposes and "Other" which, if selected, opened a field to describe the purpose/s. Thus the term "reuse" was employed at this point, in a question about the main purpose of reuse of particular data. The predefined purposes of data reuse were distinct: "Building a database for the research community", "Comparison to own research results", and "Use together with own research data". But the respondents could select each of them, as well as add other purposes, because in a project the data might be used for very different purposes.

Use of shared data in the last 2 years

Did you / your research group in the last 2 years use any data which other researchers made available through a publicly accessible digital repository or databases?

Yes	220	58.5%
No	154	41.0%
No answer	2	0.5%
	376	100%

***Table 11**: Use in the last two years of data available from publicly accessible repositories/databases.*

An astonishing number of respondents said that in the last two years they (their research group) used data made available by other researchers through a publicly accessible digital repository or databases[32]. Nearly 60% of respondents who answered the question did so, over 45% of all survey participants.

Why data was not reused

34 respondents provided comments and most (24) said relevant data was not available or posed some difficulties. Six respondents said that they did not need data others may have made available (e.g. *"we produce most data ourselves"*), while three where unsure. One respondent mentioned, *"We have used data from other institutions that are not yet publicly accessible"*.

Some statements which mention the kind of data sought, why relevant data was not available, or why available data was not relevant or difficult to use are:

- *My current research interest is limited to local/national data about STECCI [monumental medieval tombstones] which are not available in other researchers' repositories.*
- *There is no local repository for archaelogical species.*
- *There aren't any useful archaeological repositories in Croatia that I am aware of.*
- *The projects were aimed at the production of new data / the existing data is not reliable.*
- *I conducted a test to see if other researchers' survey data in the DANS archive were re-usable, but quickly ran into interpretation problems.*
- *There is no such repository tackling our research interests in the last two years.*
- *New field in research. Data re-use need common reference models for re-use.*
- *No other research group offer access to raw data.*
- *Because the used data are not digitized*
- *Mainly because of missing licensing information*
- *No such data was available for easy reuse or download.*
- *Не използвани (not usable)*
- *Lack of support*

[32] Two respondents did not answer the question but continued to fill the questionnaire after the question on the main purpose of data reuse.

4.7.3 Survey results: Purposes of data reuse

Respondents who said that in the last two years they had used data which others made available through a publicly accessible digital repository or databases were presented two requests: first they were asked to describe what kind of data and from which repository or database, then to state the main purpose of the data reuse. In this and the next section we first look into the purposes while an overview of the types and sources of data is given thereafter.

The 220 respondents who had declared that they (their research group) used data made publicly available by others were asked "*What was the main purpose of the data reuse?"* and presented a table with three predefined purposes and "Other - please specify" to describe other purposes of the data reuse. Multiple answers to the question were possible.

Purposes of data reuse	***Was a purpose for***	***Percent of respondents*** *(220)*
Building a database for the research community	68	31%
Comparison to own research results	121	55%
Use together with own research data	139	63%
Other purpose/s	19	7%

Table 12*: Purposes of data reuse (multiple answers possible).*

All of the 220 respondents answered the question. Each ticked more than one of the three predefined purposes, and 19 also "Other" of which 18 provided some information. In summary, most respondents said that the main purposes were to compare own empirical, data-based results to those of others, and to use the data of others together with their own research data. Building a database for the research community was much less important, but 31% agreed that this was one important purpose.

"Other" purposes

Research*:* This was the main "other" purpose for seven respondents. Two statements were simply *"Research"* and *"Use the dataset for research".* The purposes for (re)using the data of the other five respondents were

- *Typology terms and methodology*
- *As test data for building our data structure*
- *To test whether DANS survey datasets contain enough metadata to allow re-use without requiring further information from the depositors.*
- *Test algorithms and approaches on independent data sets*
- *Ancient vases imagines detection*

Education*: "Educational", "Teaching", "Museum guidance material".*

Heritage protection*: "Protection of archaeological heritage"* and *"Management and protection of archaeological heritage"*

Other purposes, perhaps related to one of the above: *"Search for information", "Comparison with commercial data", "Documentations", "Own database", "Addition the information of archaeological sites", "Enable readers to view narrative and underlying data together".*

4.7.4 Analysis of purposes and examples

The main purposes of researchers for (re)using data made available by others arguably were (1) to use them for comparing their research results to those reported by others, and (2) to use available data together with own research data. The first purpose can also lead to doubt that the results presented by others are sound. In addition, researchers and other (re)users may have various other purposes for (re)using data made available by others. Research-related purposes that surfaced in our survey were to use the data as test data (e.g. "test algorithms and approaches"), as a conceptual resource (e.g. "typology terms", "data structure"), or to support the narrative of a publication with data (e.g. charts) published by others.

Comparison to own research results

Comparing own research results to those of others is an important part of research. Researchers compare their results to what others found, and how well this is supported by their data. Regarding the data, often they only had available what is presented in a research paper or report, e.g. summary tables or charts which present statistical results. In addition, there could be supplementary material that provides more detailed documentation of the data which underpin the presented research results.

Proponents of sharing and reusing open or FAIR research data believe the full underlying data should be available for others to consult and/or to use for further research. One important reason for requesting the full data is that reported results may appear questionable, and replication of the data analysis could result in rejection or corroboration of the results.

The ARIADNEplus 2019 survey did not ask researchers about these activities specifically, and little of the information we have from other questions points in this direction. Investigating the reasons of researchers to share or not share their data, Fecher et al. (2015) found that only 12% of their 1,400 survey respondents were concerned that others could criticize or falsify their work, while 80% worried that others could use their data for publications before them. The ARIADNEplus respondents could have chosen to add questioning research results of others under "Other" purposes, but did not.

Asked why during the last two years they did not use data made available through a publicly accessible digital repository or database, most said that such resources were not available, or that there was some relevant data but accessing and using it proved to be difficult. The latter included issues such as missing licensing information, difficulty to reuse or even download, lack of support, among others.

One respondent said that projects carried out by the research group "*were aimed at the production of new data / the existing data is not reliable*", but this concerned the base of available research data in general, not those of particular projects. Others said that there was no repository with data relevant for their research interests, or that they are working in a new field of research, without common reference models for reuse. While these are only some statements, it seems reasonable to assume that most researchers are primarily interested in making progress in their own research, not to reproduce or replicate work of others, i.e. doing the same or trying to falsify them.

Use of shared data with own research data

Use of data publicly shared by others together with own data for research purposes is actual data reuse. Descriptions of respondents which include more than one data type or source suggest that there are some patterns of frequent data reuse in archaeology.

Patterns

Based on typical examples given by respondents at least three patterns of reusing different data types or sources, with implied use together with own research data, could be distinguished.

Information on sites (maps, distribution) and fieldwork reports:

- *"Results of fieldwork and prospection (Archaeological Map of the CR). Systematic registers of specific site types and GIS data (Longwood project, CZ_Retro, national registers etc.)"*;
- *"Excavation reports and underlying data (easy.dans.knaw.nl), various national maps (PDOK plugin in QGIS), national elevation data (www.ahn.nl), national dendrochronological data (dendro.dans.knaw.nl)"*;
- *"archaeological data - Information System on Archaeological Data, maps - State Administration of Land Surveying and Cadastre"*;
- *"data published online by a web platform managed by the local superintendence which published excavation reports and topographic data"*;
- *"General data about sites in Romania and archaeological reports, provided by the National Institute of Heritage"*;
- *"grey data, pictures, maps - Digital Archive of Institute of Archeology (Academy of Sciences and Arts). Acheological Map of Inst of Czech Republic, GIS projects of Historical Landscape"*;
- *"Site and monuments database, online archaeological reports"*.

Cartographic, GIS and LiDAR data:

- *"National LiDAR and cartography"*;
- "GIS, LiDAR, descriptions of archaeological monuments";
- *"GIS and LIDAR data"*;
- *"Lidar in OpenTopography, GIS in NGO web sites"*;
- *"mainly satellite data and maps"*.

Databases/catalogs of different artefacts:

- *"Portable Antiquities Scheme; Joconde; Artefacts"*;
- *"Coin and ceramic catalogs"*;
- *"Coins and ceramics"*;
- *"Databases of inscriptions and coins"*;
- *"numismatic collections, museum collections"*.

Building a database for the research community

Building a database for the research community was certainly not the main purpose of respondents who (re)used data made available by others, but still 61 of 220 respondents (31%) said it was at least an important purpose. The 61 respondents were from all organisational categories distinguished as part of the survey demographics: 36 were from a university or public research organisation, 11 from a governmental institution, nine from a museum, nine from a private company or research institute, and three from another category (local municipality, NGO, and one not given).

Comparing their shares in this group (61) and of the total of all survey participants with known organisational background (482), the percentages from a university/public research organisation are

exactly the same (53%), governmental institution nearly the same (16% here and 15% in the total), while museum was lower, 13% and 19%, respectively. The percentage of private company or research institute is higher, 13% compared to 8% in the total, but representatives of other categories were too few in both samples to compare.

Examples

The respondents were not asked to describe how they reused the data for stated purposes. Therefore not many descriptions of how respondents in the last two years reused data for "building a database for the research community" are available. The small number of examples could also be interpreted that some respondents had the intention to later make a database created in a project accessable to others as well.

Some examples which suggest that one of the purposes of data reuse was "building a database for the research community" are:

- *"The MedRadiocarbon database(https://github.com/ahb108/MedRadiocarbon) was used to assess the current status of knowledge about radiocarbon dated contexts in the Liguria region, for which even our office (Soprintendenza) had no complete listing"*
- *"The data collected and published in the National Chronicled of Research in Romania was used by our institution in order to develop the National Repertory fo Archaeology database"*
- *"Developing a digital resource that links to digital data"*
- *"external links to digital (photographic) archives like Ubi erat lupa, British School at Rome, Corpus Inscriptionum Latinarum, U.S. epigraphy project"*;
- *"public domain or CC0 images (pictures, pages of book) with URI (Gallica ; HAL and Medihal ; archive.org...). Thesauri alignments (Getty, French Ministry of Culture vocabularies...) ; References to publications (URI) and persons (ISNI). Commitment in the building of the Cultural Heritage graph (web of data): links with special domains repositories"*;
- *"Numismatic data from various museums and collections via LOD [Linked Open Data] resources such as OCRE, CRRO etc."*

The first two examples are from governmental institutions, the others are examples of databases (also) providing links to several data sources or items of other providers, particularly using a Linked Open Data (LOD) approach. Other examples are not as clear, e.g., *"Augmenting existing finds data to the new finds database"*.

The following is an example of a research institution that makes content from their projects available as a "database" for use also by others, *"The institute has a website through which most of the projects' results are regularly uploaded and accessible, e.g. coins, artefacts, photographic archives. The users are also able to download the specific software which will help them see the images (e.g. RTI viewer)"*.

4.7.5 Types and sources of data

The survey respondents were asked whether or not in the last two years they (their research group) used data which other researchers had made available through a publicly accessible digital repository or database. Those who said yes (220) were then asked to briefly describe in a free-text field what kind of data and from which repository or database it was derived. 188 respondents provided descriptions.

The descriptions included the source and type of data (some different ones), only the source(s) or only the type(s) of data. The descriptions also vary regarding the detail of description. For example, some respondents mentioned only one or two data types, e.g. "14C data", "genetic data", "find data", "coins and ceramics" or "GIS and LIDAR data".

Many repondents gave only the name of a source, e.g. "AMČR" [Archeological Map of Czech Republic] or "Archaeological Map of Bulgaria", which could include a variety of data types, as in statements such as *"several sources from the ADS (U York)"* or *"all sorts of data from ADS"*. Other respondents elaborated on data which they (re)used in greater detail, for example, *"Textual and geodesic data from the national finds and monuments register, Fund og Fortidminder (both the public and password protected versions). Textual information from reports produced by The National Museum of Denmark. Images and textual data from Swedish Rock Art Research Archives. National mapping and LIDAR datasets from the Danish and Norwegian Geodetic agencies."*

In total, the descriptions contained around 250 mentions of various types of data. Figure 7 presents a word cloud of types mentioned. In order to produce it, the mentions of data had to be extracted from the descriptions, separated (e.g. coin and ceramic catalogs, became two catalogs), harmonised (e.g. spelling), and grouped together. Many types in the word cloud are the originals, for example, 3D models, chemical analyses, geoarchaeological data or genetic data, each mentioned by only one respondent while, for example, 14C data or GIS data by many more. Others had to be summarised with more common terms used by respondents (e.g. images instead of pictures or photos). Furthermore, in some cases where respondents referred to sources, e.g. coin catalogs or museum collections, it seemed more appropriate to include these in the word cloud rather than individual items (e.g. coin-catalog instead of coin-data). Moreover, many entries had to be hyphenated to prevent separation of words by the word cloud generation tool WordItOut.

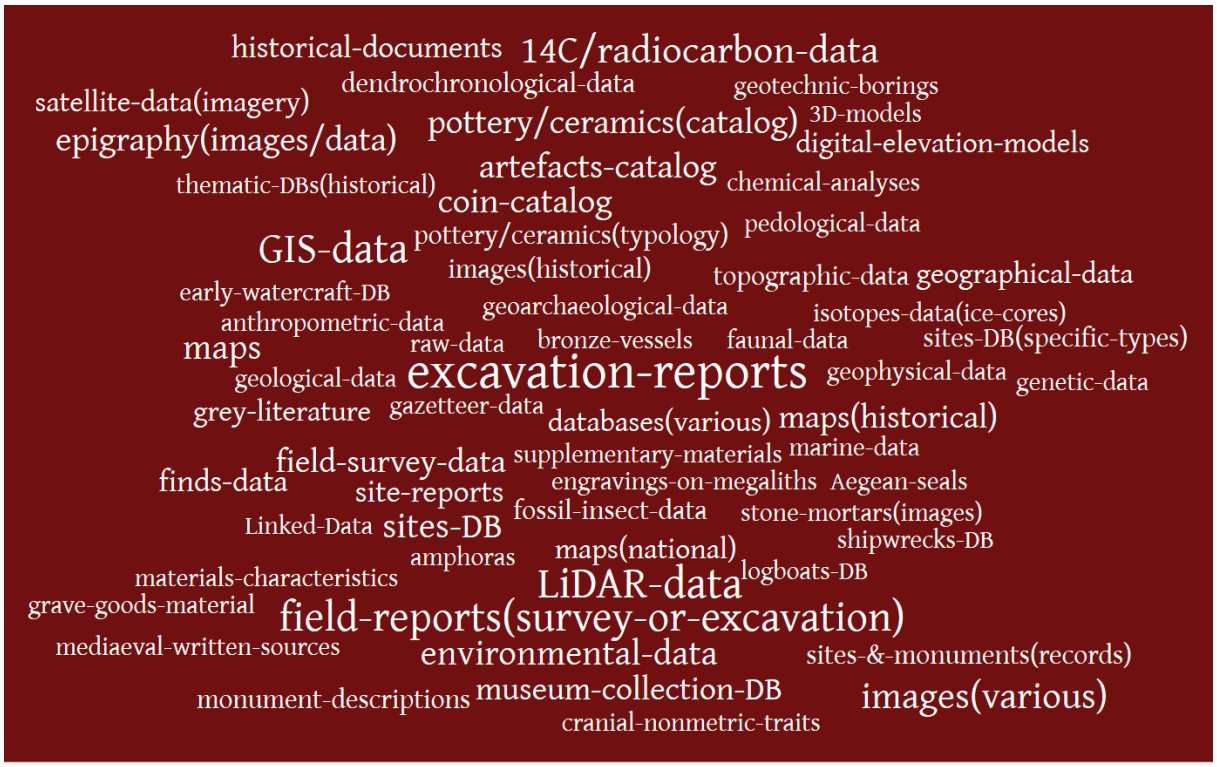

***Figure 7**: Word cloud of 50+ types of data, data sources and specific research objects.*

Finally, around 250 mentions of different data which 188 respondents had (re)used were reduced to a more digestable overview of around 50 types of data or data sources. In the word cloud were also some specific objects or parts thereof included, e.g. Aegean seals, amphoras, bronze vessels, grave foods material, stone mortars, on which respondents (re)used information or data. Each of these has been mentioned only by one respondent. Some other single mentions are addressed after a closer look into the large groups of data types or sources.

Groups of data types or sources

- *Fieldwork reports/data:* This group comprises excavation-reports (18), field-reports (survey-or-excavation) (12) and field-survey-data (6). Respondents often mentioned fieldwork reports without distinguishing between survey and excavation reports. Overall excavation reports are certainly one of the most often consulted sources because these usually contain information for different subject experts.
- *Maps and LiDAR data:* what was not immediately clear from the word cloud was the importance of maps for archaeological researchers. Respondents mentioned historical maps (5), national maps (2) and maps generally (8), the latter two often together with LiDAR data (14) and/or digital-elevation-models (3). Cartographic material and LiDAR data are provided by national mapping agencies; LiDAR data and satellite imagery (3 mentions) are sourced from public or commercial remote sensing organisations.
- *GIS and databases of archaeological sites:* Several respondents mentioned GIS data (12) without making clear if from a national/regional service or archaeological projects. Also often unclear remained mentions of databases of sites generally (8) or specific site types (2). Records of sites & monuments appeared only twice, however very likely some of the sites databases are national registries.
- *Cultural artefacts:* Most often mentioned were pottery/ceramics catalogs (9), epigraphical images/data (7) and coin catalogs (7). Furthermore, artefact catalogs (7), and museum collection databases (5). Most of the various mentions of images (9) also relate to cultural artefacts.
- *14C/radiocarbon data:* 14C data were mentioned 10 times; dendrochronological data by two respondents.
- *Environmental data:* This large group of data types was mentioned by eight respondents generically, while also paleoenvironmental data (e.g. fossil insect data, faunal data), pedological data (soils), and isotope data from ice-cores can be included here.

Among the entries in the word cloud are 29 which were mentioned only by one respondent. Some examples among those not already covered are: the large group of marine data; geoarchaeological data (general), geological and geophysical data; chemical analyses and characteristics of materials; genetic data and anthropometric data. Also Linked Data was mentioned explicitely only by one respondent, as were 3D models.

This does of course not mean that these types of data are less important. The reference sample is 188 statements of survey respondents who said that in the last 2 years they (their research group) used data which other researchers had made available through a publicly accessible digital repository or database.

4.7.6 Overview of data sources

Listed are 96 sources from which respondents (re)used data for their research or other purposes. The list does not include journals or platforms for sharing articles. Mentioned platforms are HAL - Hyper Articles en Ligne (2 mentions), Hrčak, the Croatian scientific and professional journals portal (2), ResearchGate (3), and Academia.edu (7). The data resources are listed in alphabetical order and the descriptions include the number of mentions (if more than one) and examples of data or data sources where mentioned by respondents. If differently named data sources from one repository or information system were distinguished, the list would increase to well over 100 sources:

- Actueel Hoogtebestand Nederland, 2 mentions [LiDAR data, national elevation data]

- AMČR - Archeological Map of Czech Republic, 9 mentions (examples: overview of sites, GIS projects of Historical Landscape)
- Archaeological Map of Bulgaria - National Archaeological Information System, 9 mentions [geographic distribution of sites, fieldwork reports]
- Archaeological Survey of Ireland
- Archaeology Data Service - ADS (7 mentions) (examples: GIS data, survey and excavation data, 3D models, Linked Data, several sources)
- ARIADNE, 4 mentions [data accessible through ARIADNE used for research purposes, environmental data)
- Artefacts : Encyclopédie en ligne des petits objets archéologiques, 2 mentions [small finds data]
- ARUP-CAS, Institute of Archaeology, Czech Republic [digital archive]
- Banque du Sous-Sol du Bureau de Recherches Géologiques et Minières, BRGM, France
- Bayerischen Staatsbibliothek München, Monumenta Germaniae Historica [mediaeval written sources]
- BRGM, the French geological survey [InfoTerre: geological database]
- British Museum [coin collection online]
- British School at Rome [epigraphy images]
- BugsCEP Coleopteran Ecology Package [fossil insect data]
- c14.sk - Archaeological Chronometry in Slovakia
- CalPal - Radiocarbon Calibration Online [used for 14C data]
- Central Institute for Catalogue and Documentation, Italy [artefact catalogue]
- Centre for Medieval Studies, Prag, Czech Republic [mediaeval written sources online]
- Citeres - Cités, Territoires, Environnement et Sociétés, Université de Tours, France
- Coinage of the Roman Republic Online - CRRO [numismatic collections data]
- Corpus der Minoischen und Mykenischen Siegel online
- Corpus Inscriptionum Latinarum [epigraphy images]
- CPAT Regional Historic Enivronment Record (Clwyd-Powys Archaeological Trust, UK)
- Cranial Nonmetric Traits Database, Queen's University
- CZ_RETRO (Association for the Renewal of the Village and the Small Town, Czech Republic) [settlements database]
- DAI Arachne [information about Aegean Seals]
- Danish Geodata Agency [national maps and LiDAR data]
- DANS, dendro.dans.knaw.nl (national dendrochronological data)
- DANS, e-depot for Dutch Archaeology, 4 mentions [e.g. Zakynthos Archaeology Project dataset; excavation reports and underlying data]
- Domitilla Catacomb Project Database [topographic data]
- Early Watercraft, Computer Vision Laboratory, University of Ljubljana, Slovenia [fluvial ports]

- Epigraphic Database Roma
- European Soil Database [pedological maps]
- European Space Agency (satellite imagery)
- Europeana, 2 mentions [museum collections, WW1 materials]
- FASTI Online [sites information, excavation reports]
- Fund og Fortidminder, Denmark, national finds & monuments register, 2 mentions [e.g. excavation reports]
- Global Land Cover Facility - GLCF, University of Maryland (access to satellite imagery)
- Goldman Osteometric Data Set [anthropometric data]
- Heidelberg Database [epigraphy data]
- HumaNum, la TGIR des humanités numeriques [surveys of schematic engravings on megaliths]
- Hungarian National Museum, Archaeology Database, 3 mentions [sites, excavation reports]
- icéramm.fr [ceramics typology]
- INRAP, Dolia database, 4 mentions [e.g. rapports de fouilles]
- Instituto de Antropología de Córdoba, Argentina [various]
- Instituto Geográfico Nacional de la República Argentina [digital elevation models]
- Israel Antiquities Authority, Archaeological Division [repository]
- Joconde, Collections des musées de France
- Kartverket - Norwegian Mapping and Cadastre Authority [national maps and LiDAR data]
- Kramerius - Digital Library of the Czech Academy of Sciences [old professional magazines]
- Library of Congress [images]
- Logboats from Europe [database of files and images]
- Long-term woodland dynamics in Central Europe project (2012-2016, project funded by the European Research Council)
- mapire.eu, 2 mentions [historic maps]
- MédiHal - Archive ouverte de photographies et d'images scientifiques (CCSD-CNRS, France)
- MEDIN - Marine Environmental Data and Information Network [marine data]
- MedRadiocarbon - Collection of radiocarbon dates from Mediterranean regions (available on GitHub)
- Ministry of Culture, Catalogue of the Listed Archaeological Sites and Monuments of Greece
- Ministry of Culture, Czech Republic [information on museum collections]
- Nabunken - Comprehensive Database of Archaeological Site Reports in Japan, 3 mentions [e.g. excavation reports archive]
- National Heritage Institute, Department of Archaeology, State Archaeological List of the Czech Republic
- National Institute of Heritage, National Archaeological Repository of Romania, 4 mentions [list of annual reports of excavations, excavations reports, sites spatial distribution]

- National Library of France - Gallica digital library [images]
- National Library of Scotland [historic maps]
- National Museum of Denmark [survey and excavation reports)
- National Research Centre on Human Evolution - CENIEH [repository]
- Naxos Project [a dataset made available publicly]
- Neotoma [environmental data]
- NOAA - National Oceanic and Atmospheric Administration, USA, 2 mentions [environmental geodata, ice core isotope data]
- oldmaps.geolab.cz [historic maps]
- Online Coins of the Roman Empire - OCRE [database of American and European numismatics collections]
- Open Context (Alexandria Archive Institute)
- Open Geospatial Consortium
- OpenTopography [LiDAR data]
- Portuguese Archaeological Institute, Endovelico database [information about archaeological surveys]
- Publieke Dienstverlening op de Kaart, Netherlands [national maps]
- Radiocarbon CONTEXT database, University of Cologne, Germany
- real-mta [information on archaeological sites]
- SEAD - Strategic Environmental Archaeology Database, Sweden [fossil insect data]
- Slovenian Environment Agency (ARSO), Geoportal [LiDAR data]
- Smithsonian Institution [images]
- Social Archaeology of Early Iron Age and Early Archaic Greece project (University of Thessaly, Volos, Greece)
- Swedish Rock Art Research Archives (SHFA)
- tDAR – The Archaeological Record, 2 mentions [ceramic data, faunal data]
- Ubi erat lupa [epigraphy images]
- UK Ordnance Survey [maps, LiDAR data]
- UK Portable Antiquities Scheme
- United States Geological Survey (USGS), 2 mentions [satellite imagery, digital elevation models]
- University of Florence, DBAS@egeanLab, Aegean Collections of the National Archaeological Museum of Florence
- University of Oxford, Beazley Archive, 2 mentions
- University of Oxford, Oxford Economic Project, 2 mentions [Coin Hoards of the Roman Empire, Shipwrecks database]
- University of Tokyo, Historiographical Institute, Database on Japanese Ancient Documents
- Urban Brussels [city administration: excavation reports]

- US Epigraphy Project (Center For Digital Scholarship, Brown University, USA)
- ZRC-SAZU - Institut of Archaeology, Arkas database [archaeological sites in Slovenia]
- ZRC-SAZU - Institut of Archaeology, Zbiva database [information about grave goods materials]

4.7.7 Summary and suggestions

Summary of main results

Why reuse is important

Sharing data is important but without (re)use the benefits associated with open data sharing would not materialise. There are many good arguments for making data available, for instance, that reported research results can be scrutinized and duplicative data collection prevented. Particularly strong however is the argument that reuse of data, for example to investigate new research questions, allows exploitation of previous investment. Preserved data that is being reused gains in value, otherwise it might be perceived only as a cost factor.

"Return on investment" expected by research funders explains much of the increasing pressure on researchers to share their data from publicly funded research for reuse. It is also very important for repositories to document not only downloads but actual reuse. Metadata with rich context information is essential for reusing data, as is a license that clearly states what users are allowed to do with the data.

Results for reuse

Results of the ARIADNEplus survey confirm that archaeological researchers often (re)use available data and allow some insights about what and how. The survey participants were asked, *"Did you / your research group in the last 2 years use any data which other researchers made available through a publicly accessible digital repository or databases?".* An astonishing number of 220 respondents said they did and also briefly described the data types and/or the sources.

In comments 34 respondents also gave reasons why they did not (re)use other's data. Most said relevant data was not available or posed some problems, difficulty to access or use, missing licensing information, lack of support, among others. Some also said that they did not need data from other researchers.

Main purposes of data reuse

The 220 respondents were asked "*What was the main purpose of the data reuse?"*, and three predefined purposes and the option "Other - please specify" offered. Building a database for the research community was a purpose for 31%, comparison to own results for 55%, and use together with own research data for 63% (multiple answers were possible). Few mentioned other purposes for the data (re)use, for example, to use it as test data (e.g. "test algorithms and approaches") or as a conceptual resource (e.g. "typology terms", "data structure").

Patterns

188 respondents provided descriptions of data and/or sources used, often only one or two data types (e.g. *"14C data"*, *"genetic data"*, *"find data"*, *"coins and ceramics"*), one source, e.g. *"AMČR"* [Archeological Map of Czech Republic] or "Archaeological Map of Bulgaria", or general statements such as *"several sources from the ADS (U York)".*

An analysis of descriptions including two or more data types or sources suggests three patterns of (re)use of available with own data which could be quite frequent in archaeology. Most often reused together were

- Information on sites (maps, distribution) and fieldwork reports (e.g. *"Site and monuments database, online archaeological reports"*),
- Cartographic, GIS and LiDAR data (e.g. *"National LiDAR and cartography"*; *"GIS, LiDAR, descriptions of archaeological monuments"*),
- Databases/catalogs of different artefacts (e.g. *"Coin and ceramic catalogs"*; *"Databases of inscriptions and coins"*).

The report chapter on data reuse includes an overview of groups of data types or sources; a word cloud of over 50 types of data, data sources and specific research objects; and a list of 96 named sources from which respondents sourced data.

Suggestions for ARIADNEplus

- ARIADNEplus should promote and support data reuse so that the investment in the collection of archaeological data can be exploited further for research, education and other purposes. Ways to enable with the ARIADNEplus infrastructure and services easy and effective reuse of data should be investigated.
- Different purposes and forms of data reuse should be considered to understand better actual practices of data reuse in archaeology so that these can be supported effectively.

4.8 Readiness to share data

4.8.1 Background

Survey respondents were asked if in the last two years they or their research group used data other researchers made available through a publicly accessible digital repository or databases. A high percentage of respondents who answered the question (58.5%) said that they did so. Many respondents also described the source and/or type of the data they (re)used. This result came as a surprise because the answers and comments regarding barriers to share data painted a darker picture of data sharing in the archaeological sector (see *Section 4.6*).

In the survey one question on the readiness of archaeologists to share data through publicly accessible repositories or databases was placed after the questions on data reuse. The idea was to relate respondents' evaluation of the readiness to the level of recent reuse of available data and other results of the survey. The survey participants were asked if they thought that the readiness increased or did not increase in the last five years. Like the result for reuse of data the one on readiness shows an overall brighter horizon for open data sharing in archaeology.

In the survey template "readiness" was intentionally not defined to capture respondents feelings in this regard. The general understanding of readiness is that it comprises both willingness and capability to do something, in our case to share data through publicly accessible repositories or databases. Regarding willingness to do this, important factors are whether or not there is a mandate to do it, if peers do it, and perception of own benefit; capability depends on own skills, training, available support, among others.

4.8.2 Survey results

Do you agree to the following statement: "In the last 5 years the readiness of archaeologists to share data through publicly accessible repositories or databases increased"?

Yes	313	83.2%
No	62	16.5%
No answer	1	0.3%
	376	100%

***Table 13**: Increase of the readiness to share data in the last 5 years.*

The percentage for "yes" of 83.2% shows a strong agreement. Of the 62 respondents (16.5%) who said "no" one did not answer the question, but commented *"No. The cause is the advancement of digital technology, not the process of publishing scientific results"*. The respondent is added to those who said "no". One respondent did not answer the question but continued to fill the questionnaire.

The next table allows the level of agreement to this question to be related to other results of the survey. In the ARIADNE 2013 survey it was around 50% and this year around 65% of respondents who answered the question said that they make data available through an accessible repository in all/most, many or at least a few projects. In this survey, June–August 2019, 58.5% also said that in the last two years they used data which other researchers shared through publicly accessible repositories or databases. Furthermore, 83.2% of respondents thought that in the last 5 years the readiness of archaeologists to make data available in this way increased.

Makes own project data available in an accessible repository Yes: = from all/most to at least a few projects; 2019: N=437-451, 2013: N=516-521		
	2013	2019
Yes	50%	65%
No	50%	35%
N/A	--	--

In the last 2 years used data made publicly available by others (accessible repository/DB) N=376	
2019	
Yes	58.5%
No	41.0%
N/A	0.5%

In the last 5 years the readiness to make data publicly available increased (accessible repository/DB) N=376	
2019	
Yes	83.2%
No	16.5%
N/A	0.3%

***Table 14**: Comparison of data sharing (2013/2019, data (re)use, and increase of readiness to share.*

These survey results suggest that the prospects for open data sharing in archaeology look good. But several of the respondents who said "yes", there was an increase in readiness, added comments that currently the situation is not as good as one might have expected.

Comments received

38 respondents added a comment to their evaluation, 28 who said "yes", 10 who said "no". Several respondents thought that there is now a higher awareness that data should be made available than five years ago but not much increase in readiness to do so. Furthermore, respondents felt that the increase takes place only slowly. More has to be done to foster data sharing:

- *Awareness has grown, even if nothing much is being done to increase the actual sharing...*
- *There is a higher awareness that data should be shared.*
- *Globaly speaking, I think that's the case. But there's still a long way to go for many archaeologists*
- *I witness this trend through several COST networking projects*
- *It increases but not enough. Still fear to have the data stolen.*
- *It did not increase much, though.*
- *To a limited extent yes*
- *But very slowly and uncomplete*
- *But only marginally*
- *[Yes] Although this has had little practical consequence so far...*
- *[Yes] However, there are still a few results.*
- *[No] Theoretically yes, practically no*

Three respondents said "yes" (it increased) but were not fully sure, for example, *"probably, I don't know certainly..."*. One said "no" but was also not sure.

Some respondents described the state of affairs at the national level:

- *Not necessarily in France, but clearly in Belgium (my homeland) & elsewhere*
- *[No] definitely not locally* (a response from Malta)
- *Probably yes, especially in the system of Archaeological map of Czech Republic*

- *In Sweden mostly due to the OA to research data the government has decided on, which shall be in action no later than 2026.*
- *It is slowly penetrating even in Italy*
- *In Hungary there is also an increase but very slowly. There are cases when archaeologists (usually younger generation) would provide their documentations to our database but their boss does not let them.*
- *In Romania more archaeologists are interested to publish data in the National Archaeological Repository, not necessarily with the aim to share their knowledge but for the reason of protection the archaeological sites of the looters. The archaeological sites recorded in the National Archaeological Repository can not be the subject of an activity of metal detection.*

Comments which addressed the situation in their research domain or the archaeological research community in general were:

- *[Yes] This is particularly so for the numismatic community*
- *Yes, but not too much, the archaeological research groups have no culture of open access to data, unlike other research groups, as biology or microbiology or DNA data.*
- *The Marwick et al paper has helped the profile tremendously and is being cited widely [Ben Marwick et al., Open Science in Archaeology, in: The SAA Archaeological Record, September 2017]*

Respondents who said “no” gave the following reasons:

- *The awareness of what is required to make data available for sharing has increased but the concerns and hesitation to share still lingers.*
- *Unfortunately, there is still a lot of protectionism and competition.*
- *There is still a reluctance to share data amongst many archaeologists – particularly academics.*
- *The data sets themselves are still being viewed as proper scientific work rather then what we actually do with that data.*

But also respondents who said “yes” mentioned issues that needed to be addressed:

- *It still is a little bit difficult to explain to researchers why this is important... until the day they see the enormous advantage within their own research, often for needs of comparison*
- *Even though archaeologists are likely to share their published data, some of them have concerns about unfair use of their data.*
- *Within reason and depending on the source. There is still a tendency to complete your own research first, publish it and then release data afterwards. There have also been a number of cases of academics using online accessible data and not acknowledging the source, which have resulted in cautionary tales from the archaeologists involved.*
- *Students and future researchers don’t learn new or “best” practices enough. It is high time to make further progress.*
- *As cited in literature, there is an age/profile difference - early career and close to retirements researchers share most. Mid-career less so.*

Technology: Respondents also considered the role digital technology play in increasing readiness

- One respondent saw no increase in the last 5 years but a positive effect of technology, "*The cause is the advancement of digital technology, not the process of publishing scientific results".* Another respondent felt that there was an increase also highlighted the role of technology, "*Think, that it depends on technical possibilities too (they have increased in the last 5 years)".*
- One elaborated comment addressed issues in the application of technology, particularly regarding a productive collaboration between technical and domain experts: *"Interest has increased but the pratices are often weak or awkward. Except for fields deeply linked with computing (Geomatics...), people don't change easily their practices and are not so helped (lack of IT services); especially concerning databases or repositories on Cultural Heritage or on excavations. Funding is rather scarce in SHS and in a relatively short term for archaelogy. Their is also a confusion between information and computer specialists that leads to unsuitable applications; also between a domain specialist and an information specialists who don't really know or understand the domain: inappropriate terms/concept in thesauris or unsuitable modeling (problems with human resources management). Many applications are made 'to be made' but not for a long-term use".*

4.9 ARIADNEplus data search & access services

4.9.1 Background

This section reports results for three closely related survey questions on data search and access. The survey respondents were asked to evaluate (1) the online availability of different types of archaeological data, and (2) how helpful it would be to discover and access via the ARIADNEplus portal (2.1) integrated data collections or datasets, and (2.2) integrated data items of the different types of data.

The questions 2.1 and 2.2 concern the difference between so-called "collection-level" and "item-level" data access. The difference can be explained looking into the accessible ARIADNE datasets: At present over 1.9 million data records are integrated in the ARIADNE catalogue and portal. These provide access to about 3.7 million data items, because in many cases one record describes and directs the portal user to data sets of hundreds or thousands of items of fieldwork archives, artefact databases, entries of scientific databases, including dendrochronology data, for instance. Thus there are data collections from which each item can be found directly on the portal while in other cases only indirectly by following a link in the record of the collection (or database) served by the portal.

This difference between item-level access and collection-level access is due to the technical setup of some data collections which make it difficult to provide records of single items. In other cases it is preferable to provide access at a higher level, e.g. the description of a collection or database in a repository, rather than individual items without contextual information. Therefore, in ARIADNEplus for each new collection or update of existing data collections in the ARIADNE catalogue, the best integration approach will be defined taking account of the content and technical setup of the collection.

4.9.2 Mapping of types of data

ARIADNEplus will incorporate data from a wider range of archaeological research domains, and aims to integrate more datasets at item-level to provide advanced semantic data search for finding data items based on semantically defined relations. Additional data types include research data from environmental archaeology, maritime and underwater archaeology, bio-archaeology, inorganic materials studies, epigraphy, among others. This requires standardised description of records of

different types of data by the providers, based on application profiles for data records jointly developed by domain researchers, data managers and vocabulary/terminology experts.

The data types listed in the survey questions generally correspond to the types of datasets of project partners will be integrating in the ARIADNEplus dataset catalogue and portal. But for the survey in some cases a more general or detailed description was used. Table 15 presents the mapping of the lists of data types.

Survey list of data types	***ARIADNEplus list of data types***
Satellite or airborne remote sensing data (e.g. LiDAR)	*Remote sensing data (general)*
Sites & monuments databases or inventories	*Spatio-temporal data for sites & monuments* *Monuments and sites inventories* *Standing Structures*
National GIS data and maps	*GIS data (general)*
Environmental archaeology datasets	*Environmental archaeology* *Palaeo-environments*
Maritime and underwater archaeology data	*Maritime and underwater archaeology*
Field survey/prospection data	*Fieldwork: prospection/survey, incl. metal detector surveys*
Excavations data (e.g. excavation archive)	*Fieldwork: excavations*
Unpublished fieldwork reports (‘grey literature’)	*Fieldwork reports (‘grey literature’)*
Radiocarbon, dendrochronology and other dating data	*Dating (different methods)*
Scientific data/analysis of inorganic remains	*Inorganic materials studies*
Scientific data/analysis of biological remains	*Bio-archaeology* *Human palaeo-biology and anthropology* *Ancient DNA*
Artefact/finds databases or image collections	*Archaeological finds (e.g. museum collections)*
Inscriptions, coins or other special databases	*Inscriptions, coins or other special databases*

***Table 15**: Mapping of data types in the survey and ARIADNEplus data types.*

4.9.3 Results: Online accessibility of different types of data

The survey participants were asked to rate the online accessibility of the different types of data which they (their research group) may need to prepare to carry out a project. The results are presented in Table 16.2. The types of data are listed as in the survey questionnaire, not according to the rating of their accessibility. The rating is indicated as follows:

Rating of accessibility	***Colour***	***Percentage of "very good" and "good" (actual)***
High		53.8% – 65.4%
Middle		39.3% – 48.3%
Low		22.3% – 36.2%

***Table 16.1**: Scheme for the evaluated accessibility of different types of data.*

How would you rate the online accessibility of these types of data, which you/your research group may need to prepare or carry out a project? (1) very good, (2) good, (3) fair, (4) poor:

	N	Very good	Good	Fair	Poor
Satellite or airborne remote sensing data (e.g. LiDAR)	355	**100** (28.2%)	**91** (25.6%)	**79** (22.3%)	**85** (23.9%)
Sites and monuments databases or inventories	367	**112** (30.5%)	**106** (28.9%)	**96** (26.2%)	**53** (14.4%)
National GIS data and maps	364	**133** (36.5%)	**105** (28.8%)	**78** (21.4%)	**48** (13.2%)
Environmental archaeology datasets	356	**55** (15.4%)	**74** (20.8%)	**128** (36.0%)	**99** (27.8%)
Maritime and underwater archaeology data	336	**24** (7.1%)	**51** (15.2%)	**104** (31.0%)	**157** (46.7%)
Field survey/prospection data	357	**67** (18.8%)	**87** (24.4%)	**118** (33.0%)	**85** (23.8%)
Excavations data (e.g. excavation archive)	362	**80** (22.1%)	**93** (25.7%)	**101** (27.9%)	**88** (24.3%)
Unpublished fieldwork reports ('grey literature')	364	**70** (19.2%)	**73** (20.1%)	**92** (25.3%)	**129** (35.4%)
Radiocarbon, dendrochronology and other dating data	353	**46** (13.0%)	**67** (19.0%)	**134** (38.0%)	**106** (30.0%)
Scientific data/analysis of inorganic remains	342	**32** (9.4%)	**64** (18.7%)	**115** (33.6%)	**131** (38.3%)
Scientific data/analysis of biological remains	350	**31** (8.9%)	**71** (20.3%)	**123** (35.1%)	**125** (35.7%)

Artefact/finds databases or image collections	362	**76** (21.0%)	**99** (27.3%)	**105** (29.0%)	**82** (22.7%)
Inscriptions, coins or other special databases	345	**52** (15.1%)	**92** (26.7%)	**117** (33.9%)	**84** (24.3%)

***Table 16.2**: Rating of the online accessibility of different types of data (N = 336-367).*

Rating of online accessibility

Based on the ratings, three groups of data types, each with two data types or sources, can be distinguished. The basis for this clustering is the percentage of "very good" + "good" rating of the accessibility for the different data types. In the presentation that follows, this percentage is given (the percentage for "very good" is added in brackets).

(1) Rating of accessibility – High

In this group are (1.1) National GIS data & maps and Satellite or airborne remote sensing data, and (1.2) Sites and monuments databases or inventories.

- (1.1) *National GIS data and maps:* 65.4% (36.5% "very good"); *Satellite or airborne remote sensing data (e.g. LiDAR):* 53.8% (28.2%). – These are data which archaeologists do not produce themselves. Maps and LiDAR data they can get from national agencies (e.g. Danish Geodata Agency or Kartverket in Norway), satellite data and imagery from the European Space Agency and other providers.
- (1.2) *Sites and monuments databases or inventories:* 59.4% (30.5%). – Country-wide inventories/databases are usually maintained by national heritage authorities, while other databases of sites are developed by projects for their research questions.

(2) Rating of accessibility – Middle

In this group are (2.1) data and documentation from fieldwork, and (2.2) databases and catalogs of various artefacts.

- (2.1) *Excavation data (e.g. excavation archive):* 47.8% (22.1%), *Field survey/prospection data:* 43.2% (18.8%), *Unpublished fieldwork reports ('grey literature'):* 39.3% (19.2%). – These archaeologists can search in online databases/repositories of national heritage authories, archaeological data repositories (e.g. ADS in the UK, eDNA in the Netherlands), on websites such as Fasti Online (fieldwork reports), or websites of archaeological projects.
- (2.2) *Artefact/finds databases or image collections:* 48.3% (21%) and inscriptions, coins or other special databases: 41.7% (15.1%). – The rating of these data types and sources is also good because there are many online catalogs and databases of museum and other special collections as well as sources such as Artefacts (small finds in Europe) or the Portable Antiquities Scheme in the UK. The epigraphy research community is particularly active in building databases of inscriptions.

(3) Rating of accessibility – Low

In this group are (3.1) Dating and scientific analysis data, and (3.2) Environmental archaeology data and maritime & underwater archaeology data. These have a low percentage of "very good" and "good" combined (22.3–36.2%). Except for environmental data, also the share of "very good" is the smaller part of it, e.g. Scientific data/analysis of inorganic remains 9.4% of 28.1%, biological remains 8.9% of 29.2%. For most of the data types in the other groups the shares of "very good" and "good" are roughly equal.

- (3.1) Dating and scientific analysis data: *Radiocarbon, dendrochronology and other dating data:* 32.0% (13%); *Scientific data/analysis of biological remains:* 29.2% (8.9%); *Scientific data/analysis of inorganic remains:* 28.1% (9.4%). – It appears that these fields of research have a lot of work ahead to be on a par with others regarding the accessibility of their data.
- (3.2) *Environmental archaeology data:* 36.2% (15.4%), *maritime and underwater archaeology data:* 22.3% (7.1%). – These are not single data types, but fields of research. Respondents from the domain of maritime and underwater archaeology are very likely under-represented in the survey sample, therefore the online accessibility of data in this domain may not be adequately represented.

Comparison 2013/2019

For six data types it is possible to compare the results of the ARIADNE/plus 2013 and 2019 surveys. Others could not be included in the comparison because their description was changed significantly (e.g. the too general "GIS data" became "National GIS data and maps"). Other data types were not included in the 2019 survey because the methods to process them were clearly not relevant for most archaeologists (e.g. model-based computing and data mining[33]), while more relevant new ones needed to be include in the list.

ARIADNE survey 2013 *How would you rate the online accessibility of these types of data?* N = 520-540	*Very or Rather good*
Satellite & airborne remote sensing data	46.2%
Prospection & field survey data	34.8%
Excavation data	41.9%
Radiocarbon & dendrochronology data	35.3%
Data from material & biological analysis	34.0%

ARIADNEplus survey 2019 *How would you rate the online accessibility of these types of data?* N = 336-367	*Very or Rather good*
Satellite or airborne remote sensing data (e.g. LiDAR)	53.8%
Field survey/prospection data	43.2%
Excavations data (e.g. excavation archive)	47.8%
Radiocarbon, dendrochronology and other dating data	32.0%
Scientific data/analysis of inorganic remains	28.1%
Scientific data/analysis of biological remains	29.2%

***Table 17**: Comparison of the rating of six data types in the ARIADNE/plus 2013 and 2019 surveys (source of the 2013 ratings: ARIADNE 2014: 82).*

The comparison shows that respondents in 2019 perceived the accessibility of some data types as better or less good than the respondents 2013:

Better than 2013:

- Satellite & airborne remote sensing data – 2013: 46.2%, 2019: 53.8%; difference: +7.6%
- Prospection & field survey data – 2013: 34.8%, 2019: 43.2%; difference: +8.4%
- Excavation data – 2013: 41.9%, 2019: 47.8%; difference: +5.9%

[33] Use of such methods require Data Science skills. In the 2019 survey of the respondents how answered the question on training needs 90.3% said training for developing Data Science skills would be very helpful or helpful, 60.7% considered it as very helpful (see *Section 4.11*).

Less good than 2013:

- Radiocarbon & dendrochronology data – 2013: 35.3%, 2019: 32%; difference: -3.3%
- Data from material & biological analysis – 2013: 34%, 2019: inorganic/material: 28.1%, biological: 29.2%, difference on average -5.35%.

The results perhaps signal that the 2019 respondents perceived that the accessibility of remote sensing data, prospection & field survey and excavation data improved, while in comparison the accessibility of dating and scientific analysis data has stagnated or even appears as worse.

4.9.4 Results: ARIADNEplus portal for data discovery & access

The ARIADNEplus data portal will allow researchers and other users discover and access a wider range of data types from datasets or collections of European and international providers, than the initial ARIADNE portal. Therefore one objective of the survey was to identify if respondents perceive support by the portal to discover and access more helpful for some data types than for others. Data types for which help is more appreciated could then be prioritised regarding mobilisation and integration in the ARIADNEplus data catalogue and portal.

The survey participants were asked to rate how helpful for their research support by the portal would be to discover and access different types of data. The same list as in the question on online accessibility was used for this investigation. The results are presented in Table 18.2. The types of data are listed as in the survey questionnaire, not according to the rating of how helpful support by the portal to discover and access them would be. The rating is indicated as follows:

Rating	***Colour***	***Percentage of "Very helpful" and "helpful" (actual)***
High		90.4% – 93.2%
Middle		84% – 89.2%
Low		57.8% – 76.5%

Table 18.1*: Scheme for the evaluated helpfulness of portal support for discovering and accessing datasets or collections of different types of data ("collection-level access").*

Survey results

The ARIADNEplus portal will support discovery and access of a range of archaeological data types from national and international providers. To what extent would your research/the research of your organisation benefit from being able to access datasets or collections of the following types: would this be (1) very helpful, (2) helpful, (3) less helpful, or (4) not helpful for your work?:

	N	Very helpful	Helpful	Less helpful	Not helpful
Satellite or airborne remote sensing data (e.g. LiDAR)	351	**253** (72.1%)	**60** (17.1%)	**26** (7.4%)	**12** (3.4%)
Sites and monuments databases or inventories	355	**257** (72.4%)	**74** (20.8%)	**20** (5.6%)	**4** (1.1%)
National GIS data and maps	355	**258** (72.7%)	**72** (20.3%)	**21** (5.9%)	**4** (1.1%)
Environmental archaeology datasets	347	**218** (62.8%)	**86** (24.8%)	**29** (8.4%)	**14** (4.0%)
Maritime and underwater archaeology data	341	**120** (35.2%)	**77** (22.6%)	**72** (21.1%)	**72** (21.1%)
Field survey/prospection data	350	**222** (63.4%)	**85** (24.3%)	**35** (10.0%)	**8** (2.3%)
Excavation data (e.g. excavation archive)	353	**251** (71.1%)	**72** (20.4%)	**22** (6.2%)	**8** (2.3%)
Unpublished fieldwork reports ('grey literature')	353	**228** (64.6%)	**83** (23.5%)	**36** (10.2%)	**6** (1.7%)
Radiocarbon, dendrochronology and other dating data	349	**213** (61.0%)	**80** (22.9%)	**40** (11.5%)	**16** (4.6%)
Scientific data/analysis of inorganic remains	345	**171** (49.6%)	**93** (26.9%)	**59** (17.1%)	**22** (6.4%)
Scientific data/analysis of biological remains	348	**166** (47.7%)	**99** (28.4%)	**59** (17.0%)	**24** (6.9%)
Artefact/finds databases or image collections	353	**237** (67.1%)	**82** (23.2%)	**27** (7.6%)	**7** (2.0%)
Inscriptions, coins or other special databases	347	**186** (53.6%)	**75** (21.6%)	**60** (17.3%)	**26** (7.5%)

***Table 18.2**: Results for collection/dataset-level discovery & access on the ARIADNEplus portal for different types of data (N = 341-355).*

Comparison to the rated accessibility of data types

The results of a comparison of the rated online accessibility of the data types and helpfulness of portal support for discovering and accessing datasets or collections of such data are:

- Help would be highly appreciated for all data types, except maritime and underwater archaeology data (for which there was no large "lobby group"among the respondents).
- Surprisingly, the appreciation of support was lower for data types for which the accessibility was evaluated as insufficient. Instead many respondents would appreciate more support for already more accessible data types.
- An explanation for this pattern could be that, like in the case of maritime and underwater archaeology data, but not as pronounced, there were fewer respondents interested in better access to these data types.
- While this may explain the pattern of response, a mapping of the responses for accessibility and support for discovery and access reveals some interesting cases in-between (see Table 19).

		Discovery & access via the ARIADNEplus portal: Collection-Level (C-L) Rating of the helpfulness of support by the ARIADNEplus portal to discover and access datasets or collections of the data types Rating: % of Very helpful + Helpful (N = 341-355)		
		High	**Medium**	**Low**
Online Accessibility (OA) Rating of the online accessibility of the data types Rating: % of Very good + Good (N = 336-367)	**High**	Sites and monuments databases or inventories [OA:59.4%/C-L:93.2%] National GIS data & maps [OA:65.4%/C-L:93%]	Satellite or airborne remote sensing data (e.g. LiDAR) [OA:53.8%/C-L:89.2%]	
	Medium	Excavation data (e.g. excavation archive) [OA:47.8%/C-L:91.5%] Artefact/finds databases or image collections [OA:48.3%/C-L:90.4%]	Unpublished fieldwork reports ('grey literature') [OA:39.3%/C-L:88.1%] Field survey/ prospection data [OA:43.2%/C-L:87.7%]	Inscriptions, coins or other special databases [OA:41.7%/C-L:75.2%]
	Low		Environmental archaeology datasets [OA:36.2%/C-L:87.6%] Radiocarbon, dendrochronology and other dating data [OA:32%/C-L:84%]	Scientific data/analysis of biological remains [OA:29.2%/C-L:76.1%] Scientific data/analysis of inorganic remains [OA:28.1%/C-L:76.5%] Maritime and underwater archaeology data [OA:22.3%/C-L:57.8%]

***Table 19**: Mapping of the ratings of online accessibility of data types and portal support to discover and access datasets/collections of such data.*

Prioritisation of data types

- Respondents appreciated most portal support for discovery and access or Sites and monuments databases or inventories, National GIS data & maps, and Satellite or airborne remote sensing data (e.g. LiDAR). However, the online accessibility of these data types was rated much better than of other data types.
- The rationale for ARIADNEplus could hardly be to prioritise support for data types which are already much better accessible than others. The fact that the better accessible types are being provided by national mapping and heritage authorities adds to considering prioritisation of other data types.
- Clear candidates are the other data types with high or medium appreciation of portal support and currently only medium or low online accessibility. Ranked according to the appreciation of support (high before medium) and level of accessibility (low before medium) these are:
    - Excavations data (e.g. excavation archive) [OA:47.8%/C-L:91.5%]
    - Artefact/finds databases or image collections [OA:48.3%/C-L:90.4%]
    - Radiocarbon, dendrochronology and other dating data [OA:32%/C-L:84%]
    - Environmental archaeology datasets [OA:36.2%/C-L:87.6%]
    - Unpublished fieldwork reports ('grey literature') [OA:39.3%/C-L:88.1%]
    - Field survey/prospection data [OA:43.2%/C-L:87.7%]
- Taking account of under-rating by respondents of data types of domains with fewer survey participants, fairness would demand the following as the next candidates to prioritise:
    - Maritime and underwater archaeology data [OA:22.3%/C-L:57.8%]
    - Scientific data/analysis of inorganic remains [OA:28.1%/C-L:76.5%]
    - Scientific data/analysis of biological remains [OA:29.2%/C-L:76.1%]
    - Inscriptions, coins or other special databases [OA:41.7%/C-L:75.2%]

Some of these data types may also require more FAIRness to enable their integration in the ARIADNEplus dataset catalogue and portal.

4.9.5 Results: ARIADNEplus item-level data search & access

ARIADNEplus aims to integrate a wider range of data types from different archaeological domains of research, and integrate them as far as possible at item-level to provide advanced semantic data search for finding data items based on semantically defined relations between them. Therefore one objective of the survey was to investigate if respondents perceive portal support of item-level discovery and access more helpful for items of some data types than for others.

The survey participants were asked to rate for which data types they would find it helpful for their research to be able to search items within datasets integrated from multiple sources. The same list as in the previous question on support for data discovery and access at dataset/collection-level was used for this investigation. The results are presented in Table 20.2. The types of data are listed as in the survey questionnaire, not according to the rating of how helpful support for item-level search by the portal would be. The rating is indicated as follows:

Rating	***Colour***	***Percentage of "Very helpful" and "helpful" (actual)***
High		88.3% – 95.7%
Middle		83.1% – 85.8%
Low		58.7% – 79.5%

Table 20.1*: Scheme for the evaluated helpfulness of portal support for searching and accessing data items within datasets integrated from multiple sources ("item-level access").*

Survey results

The ARIADNEplus portal will also investigate where it might be useful to provide online integrated access to data sets at a more granular (or "item") level. Therefore we want to know where you would find it useful to be able to search online within datasets integrated from multiple sources? To what extent would your research benefit from such integrated access to the following data: would this be:

	N	Very helpful	Helpful	Less helpful	Not helpful
Satellite or airborne remote sensing data (e.g. LiDAR)	332	**200** (60.2%)	**76** (22.9%)	**36** (10.8%)	**20** (6.0%)
Sites and monuments databases or inventories	335	**233** (69.5%)	**75** (22.4%)	**20** (6.0%)	**7** (2.1%)
National GIS data and maps	332	**219** (66%)	**74** (22.3%)	**30** (9.0%)	**9** (2.7%)
Environmental archaeology datasets	324	**177** (54.6%)	**99** (30.6%)	**31** (9.6%)	**17** (5.2%)
Maritime and underwater archaeology data	317	**96** (30.3%)	**90** (28.4%)	**69** (21.8%)	**62** (19.6%)
Field survey/prospection data	325	**198** (60.9%)	**81** (24.9%)	**37** (11.4%)	**9** (2.8%)
Excavations data (e.g. excavation archive)	333	**230** (69.1%)	**74** (22.2%)	**22** (6.6%)	**7** (2.1%)
Unpublished fieldwork reports ('grey literature')	329	**205** (62.3%)	**84** (25.5%)	**31** (9.4%)	**9** (2.7%)
Radiocarbon, dendrochronology and other dating data	324	**184** (56.8%)	**87** (26.9%)	**36** (11.1%)	**17** (5.2%)
Scientific data/analysis of inorganic remains	322	**145** (45%)	**111** (34.5%)	**46** (14.3%)	**20** (6.2%)
Scientific data/analysis of biological remains	323	**139** (43%)	**111** (34.4%)	**51** (15.8%)	**22** (6.8%)
Artefact/finds databases or image collections	326	**212** (65%)	**80** (24.5%)	**27** (8.3%)	**7** (2.1%)
Inscriptions, coins or other special databases	325	**177** (54.5%)	**73** (22.5%)	**55** (16.9%)	**20** (6.2%)

***Table 20.2**: Results for item-level search & access on the ARIADNEplus portal for different types of data (N = 317-335).*

The percentages for “very helpful” + “helpful” are included in Table 21, together with the ratings of the previous two questions. A comparison of the results for “collection-level” and “item-level” shows:

- The same data types were rated much lower than others, Maritime and underwater archaeology data *et al*.
- Among the top rated data types, Satellite or airborne remote sensing data (e.g. LiDAR) and National GIS data and maps were rated significantly lower, very likely because respondents could not see how such data might be searched at item-level. Such searches could be on data objects represented in GIS or shown in LiDAR imagery.
- Fieldwork reports (‘grey literature’) and Artefact/finds databases or image collections were rated significantly higher, indeed, the latter ranked on top. The other data types on the list got about the same ratings
- The rating overall shows Artefact/finds databases or image collections, Excavations data (e.g. excavation archive) and, somewhat less, Unpublished fieldwork reports (‘grey literature’) closer to the generally top-ranked data types.

	Online accessibility % of Very good and Good (N = 336-367)	**Collection-level access** % of Very helpful + Helpful (N = 341-355)	**Item-level access** % of Very helpful + Helpful (N = 317-335)	**Rating overall**
Satellite or airborne remote sensing data (e.g. LiDAR)	53.8	89.2	83.1	++
Sites and monuments databases or inventories	59.4	93.2	91.9	+++
National GIS data and maps	65.4	93.0	88.3	+++
Environmental archaeology datasets	36.2	87.6	85.2	++
Maritime and underwater archaeology data	22.3	57.8	58.7	+
Field survey/prospection data	43.2	87.7	85.8	++
Excavations data (e.g. excavation archive)	47.8	91.5	91.3	+++
Unpublished fieldwork reports ('grey literature')	39.3	88.1	87.8	+++
Radiocarbon, dendrochronology and other dating data	32.0	84.0	83.6	++
Scientific data/analysis of inorganic remains	28.1	76.5	79.5	+
Scientific data/analysis of biological remains	29.2	76.1	77.4	+
Artefact/finds databases or image collections	48.3	90.4	95.7	+++
Inscriptions, coins or other special databases	41.7	75.2	76.9	+

***Table 21**: Overview of the results for online accessibility in general and access at collection-level and item-level via the ARIADNEplus data portal*

4.9.6 Summary and suggestions

Summary of main results

This section of the report presents and discusses the results for three closely related survey questions on data search and access: online availability of the different types of archaeological data that is part of the focus of ARIADNEplus, and how helpful it would be to discover and access it via the ARIADNEplus portal at both the collection level and item level.

ARIADNEplus will incorporate data from a wider range of archaeological research domains than ARIADNE, including environmental archaeology, maritime and underwater archaeology, biological and inorganic materials studies, radiocarbon, dendrochronology and other dating data methodologies, among others. Furthermore, the project aims to integrate more datasets at item-level to provide advanced semantic data search to find data items based on semantically defined relations.

Online accessibility of data types

Survey respondents rated the current availability of the different types of data ARIADNEplus aims to mobilise and integrate into the dataset catalogue and portal. The analysis of the results showed:

- good availability: Archaeological sites and monuments data (usually provided by heritage authorities), National GIS data and maps (from mapping agencies), and Satellite or airborne remote sensing data (in Europe offered freely by the European Space Agency);
- less good availability: Data and documentation from fieldwork (excavation, field survey/ prospection, fieldwork reports), and Databases and catalogs of various artefacts (e.g. museum collections);
- poor availability: dating data (e.g. dendrochronology, radiocarbon) and scientific data/analysis of biological and inorganic remains). Also the availability of environmental archaeology and maritime & underwater archaeology data was perceived as poor.

ARIADNEplus portal for data discovery & access

One particularly important objective of the survey was to identify if respondents perceive support by the ARIADNEplus portal to discover and access more helpful for some data types than for others. Data types for which help is more appreciated could then be prioritised regarding mobilisation and integration in the ARIADNEplus data catalogue and portal.

A comparison of the online accessibility rating of the data types and helpfulness of portal support for discovering and accessing datasets or collections of such data showed:

- surprisingly, the appreciation of support was lower for data types for which the accessibility was evaluated as insufficient;
- respondents were most appreciative of portal support for discovery and access of Sites and monuments databases or inventories, national GIS data & maps, and satellite or airborne remote sensing data (e.g. LiDAR, although the online accessibility of these data types was rated much better than that of other data types;
- The rationale for ARIADNEplus should not be to prioritise support for data types which are already much more accessible than others. The fact that the more accessible types are being provided by national mapping and heritage authorities indicated ARIADNEplus should prioritise other data types.

Suggestions for ARIADNEplus

The survey results tentatively suggest the following priorisation of data types for mobilisation and integration in the ARIADNEplus portal:

- Data types with high or medium appreciation of portal support, and currently medium or low online accessibility. These types are ranked according to the appreciation of support and level of accessibility:
    - Excavation data (e.g. excavation archive)
    - Artefact/finds databases or image collections
    - Radiocarbon, dendrochronology and other dating data
    - Environmental archaeology datasets
    - Unpublished fieldwork reports
    - Field survey/prospection data
- Subject-based data types of the following domains:
    - Maritime and underwater archaeology data
    - Scientific data/analysis of inorganic remains
    - Scientific data/analysis of biological remains
    - Inscriptions, coins or other special databases

    Some of these data types may also require more FAIRness to enable their integration in the ARIADNEplus dataset catalogue and portal.

Item-level access

The survey participants also rated which data types they would find helpful for their research if able to search items within datasets integrated from multiple sources. The results do not add much to the evaluation above, except that artefact/finds databases or image collections were ranked highest.

4.10 ARIADNEplus services for researchers and data managers

4.10.1 Background

A wide range of new or enhanced services for researchers and data managers is foreseen to be provided on the D4Science platform for virtual research environments. Therefore an important goal of the survey was to find out which ones the respondents from the ARIADNEplus user communities perceive as particularly helpful and could be prioritised in the service development.

The services in question are for end-users, comprising researchers (archaeologists, laboratory-based scientists and others) as well as data managers (repositories, databases). "Back-office" services, those which run the service provision platform and others which end-users do not use directly, were not included in the survey template.

The services survey respondents have been asked about will be prepared in WP15 - Innovative Services for Users. The description of the services in the WP15 work plan could of course not be used in the survey template. Each service had to be concisely described as a survey item stating what the users could do with a service (e.g. *Annotate images (e.g. artefact or laboratory images) and link them with other content*). The section that follows presents this "mapping" of planned services and survey items and briefly describes the services.

4.10.2 Mapping of planned services and survey items

The survey participants have been asked to indicate how helpful different listed services would be for their research or data management. The mentioned services correspond to existing and planned new or enhanced services. Under the latter two services have been included that are technically very demanding and not on the work plan of the project, hence potential future service.

Existing services

The following services are already available on the data portal.

- Register a dataset in a portal that allows searching data from many providers
- Discover & access archaeological data stored in repositories in different European & other countries
- Multi-lingual search for archaeological data
- Spatially and/or chronologically defined search options

These services are not included in the list of new or advanced innovative services for end-users (ARIADNEplus WP15). The services are available however some improvements are possible and planned or already prepared (e.g. enrichment of the catalogue model of the dataset registry).

The services have been included in the survey to find out how interested respondents are to make datasets available for search and access via the data portal. Regarding the options for data search ARIADNEplus aims to implement semantic search based on metadata and vocabularies in Linked Data formats.

New and enhanced services

ARIADNEplus does not start from scratch but has developed a solid base of already available services on the data portal as well as others that are not yet directly implemented in the portal (see below). Furthermore, there are open source tools available and considered for some of the services planned.

The overall goal is to enable innovative and effective ways of carrying out data-based research in archaeology. The general approach of ARIADNEplus is to provide a range of useful services for this in Cloud-based virtual research environments (VRE) based on D4Science VRE platform.

Search and visualize geo-spatial / GIS datasets

The services summarised with this survey item will be supported by D4Science Cloud-based Geoserver, which already has many of the required functionalities built-in. The GIS services comprise the usual services present in GIS systems, for example buffer definition, layer selection, proximity, viewshed analysis and so on.

The services and tools of the Geoserver will allow to integrate archaeological geo-information provided by partners and support the GIS functionalities of the space-time services, e.g. the spatially and/or chronologically defined searches mentioned under the existing but to improve services. In order to allow this the Geoserver will also rely on gazetteers and time periods vocabularies developed within the project.

Visual content services

Services for visualisation and manipulation of archaeological imagery developed by the CNR-ISTI Visual Computing Laboratory have already been a well developed part of the ARIADNE service portfolio. These and new visual content services will be enhanced in ARIADNEplus building on advances in recent projects of partners such as VisualMedia EOSCpilot Science Demonstrators[34].

Display and manipulate visual data objects (e.g. RTI images, 3D models, LiDAR data)

This survey item covers services that allow fast and efficient online rendering and manipulation of advanced forms of visual content such as Reflectance Transformation Imaging (RTI), 3D models and LiDAR imagery. These services are already Cloud-based but will be adapted to specific needs of archaeologists, and integrated in the ARIADNEplus service interface (in development).

Link and present together visual media (e.g. a 3D model) and related documentation

This survey item stands for a toolkit that allows effective organisation of visual content from archaeological research (e.g. 3D models of artefacts or monuments), including to link it with other research documentation, and visualise how they relate to each other. The toolkit comprises of tools developed by the CNR-ISTI Visual Computing Laboratory. Besides some foreseen improvements the toolkit will be ported on the ARIADNEplus Cloud environment and adapted to the service interface.

Visualize in 3D the layers of an excavation and the related documentation

This survey item covers visual documentation of archaeological excavations with a focus on 3D documentation, specifically visualisation in 3D of the excavation layers and related documentation. It will build on the Ephemera service developed by ARIADNEplus partner CYI-STARC[35]. As with other services porting on the ARIADNEplus Cloud environment and adaption to the service interface is required.

Annotate images (e.g. artefact or laboratory images) and link them with other content

This survey item covers the annotation and linking of research images with other content such as protocols, documentation or published articles. The foreseen end-user services will typically support semi-automatic annotation and linking of a smaller number of content items, using relevant vocabularies (e.g. for cultural content such as inscriptions). One specialised tool that is foreseen to be

[34] http://eoscpilot.eu/science-demonstrators

[35] http://ephemera.cyi.ac.cy

included is the DAP tool[36] for archaeological objects which contain written/symbolic information. Among other functionalities it allows annotating in a CIDOC-CRM compliant way images of such objects. The tool has been developed by the ARIADNEplus partners Archaeological Museum of Zagreb (AMZ) and CNR-ISTI Visual Computing Laboratory. Porting of this tool on the ARIADNEplus Cloud environment and adaption to the service interface is required.

Post a picture of an object and get suggestions for similar ones

This survey item stands for image recognition, comparison and retrieval services. Such services are technically very demanding and not on the work plan of the project, but may be attempted within the technical development horizon, which extends beyond the current ARIADNEplus project.

Textual content services

Annotate texts (e.g. fieldwork or laboratory reports) and link them with other content

This survey item stands for end-user services that support semi-automatic annotation and linking of a smaller number of documents such as archaeological reports and other content. One candidate for developing such services is the open source annotation system Pundit[37] that has been used by several digital humanities projects. In ARIADNEplus the services will need to support using archaeological and other scientific vocabularies (e.g. scientific work such as material analyses).

Identify & extract information from textual sources (e.g. a document repository) to produce metadata

This survey item covers a service that is based on the previous ARIADNE text mining and NLP tool, further developed into TEXTCROWD, a Cloud-based NLP tool created as a Science demonstrator in the EOSCpilot EU project[38]. Further development will include extension regarding the NLP functionality and languages (so far English, Italian, Dutch), and porting on the ARIADNEplus Cloud environment as well as adaptation to the service interface.

Process many documents (using Natural Language Processing) to find those on certain topics or specific information they may contain

This survey item stands for text analysis and recommendation services provided online on top of large document repositories. Such services are technically very demanding and not on the work plan of the project. As in the case of the advanced image recognition, comparison and retrievel services (see above), the technical development horizon extends beyond the current ARIADNEplus project.

Data vocabularies, mapping and linking

The following three services are not part of the ARIADNEplus WP15 services for end-users such as archaeologists and laboratory-based scientists. The services are mainly intended for data managers (repositories, databases) to enhance and employ vocabularies with the goal to link and integrate own and other datasets based on Linked Data standards and technologies.

Align own vocabulary terms with international thesauri (e.g. Getty Arts & Architecture Thesaurus)

This survey item stands for the Vocabulary Matching Tool developed by the Hypermedia Research Group of ARIADNEplus partner University of South Wales. The tool allows aligning own vocabulary (term list, thesaurus) with common vocabulary, in ARIADNE particularly the Getty Arts & Architecture

[36] http://tss.isti.cnr.it/dap

[37] http://thepund.it

[38] http://eoscpilot.eu/science-demonstrators

Thesaurus, but also other thesauri are possible. The tool is already available as a service in the ARIADNEplus Cloud environment.

Map a database (schema) to the CIDOC-CRM extended for archaeological research data

This survey item stands for the Mapping Memory Manager (3M) system developed by ARIADNEplus partner FORTH-ICS in the ARIADNE project. It allows mapping of database schema to the CIDOC-CRM (ontology), including the extensions developed in ARIADNE for archaeological research data (e.g. excavation, standing structures, epigraphy).

Use Linked Data to interlink own and other datasets

Use of Linked Data standards and technologies is the general approach in ARIADNEplus for data integration and some of the search services. Project datasets in Linked Data formats will also be made available (e.g. via an API) to external developers for interlinking datasets. The survey items has mainly be included to investigate the interest of survey participants in using Linked Data standards and technologies.

4.10.3 Survey results

The survey participants were asked to evaluate the different services of the ARIADNEplus service portfolio described above regarding how helpful these would be for their research or data management:

"The ARIADNEplus project will provide a range of services and tools for archaeological researchers and data managers. To what extent would your research or data management benefit from the following services/tools: would they be (1) = very helpful, (2) = helpful, (3) less helpful or (4) = not helpful for your work?"

The results are presented in Table 22.2. The services are listed as in the survey questionnaire, not according to a ranking based on the expressed appreciation. The level of appreciation is indicated as follows:

Appreciation	***Colour***	***Percentage of "very helpful" and "helpful" (actual)***
High		93.5% – 96.3%
Middle		78.3% – 83.0%
Low		74.3%, 74.5%

Table 22.1*: Scheme for ranking the ARIADNEplus services based on survey responses.*

Services for researchers and data managers	***N***	***Very helpful***	***Helpful***	***Less helpful***	***Not helpful***
Register a dataset in a portal that allows searching data from many providers	336	**251** (74.7%)	**70** (20.8%)	**10** (3.0%)	**5** (1.5%)
Discover & access archaeological data stored in repositories in different European & other countries	337	**248** (73.6%)	**72** (21.4%)	**12** (3.6%)	**5** (1.5%)
Multi-lingual search for archaeological data	336	**199** (59.2%)	**80** (23.8%)	**46** (13.7%)	**11** (3.3%)
Spatially and/or chronologically defined search options	330	**269** (81.5%)	**49** (14.8%)	**9** (2.7%)	**3** (0.9%)
Search and visualize geo-spatial / GIS datasets	334	**230** (68.9%)	**82** (24.6%)	**17** (5.1%)	**5** (1.5%)
Post a picture of an object and get suggestions for similar ones	335	**181** (54.0%)	**93** (27.8%)	**50** (14.9%)	**11** (3.3%)
Display and manipulate visual data objects (e.g. RTI images, 3D models, LiDAR data)	329	**166** (50.5%)	**102** (31.0%)	**48** (14.6%)	**13** (3.9%)
Link and present together visual media (e.g. a 3D model) and related documentation	331	**165** (49.8%)	**104** (31.4%)	**47** (14.2%)	**15** (4.5%)
Visualize in 3D the layers of an excavation and the related documentation	331	**150** (45.3%)	**96** (29.0%)	**67** (20.2%)	**18** (5.4%)
Annotate images (e.g. artefact or laboratory images) and link them with other content	335	**166** (49.6%)	**109** (32.5%)	**48** (14.3%)	**12** (3.6%)
Annotate texts (e.g. fieldwork or laboratory reports) and link them with other content	333	**162** (48.6%)	**103** (30.9%)	**54** (16.2%)	**14** (4.2%)
Process many documents (using Natural Language Processing) to find those on certain topics or specific information they may contain	329	**164** (49.8%)	**101** (30.7%)	**54** (16.4%)	**10** (3.0%)
Identify & extract information from textual sources (e.g. a document repository) to produce metadata	331	**170** (51.4%)	**94** (28.4%)	**55** (16.6%)	**12** (3.6%)
Align own vocabulary terms with international thesauri (e.g. Getty Arts & Architecture Thesaurus)	332	**165** (49.7%)	**95** (28.6%)	**49** (14.8%)	**23** (6.9%)
Map a database (schema) to the CIDOC-CRM extended for archaeological research data	318	**127** (39.9%)	**110** (34.6%)	**61** (19.2%)	**20** (6.3%)
Use Linked Data to interlink own and other datasets	320	**165** (51.6%)	**100** (31.3%)	**45** (14.1%)	**10** (3.1%)

Table 22.2*: Services for researchers and data managers: overview of survey results (N = 318-337).*

Comments received

The survey participants were invited to provide "Comments (e.g. ideas for other useful services ARIADNEplus could provide)". Twelve respondents provided comments that are listed below. One respondent suggested *"A click-able site list of all sites accessible in Ariadne?"*, which ARIADNEplus could provide easily based on the intended GIS services. Another respondent considered *"Annotate 3D models and link them with other content"*, a service that is foreseen in ARIADNEplus. Other respondents confirmed that all or particular services on the list would be very useful to have.

One respondent said that some services on the list needed to be described so that non-experts can understand how they would work. While this is hardly possible in an online survey it will be necessary to provide practical information and guidance when the services are made available.

Regarding the use of the CIDOC-CRM ontology in ARIADNEplus services, one respondent had a strong opinion about the CIDOC-CRM (not to use it), while two others had never heard of it. The comments of one respondent deserve special attention and are addressed after the list of other comments.

Comments:

- *All sound great! Being able to layer data and also to compare with other contexts would be really useful. Including a number of search possibilities would be great - chronological, regional, artefact form, feature or building type, keyword searchable ...*
- *Really, this would be heaven :-)*
- *A click-able site list of all sites accessible in Ariadne?*
- *Annotate 3D models and link them with other content*
- *I find Linked Data particularly important*
- *The online archives of Corinth and Athens agora, as well as the EFA Archives, are good examples of similar platforms*
- *The 3D visualizations would only be useful if there was a way to export from the ARIADNEplus*
- *In some cases at least it seems necessary to provide explanation of how the services work, understandable for the non-experts*
- *Avoid the CIDOC-CRM! Way too complex. Has not solved any real-world problems.*
- *I have no idea what CIDOC-CRM is*
- *What is CIDOC-CRM?*

One respondent was very critical about some of the services ARIADNEplus intends to provide:

- *Where I have given 4, some of these functions are best served using existing, desktop software, or are outside of my research field. I see no purpose in ARIADNE providing 3D or LiDAR manipulating services. The suggested vocabularies are useless for environmental or other scientific data. Other tools exist for extracting from text sources and tagging map or image parts (e.g. RECOGITO). I see no point in NLP parsing of reports for certain topics, humans can do this fast and more accurately.*

One critique is that for some functions there are desktop or online tools available. Which functions are not mentioned by the respondent but the approach of ARIADNEplus generally is to provide tools/functions as web-services of Cloud-based virtual research environments. This is a very different approach than just to provide a tool online. However, the project will have to convince users about its advantages over stand-alone desktop and online tools. The respondent doubts that services for manipulating 3D or LiDAR are useful (to be demonstrated in ARIADNEplus), and thought that humans

can discover better topics in reports without a digital tool, which is not the case when a large number of reports is considered (as in ARIADNEplus). Rightly, the respondent found that the explicitely named vocabularies (Getty AAT, CIDOC-CRM) are not intended for environmental or other scientific data specifically.

4.10.4 Summary and suggestions

Summary of main results

The survey results for the ARIADNEplus services for researchers and data managers can be summarised as follows:

Services which ARIADNE/plus already provides: A very encouraging survey result is that respondents appreciated such services most (and may have already used them). These services are:

- Register a dataset in a portal that allows searching data from many providers
- Discover & access archaeological data stored in repositories in different European & other countries
- Spatially and/or chronologically defined search options

"Register a dataset in a portal that allows searching data from many providers" is the second highest ranked of all services listed (behind the not yet available to *"Search and visualize geo-spatial/GIS datasets"*). This signals a high interest of survey respondents to make datasets available for search and access via the ARIADNEplus data portal.

Multi-lingual search: This service is also already available on the data portal but was appreciated significantly less than the ones above. This does not speak against support of multi-linguality when portal users search for particular subjects because it ranks much higher than other services on the list.

Top on the list of new services – Search and visualize geo-spatial/GIS datasets: These services are the highest ranked among the new services, and are part of the plan of services ARIADNEplus will implement. It is worth noting that in the ARIADNE 2013 survey "A portal that makes it more convenient to search for archaeological data stored in different databases" was seen as "very helpful" for their research by 79% of respondents (very or rather helpful 96%, N=481), while "Services for Geo-integrated data" by 52% (very or helpful 81%, N=471). With the portal in place, services for geo-spatial/GIS data now are on top of appreciated new services.

Lowest on the list – Map a database (schema) to the CIDOC-CRM extended for archaeological research data: This result does not come as a surprise because the service is specifically for data managers (databases, repositories) and these make up only 20% of the survey respondents.

Also low – Visualize in 3D the layers of an excavation and the related documentation: Survey respondents also evaluated this service as less helpful for their research or data management work than others. 53% of respondents were archaeological researchers (field work) but also many of these may have seen the service as beyond their expertise or what they might use in practice.

Services in the middle range: All other services were in the middle range of appreciation, judged as "very helpful" or "helpful" by between 78.3–83% of the responses per service. Listed according to the percentages (more to less):

- Use Linked Data to interlink own and other datasets
- Annotate images (e.g. artefact or laboratory images) and link them with other content,
- Post a picture of an object and get suggestions for similar ones
- Display and manipulate visual data objects (e.g. RTI images, 3D models, LiDAR data)

- Link and present together visual media (e.g. a 3D model) and related documentation
- Process many documents (using Natural Language Processing) to find those on certain topics or specific information they may contain
- Identify & extract information from textual sources (e.g. a document repository) to produce metadata
- Annotate texts (e.g. fieldwork or laboratory reports) and link them with other content
- Align own vocabulary terms with international thesauri (e.g. Getty Arts & Architecture Thesaurus)

"Future" services: Two services were suggested that are technically very demanding and not included in the work plan of the project, hence potential future service. These are text analysis and recommendation provided online on top of large document repositories *("Process many documents (using NLP) to find those on certain topics or specific information they may contain")* and image recognition, comparison and retrieval *("Post a picture of an object and get suggestions for similar ones").* The latter service was seen a bit more interesting.

General evaluation

The two lowest ranked services still were considered as very helpful or helpful by close to 75% (74.3%, 74.5%) of respondents, those in the middle range by 78.3–83%, and the ones on top by 93.5–96.3%. However, of the latter only *"Search and visualize geo-spatial / GIS datasets"* (93.5%) is not yet available on the data portal.

It is also worth noting that in the middle range, after the already available multi-lingual search (83%), first comes *"Use Linked Data to interlink own and other datasets"* (82.9%), followed by four services for working with visual content (e.g. high-resolution images, RTI images, 3D models, LiDAR), at 81.2–82%. This signals a high interest in visual content services, at least considerable more than for textual content.

Suggestions for ARIADNEplus

The main suggestions that can be derived from the survey results are:

- Devote special attention to the new services for search and visualisation of geo-spatial/GIS datasets.
- Prioritise also the use of Linked Data for interlinking datasets, particularly at item-level.
- Continue to enhance the existing and develop new visual content services of interest.
- Evaluate further which services for textual content are of interest to users, including services not yet considered.
- Promote further the use of CIDOC-CRM by making clear its capability to integrate research data conceptually, especially regarding the ontology extensions developed in the ARIADNE project for archaeology (e.g. excavations, standing structures, epigraphy).
- In the testing and evaluation of online tools with end-users investigate if there are any reservations against using them them as services in a Cloud-based virtual research environment instead of a stand-alone desktop or online tool.

4.11 Training needs

4.11.1 Background

ARIADNEplus training activities mainly concern needs and requirements regarding the management of archaeological research data in view of making the data generated in research projects available through a publicly accessible repository. Therefore the overall theme regarding training is capacity building for open sharing of data by archaeological researchers and research groups (e.g. data managers of large projects) as well as research organisations (e.g. managers of institutional repositories). The sections that follow provide background on the related topics Data Management Plans (DMP) and support in Research Data Management (RDM). In addition, the topic of Data Science in archaeology is addressed briefly.

Data Management Plans (DMP)

It is understood that proper research data management should start and be supported as early as possible, while researchers are working on their projects. A Data Management Plan (DMP) can provide a basis for this support. To implement a DMP is requested by ever more research funders with the goal that the data which underpins research publications will be shared after project completion.

A DMP should describe the kinds of data that will be created or collected, the methods applied, the data documentation (metadata), where and how long the data will be stored, and ways how the data could be shared and accessed. The grant applicant also has to justify why the proposed measures are considered optimal, and explain which limitations may apply, e.g. due to ethical or intellectual property issues.

Williams et al. (2017) reviewed requirements defined for DMPs by several research funders and found that there is no general and definitive list of topics that should be covered in a DMP for a research project. Rather, the study identified high variability in required or suggested topics (in total forty-three topics) and inconsistent requirement set by funders. In general, requirements focus on post-publication data sharing rather than upstream activities that lead to data quality, provide traceability or enable reproducibility. Parham et al. (2016) analysed 500 DMPs submitted to six of the seven directorates of the U.S. National Science Foundation (NSF) to see how researchers in different fields understand and interpret the NSF data management requirements and plan for managing, sharing, and archiving their data.

Critiques of mandatory DMPs argue that these generate a lot of work but there is no evidence that the DMPs have any positive effects, particulary not regarding the goal of open data (Morgan & Janke 2017; Smale et al. 2018). This outcome can of course be expected if open data sharing is not checked and reinforced. Regarding the lack of empirical evidence for positive effects of DMPs it must be noted that the formal requirement of having a DMP in place is relatively new, in some countries for some years now (in others not at all), hence providing quantitative evidence at this point is difficult. Positive effects of DMPs and data management support are highlighted in a number of cases studies on projects large and small (e.g. Burnette et al. 2016; Curdt 2019; Petters et al. 2019), but there are too few and presented in different ways to allow a meta-analysis of effects.

Our view of DMPs for research projects is that such plans should not be seen isolated from the broader picture of Research Data Management support by different actors such as universities aiming to provide support via their research libraries, domain data repositories and other supporting organisations. It is also preferable to separate the objective of open data sharing (the core objective of research funders) from the educative goal of better data management practices.

The latter should be taken care for by the research community, by stepping up the hands-on training of such practices, involving experienced researchers, data curators, repositories. DMPs can be part of this, but these are plans, which need to be revised in the course of the actual data management work. Mandatory DMPs typically include that the DMP should be revised if need may be, but that often does not happen in practice, or only if an update is formally required (e.g. at the end of a project). For DMPs being an effective instrument of good RDM such management must be promoted, supported (training), and expected by the research community. The proof of such management would be that open/FAIR data is being deposited for preservation and access in appropriate data repositories.

Research Data Management (RDM)

Research Data Management (RDM) is a large topic. Li & Eichmann-Kalwara (2019) conducted a survey of the scholarly literature on RDM since the 2000s and present different clusters within this interdisciplinary field. Larger clusters are "scientific collaboration", "research support service", and "data literacy". Topics such as "digital curation" and "information processing" appeared most frequently. They also note a sharp increase in several specific topics such as "data sharing" and "big data".

Regarding support for RDM there are a lot of introductory level books, articles and materials (e.g. Bryant et al. 2017; Corti et al. 2013; Jones et al. 2013; Pryor et al. 2013; Ray 2014). It appears that support for RDM has been mainly delegated to research libraries. Some years ago their engagement in this newly assigned role was generally low, e.g. Corrall *et al*. (2013; 88 institutions located in Australia, New Zealand, the UK and Ireland); Cox & Pinfield (2013; 81 higher education institutions in the UK); Tenopir *et al*. (2012: 221 members of the US Association of College and Research Libraries).

With often little prior knowledge in systematic data management, researchers turned to their institute or university for support in all related matters. Some had already taken steps to provide advice and support, while many others were not yet prepared to meet this new demand and responsibilities. The focus of those a step ahead was on supporting institutional policies through providing advisory and training services. Support of technical aspects of data curation such as metadata creation was rather limited, according to the mentioned surveys between 10-20%.

The engagement of research libraries may have increased but very likely remained focused on rather general support activities such as advice (e.g. how to draw up a DMP, IPR/licensing, etc.) and organising introductory level training for research students. The results of a survey by Cox et al. (2017) with 170 respondents from Australia, Canada, Germany, Ireland, the Netherlands, New Zealand, and the UK confirm this assumption.

The main challenges for the research libraries and institutional document repositories they manage are limited resources (funds, personnel) and skills gaps regarding research data curation. It can be assumed that a broad engagement of research libraries in RDM will only be possible based on a sufficiently large workforce of skilled "data librarians". However, this professional role has been perceived as "a gap in the market" (Hyams 2008), a "field undefined" (Alvaro *et al*. 2011), or an "accidental" rather than clear career choice (Pryor & Donnelly 2009).

In December 2016 the Confederation of Open Access Repositories (COAR), most of which repositories managed by university libraries, conducted a survey on RDM. The survey received 43 responses from from all continents, with the majority (25) from European institutions (Shearer & Furtado 2017). 23 of the respondents said that their repository already collects research data, while of the others most indicated that they had plans to do so in the near future. The major challenges for collecting data in their institution were engaging researchers, a lack of institutional policies for RDM, and infrastructure for storage and preservation. The COAR membership survey 2018 with 59 respondents from 23 countries found, *"The three biggest challenges related to repositories are (1) user engagement and*

getting content deposited, (2) awareness and visibility of repository, and (3) research data management" (COAR 2018).

In the United States the support for RDM has been mainly delegated to research libraries (Fearon et al. 2013, provide several examples). In Europe, the Association of European Research Libraries (LIBER) aims to promote that the libraries take up the responsibility for RDM support and develop the capacity to do so.[39]

The LIBER Working group on E-Science / Research Data Management in their recommendations for research libraries "to get started with research data management" wrote that data librarianship *"probably is a profession in itself but a lot of work regarding data services can also be done by e.g. (reskilled) information specialists"* (LIBER 2012). The group considered reskilling as a critical requirement since very few libraries could hire new, specialised staff. The need of re- or up-skilling is also emphasised in many other publications (e.g. Auckland 2012; CLIR 2013; Cox *et al*. 2012; Gow & Molloy 2014, based on a survey of the DigCurV - Digital Curator Vocational Education project, 2011-2013).

In 2015, LIBER issued a "factsheet" on the expected "new leadership role" of research libraries regarding research data (LIBER 2015). Previously their Steering Group for Scholarly Communication and Research Infrastructures Steering Committee published an number of RDM case studies on how research libraries provide RDM support (LIBER 2014). Chiarelli & Johnson (2017, for JISC, UK) present a much broad panorama of RDM actors and support activities.

RDM support in archaeology

Results of the ARIADNE 2013 survey and interviews showed that data management support for archaeological researchers by their institute was rather low. The responsibility for maintaining the data after project completion remained with the researchers, 54% with the project manager, 27% an appointed member of the research team, and 19% "other" (ARIADNE 2014: 100). Major reasons for this are the understanding that the data is "owned" by the researchers and lack of data curation resources at the institutional level.

The responsibility for scaling up data management know-how for research projects is clearly with the university departments and institutes active in archaeological research. This has been emphasised by ARIADNE partners (ARIADNE 2015: 154), for example:

"Promote and consider offering training in data management/data documentation also for researchers and PhD students (not only for institutional/project data managers), as part of basic PhD training. Fostering a culture of data sharing means we need to make researchers in general to gain knowledge about data documentation".

"*Begin training on an earlier/lower level – creation of data(-bases), structured folders, use of metadata on a file level as part of university curriculum. Or, for those who got out of university some time ago – courses. Regular updates on what is going on, good practices"*;

Still it is important to foster RDM capacity building at the institutional level, not only and maybe not primarily on data management planning by individual researchers. RDM is necessary at the institutional level in order to ensure that the necessary policies, resources and skills are in place. Archaeological research institutes will increasingly become accountable for data that is produced by the researchers they employ. This will be due to conditions such as grant contracts for research programmes, legal regulations or professional codes of conduct. The research institutes and researchers are confronted with requirements imposed by various actors, including public

[39] Currently this task is with the LIBER Research Data Management Working Group (2018-2022), https://libereurope.eu/strategy/research-infrastructures/rdm/

administrations (e.g. heritage management), research funders, professional organisations, and others. Data policies at the institutional level can allow roles and responsibilities regarding data curation and access to be set forth and monitored internally. There is a need for data managers to take care for institutional data repositories and databases as well as to support projects, which in archaeology can extend over many years.

There is no lack of guidance material for good practice data management. Among the outstanding resources for archaeologists are the Guides to Good Practice offered by the Archaeology Data Service (ADS) & Digital Antiquity[40], and the ADS DataTrain materials for post-graduate teaching on research data management in archaeology[41]; for archiving of archaeological material and data there is the ARCHES guide to good practice (2014, available in several European languages)[42]. But guides to good practice need to be implemented in actual practice, in new and on-going projects with reliable institutional support.

Need of repositories for archaeological data

Regarding the curation of data for long-term curation and access there are considerable doubts about repositories managed by university libraries doing this. Proper curation of disciplinary research data requires specialisation which is difficult to achieve for the many different disciplines present at a university. There is a lack of RDM expertise of research library staff (see above) and the already implemented document repositories are considered as not adequate for research data. Moreover, repositories of universities and others dedicated to one institution usually accept data only from affiliated researchers.

Leading examples of data repositories are specialising in research fields. In archaeology these are repositories dedicated to archaeological data, e.g. Archaeology Data Service (UK) and tDAR - The Digital Archaeological Record (Digital Antiquity, Arizona State University, USA). Or they part of data repository services for social sciences and humanities, e.g. the e-Depot for Dutch Archaeology of Data Archiving and Networked Services - DANS (Netherlands). Also the Digital Repository of Ireland (DRI) has a focus on social sciences and humanities and recently ingested a first large collection of archaeological documentation of fieldwork commissioned 2001-2016 by Transport Infrastructure Ireland (TII 2017).

However, in many European and other countries archaeologists do not yet have available an appropriate digital repository where they can safely deposit and make available their data to the research community and other users. Ideally such a repository has a national scope and is mandated by research funders for depositing data from archaeological investigations. This provides advantages in several respects, including clear orientation of all stakeholders, expertise in archiving archaeological data, cost-effectiveness of data curation and access (e.g. economies of scale), among others. From the perspective of ARIADNEplus one or only a few core repositories per country from which data records can be aggregated is of course the preferred scenario (Geser 2019a; Geser 2019b: 195-196).

Fortunately, the issue of a lack of appropriate data repositories is now being addressed by the COST Action SEADDA, the Saving European Archaeology from the Digital Dark Ages network that involves ARIADNEplus partners and institutions from other countries, including almost all European countries. SEADDA brings archaeologists and data management specialists together to share expertise, provide knowledge and training in matters of data archiving and access, and help archaeological communities to address problems in the most appropriate way within their own countries.

[40] ADS & Digital Antiquity: Guides to Good Practice, http://guides.archaeologydataservice.ac.uk

[41] ADS: DataTrain, http://archaeologydataservice.ac.uk/learning/DataTrain

[42] ARCHES - Archaeological Resources in Cultural Heritage (2014): The Standard and Guide to Best Practice in Archaeological Archiving in Europe, online, http://archaeologydataservice.ac.uk/arches/Wiki.jsp?page=Main

Data Science

In the survey question on training needs also training to develop data science skills, i.e. use of advanced data processing and analysis methods, has been included. Archaeologists generally have data processing and analysis skills because data collection is expensive and adequate skills are necessary to get as much insights as possible from statistical and other analysis of the data. Topics in quantitative archaeology include statistics, classification of objects, spatial analysis, modelling, simulation and data mining, among others.

Questions of the ARIADNE 2013 survey asked respondents about different types of data their research group would need to carry out projects, the online availability of the data, and if the research group produces such data themselves. Included were *"Data for model-based computing, simulation"* and *"Results of data mining for identifying patterns or interesting outliers ("data mining" covers various analytical techniques for discovering patterns in large data sets)".* Well over 500 respondents ranked such data and methods as the least important of the list of eleven presented (ARIADNE 2014: 79-84 and 95-96).

Reasons may have been that respondents did not expect much from the mentioned data mining or computing approaches in archaeology and/or lacked expertise to apply the required technology/ software and methods. The online availability of data resources for applying them was perceived as very low, and exemplary comments were *"Data mining is currently undervalued due to a lack of repositories"* or *"Not much available in the way of data mining - it may be important in the future"*.

Data science is often related mainly or exclusively to so-called "big data". The IT research and advisory company Gartner defines it as *"high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation"* (Gartner: Glossary, n.d.).

To the characteristics volume (large quantities of data), velocity (created in real-time) and variety (being structured, semi-structured and unstructured), suggested by the Meta Group (Laney 2001), others added "V's" such as Veracity, Viscosity, Virality, Value (Big Data Alliance, n.d.). Kitchin & McArdle (2016) discuss the initial V's and other big data traits looking into a number datasets from different domains. Al-Barashdi & Al-Karousi (2018) provide a review of research literature on different topics related to "big data".

Archaeological "big data" has only in recent years been addressed by some researchers, with regard to the volume of not yet digitised information (Wesson & Cottier 2014), its complexity and intricacy (Cooper & Green 2015), or how "big data" computing techniques might be used in archaeology, e.g. to identify relevant patterns in data that suggest new research questions (Gattiglia 2015). Practicalities and possibilities of "big data" in archaeology have been discussed in a recent conference at the McDonald Institute for Archaeological Research, University of Cambridge (brief abstracts of the contributions are available).[43]

In an article in *The European Archaeologist* Paludan-Müller (2015) suggests, *"There is no reason why archaeology should not weigh in from its privileged position for analysing big data to understand long-term developments over vast spaces, be it in the development of global power and economic structures and of empires, the shifting patterns of migration or the response of human society to changing climate"* (Paludan-Müller 2015: 7).

One example is the growth, sustainability or collapse of cities (Ortman et al. 2014). More generally, "big data" based research could allow archaeology explore long-term dynamics of human society and its interactions with the natural world (so-called Coupled Human and Natural Systems). However,

[43] Big Data in Archaeology: Practicalities and Possibilities, McDonald Institute for Archaeological Research, University of Cambridge, 27-28 March 2019, https://erikgjesfjeld.wixsite.com/big-data-archaeology

aggregated and integrated large archaeological datasets as "big data" for data mining or other advanced computing methods are not readily available.

An online survey on what archaeologists see as most important scientific challenges has been conducted by Kintigh *et al*. (2014). The input from 181 respondents has been organised and refined into 25 grand challenges, grouped in five large categories, "Human-Environment Interactions" and "Movement, Mobility, and Migration", for instance. The survey report does not address specific data technologies or forms of collaborative e-research which may be necessary to tackle the described challenges (in this regard see Kintigh *et al*. 2015).

The concern is primarily that large-scale, thoroughly documented and integrated datasets will be required for most of the challenges to apply sophisticated modelling for comparative analyses and synthetic research. For example, one challenge under the Human-Environment Interactions group is "*to join disparate efforts into a broad-based initiative that can integrate existing and new sets of archaeobiological, geomorphological, paleoenvironmental, demographic, and other relevant data to model human/environmental interactions through time".*

The Coalition for Archaeological Synthesis (CfAS)[44] fosters synthetic research in archaeology to advance science and benefit society (Altschul et al. 2017 and 2018). Currently the Coalition aims to establish collaborative synthetic projects on human migration as understood from a long-term perspective. The initiative is jointly sponsored by the Society for American Archaeology (SAA) and the European Association of Archaeologists (EAA). The projects will be progressed over 2-3 years by working groups that are expected to integrate multiple perspectives and employ a wide range of data sources to generate the envisaged explanatory insights. The first CfAS workshop to design the projects was held in the fall of 2019 in Arizona. Prof Franco Niccolucci, coordinator of ARIADNEplus attended the workshop, and participation of the project in the initiative has been agreed.

4.11.2 Survey results

In the survey question on training needs of researchers and data managers eight activities were presented. The first item on the list, "Apply open/FAIR data principles in archaeology", is the most general. It does not state a particular activity, and in practice it requires many different activities. The next items on the list were arranged as a sequence of activities from defining and implementing a Data Management Plan to depositing project data in a digital repository. Interspersed was to develop data science skills, i.e. to process and analyse with advanced methods datasets of a large archaeological project. Part of the management of such datasets would be to create metadata for them using domain vocabularies. Activities to produced metadata and use vocabularies for this purpose are the two most closely related items on the list. From the perspective of researchers one activity on the list is not part of the described sequence which is to manage a digital repository of archaeological data reveived from projects.

The survey respondents were invited to evaluate the current training needs of archaeologists regarding the mentioned activities, not if they felt needing training on any of them themselves.

When evaluating the current training needs of archaeologists regarding data management and processing, which of the following would be (1) very helpful, helpful, (3) less helpful, or (4) not helpful for your work?

	N	***Very helpful***	***Helpful***	***Less helpful***	***Not helpful***

[44] Coalition for Archaeological Synthesis (CfAS) , USA, http://www.archsynth.org

Apply open/FAIR data principles in archaeology	330	**222** (67.3%)	**90** (27.3%)	**15** (4.5%)	**3** (0.9%)
Define and implement a Data Management Plan	328	**183** (55.8%)	**109** (33.2%)	**34** (10.4%)	**2** (0.6%)
Manage datasets of a large archaeological project	329	**194** (59%)	**109** (33.1%)	**21** (6.4%)	**5** (1.5%)
Develop data science skills (use advanced data processing and analysis methods)	328	**199** (60.7%)	**97** (29.6%)	**28** (8.5%)	**4** (1.2%)
Produce metadata for archaeological datasets	330	**189** (57.3%)	**107** (32.4%)	**30** (9.1%)	**4** (1.2%)
Use domain vocabularies to describe datasets	328	**159** (48.5%)	**125** (38.1%)	**39** (11.9%)	**5** (1.5%)
Deposit project datasets in a digital repository	328	**206** (62.8%)	**97** (29.6%)	**23** (7%)	**2** (0.6%)
Manage a digital repository of archaeological data	328	**193** (58.8%)	**97** (29.6%)	**32** (9.8%)	**6** (1.8%)

***Table 23**: Results for training needs of archaeologists regarding data management and processing (N = 328-330).*

Comments of respondents: Only three respondents added a comment. One respondent did not understand what "Use domain vocabularies to describe datasets" means, while another said that implementing a DMP is a priority in an ongoing project. The third comment is, *"All very important as most archaeologists lack these skills – there is no training during their university trajectory where they learn these kinds of things".*

Indeed, a clear majority of respondents said that training of archaeologists would be very helpful or helpful regarding all eight activities on the list. The highest approval is 94.5% for training to apply open/FAIR data principles (67.3% very helpful), while the lowest, but still high approval, is 86.6% for training to use domain vocabularies to describe datasets (48.5% very helpful).

But it is worth to note that, in addition to open/FAIR data, training for three activities got a significantly higher approval than others: Deposit project datasets in a digital repository (very helpful or helpful 92.4%), Manage datasets of a large archaeological project (92.1%), and Develop data science skills (90.3%). The percentages of "very helpful" for these are 62.8%, 59.0% and 60.7%, respectively. In comparison, 57.3% of respondents evaluated as "very helful" Produce metadata for archaeological datasets, 55.8% Define and implement a Data Management Plan, and 48.5% Use domain vocabularies to describe datasets.

The lower scores for the latter three activities, in particular those related to metadata and vocabularies, did not come as a surprise. These are more important for managers of datasets of a large archaeological project and managers of repositories than researchers who primarily collect and interpret data from fieldwork. The latter make up 53% of all survey respondents, while 13% are managers of an institutional repository (or other services that provide access to data), and 7% managers of project databases. Also, regarding barriers to deposit their research data in digital repositories, 74% of respondents said that the work effort for providing the data and metadata in the required formats is a "very" or "rather" important to barrier to do so.

Training on how to deposit project datasets in a digital repository ranks high, very helpful or helpful 92.4% (very helpful 62.8%), indicating that respondents perceive it as important to increase the readiness of researchers to do so.

A remarkable result is also the high approval for training to develop data science skills, which slightly over 90% of respondents considered as very helpful or helpful. As very helpful it was considered by more respondents than five other items on the list, only data deposition and the open/FAIR data principles were appreciated more.

Regarding the result for training on how to apply open/FAIR data principles in archaeology (94.5% very or rather helpful) it should be noted that the ARIADNEplus project promotes the principles and a large number of survey respondents work at partner organisations. The awareness of the principles is very likely higher among the survey respondents than on average among researchers in archaeology as well as other disciplines.

The Figshare 2018 survey included the question *"How familiar are you with the FAIR principles in relation to open data?".* 1,239 respondents answered the question of which 60% said they had never heard of the FAIR principles before, 25% had heard of the principles but was not familiar with them, while 15% claimed being familiar with the principles. Asked *"How compliant is your data with FAIR"*, 30.7% skipped the question, 24.9% chose "neutral", while 12.3% said "very much", 7.8% "not very much", 19.8% "somewhat", and 4.5% "not at all". Researchers from different disciplines participated in the survey, e.g. biology 19.3%, medicine 14.2%, social sciences 14.1%, earth & environmental sciences 10%, engineering 6.8%, humanities 3.7% (Figshare 2018; figures for familiarity with FAIR extracted from the available dataset; for compliance and the disciplinary composition from the interactive visualisation based on the dataset).

4.11.3 Summary and suggestions

Summary of main results

Survey respondents who answered the question (around 330) thought that training on all of the listed activities would be very helpful or helpful between 86.6–94.5%. The percentages for "very helpful" ranged between 48.5–67.3%.

Regarding "very helpful" significantly less appreciation was expressed for training in how to create and implement a data management plan (DMP), manage a digital repository, produce metadata and use domain vocabularies to describe archaeological datasets. Data science skills, managing datasets of a large archaeological project, depositing project datasets in a digital repository and, above all, apply open/FAIR data principles in archaeology were scored higher.

The open/FAIR data principles are generally relevant for all participants and training on how to apply them ranked on top. Among the survey participants a higher than "average" awareness of the principles can be assumed.

That researchers are the largest group in the survey sample certainly had a considerable impact on the results. Researchers worry about additional data-related work, which explains why training regarding DMPs, metadata and vocabularies is appreciated less.

When researchers need to deposit data in a repository, the question of metadata comes up. All studies on data sharing through digital repositories, including the ARIADNE/plus surveys, found that researchers consider the effort to provide the required metadata as a barrier to open data sharing. While data repositories and users would benefit from high-quality metadata, data creators face the burden and usually prefer not to invest much effort on providing metadata.

Inconsistently, in the ARIADNEplus survey, training on data deposition appeared to be welcome, despite the (not recognised) fact that this would require dealing with metadata and vocabularies. Awareness of an increasing expectation that data from funded research projects should be deposited may have contributed to this result.

To define and implement a data management plan seems to be unappealing. But training on how to manage datasets of a large archaeological project was appreciated. Such management is necessary and the task to do it can be taken on by, or delegated to, one or two team members who are trained to do it.

Training to develop data science skills, i.e. use advanced data processing and analysis methods, promises to provide more value from the effort invested in the collection of data. It was considered as very helpful by more respondents than other five items on the list, only data deposition and the open/FAIR data principles were appreciated more.

Training to manage a digital repository of archaeological data was appreciated less. But managing a digital repository is a professional activity of a smaller group and distinct from being a researcher; only 13% of the survey respondents were managers of an institutional repository.

Suggestions for ARIADNEplus

The following suggestions take account of the ARIADNEplus plans for training, the survey results on training needs, and the background on data management plans, research data management, and data science presented in *Section 4.11.1*. Furthermore, the fact that the closely related SEADDA project focuses on data respositories for archaeology is considered.

FAIR data principles

The survey found that training for the application of open/FAIR data principles in archaeology would be appreciated most, both by researchers and data managers, and ARIADNEplus is committed to support these principles within the archaeological sector.

The project has a work package comprised of six tasks dedicated to policies and good practices for FAIR data management. The tasks include evaluating implications of the implementation of the principles in the sector; provide policy support tools such as a flexible Data Management Plan template and supporting wizards; guidelines and support on repository creation, management and quality control; guidance on how to realise FAIRness of data taking account different regulations in participating countries, IPR-related and other issues; and offering practical training material and workshops.

Significant contributions to capacity building and take-up of the FAIR principles by these activities can be expected. The background given for training needs suggests that:

- the contributions should be as practical as possible, distinct from the broad wave of general information on the FAIR principles by ever more on the FAIR data “bandwagon”;
- project partners involved in the tasks mentioned should consider what falls, at least at a general level, within the remit of other organisations, e.g. support for DMPs and research data management in general by university libraries and repositories;
- ARIADNEplus training activities on FAIR data should focus on what matters for archaeological researchers and data managers specifically.

Data Management Plans (DMP)

Survey respondents ranked data management planning much below other training opportunities. To define and implement a DMP and related activities (metadata, vocabularies) adds work, but researchers are unsure they will benefit from this additional work. The background section on DMPs

notes that requirements defined by research funders for such plans are varied and may be inconsistent. Critiques argue that DMPs generate a lot of work while there is little evidence for positive effects.

Suggestions for countering these negative perceptions of DMPs are

- provide case studies on data management planning of archaeological projects making clear the benefits for projects and researchers involved;
- consider a DMP template that covers the minimum standard requirements while focusing more on the practicalities of different archaeological projects large and small;
- in general, rather than dry information about creating DMPs consider how to help the archaeological research community step up practical training of PhD students and early-career researchers in tried and proven data management practices.

Data managers of projects and repositories

Survey respondents appreciated training on how to manage datasets of a large archaeological project, while less so on how to manage a digital repository. The latter is a professional role and only a minority of respondents were managers of an institutional repository.

Data managers of projects

Training for data managers of archaeological projects will be provided in the ARIADNEplus trans-national access (TNA) programme, specifically under the TNA themes Data Stewardship and Implementing Interoperability. These researchers and data managers will take a keen interest in developing metadata for archaeological datasets with domain vocabularies in order to manage, use and share FAIR datasets.

Managers of repositories

In matters pertaining to archaeological repositories ARIADNEplus will benefit from coordinating activities with the SEADDA project, in which many consortium partners participate. SEADDA aims to foster the development of archaeological data repositories in countries where the research community lacks an appropriate repository, while ARIADNEplus supports finding and accessing data that is being shared through existing repositories.

Therefore, ARIADNEplus could

- help developers of repository initiatives plan participation of the repositories in its research infrastructure at an early stage,
- repository projects which are more advanced might benefit from available services, for example, by using data description and mapping services for representative initial datasets.

Research Data Management (RDM)

Training provided in the TNA framework as well as tutorials and workshops partners will organise on ARIADNEplus services and tools do not scale. In order to reach a higher number of researchers and data managers with information and guidance on RDM, ARIADNEplus can

- continue to make them aware of available guides to good practice, e.g. the guides offered online by the Archaeology Data Service/Digital Antiquity;
- provide a series of webinars on FAIR archaeological research data with contributions by experts from the ARIADNEplus partnership;

- organise "train the trainer" workshops at conferences or research institutes so that research data managers can serve as disseminators of good practices, including in questions of IPR and copyrights and sensitive archaeological data.

Data science skills

Survey participants appreciated training to develop data science skills, i.e. use of advanced data processing and analysis methods, more than the other five items on the list. Such training promises to get more from the effort invested in the collection of data.

ARIADNEplus has limited capacity to raise the level of data science skills of archaeological researchers. What the project can offer is:

- raise awareness of Open Science practices related to the sharing and (re)use of FAIR data;
- support the documentation and integration of archaeological datasets based on metadata standards and domain vocabularies;
- use of ARIADNEplus Cloud-based Virtual Research Environments (VREs) for data science tasks;
- a specific activity could also be to organise a group of high-potential young researchers for a collaborative synthetic project on human migration related to the initiative of the Coalition for Archaeological Synthesis (CfAS), to explore the potential of ARIADNEplus VREs, services and datasets for comparative analyses and synthetic research.

4.12 Final comments & suggestions of respondents

At the end of the questionnaire the participants were invited to give further comments or make suggestions. Some respondents provided further details about their work, while others encouraging statements regarding the work of ARIADNEplus, e.g. *"Setting up such an information network is a great idea"* or *"Thanks for doing this job".*

One respondent thought that it will be challenging: "*It will be very hard to implement all those things that you want on such a scale because they depend on lots of factors, e.g. knowledge of data management of archaeologist, different workflows, different data formats in use and making them future proof, different data collection principles, time consuming data preparation for online repository,... Good luck!"*

Suggestions on how to address the challenge were:

- *"How to convince researchers it is in their scientific interest to share data?"*
- *"You have work with research group and researchers more than with institutions if you will like to obtain better results."*
- *"The success of the project depends on the synergy of the organizations that are not involved in it"*
- *"It could be useful to open a forum for online suggestions and to organize meetings dedicate to this questions with all the institutions"*
- *"Making field data accessible is important, providing the means to publish this data is even more important"*
- *"Better communication with potential users in preparing data collection projects and reflecting on their practical experience".*
- *"Our research group would need more economic funds and more human resources to enter data and images preserved in our photo and document archives"*

Suggestions for the ARIADNEplus services or tools:

- *"Consider decentralized approaches (not one portal that fits all), foster collaboration between local, regional, national infrastructures"*
- *"Are you really sure, that English is the only one best language of international (European) historical sciences (i.e. archaeology) platform? Für interdisziplinäre Arbeit im mittel-europäischen Bereich finde (nicht nur) ich besser Deutsch / or bilinqual"*
- *"National maps are being collected by INSPIRE - any overlap?"*
- *"Within french MASA Consortium, we are working on OpenArchaeo, a user-friendly web portal to query different triplestores dataset, and OpenTermAlign, a web interface to align an unstructured vocabulary with a standardized vocabulary. The MASA consortium suggests that these two applications join the tools of the ARIADNEplus infrastructure."*

Other interesting statements

- *"There is no archaeological repository in Turkey. Answers shows my knowledge about archaeological research data in Turkey."*
- *"Archaeology is an interesting context for knowledge management research enabling me to discover interesting spacial, temporal and relational patterns from data."*

5 Outlook: Next user needs survey phase

This report will be updated at the end of third project year (in month 34 of 48) with results of studies on important specific topics identified in the online survey and thereafter in the communication with members of the ARIADNEplus community and other stakeholders. Important topics may arise from the extended ARIADNEplus community network or relevant new EU policies or programmes.

Specific topics identified in the first survey round regarding the sharing and reuse of archaeological datasets include

- the need to understand better different purposes and practices of data reuse (and lack of reuse) in domains of archaeological research so that reuse can be fostered and supported effectively,
- the important role of data citation and potential ARIADNEplus services for citation tracking and metrics,

Special attention will be devoted to the services of the ARIADNEplus e-infrastructure. Important topics here concern the overall approach rather than particular services. The overall approach includes

- Cloud-based provision of end-user services on the D4Science platform,
- Use of Virtual Research Environments (VREs) which combine user services,
- Linked Data to integrate and exploit knowledge and data resources.

Development and adoption of services and tools from related technological markets, open source as well as commercial, need to be considered regarding the ARIADNEplus setup as well as particular services or tools. For Cloud-based platforms and services the main reference of course is the European Open Science Cloud.

The envisaged Cloud-based VREs, service bundles or particular services on the D4Science platform will require evaluation from the perspective of intended user groups, involving group members, who can also propose VREs for their research. In-depth studies on particular technology markets will be conducted alongside the design and development of the VREs.

Therefore, in the second round of the user needs survey a combination of special studies and user involvement methods (e.g. focus groups or interviews) will be employed.

The continued reviewing of strategic European frameworks, technological developments and user segments of the ARIADNEplus services will help keeping the project up-to-date and indicating to the project bodies any necessary adjustment to the project planning, where necessary.

6 References

Adamick J. & Reznik-Zellen R. (2010a): Representation and Recognition of Subject Repositories. In: D-Lib Magazine, 16 (9/10), http://www.dlib.org/dlib/september10/adamick/09adamick.html

Adamick J. & Reznik-Zellen R. (2010b): Trends in Large-Scale Subject Repositories. In: D-Lib Magazine, 16 (11/12), http://www.dlib.org/dlib/november10/adamick/11adamick.html

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

Al-Barashdi H. & Al-Karousi R. (2018): Big Data in academic libraries: literature review and future research directions. In: Journal of Information Studies & Technology, 2018.2, https://doi.org/10.5339/jist.2018.13

Aloia N., Binding C., Cuy S., Doerr M., et al. (2017): Enabling European Archaeological Research: The ARIADNE E-Infrastructure. In: Internet Archaeology 43, https://doi.org/10.11141/ia.43.11

Altschul J.H., Kintigh K.W., Klein T.H, Doelle W.H., Hays-Gilpin K.A., et al. (2017): Fostering synthesis in archaeology to advance science and benefit society. In: PNAS, 114(42): 10999-11002, http://www.pnas.org/content/pnas/114/42/10999.full.pdf

Altschul J.H., Kintigh K.W., Klein T.H, Doelle W.H., Hays-Gilpin K.A., et al. (2018): Fostering Collaborative Synthetic Research in Archaeology. In: Advances in Archaeological Practice, 6(1): 19-29, https://doi.org/10.1017/aap.2017.31

Alvaro E., Brooks H., Ham M., Poegel S. & Rosencrans S. (2011): E-science librarianship: Field undefined. In: Issues in Science and Technology Librarianship, 66(11), http://www.istl.org/11-summer/article1.html

Archaeology Data Service & Digital Antiquity: Guides to Good Practice, http://guides.archaeologydataservice.ac.uk

Archaeology Data Service: DataTrain, http://archaeologydataservice.ac.uk/learning/DataTrain

ARCHES - Archaeological Resources in Cultural Heritage (2014): The Standard and Guide to Best Practice in Archaeological Archiving in Europe, online, http://archaeologydataservice.ac.uk/arches/Wiki.jsp?page=Main

ARIADNE (2014): First Report on Users' Needs, D2.1, April 2014, http://legacy.ariadne-infrastructure.eu/resources-2/deliverables/

ARIADNE (2015): Second Report on Users' Needs. Project deliverable D2.2, February 2015, http://legacy.ariadne-infrastructure.eu/resources-2/deliverables/

Armbruster C. & Romary L. (2010): Comparing Repository Types. In: International Journal of Digital Library Systems, 1(4): 61-73; manuscript: http://arxiv.org/ftp/arxiv/papers/1003/1003.4187.pdf

Assante M., Candela L., Castelli D. & Tani A. (2016). Are Scientific Data Repositories Coping with Research Data Publishing? In: Data Science Journal, 15, p.6, http://doi.org/10.5334/dsj-2016-006

Auckland, Mary (2012): Re-Skilling for Research: An Investigation into the Role and Skills of Subject and Liaison Librarians Required to Effectively Support the Evolving Information Needs of Researchers. London: Research Libraries UK, http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf

Austin T. & Mitcham J. (2007): Preservation and Management Strategies for Exceptionally Large Data Formats: 'Big Data'. English Heritage / Archaeology Data Service, 28 September 2007, https://archaeologydataservice.ac.uk/resources/attach/bigdata_final_report_1.3.pdf

Baynes, Grace (2018): From Green Shoots to "Grassroots": How Can We Accelerate Data Sharing?, pp. 14-17, in Figshare (2018): The State of Open Data 2018. October 2018, https://doi.org/10.6084/m9.figshare.7195058

Beagrie N. and Houghton J. 2013 The value and impact of the Archaeology Data Service. A study and methods for enhancing sustainability, Charles Beagrie Ltd and Victoria University, September 2013, http://repository.jisc.ac.uk/5509/

Beck A. & Neylon C. (2012): A vision for Open Archaeology. In: World Archaeology, 44(4): 479-497, http://www.tandfonline.com/toc/rwar20/44/4#.VEPIOskucyg

Beck, Anthony (2013): Opening access to heritage resources: risk, opportunity or paradigm shift?, pp. 40-47, in: Archäologie und Informationssysteme, Arbeitshefte 42, http://www.landesarchaeologen.de/fileadmin/Dokumente/News/NI_Archaeologie_und_Informationssysteme_Arbeitsheft42.pdf#page=4

Berners-Lee, Tim (2006[2010]): Linked Data; [added 2010] Is your Linked Open Data 5 Star? w3.ogr, http://www.w3.org/DesignIssues/LinkedData.html

Big Data Alliance (n.d.): What is Big Data?, http://www.bigdata-alliance.org/what-is-big-data/

Big Data in Archaeology: Practicalities and Possibilities, McDonald Institute for Archaeological Research, University of Cambridge, 27-28 March 2019, https://erikgjesfjeld.wixsite.com/big-data-archaeology

Bishop, Dorothy (2015): Who's afraid of Open Data: Scientists' objections to data sharing don't stand up to scrutiny. In: London School of Economics, Impact Blog, 16 December 2015, https://blogs.lse.ac.uk/impactofsocialsciences/2015/12/16/whos-afraid-of-open-data-dorothy-bishop/

Borgman, Christine (2015): If the Data Sharing is the Answer, what is the Question? In: ERCIM News 100, January 2015, https://ercim-news.ercim.eu/en100/special/if-data-sharing-is-the-answer-what-is-the-question

Borgman, Christine L. (2010): Research Data: Who will share what, with whom, when, and why? Fifth China – North America Library Conference 2010, Beijing, 8-12 September 2010, http://works.bepress.com/borgman/238

Borgman, Christine L. (2012): The Conundrum of Sharing Research Data. In: Journal of the American Society for Information Science and Technology, 63(6): 1059-1078, http://dx.doi.org/10.1002/asi.22634

Bryant R., Lavoie B. & Malpas C. (2017): The Realities of Research Data Management. Dublin, Ohio, USA: OCLC Research, https://doi.org/10.25333/C3PG8J

Burnette M.H., Williams S.C. & Imker H.J. (2016): From Plan to Action: Successful Data Management Plan Implementation in a Multidisciplinary Project. In: Journal of EScience Librarianship, 5(1), 6, https://doi.org/10.7191/jeslib.2016.1101

Burton A., Koers H., Manghi P., Stocker M., et al. (2017): The Scholix Framework for Interoperability in Data-Literature Information Exchange. In: D-Lib Magazine, 23(1/2), https://doi.org/10.1045/january2017-burton

CAA International - Computer Applications and Quantitative Methods in Archaeology (CAA), http://caa-international.org

Callaghan, Sarah (2019): Research Data Publication: Moving Beyond the Metaphor. In: Data Science Journal, 18(1), p.39, http://doi.org/10.5334/dsj-2019-039

Candela L., Castelli D., Manghi P. & Tani A. (2015): Data Journals: A Survey. In: Journal of the Association for Information Science and Technology, 66 (1): 1747-62, doi:10.1002/asi.23358; https://zenodo.org/record/18377/files/DataJournalsSurvey.pdf

Carpenter, Todd (2017): What Constitutes Peer Review of Data? A Survey of Peer Review Guidelines. In: The Scholarly Kitchen (Society for Scholarly Publishing), 11 April 2017, https://scholarlykitchen.sspnet.org/2017/04/11/what-constitutes-peer-review-research-data/

Center for Knowledge Infrastructures, UCLA Department of Information Studies, https://knowledgeinfrastructures.gseis.ucla.edu

Chavan V. & Penev L. (2011): The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics, Vol.12, Suppl 15: S2, http://doi.org/10.1186/1471-2105-12-S15-S2

Chiarelli A. & Johnson R. (2017): Making the case for Research Data Management - Case Studies. JISC, June 2017, http://doi.org/10.5281/zenodo.817936

CLIR - Council on Library and Information Resources (2009): Working Together or Apart: Promoting the Next Generation of Digital Scholarship. Washington: CLIR, http://www.clir.org/pubs/reports/pub145/pub145.pdf

CLIR - Council on Library and Information Resources (2013): Research Data Management. Principles, Practices, and Prospects. DataRes project report. Washington: CLIR, http://www.clir.org/pubs/reports/pub160

Coalition for Archaeological Synthesis (CfAS), USA, http://www.archsynth.org

COAR - Confederation of Open Access Repositories (2018): Results of COAR membership survey. COAR website, 4 October 2018, https://www.coar-repositories.org/news-updates/results-of-coar-membership-survey/

Cooper A. & Green C. (2015): Embracing the complexities of "Big Data" in archaeology: the case of the English Landscape and Identities Project. In: Journal of Archaeological Method and Theory, 22(1); https://ora.ox.ac.uk/objects/uuid:7253bc04-93ca-451f-814c-d57e7ddccb02

Corrall S., Kennan M.A. & Afzal W. (2013): Bibliometrics and research data management: emerging trends in library research support services. In: Library Trends, 61(3), http://muse.jhu.edu/login?auth=0&type=summary&url=/journals/library_trends/v061/61.3.corrall02.html

Corti L., Van den Eynden V., Bishop L. & Woollard M. (2013): Managing and sharing research data: A guide to good practice. London: SAGE

Costa S., Beck A., Bevan A. & Ogden J. (2012): Defining and advocating Open Data in archaeology, pp. 449-456, in: Proceedings of CAA 2012 Southampton, Amsterdam University Press, http://dare.uva.nl/cgi/arno/show.cgi?fid=516092

Costello, Mark J. (2009): Motivating Online Publication of Data. In: BioScience, 59(5): 418-427, https://doi.org/10.1525/bio.2009.59.5.9

Cousijn H., Feeney P., Lowenberg D., Presani E. & Simons N. (2019): Bringing Citations and Usage Metrics Together to Make Data Count. In: Data Science Journal, 18(1), p.9, http://doi.org/10.5334/dsj-2019-009

Cousijn H., Kenall A., Ganley E., Harrison M., et al. (2018): A data citation roadmap for scientific publishers. In: Scientific Data, Volume 5, Article number: 180259, https://doi.org/10.1038/sdata.2018.259

Cox A., Kennan M.A., Lyon L. & Pinfield S. (2017): Developments in research data management in academic libraries: Towards an understanding of research data service maturity. In: Journal of the Association for Information Science and Technology, 68(9): 2182‑2200, https://doi.org/10.1002/asi.23781

Cox A.M, Verbaan E. & Sen B. (2012): Upskilling liaison librarians for research data management. In: Ariadne, Issue 70, http://www.ariadne.ac.uk/issue70/cox-et-al

Cox A.M. & Pinfield S. (2013): Research data management and libraries: Current activities and future priorities. In: Journal of Librarianship and Information Science, June 2013, http://eprints.whiterose.ac.uk/76107/7/WRRO_76107.pdf

Curdt, Constanze (2019): Supporting the Interdisciplinary, Long-Term Research Project 'Patterns in Soil-Vegetation-Atmosphere-Systems' by Data Management Services. In: Data Science Journal, 18(1), https://doi.org/10.5334/dsj-2019-005

D4Science (Institute of Information Science and Technologies, ISTI-CNR, Pisa, Italy), https://www.d4science.org

Dallmeier-Tiessen S., Darby R., Gitmans K., Lambert S., et al. (2014): Enabling Sharing and Reuse of Scientific Data. In: New Review of Information Networking, 19(1): 16-43; http://centaur.reading.ac.uk/36863/1/ODE-NRIN-Paper_accepted.pdf

Danish Agency for Science and Higher Education (2018): Preliminary analysis: Introduction of FAIR data in Denmark. Study by Oxford Research and Højbjerre Brauer Schultz. March 2018, https://ufm.dk/en/publications/2018/preliminary-analysis-introduction-of-fair-data-in-denmark

Data Archiving and Networked Services (DANS): e-Depot Dutch Archaeology (eDNA), http://www.edna.nl

Digital Repository of Ireland (DRI), https://www.dri.ie

Digital Science (Holtzbrinck), https://www.digital-science.com

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-use of Public Sector Information (PE/28/2019/REV/1), http://data.europa.eu/eli/dir/2019/1024/oj

Directory of Open Access Repositories (OpenDOAR), https://v2.sherpa.ac.uk/opendoar/

DISCO - Discovering the Archaeologists of Europe 2014 Transnational Report, York Archaeological Trust, November 2014. Available http://www.discovering-archaeologists.eu/national_reports/2014/transnational_report.pdf

DRYAD (repository), https://datadryad.org

e-InfraNet (2013): "Open" as the default modus operandi for research and higher education. Prepared by L. van der Vaart, et al. e-InfraNet (ERA-NET) project, https://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2013/e-InfraNet-Open-as-the-Default-Modus-Operandi-for-Research-and-Higher-Education.pdf

e-IRG - e-Infrastructures Reflection Group (2013): White Paper 2013. The Hague, The Netherlands, http://e-irg.eu/white-papers

e-IRG (2018): The EOSC Stakeholder Forum Shaping the future of the European Open Science Cloud (EOSC). In: e-IRG Magazine, Spring 2018, p.3, http://e-irg.eu/documents/10920/12770/e-IRG+Magazine+2018-1+Online+version+%284+Mbyte%29.pdf

Elsevier: Data in Brief (Journal) https://www.journals.elsevier.com/data-in-brief/

Enke N., Thessen A., Bach K., Bendix J., Seeger B. & Gemeinholzer B. (2012): The user's view on biodiversity data sharing - Investigating facts of acceptance and requirements to realize a

sustainable use of research data. In: Ecological Informatics, 11, 25–33. https://doi.org/10.1016/j.ecoinf.2012.03.004

EOSC Executive Board (2019): European Open Science Cloud (EOSC) Strategic Implementation Plan. European Commission, DG Research and Innovation, Brussels, June 2019, http://dx.doi.org/10.2777/202370

EOSC Portal, https://www.eosc-portal.eu

EOSC Secretariat, https://www.eoscsecretariat.eu

EOSCpilot Science Demonstrators, http://eoscpilot.eu/science-demonstrators

European Association of Archaeologists (EAA), https://www.e-a-a.org

European Commission (2013): Digital science in Horizon 2020. Concept paper, DG Connect, https://ec.europa.eu/digital-single-market/en/open-science

European Commission (2014): Public consultation on 'Science 2.0': Science in Transition. European Commission, DG Connect, http://ec.europa.eu/research/consultations/science-2.0/consultation_en.htm

European Commission (2016a): Communication on European Cloud Initiative - Building a competitive data and knowledge economy in Europe. COM/2016/0178 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016DC0178

European Commission (2016b): Open Innovation, Open Science, Open to the World – a vision for Europe. European Commission. DG Research and Innovation. Luxembourg: Publications Office of the European Union, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=16022

European Commission (2016c): European Charter for Access to Research Infrastructures. Principles and Guidelines for Access and Related Services. European Commission, DG Research & Innovation. Luxembourg: Publications Office of the European Union, https://ec.europa.eu/research/infrastructures/pdf/2016_charterforaccessto-ris.pdf

European Commission (2016d). Guidelines on FAIR Data Management in Horizon 2020. Version 3.0, European Commission, DG Research and Innovation, 26 July 2016, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Commission (2018a): Commission Staff Working Document - Implementation Roadmap for the European Open Science Cloud. SWD(2018) 83 final, Brussels, 14.3.2018, https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf#view=fit&pagemode=none

European Commission (2018b): Commission Recommendation (EU) 2018/790 of 25 April 2018 on Access to and Preservation of Scientific Information, C/2018/2375, http://data.europa.eu/eli/reco/2018/790/oj

European Commission (2018c): Commission Staff Working Document - Accompanying the document: Commmission Recommendation on access to and preservation of scientific information, C(2018)2375 final. - SWD(2018) 123 final, Brussels, 25 April 2018, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51637

European Commission (2018d): Second High Level Expert Group on the European Open Science Cloud: Prompting an EOSC in practice. Final Report and Recommendation, November 2018, Luxembourg: Publications Office of the European Union, http://dx.doi.org/10.2777/112658

European Commission (2018e): Commission launches European Open Science Cloud (23 November 2018), https://ec.europa.eu/info/news/commission-launches-european-open-science-cloud-2018-nov-23_en

European Commission: Digital Single Market - The European Cloud Initiative, https://ec.europa.eu/digital-single-market/en/%20european-cloud-initiative

European Commission: Digital Single Market policy: Open Science, http://ec.europa.eu/digital-agenda/en/open-science

Expert Group on FAIR Data (2018): Turning FAIR into reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data, November 2018, Luxembourg: Publications Office of the European Union,http://dx.doi.org/10.2777/1524

FAIMS - Federated Archaeological Information Management System (Australia), https://www.fedarch.org

Faniel I., Kansa E., Whitcher-Kansa S., Barrera-Gomez J. & Yakel E. (2013): The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. JCDL 2013 Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, 295-304 (preprint), http://www.oclc.org/content/dam/research/publications/library/2013/faniel-archae-data.pdf

Fearon D., Gunia B., Pralle B.E., Lake S. & Sallans A.L. (2013): Research Data Management Services. SPEC Kit 334. Washington, DC: Association of Research Libraries, https://doi.org/10.29242/spec.334

Fecher B. & Friesike S. (2014): Open Science: One Term, Five Schools of Thought. In: Bartling S. & Friesike S. (eds.): Opening Science. The Evolving Guide on How the Web is Changing Research, Collaboration and Scholarly Publishing, http://book.openingscience.org

Fecher B., Friesike S., Hebing M., Linek S. & Sauermann A. (2015): A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing. DIW Berlin, Discussion Paper, No. 1454, February 2015, http://www.hiig.de/wp-content/uploads/2015/02/dp1454.pdf

Fenner M., Lowenberg D., Jones M., Needham P., et al. (2018): Code of practice for research data usage metrics release. In: PeerJ Preprints, 6(e26505v1), https://doi.org/10.7287/peerj.preprints.26505v1

Figshare (2016): The State of Open Data 2016, https://dx.doi.org/10.6084/m9.figshare.4036398; Data: https://figshare.com/articles/Open_Data_Survey/4010541

Figshare (2017): The State of Open Data 2017, https://figshare.com/articles/_/5481187; Interactive visualisation: https://knowledge.figshare.com/articles/item/state-of-open-data-2017; Dataset: https://figshare.com/articles/State_of_Open_Data_survey_2017/5480710

Figshare (2018): The State of Open Data 2018, https://doi.org/10.6084/m9.figshare.7195058; Interactive visualisation, https://knowledge.figshare.com/articles/item/state-of-open-data-2018; Dataset: https://ndownloader.figshare.com/files/13320101;

Franck, Gwen (2019): Insights into the Economy of Open Scholarship: A Collection of Interviews. Knowledge Exchange, May 2019, http://doi.org/10.5281/zenodo.2840171

Frank R., Kriesberg A., Yakel E. & Faniel I.M. (2015): Looting hoards of gold poaching spotted owls: Data confidentiality among archaeologists & zoologists. In: Proceedings of the Association for Information Science and Technology, 52(1): 1-10, http://doi.org/10.1002/pra2.2015.145052010037

Frosini L., Bardi A., Manghi P. & Pagano P. (2018): An aggregation framework for digital humanities infrastructures: The Parthenos experience. In: CIRES*it*, 8(1): 33-50. http://doi.org/10.2423/i22394303v8n1p33

Gartner Glossary (n.d.), http://www.gartner.com/it-glossary/big-data

Gattiglia, Gabriele (2015): Think big about data: Archaeology and the Big Data challenge. In: Archäologische Informationen, Band 38: Fokus: Open Access & Open Data, early view: http://www.dguf.de/fileadmin/AI/ArchInf-EV_Gattiglia.pdf

GenBank, http://www.ncbi.nlm.nih.gov/genbank/

Geser, Guntram (2019a). ARIADNEplus and community data repositories. Innovative solutions for sharing open archaeological data. In: CHNT'24 - International Conference on Cultural Heritage and New Technologies, Vienna, 2019, https://www.chnt.at/wp-content/uploads/ARIADNEplus-and-community-data-repositories.pdf

Geser, Guntram (2019b). Innovation and Impact of the ARIADNE Initiative, pp. 187-199, in: Richards J. & Niccolucci F. (eds.): The ARIADNE Impact. Budapest: Archaeolingua, http://doi.org/10.5281/zenodo.3476712

GO FAIR, https://www.go-fair.org

Gow A. & Molloy L. (2014): Ahead of the CurV: Digital Curator Vocational Education. In: C. Mills, M. Pidd & E. Ward (eds.): Proceedings of the Digital Humanities Congress 2012. Studies in the Digital Humanities. Sheffield: HRI Online Publications, http://www.hrionline.ac.uk/openbook/chapter/dhc2012-gow

Grabus S. & Greenberg J. (2019): The Landscape of Rights and Licensing Initiatives for Data Sharing. In: Data Science Journal, 18(1), p.29, http://doi.org/10.5334/dsj-2019-029

Harley D., Acord S., Earl-Novell S., Lawrence S. & King C.J. (2010): 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

Huang X., Hawkins B.A. & Qiao G.(2013): Biodiversity Data Sharing: Will Peer-Reviewed Data Papers Work? In: BioScience, 63(1): 5-6, https://doi.org/10.1525/bio.2013.63.1.2

Huang X., Hawkins B.A., Lei F., Miller G.L., et al. (2012): Willing or unwilling to share primary biodiversity data: results and implications of an international survey. In: Conservation Letters, 5(5): 399-406, http://doi.org/10.1111/j.1755-263X.2012.00259.x

Huggett, Jeremy (2017): Citing Data Reuse. In: Introspective Digital Archaeology (blog), 23 May 2017, https://introspectivedigitalarchaeology.wordpress.com/2017/05/23/citing-data-reuse/

Huggett, Jeremy (2018): Reuse Remix Recycle: Repurposing Archaeological Digital Data. In: Advances in Archaeological Practice, 6(2): 93-104, doi: 10.1017/aap.2018.1; final accepted version, http://eprints.gla.ac.uk/161510/1/161510.pdf

Hyams, Elspeth (2008): Data librarianship: a gap in the market. In: Update magazine, 7(6), http://www.era.lib.ed.ac.uk/handle/1842/2499

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

IANUS: Data portal, http://datenportal.ianus-fdz.de

Ivanović D., Schmidt B., Grim R. & Dunning A. (2019): FAIRness of Repositories & Their Data. A Report from LIBER's Research Data Management Working Group. Retrieved from https://doi.org/10.5281/zenodo.3251593

Jacoby, JoAnn (2010): Share and share alike? Data-sharing practices in different disciplinary domains, pp. 77-94, in: Witt S.W. & Rudasill L.M. (eds.): Social Science Libraries. Interdisciplinary Collections, Services, Networks. De Gruyter, https://doi.org/10.1515/9783110232158.69

Jones S., Pryor G. & Whyte A. (2013): How to Develop Research Data Management Services - a guide for HEIs. DCC How-to Guides. Edinburgh: Digital Curation Centre, http://www.dcc.ac.uk/resources/how-guides

Kansa E., 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, Eric (2012): Openness and archaeology's information ecosystem. In: World Archaeology, 44(4): 498-520; preprint, http://escholarship.org/uc/item/4bt04063

Kansa, Eric (2014): Open Context and Linked Data. ISAW Papers 7.10, http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/kansa/

Kansa, Eric (2015): Contextualizing Digital Data as Scholarship in Eastern Mediterranean Archaeology. In: CHS Research Bulletin, 3(2), http://nrs.harvard.edu/urn-3:hlnc.essay:KansaE.Contextualizing_Digital_Data_as_Scholarship.2015

Kervin K.E, Michener W.K. & Cook R.B. (2013): Common Errors in Ecological Data Sharing. In: Journal of eScience Librarianship, 2(2): e1024, https://doi.org/10.7191/jeslib.2013.1024

Kim, Youngseek (2017): Fostering scientists' data sharing behaviours via data repositories, journal supplements, and personal communication methods. In: Information Processing & Management, 30(4): 871-885, https://doi.org/10.1016/j.ipm.2017.03.003

Kintigh K., Altschul J., Kinzig A.P., Limp F., et al. (2015): Cultural Dynamics, Deep Time, and Data: Planning Cyberinfrastructure Investments for Archaeology. In: Advances in Archaeological Practice, 3(1): 1-15, https://www.cni.org/wp-content/uploads/2015/02/archaeology-CI.pdf

Kintigh K.W., Altschul J.H., Beaudry M.C., et al. (2014): Grand challenges for archaeology. In: American Antiquity, 79(1): 5-24, http://www2.lawrence.edu/fast/PEREGRIP/Publications/2014AmAntiqGrandChallenges.pdf

Kitchin R. & McArdle G. (2016): What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. In: Big Data & Society, 3(1), https://doi.org/10.1177/2053951716631130

Kramer D.B., Hartter J., Boag A.E., Jain M., Stevens K., Nicholas K.A., McConnell W.J. & Liu J. (2017): Top 40 questions in coupled human and natural systems (CHANS) research. In: Ecology and Society 22(2):44, https://doi.org/10.5751/ES-09429-220244

Kratz J. & Strasser C. (2014): Data publication consensus and controversies. F1000Research 3: 94, http://dx.doi.org/10.12688/f1000research.4518

Kratz J. & Strasser C. (2015a): Researcher Perspectives on Publication and Peer Review of Data. In: PLOS ONE, 10(4), e0123377, https://doi.org/10.1371/journal.pone.0117619

Kratz J. & Strasser C. (2015b): Making data count. In: Scientific Data, Vol. 2, 150039, http://doi.org/10.1038/sdata.2015.39

Lake, Mark (2012): Open Archaeology. In: World Archaeology, 44(4: special issue), 471-478, http://dx.doi.org/10.1080/00438243.2012.748521

Laney, Doug (2001): 3D data management: Controlling data volume, velocity and variety. Meta Group, 6 February 2001, http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lawrence B., Jones C., Matthews B., Pepler S. & Callaghan S. (2011): Citation and peer review of data: moving towards formal data publication. In: International Journal for Digital Curation, 6(2): 4-37, https://doi.org/10.2218/ijdc.v6i2.205

LeClere, Felicia (2010): Too many researchers are reluctant to share their data. In: The Chronicle of Higher Education, 3 August 2010, http://chronicle.com/article/Too-Many-Researchers-Are/123749

Lee D.J. & Stvilia B. (2017): Practices of research data curation in institutional repositories: A qualitative view from repository staff. In: PLoS ONE 12(3): e0173987, https://doi.org/10.1371/journal.pone.0173987

Li K., Greenberg J. & Dunic J. (2019): Data objects and documenting scientific processes: An analysis of data events in biodiversity data papers. arXiv:1903.06215, 14 March 2019, https://arxiv.org/abs/1903.06215

Li Z. & Eichmann-Kalwara N. (2019): Mapping the Scholarly Literature Found in Scopus on "Research Data Management": A Bibliometric and Data Visualization Approach. In: Journal of Librarianship and Scholarly Communication, 7(1), eP2226, http://doi.org/10.7710/2162-3309.2266

LIBER - Association of European Research Libraries (2012): Ten recommendations for libraries to get started with research data management. Birte Christensen-Dalsgaard et al., final report of the Working group on E-Science / Research Data Management, July 2012, http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf

LIBER - Association of European Research Libraries (2014): Research Data Management Case Studies. LIBER Scholarly Communication and Research Infrastructures Steering Committee, June 2014, https://libereurope.eu/strategy/previous/2013-2017/strategic-direction-1-enable-open-science/wg-scientific-information-infrastructures/research-data-management-case-studies/

LIBER - Association of European Research Libraries (2015): Libraries & Research Data: Towards a new leadership role, https://libereurope.eu/wp-content/uploads/2015/11/LIBER-Libraries-and-Research-Data-factsheet1.pdf

Make-Data-Count (2018): Implementing the COUNTER Code of Practice for Research Data in Repositories. Available at Github, https://github.com/CDLUC3/Make-Data-Count/blob/master/getting-started.md

Mappa (repository, University of Pisa, Italy), https://www.mappaproject.org

Marwick B. & Pilaar Birch S.E. (2018): A Standard for the Scholarly Citation of Archaeological Data as an Incentive to Data Sharing. In: Advances in Archaeological Practice, 6(2): 125-143, https://doi.org/10.1017/aap.2018.3

Marwick, Ben et al. (2017): Open Science in Archaeology. In: The SAA Archaeological Record, September 2017, pp. 8-13; https://www.researchgate.net/publication/320068431

Masum H., Rao A., Good B.M., Todd M.H., et al. (2013): Ten Simple Rules for Cultivating Open Science and Collaborative R&D. In: PLoS Computational Biology, 9(9): e1003244. https://doi.org/10.1371/journal.pcbi.1003244

McManamon F.P., Kintigh K.W., Ellison L.A., and Brin A. (2017): tDAR: A cultural heritage archive for twenty-first-century public outreach, research, and resource management. In: Advances in Archaeological Practice, 5, Special Issue 3, 238-249, https://doi.org/10.1017/aap.2017.18

Moedas, Carlos (2015): Open Innovation, Open Science, Open to the World. Speech at the 'Opening up to an ERA of Innovation' Conference, Brussels, 22 June 2015, http://europa.eu/rapid/press-release_SPEECH-15-5243_en.htm

Moedas, Carlos (2016): Forword by Commissioner Carlos Moedas, p.4, in: Realising the European Open Science Cloud: First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud. European Commission, Brussels, http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

Mons B., Neylon C., Velterop J., Dumontier M., da Silva Santos L.O.B. & Wilkinson M. (2017): Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. In: Information Services & Use, 37(1): 49-56, http://content.iospress.com/articles/information-services-and-use/isu824

Mons, Barend (2016): We need 500.000 respected data stewards to operate the European Open Science Cloud: Interview with Barend Mons. In: e-IRG news blog, 4 May 2016, http://e-irg.eu/news-blog/-/blogs/we-need-500-000-respected-data-stewards-to-operate-the-european-open-science-cloud

Mooney, Hailey (2016): Scholarly Communication and Data, pp. 195-218, in: Kellem L.M. & Thompson K. (eds.): Data Librarianship: The academic data librarian in theory and practice. Chicago: Association of College and Research Libraries. Available at http://hdl.handle.net/2027.42/120919

Morgan H. & Janke A. (2017). DMRs, making DMPs relevant again. In: ANDS Central weblog, 22 May 2017, http://andscentral.blogspot.com/2017/05/dmrs-making-dmps-relevant-again.html

Muñoz, Trevor (2013): Data Curation as Publishing for the Digital Humanities. In: Journal of Digital Humanities, 2(3), http://journalofdigitalhumanities.org/2-3/data-curation-as-publishing-for-the-digital-humanities/

Nelson, Bryn (2009): Data sharing: Empty archives. In: Nature 461: 160-163, http://doi.org/10.1038/461160a

Niccolucci F. & Richards J. (2019): ARIADNE and ARIADNEplus, pp. 7-25, in: The ARIADNE Impact. Budapest: Archaeolingua, http://doi.org/10.5281/zenodo.3476712

Nicholas D., Watkinson A., Jamali H.R., Herman E., et al. (2015): Peer review: still king in the digital age. In: Learned Publishing, 28(1): 15-21, https://doi.org/10.1087/20150104

Ogier A.L., Brown A.M., Petters J., Hilal A. & Porter N. (2018): Enhancing Collaboration Across the Research Ecosystem: Using Libraries as Hubs for Discipline-Specific Data Experts. In: Practice and Experience in Advanced Research Computing. DOI: https://doi.org/10.1145/3219104.3219126

Open Context (Alexandria Archive Institute, USA), https://opencontext.org

Open Knowledge Foundation: The Open Definition, https://opendefinition.org

Open Research Data Task Force (2018): Case Studies, July 2018, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775379/Case-studies-ORDTF-July-2018.pdf

Ortman S.G., Cabaniss A.F.H., Sturm J.O. & Bettencourt L.M.A. (2014): The Pre-History of Urban Scaling. In: PLoS ONE, 9(2): e87902, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0087902

Paludan-Müller, Carsten (2015): The European Association of Archaeologists. Some thoughts for the future of our Association. In: The European Archaeologist, Issue 43: 5-9, https://www.e-a-a.org/EAA/Publications/TEA/Archive/EAA/Navigation_Publications/TEA_content/Archive.aspx#45

Pampel H., Vierkant P., Scholze F., Bertelmann R., et al. (2013): Making Research Data Repositories Visible: The re3data.org Registry. In: PLoS One, 8(11): e78080, http://doi.org/10.1371/journal.pone.0078080

PANGAEA - Data Publisher for Earth & Environmental Science, http://www.pangaea.de

Parham S.W., Carlson J., Hswe P., Westra B. & Whitmire A. (2016). Using data management plans to explore variability in research data management practices across domains. International Journal of Digital Curation, 11(1), 53–67, https://doi.org/10.2218/ijdc.v11i1.423

PARSE.Insight (2009): Insight into digital preservation of research output in Europe. Project deliverable D3.4: Survey Report, 9 December 2009, https://libereurope.eu/wp-content/uploads/2010/01/PARSE.Insight.-Deliverable-D3.4-Survey-Report.-of-research-output-Europe-Title-of-Deliverable-Survey-Report.pdf

Parsons M.A. & Fox P.A. (2013): Is data publication the right metaphor? In: Data Science Journal, 12, pp.WDS32–WDS46, http://doi.org/10.2481/dsj.WDS-042

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

Pasquetto I.V., Randles B.M. & Borgman C.L. (2017): On the Reuse of Scientific Data. In: Data Science Journal, 16.8, http://doi.org/10.5334/dsj-2017-008

Pearce N. & Smith A.H. (2011): Data sharing: not as simple as it seems. In: Environmental Health, 10:107, http://doi.org/10.1186/1476-069X-10-107

Perrino T., Sperling A., Beardslee W., Sandler I. et al. (2013): Advancing Science Through Collaborative Data Sharing and Synthesis. In: Perspectives on Psychological Science, 8(4): 433-444, https://doi.org/10.1177/1745691613491579

Petters J.L., Brooks G.C., Smith J.A. & Haas C.A. (2019). The Impact of Targeted Data Management Training for Field Research Projects – A Case Study. In: Data Science Journal, 18(1), 43, http://doi.org/10.5334/dsj-2019-043

Pierce H.H., Dev A., Statham E. & Bierer B.E. (2019): Credit data generators for data reuse. In: Nature, 570:7759, 30-32, http://doi.org/10.1038/d41586-019-01715-4

Piwowar H., Priem J., Larivière V. et al. (2018): The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. In: PeerJ, 6: e4375, http://doi.org/10.7717/peerj.4375

Piwowar H.A. & Vision T.J. (2013): Data reuse and the open data citation advantage. In: PeerJ, 1:e175, http://dx.doi.org/10.7717/peerj.175

Popkin, Gabriel (2019): Data sharing and how it can benefit your scientific career. In: Nature, 13 May 2019, https://www.nature.com/articles/d41586-019-01506-x

PricewaterhouseCoopers EU Services (2018a): Cost of not having FAIR research data. Cost-Benefit analysis for FAIR research data. March 2018, http://doi.org/10.2777/02999

PricewaterhouseCoopers EU Services (2018b): Policy Recommendations. Cost-Benefit analysis for FAIR research data. May 2018, http://doi.org/10.2777/706548

PricewaterhouseCoopers EU Services (2018c): Cost-Benefit analysis of becoming FAIR. European Commission, DG for Research and Innovation, May 2018. Guidebook: https://ec.europa.eu/info/files/cost-benefit-analysis-becoming-fair-guidebook_en; calculation tool: https://ec.europa.eu/info/files/cost-benefit-analysis-becoming-fair_en

Pryor G. & Donnelly M. (2009): Skilling up to do data: whose role, whose responsibility, whose career? In: International Journal of Digital Curation, 4(2): 158-170, http://ijdc.net/index.php/ijdc/article/view/126

Pryor G., Jones S. & Whyte A. (Eds, 2013): Delivering research data management services: Fundamentals of good practice. London: Facet.

Pryor, Graham (2009): Multi-scale Data Sharing in the Life Sciences: Some Lessons for Policy Makers. In: International Journal of Digital Curation, 4(3), https://doi.org/10.2218/ijdc.v4i3.115

Ray, Joyce (ed., 2014): Research Data Management: Practical Strategies for Information Professionals. Purdue University Press. Available at: https://www.cni.org/wp-content/uploads/2013/10/Research-Data-Mgt-Ch19-Lynch-Oct-29-2013.pdf

RIN - Research Information Network & Key Perspectives (2008): To Share or not to Share: Publication and Quality Assurance of Research Data Outputs. Main Report and Annex. RIN: London, http://eprints.soton.ac.uk/266742/1/Published_report_-_main_-_final.pdf

RIN - Research Information Network (2009): Patterns of information use and exchange: case studies of researchers in the life sciences. London: RIN & British Library, http://www.rin.ac.uk/system/files/attachments/Patterns_information_use-REPORT_Nov09.pdf

RIN & NESTA - Research Information Network & National Endowment for Science, Technology, and the Arts (2010): Open to All? Case Studies of Openness in Research. Report prepared by the Digital Curation Centre. London: RIN, http://apo.org.au/files/Resource/nesta-rin_open_science_v01.pdf

Rouder, Jeffrey N. (2015): The What, Why, and How of Born-Open Data. In: Behavior Research Methods, 48/3): 1062-69, https://doi.org/10.3758/s13428-015-0630-z

Rusbridge, Chris (2007): Create, curate, re-use: the expanding life course of digital research. Educause Australasia 07 conference paper, https://www.era.lib.ed.ac.uk/handle/1842/1731

Rusbridge, Chris (2008): Tomorrow, and tomorrow, and tomorrow: poor players on the digital curation stage, pp. 207-217, in: Earnshaw R. & Vince J. (eds.): Digital Convergence – Libraries of the Future. Springer; manuscript available at: https://www.era.lib.ed.ac.uk/handle/1842/2150

Sands, Rob (2009): Artefact Studies, Artefact Data and Information Technology: A point in time. University College Dublin, School of Archaeology, November 2009.

Sayogo D.S & Pardo T.A. (2013): Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. In: Government Information Quarterly, 30: 19-31, https://doi.org/10.1016/j.giq.2012.06.011

Schmidt B., Gemeinholzer B. & Treloar A. (2016): Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. In: PLoS ONE, 11(1): e0146695, https://doi.org/10.1371/journal.pone.0146695

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

SEADDA - Saving European Archaeology from the Digital Dark Age SEADDA (COST Action), https://www.seadda.eu

Shearer K. & Furtado F. (2017): COAR Survey of Research Data Management: Results. Confederation of Open Access Repositories (COAR), 24 January 2017, https://www.coar-repositories.org/files/COAR-RDM-Survey-Jan-2017.pdf

Silvello, Gianmaria (2018): Theory and practice of data citation. In: JASIST, 69(1): 6-20, doi: 10.1002/asi.23917; https://www.researchgate.net/publication/317400531

Smale N., Unsworth K., Denyer G. & Barr D. (2018). The History, Advocacy and Efficacy of Data Management Plans. BioRxiv, 17. October 2018, https://www.meta.org/papers/the-history-advocacy-and-efficacy-of-data/990443499

Sobotkova, Adela (2013): The Use of Information Technology in Australian Archaeology: the FAIMS Digital Data Survey Report. FAIMS project, 25 November 2013, https://fedarch.org/documents/DigitalDataSurveyReport.pdf

SPARC Europe & Digital Curation Centre (2019): An Analysis of Open Science Policies in Europe. August 2019, https://zenodo.org/record/3379705

SPARC Europe (2017): The Open Data Citation Advantage. Briefing Paper, 28 February 2017, https://sparceurope.org/download/1233/

SPARC Europe: The Open Access Citation Advantage Service, https://sparceurope.org/?page_id=978

Springer Nature (2018). Practical Challenges for Researchers in Data Sharing. Report prepared by D. Stuart, G. Baynes, I. Hrynaszkiewicz et al., March 2018, https://doi.org/10.6084/m9.figshare.5975011.v1

Strasser, Charly (2013): Closed Data… Excuses, Excuses. In: UC3- University of California Curation Center, 24 April 2013, https://uc3.cdlib.org/2013/04/24/closed-data-excuses-excuses/

Swedish National Data Service (SND, Sweden), http://snd.gu.se

tDAR - The Digital Archaeological Record (Digital Antiquity, Arizona State University, USA), http://www.tdar.org

Tenopir C., Allard S., Douglass K., Aydinoglu A.U. *et al*. (2011): Data Sharing by Scientists: Practices and Perceptions. In: PLoS ONE, 6(6): e21101, 29 June 2011, http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101

Tenopir C., Birch B. & Allard S. (2012): Academic libraries and research data services. Association of College & Research Libraries, http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Tenopir C., Dalton E.D., Allard S., Frame M., et al. (2015): Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. In: PLoS ONE, 10(8): e0134826, 26 August 2015, https://doi.org/10.1371/journal.pone.0134826; data in S1 Appendix, https://doi.org/10.1371/journal.pone.0134826.s001

The Royal Society (2012a): Science as an Open Enterprise. Final Report, June 2012, https://royalsociety.org/policy/projects/science-public-enterprise/Report/

TII - Transport Infrastructure Ireland (2017): Digital heritage collections launch. TII website, 19 June 2017, https://www.tii.ie/news/archaeology/tii-digital-heritage-coll/

Tsoukala V., Adoojan M. & Dechamp J.F. (eds., 2018): Access to and Preservation of Scientific Information in Europe. Report on the implementation of Commission Recommendation C(2012) 4890 final. European Commission, DG Research and Innovation, May 2018, http://doi.org/10.2777/642887

van de Sandt S., Dallmeier-Tiessen S., Lavasa A. & Petras V. (2019): The Definition of Reuse. In: Data Science Journal, 18(1), p.22, http://doi.org/10.5334/dsj-2019-022

Veletsianos G. & Kimmons R. (2012): Assumptions and Challenges of Open Scholarship. In: International Review of Research in Open and Distributed Learning, 13(4): 166-189, https://doi.org/10.19173/irrodl.v13i4.1313

Wallis J.C., Rolando E. & Borgman C.L. (2013): If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. In: PLoS ONE, 8(7): e67332, http://dx.doi.org/10.1371/journal.pone.0067332

Ward C., Freiman L., Jones S., Molloy L. & Snow K. (2011): Making Sense: Talking Data Management with Researchers. In: International Journal of Digital Curation, 6(2): 265-273, http://www.ijdc.net/index.php/ijdc/article/view/197

Wesson C.B. & Cottier J.W. (2014): Big Sites, Big Questions, Big Data, Big Problems: Scales of Investigation and Changing Perceptions of Archaeological Practice in the Southeastern United States. In: Bulletin of the History of Archaeology, 24:16, http://www.archaeologybulletin.org/article/view/bha.2416

Wilkinson M. D. *et al*. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: Scientific Data 3, 15 March 2016, http://www.nature.com/articles/sdata201618

Williams M., Bagwell J. & Zozus M. (2017): Data management plans: the missing perspective. In: Journal of Biomedical Informatics, Vol.71, 130-142, https://doi.org/10.1016/j.jbi.2017.05.004

Wilshusen R.H., Heilen M., Catts W., de Dufour K. & Jones B. (2016): Archaeological Survey Data Quality, Durability, and Use in the United States. Findings and Recommendations. In: Advances in Archaeological Practice, 4(2): 106-117, 10.7183/2326-3768.4.2.106, https://www.researchgate.net/publication/303463999

Wilson A.T. & Edwards B. (eds., 2015): Open Source Archaeology. Ethics and Practice. De Gruyter Open, http://www.degruyter.com/viewbooktoc/product/460080

Wittenberg J., Sackmann A. & Jaffe R. (2018): Situating Expertise in Practice: Domain-Based Data Management Training for Liaison Librarians. In: The Journal of Academic Librarianship, 44(3): 323-329, https://doi.org/10.1016/j.acalib.2018.04.004

ZENODO (CERN, repository), https://zenodo.org

Zilioli M., Lanucara S., Oggioni A., Fugazza C. & Carrara P. (2019): Fostering Data Sharing in Multidisciplinary Research Communities: A Case Study in the Geospatial Domain. In: Data Science Journal, 18(1), p.15, http://doi.org/10.5334/dsj-2019-015

Zimmerman, Ann (2008): New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. In: Science, Technology & Human Values, 33(5): 631-652, http://scientificdatasharing.com/wp-content/uploads/2011/02/Zimmerman_New-knowledge-from-old-data_2008.pdf