



## **ARIADNE First report on Users' Needs – summary**

The ARIADNE project (“Advanced Research Infrastructure for Archaeological Dataset Networking in Europe”) carried out research to understand users’ requirements with regard to archaeological research data to inform the infrastructure and services being developed.

Following a series of pilot interviews, ARIADNE ran two online surveys, one targeted at archaeological researchers and directors of research institutions, and the second targeted the managers or directors of data repositories. 692 researchers and 52 repository managers completed the questionnaire.

This document provides a summary of the main findings of the survey. The full report is freely available online at <http://www.ariadne-infrastructure.eu/Resources>.

### **Context**

In recent decades, research has become increasingly collaborative, distributed and data-intensive. There is the expectation that with large integrated datasets, e-infrastructure and tools, new scientific questions can be tackled.

In developing the ARIADNE infrastructure and services it is important for researchers, data managers and technology experts to be involved in defining what kind of e-infrastructures, tools and services fit their requirements.

The foundation of new knowledge in archaeology is mainly research projects, for example, a research excavation. The wider context includes national heritage agencies, local government curators and commercial archaeology services. Sophisticated digital tools and scientific equipment and techniques are used, with archaeologists producing much of their data themselves, in surveys, excavations, laboratory measurements and analysis of physical and biological finds. Other data is also used, such as remote sensing and imaging data, and there are a multitude of directories, catalogues, bibliographies, reference collections, text and image corpora, and digital editions relevant for archaeological researchers.

### **Deposition and sharing of research data**

Research data is collected, archived and maintained in a diverse data landscape. There are project-level repositories or databases, institutional repositories (research centres, museums and other), data centres, subject and domain repositories.

The survey explored how researchers deposit (and possibly) share the *new data* they produce. The answers indicated that data may not only be scattered across different institutional databases, but a good deal remains on researchers’ computers. Nearly half of respondents store data (at least temporarily) on their own computer for all or most projects, and about a third in a shared project archive or institutional server.

Three main challenges were identified for depositing and sharing data in an online repository:

- There is a perceived lack of professional recognition and reward for sharing the data;
- The work effort required to prepare data for deposit in a repository;
- In some cases there is a lack of suitable available repositories.



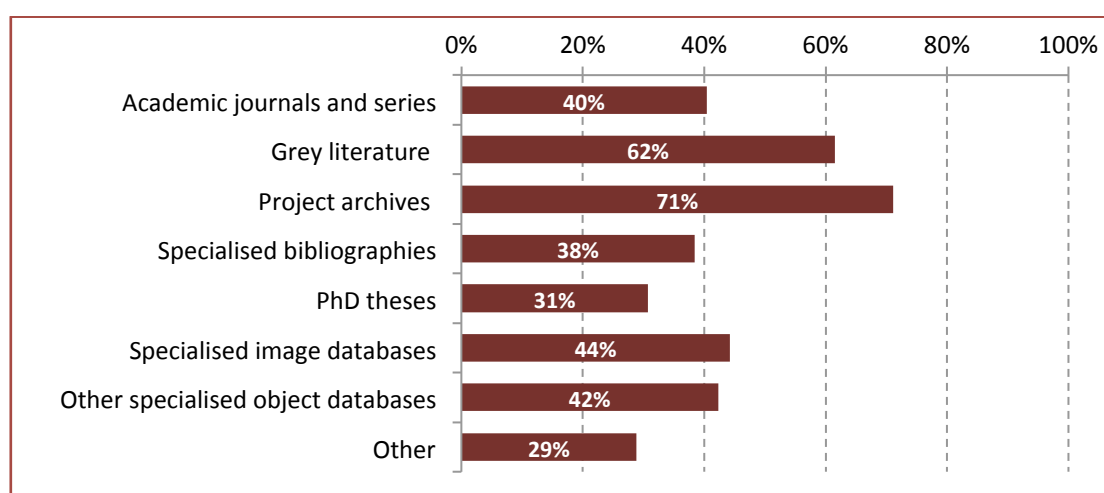
In general the survey found that levels of awareness and use of repositories is low, and there is little incentive to share data. The results suggest that most digital repositories still have, to a large extent, a national use context and user base.

### Use of research data

Unexpectedly, the ARIADNE survey found that *there is no single most important source of data* - it is the quality of the data that matters to researchers and not the source.

When asked, 75% of researchers said excavation data was “very important” for their research. Also rated as “very important” were GIS data, data from material or biological analysis and field surveys. Remote sensing, laser scanning, site management data, corpus studies, model based computing were also rated as “rather important”.

The survey found considerable differences in the portfolios of the individual repositories. Most hold project archives and grey literature; these are regarded as the *most important* in terms of user demand. Specialised bibliographies were also considered important.



“Which of the following collections does your data repository hold?”

Excavation data (73%), GIS data (68%) and data from material/biological analysis (60%) are the types most commonly held.

### Online availability

Few researchers taking part in the survey felt the online availability of research data is fully satisfactory. Satellite and airborne remote sensing data, excavation data and GIS data appear to be the most advanced in terms of online access.

The main challenges that confront researchers in their daily research when searching for data, are:

- **Data transparency:** 90% of respondents agreed that they often “did not know what is available, because research data are stored in so many different places and databases”.
- **Data accessibility:** when data is available, access is too complicated – there are restrictions on access to specific groups, costs can be a major problem for online research, access to a single item can require a full subscription, or there is a lack of time to process the relevant literature.
- **Data quality:** the available data are often not complete or not well organised.



About one third of the repository managers reported an “open access” policy (no registration required) for all, most or at least some data sets, while another third grant access only on request (and not necessarily to all or most data), others restrict access to specific communities only.

“Language” was not seen as a major issue by either researchers or repository managers. Only about a third of the researchers felt that it was an important issue. About half of the repository managers believed that metadata in English would be sufficient.

### Challenges for repository managers

The major challenge with which repository managers are confronted in their daily work is ensuring metadata quality; while 90% of managers see this as their major challenge only a few researchers mentioned insufficient quality of metadata. But the work effort involved in preparing metadata was a barrier to deposit for 80% of researchers. Managing a rising number of data sets was an important challenge for more than 80% of data managers. The fast pace of technical innovation and changes in the regulatory framework were an important challenge for some repositories but not for others – about 60% said these issues were at least “rather important”, 40% said they were not a relevant concern.

Repository managers perceived two user needs as being increasingly important:

- convenience in the use of repositories;
- individual service and guidance.

### Expectations for ARIADNE

The project aims to provide additional functionality in a data workflow where:

1. data is produced and managed by research projects or other activities, e.g. heritage management,
2. content, data and metadata is deposited in institutional repositories,
3. metadata from several repositories is collected by data centres or subject/domain-based repositories, and search services are provided.

ARIADNE will not replace existing repositories, services or tools, but will provide additional functionality, such as enabling cross-search and other services for the archaeological research community.

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

**Improved data transparency:** ARIADNE should provide a broad overview of existing data resources, beyond the partners' resources. An easy way of “registering” data resources may be required.

**Capability of cross-searching data repositories:** This is one of the main advantages users of the ARIADNE e-infrastructure and services will expect.

**Improved conditions of access:** A research policy objective, which the ARIADNE project can support by promoting open access principles.

**Filtering “useful” and re-useable resources:** The data that are accessible online are sometimes not as useful as they could be for various reasons (data structure, not up to date, incomplete or lacking in important details). A lot of data are not in re-usable formats (such



as data tables in PDFs) or not available under an adequate license. Users of the ARIADNE portal would benefit from a means of assessing the usefulness of the data.

**Portal service:** Respondents suggested that ARIADNE should establish a new portal for data search with added value. Improved overview, cross-searching and filtering of data resources would add value but further requirements are likely to be identified as users discover “useful” data(sets).

**Services for repositories and other websites:** ARIADNE should also be seen as a service for data repositories, other websites and for specific communities of practitioners. ARIADNE might help enrich underlying repositories, for instance by suggesting (and providing) links to similar or complementary collections or items held by other repositories. RSS feeds on available new data relevant for particular subjects or geographic regions and an Application Programming Interface (API) would be useful for website and application developers.

## Conclusions

ARIADNE has an opportunity to create real value for users and that the project is highly relevant in terms of addressing user needs. There is high potential for impact if services can deliver improvement in any of the areas where the research community expressed a particular need for improvements:

- **Transparency** of the available research data (it is difficult to know which data actually exists, due to the enormous fragmentation of repositories and data sets in the field) and in **data accessibility**.
- Major **barriers to accessibility** include costs (e.g. for obtaining licences to use pictures, for subscription fees) and the fact that relevant literature and data is often kept in private collections.
- **Data and metadata quality** are particular concerns from the perspective of data managers. When searching data, it is the quality of data that matters not the source.
- The two major **barriers for sharing data** with other researchers are: a perceived lack of recognition for sharing and the additional work effort for preparing the data for deposit.

ARIADNE will provide a platform to raise awareness of digital repositories. The project can help improve the overview of available research data and will work to promote open sharing of data in the archaeology sector in Europe. Going beyond advocacy, this requires advice and support in data management, effective metadata generation, licensing, data citation standards and building recognition of data sharers.

## Acknowledgement

We are very grateful to all the respondents who took part in the survey and interviews for contributing their time and insight, and for helping to ensure that the work carried out by the project is well aligned with the needs of the sector.

ARIADNE (“Advanced Research Infrastructure for Archaeological Dataset Networking in Europe”) is funded under the European Community's Seventh Framework Programme (FP7-INFRASTRUCTURES-2012-1).

## Reference

ARIADNE - First Report on Users' Needs. Project deliverable D2.1, April 2014, forthcoming:  
<http://www.ariadne-infrastructure.eu/Resources>