

ARIADNEplus and community data repositories

Innovative solutions for sharing open archaeological data

Guntram GESER, Salzburg Research, Austria¹

Keywords: Research Infrastructure: Repositories: Open Data: Business Models

Issues and context

Archaeology should be a field leading in data preservation, sharing and integration. Most archaeological fieldwork cannot be repeated and the digital record of excavations and other investigations is fragile but needed as evidence and basis for further research, comparative analysis and broad synthesis. However, 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. Digital infrastructure for finding and accessing data of repositories in different countries has only recently been established by ARIADNE, the Advanced Research Infrastructure for Archaeological Data Networking in Europe. The issue of a lack of appropriate data repositories 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 infrastructure. ARIADNEplus will update existing and incorporate additional datasets in the ARIADNE catalogue, including also data from scientific analyses, and provide new data services and tools. Expansion of the pool of datasets in the years to come will depend on accessible repositories across Europe (and beyond) richly filled by the research communities.

National-level data repositories as the most effective approach

Many European countries lack a state-of-the-art digital repository where archaeologists can deposit their data for long-term preservation and make it available to the research community. The optimal solution is building and mandating deposition of the data in a national-level repository. It is the most effective approach in several respects, including clear orientation of all stakeholders, formation of a trusted centre of expertise, guidance and support, cost-effectiveness of data curation and access (e.g. economies of scale). The alternative, particularly in large countries, is a scenario in which many institutions aim to build their own data repositories, with a lot of duplication of effort, implementation of different standards, and competition for scarce funding. Moreover, repositories dedicated to one institution usually accept data only from affiliated researchers.

References for national-level archaeological data repositories exist, for example the ARIADNEplus partners Archaeology Data Service (UK) and the E-Depot for Dutch Archaeology of Data Archiving and Networked Services - DANS (Netherlands). In Germany, unfortunately the development of the IANUS Research Data Centre for Archaeology and Classical Studies, funded 2011-2017 by the German Research Foundation, has been discontinued. In the United States, Digital Antiquity at the Arizona State University (also a partner in ARIADNEplus) aspires to provide a national-level repository with tDAR, The Digital Archaeological Record (McManamon et al., 2017).

The benefits of such repositories stem from their role as reliable central hubs for information and data resources which make research easier, faster and cheaper. In the case of the Archaeology Data Service (ADS) the increase in research efficiency of the users has been calculated to be worth at least 5 times the costs of operation; including other benefits £ 1 invested in ADS yields up to £ 8.30 return of investment (Beagrie and Houghton, 2013; on the development of the ADS see Richards 2017).

There are many advantages of preserved data that can be found and accessed easily in one place. For example, it can prevent unnecessary replication of work, allow verification of research integrity, promote collaboration, and combination and analysis of data to address new research questions. The Keeping

¹ Author's address: Guntram Geser, Salzburg Research, Jakob Haringer Straße 5/3, 5020 Salzburg, Austria; guntram.geser@salzburgresearch.at.

Research Data Safe (KRDS) Benefits Analysis Toolkit considers over 30 benefits for researchers, institutions and society (Charles Beagrie Ltd., 2011).

Regarding business models for sustainable data repositories for archaeology the largest part of the costs will typically be covered directly or indirectly from public funds, but income from deposit charges of archives of developer-led projects and other sources (e.g. grants of private foundations) could be a significant part of a mixed model (OECD 2017). Looked at from the perspective of individual research projects the cost of data preservation for long-term access is only a fraction of the total project costs, 1-3% depending on the type of investigation and data generated. These percentages are for project archives that require much curatorial support; “self-service” deposit of some files costs much less, e.g. tDAR charges \$10 per individual file (up to 10 MB), discounted to \$5 per file for purchase of 100 or more deposits (McManamon et al., 2017).

The scenario from the perspective of ARIADNEplus

The objective of the ARIADNEplus data infrastructure is to allow researchers and other users discover and access data held and shared by repositories across Europe (and beyond). From this perspective ideally one or only few repositories per country from which data records can be aggregated is of course the preferred scenario. A proliferation of repository building projects and dispersion of archaeological data resources would make the tasks required to integrate such resources much more difficult. The tasks include a significant amount of effort for support in the preparation of data records so that advanced data search and access methods (e.g. based on Linked Data) can be applied on the aggregated pool of metadata.

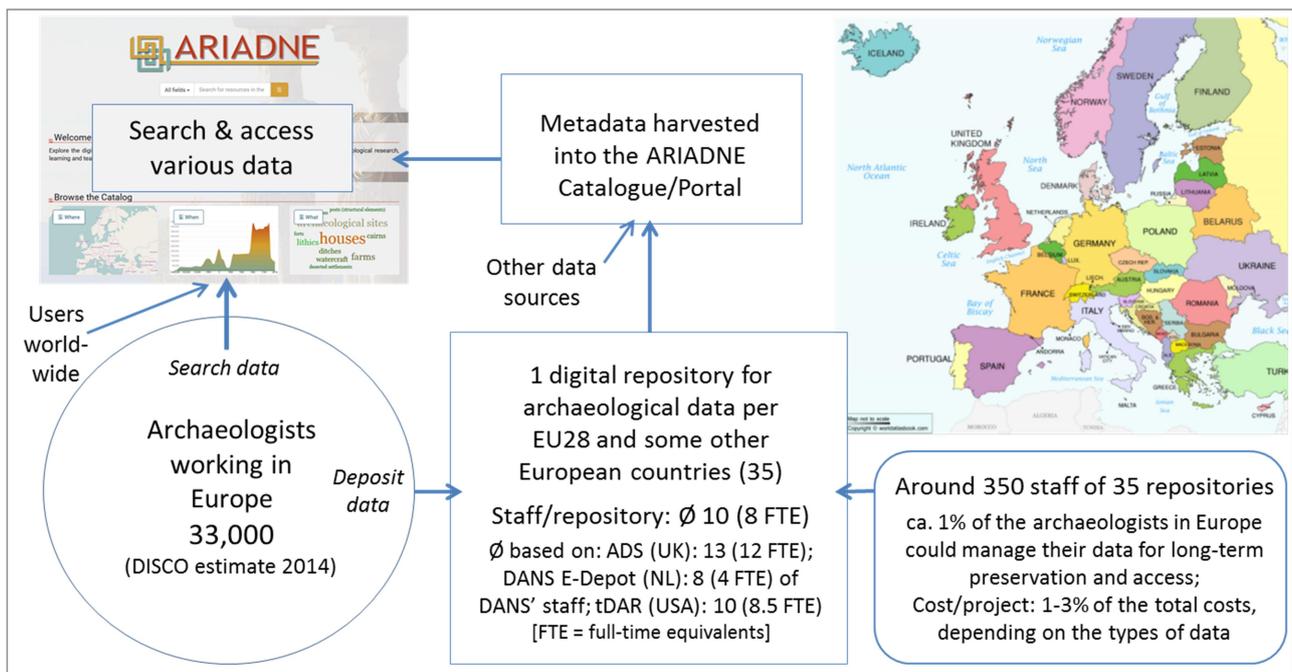


Fig. 1. Scenario of data preservation and access in Europe based on national-level repositories (© Salzburg Research).

Figure 1 illustrates a scenario in which each European country (the EU28 and others) would have but one mandated national-level archaeological data repository from which ARIADNEplus harvests the metadata of datasets to feed the data search portal. The scenario shows that a small number of repository staff could acquire, preserve and provide access to valuable data from the work of the archaeologists in Europe. 35 repositories would require in total only about 300 staff, which is around 1% of the estimated 33,000 archaeologists working in Europe (DISCO, 2014). The scenario of course does not exclude some division of work between repositories, for example, between repositories for long-term preservation and access (the focus of this paper) and repositories dedicated to particular research fields or themes in which the research community regularly updates existing and adds new datasets, research reports, etc. Such dedicated repositories may have an international scope and be maintained by institutes leading in the respective research field or theme.

Open research data require mandates and support by research funders

Researchers often share data, but mainly with project collaborators and other trusted colleagues. Therefore a lot of valuable data, although funded publicly, is not available to the research community and other potential users. In a survey of 1560 academic researchers of different disciplines 58% said that they shared data with researchers they know personally while only 13% made data publicly available (Fecher et al., 2015). A lack of academic recognition and reward, fear that data might be misinterpreted or misused, and the additional work required to prepare data for use by others (e.g. data description) are strong barriers to sharing data through an accessible repository. Therefore 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). Thereby, instead of being inaccessible and eventually lost, valuable research data can be preserved and become available for further research, education and other uses.

Growing an open data culture in archaeology

Archaeological data repositories and the ARIADNEplus data infrastructure will flourish only within a research culture that values preservation, sharing and reuse of data. Archaeology should be a field leading in open (or FAIR) data because academic as well as preventive archaeology are conducted in the public interest in archaeological heritage and knowledge. The fact that much of the creation of its data results from the destruction of primary evidence makes preserving and open sharing of the digital record all the more critical. However, many archaeologists around Europe are not yet well equipped and supported for archiving and sharing open data. As the matter is complex, strong leadership regarding data policies (mandates, funding), state-of-the-art repositories, training and support is necessary.

Acknowledgement

The ARIADNEplus project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823914.

References

- Beagrie, N. and Houghton, J. (2013). The value and impact of the Archaeology Data Service. A study and methods for enhancing sustainability. Final Report, Charles Beagrie Ltd. and Victoria University. Available at <http://repository.jisc.ac.uk/5509/> (Accessed: 30 June 2019).
- Charles Beagrie Ltd. (2011). KRDS Benefits Analysis Toolkit. Available at <https://beagrie.com/krds-i2s2.php> (Accessed: 30 June 2019).
- DISCO - Discovering the Archaeologists of Europe (2014). Transnational Report. York Archaeological Trust. Available at http://www.discovering-archaeologists.eu/national_reports/2014/transnational_report.pdf (Accessed: 30 June 2019).
- 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. Available at <http://www.hiig.de/wp-content/uploads/2015/02/dp1454.pdf> (Accessed: 30 June 2019).
- McManamon, F.P., Kintigh, K.W., Ellison, L.A. & Brin, A. (2017). 'tDAR: A cultural heritage archive for twenty-first-century public outreach, research, and resource management', *Advances in Archaeological Practice*, 5 (Special Issue 3), pp. 238-249, DOI: 10.1017/aap.2017.18.
- OECD (2017). Business models for sustainable research data repositories. OECD Science, Technology and Innovation Policy Papers, No. 47, December 2017. DOI: 10.1787/302b12bb-en.
- Richards, J.D. (2017). 'Twenty years preserving data: A view from the UK', *Advances in Archaeological Practice*, 5 (Special Issue 3), pp. 227-237, DOI: 10.1017/aap.2017.11.