What is an archaeological research infrastructure and why do we need it?

Aims and challenges of ARIADNE

Edeltraud ASPÖCK1 | Guntram GESER2

1 Austrian Academy of Sciences, OREA Institute for Oriental and European Archaeology, Vienna, Austria | 2 Salzburg Research, Salzburg, Austria

Abstract: Research infrastructures are facilities, resources and services used by the scientific community to conduct research. They include major research instruments, laboratories, databases, libraries and archival collections. Research infrastructures require digital infrastructure and services for information storage, management, processing, publication and access. In archaeology, data centres and repositories where researchers can deposit and access documentation and data from excavations, field surveys, laboratory analysis, etc. already exist in several European countries. The EU FP7 project ARIADNE (www.ariadne-infrastructure.eu) will set up a common interface to a digital infrastructure integrating archaeological information and datasets held by data centres, institutional and other repositories which are distributed across Europe. This paper addresses the following questions: What is an e-infrastructure for research? How will ARIADNE integrate digital resources on the European level? What are major challenges of the project? Beside the considerable semantic and technical challenges, particularly the need to facilitate a culture of collaboration and data sharing among a multi-disciplinary research community is emphasised.

Keywords: e-infrastructure, repositories, archaeological datasets, integration, data sharing

Introduction: What is an e-infrastructure for research?

Research infrastructures include major scientific equipment and sets of instruments, e.g. laboratories, remote sensing instruments, a research vessel etc., and knowledge-based resources such as scientific databases as well as content collections held by libraries, archives and museums. Research infrastructures require technologies for data collection, storage, management, access, processing and sharing, as well as research communication and collaboration tools. This e-infrastructure should allow easy and controlled access to integrated digital resources and tools, and to set up virtual research communities, i.e. research collaborations formed across geographical, disciplinary and organisational boundaries. E-research infrastructures or “cyberinfrastructure” (the term used in the USA) are understood as important drivers of innovative scientific research that has become increasingly collaborative, distributed and data-intensive. In recent years, special emphasis has also been placed on “big data” and “data driven” research based on large integrated datasets. Overall there is the expectation that with such data, e-infrastructure and tools new scientific questions can be tackled in international and interdisciplinary collaboration (RIDING THE WAVE 2010).
In Europe, the European Strategy Forum on Research Infrastructures (ESFRI) was launched in 2002 to support a coherent approach to policy-making on pan-European infrastructures, which are perceived as one of the pillars of the European Research Area (cf. their Roadmap reports since 2006, e.g. ESFRI 2011). In the United States, the National Science Foundation (NSF) Advisory Panel on Cyberinfrastructure played a similar role (NSF 2003, NSF 2007). Both have stimulated various commissions and surveys to identify relevant existing and to promote new research infrastructures in different disciplines, including the humanities (e.g. ACLS 2006; HERA-NET 2006; AVROSS 2007; ESF 2011).

For example, the international AVROSS (2007) survey on e-research projects in the social sciences and humanities (218 projects) includes some results on archaeology: respondents from this domain appeared relatively more experienced in such projects, having started already in the 1990s and they were involved in projects more often than other respondents. While most projects used communication and collaboration tools (80%) and distributed data (77%), two elements were most present in archaeological projects: Innovative data collection methods with 68.8% (25.3% in all projects), and 3D objects and virtual environments with 41.7% (15.7% in all projects). Though the archaeological projects typically were not large-scale, the average initial funding was just 148,000 € and the average duration 18 months, compared to 335,000 € and 36 months in the overall sample.

In Europe, major funding of e-research infrastructures has been made available through the Infrastructure and ICT strands of the EU Framework Programmes as well as national research funding agencies. On the ESFRI Roadmap are the large-scale, European-level projects, including infrastructures for the social sciences and humanities (e.g. DARIAH - Digital Research Infrastructure for the Arts and Humanities; CLARIN - Common Language Resources and Technology Initiative; CESSDA - Council of European Social Science Data Archives). The ESFRI scheme includes support of the design, implementation and upgrading of infrastructures for which several countries are committed to contribute funds to ensure sustainability.

**ARIADNE – Advanced Research Infrastructure for Archaeological Dataset Networking in Europe**

ARIADNE is an “Integrating Activity” project funded under the Research Infrastructures strand of the European Union’s Seventh Framework Programme (FP7-Infrastructures-2012-1, contract no. 313193). The project started in February 2013 and will run for four years. It involves 24 partners from 16 European countries who develop a common data infrastructure and related tools, work on knowledge organisation and integration (e.g. use of CIDOC CRM, thesauri, linked data), involve other stakeholders in Europe and beyond, and offer training, good practice guides, and access to their research facilities both on-site and online (RICHARDS 2012; NICCOLUCCI and RICHARDS 2013).

The overall goal of ARIADNE is to overcome the fragmentation of archaeological data resources and to foster a research culture that values sharing and re-use of data, going beyond organisational and national boundaries. In archaeology, some national data centres, institutional repositories and various other online resources already exist or are under development. Some examples, which are part of the ARIADNE initiative, are
Aspöck | Geser – What is an archaeological research infrastructure?

• the UK-based Archaeology Data Service (ADS), the first and largest archaeological data centre which curates and provides access to over 700 research archives and about 20,000 fieldwork reports;
• the e-depot for Dutch archaeology (EDNA), managed by Data Archiving and Networked Services (DANS), contains 3,000 datasets (including photographs, data tables, GIS data, etc.) and 18,500 archaeological reports; DANS together with institutes in other European countries also manages the Digital Collaboratory for Cultural Dendrochronology (DCCD);
• the Fasti Online database of archaeological excavations since the year 2000, managed by the International Association of Classical Archaeology: excavations primarily in Italy, Bulgaria, Romania, Macedonia and Albania (some 12,000 excavation reports and site summaries);
• the ARACHNE online database of the German Archaeological Institute (DAI) and the Institute of Classical Archaeology in Cologne (over one million images of finds, architecture and excavations with meta information and digitised historical literature); ARIADNE partner DAI coordinates also the development of the German IANUS archaeological data centre;
• the Swedish National Data Service (SND) curates archaeological data since 2011, currently some 400 archaeological surveys, the Swedish Rock Art Archive material, and some thematic databases, but expecting a tenfold increase in data within the next few years; SND also involves the Strategic Environmental Archaeology Database (SEAD), Sweden;
• the Italian Ministry for Cultural Assets and Activities (MiBAC) holds archaeological records throughout Italy (about 2 million records) and aims to digitise and make records and other resources (e.g. GIS data) more widely accessible.

These are but a few resources of the widely distributed European resources the ARIADNE project aims to integrate and make accessible through a virtual infrastructure. The project of course builds on previous and on-going collaborative efforts and experiences of partners, e.g. ARENA - Archaeological Records of Europe - Networked Access, EPOCH – Excellence in Processing Open Cultural Heritage, CARARE - Connecting Archaeology and Architecture in Europeana, and DARIAH - Digital Research Infrastructure for the Arts and Humanities. The project intends also to link up, where appropriate, with other archaeological e-infrastructure initiatives such as Open Context (USA), tDAR - The Digital Archaeological Record (USA), FAiMS (Australia) and Sustainable Archaeology (Canada). ARIADNE invites associated partners and projects wishing to make available datasets or collaborate in training or other events. Researchers can also take part in ARIADNE’s nine special interest groups and summer schools.

The ARIADNE data infrastructure
The core activity of ARIADNE is to develop an e-infrastructure for archaeological dataset networking in order to leverage the sharing and collaborative use of available data across organizational and national boundaries. It is worth first to distinguish this effort from other types of e-infrastructures:
Virtual Research Environments (VREs)

VREs is an umbrella category of research environments that centre on e-research functions which are selected and combined according to the needs of specific research communities. The combination can be centred more on networking and information exchange, collaborative digital collection formation and usage, offering programming and data processing tools (BOS et al. 2007; CARUSI and REIMER 2010). To provide but a few examples of VREs in archaeology: BoneCommons promotes communication and collaboration among the zooarchaeological community; Virtual Research Environment for Archaeology (VERA) is a collaborative working environment of the Silchester Town Life Project that needed to enhance the flow of information from excavation and data analysis to the publication of research results (RAINS 2011). Alison BABEU (2011) provides many examples of VREs that have been developed for “digital classists”. The recent trend to use Web 2.0 tools and services for VREs in archaeology, e.g. for research networking and knowledge exchange, is discussed in a recent publication on “Archaeology 2.0”. (KANSA et al. 2011)

Distributed Computing Infrastructure (DCI)

A Distributed Computing Infrastructure provides a Scientific Gateway to Grid and/or Cloud resources and software applications for working with research data, e.g. data processing, storage and transfer. Leading promoters of DCIs are the European Grid Infrastructure (EGI), European Middleware Initiative (EMI) and European Desktop Grid Initiative (EDGI). User groups can share data resources and computing applications (“virtual machines”) and, thereby, form a virtual research community. The main user groups are in the natural sciences and engineering disciplines while archaeology and other humanities researchers seldom deploy Grid-based datasets and processing. The major impediments arguably are the diverse and complex types of datasets, lack of consistent data structures, incomplete, isolated and often not openly available data sources. (cf. HEDGE 2009)

Data Infrastructures

A Data Infrastructure centres on making distributed research data accessible and (re-)useable which involves preparation, registration, networking and exchange of data, including the required protocols and standards, for example controlled vocabulary (thesauri, ontologies). Such an infrastructure can be generic, i.e. used across many domains (e.g. spatial data infrastructure), or focused on aggregating and making accessible data of one or a few related disciplines. The ARIADNE e-infrastructure basically is a data infrastructure, but will also include some useful tools for data mining, annotation and other research purposes. A broad notion of “data” is used which includes content (e.g. papers, reports, images, maps, etc.), various datasets, and metadata records of content and datasets.

Elements of the ARIADNE e-infrastructure and services framework

ARIADNE’s e-infrastructure and services are developed and provided on top of existing institutional infrastructures of research groups and institutions (e.g. institutional or subject-/domain-based repositories and portals). Therefore the framework is based on a data workflow scheme that distinguishes four levels and respective data management and user communities:
• Level 1: Research projects: Data is in the first place produced and managed by research projects (i.e. field surveys, excavations, laboratory analysis, etc.) or in the context of other regular work (e.g. heritage management);
• Level 2: Research institutions (as the stable “umbrella” of projects): The data produced at Level 1 is deposited with metadata in institutional repositories or higher-level, national / international data centres or repositories, including “self-archived” reports, publications, etc. as well as various datasets;
• Level 3: National data centres and international subject or domain-based repositories: In such centres and repositories research projects and individual researchers deposit data according to common practice or (increasingly) based on funding contracts that require such deposits;
• Level 4: ARIADNE: Provides a data registration facility for deposited data (at Level 2 and 3) and discovery and access services based on a common metadata pool, thesauri and ontologies (especially, CIDOC CRM) which enable the e-infrastructure to integrate semantically the available data resources.

Thus, ARIADNE does not replace any existing infrastructures but provides integrating services on top of them (e.g. cross-search of different repositories). Thereby ARIADNE makes currently isolated data resources better discoverable, accessible and (re-)useable for the archaeological research community in Europe and beyond.

Major challenges for archaeological data sharing and access

The overall goal of making archaeological data better accessible faces conceptual, technical and socio-cultural challenges of which we highlight three:

Challenge 1: Serving a multi-disciplinary field of research

Archaeology is a multi-disciplinary field of research that spans several domains and specialties of the humanities. Among others, this includes cultural anthropology, classics, ancient/historical geography, medieval studies, epigraphy and iconography. Moreover the field includes the application of natural sciences methods to archaeological research questions (e.g. physical, chemical and biological sciences, zooarchaeology, archaeobotany, etc.). There has been much debate about the specific e-research requirements of the humanities, including the issue of different perspectives of technical developers and humanities researchers. (BORGMAN 2009; ANDERSON et al. 2010; WOUTERS and BEAULIEU 2006; SVENSSON 2010) However, we assume a high level of affinity of archaeologists for ICT-supported research and a strong common ground of archaeology practitioners and technology developers within ARIADNE. But designing the ARIADNE e-infrastructure and services to serve the data management and sharing practices of the extensive and multi-disciplinary archaeological research community will still be a considerable challenge of the project. A major role for success will play active stakeholder involvement, training offers, and guidance for the preparation, registration and publication of legacy and new types of data.
Challenge 2: Bridging different documentation practices and semantics

Archaeological data and metadata are produced in different formats, standards and languages. They are organised according to different concepts, but for the purposes of the ARIADNE e-infrastructure and services interoperability at machine level should be achieved. As NICCOLOCCI and RICHARDS (2013) explain:

*The ultimate project goal of pursuing the interoperability of archaeological datasets, i.e. the capability of different information systems to communicate at machine level, first of all requires resolving differences in encoding and representation. Such standardization – which is indispensable for enabling machines to communicate with each other – will be achieved through mapping the diverse data structures to a common paradigm. (…) The project does not aim to create an illusory overarching archaeological ontology. Rather, it will proceed by clustering datasets by similarity of scope and structure into wider aggregations, stopping when further integration would be meaningless for diversity of scope or useless for diversity of content. Altogether, these clusters will form a progressively growing integrated data infrastructure responding to the diverse needs of researchers with a homogeneous interface. There still remains the need of achieving semantic interoperability, i.e. communicating across datasets with shared meanings. Here the lead reverts to archaeologists who will need to state explicitly – or at least with as little ambiguity as possible – the meaning of the concepts which the stored data refer to.*

The biggest challenge here arguably is achieving semantic interoperability, i.e. shared meaning of a multitude of archaeological concepts, which is required to semantically integrate and link the different and widely distributed data resources so that they can be cross-searched and accessed. Concerning interoperability across different platforms and systems Linked Data methods and technologies are foreseen to play a major role.

Challenge 3: Fostering a culture of data sharing based on open access principles

In order to promote the archaeological enterprise in Europe and beyond, commitment to a culture of data sharing is necessary. Making data available should be based on mutual trust and open data principles and licenses. Representative surveys on data sharing practices across all disciplines show that after the completion of research projects most data remains locked away, residing on PCs, storage devices, and restricted access servers. Only some 6-8% of researchers make data openly accessible in a community repository. (cf. PARSE 2009; SCIENCE 2011) The current data practices clearly run against what advocates of open data sharing would like researchers to do. Proper management to prevent data loss and the value of the data for re-use by the community are often not considered. Researchers still perceive more obstacles than incentives for open data publication, including lack of academic reward, additional effort for preparing data and metadata, concerns that data might be misused, and more. (BORGMAN 2010; PRYOR 2009; RIN 2008; RIN and NESTA 2010)

The current extension of open access mandates for publicly funded research from papers to research data, available trusted data repositories, and initiatives to ensure that data sharers receive the credit they deserve (e.g. data citation and tracking mechanisms) are crucial to promote the required culture of data sharing. The ARIADNE project can contribute significantly to this development in the domain of archaeological research.
Conclusions and outlook
The overall goal of ARIADNE is to overcome the fragmentation of archaeological data resources through providing an e-infrastructure that integrates archaeological information (e.g. research reports and publications) and datasets held by data centres, institutional and other repositories that are distributed across Europe. The aim of the e-infrastructure is to support the sharing and collaborative use of available data across organizational and national boundaries.
ARIADNE will not replace any existing infrastructures, e.g. data centres, institutional and subject/domain-based repositories, but provides integrating services on top of them. Thereby ARIADNE will make currently isolated data resources better discoverable, accessible and re-useable for the archaeological research community. Virtual research environments (VREs), which support the e-research needs of specific archaeological research communities, may use ARIADNE’s services to discover and build on available data resources. Providing access also to data processing services (e.g. Grid or Cloud based computing) is currently not intended as the project in the first place centres on the integration, linking, and access to data resources. Major challenges of ARIADNE are to serve an extensive and multidisciplinary field of research, to achieve semantic interoperability among different data resources of this field and, particularly, to promote open sharing of data to stimulate innovative collaborative research.

We believe that ARIADNE will affect the archaeological discipline in several positive ways. One of the most important ones is community building – the networking and joint research activities of the project will lead to a stronger integration of the European community of archaeologists. ARIADNE invites associated partners and projects wishing to make available datasets or collaborate in special interest groups. Also, the project will provide particular support for a younger generation of archaeologists, e.g. with summer schools.

Acknowledgement
The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 313193.

References
ARACHNE, http://arachne.uni-koeln.de
Archaeology Data Service (ADS), http://archaeologydataservice.ac.uk

BoneCommons, http://alexandriaarchive.org/bonecommons/


CESSDA - Council of European Social Science Data Archives, http://www.cessda.org


CLARIN - Common Language Resources and Technology Initiative, http://www.clarin.eu

DARIAH - Digital Research Infrastructure for the Arts and Humanities, http://www.dariah.eu

Data Archiving and Networked Services (DANS), http://www.dans.knaw.nl/en

Digital Collaboratory for Cultural Dendrochronology (DCCD), http://dendro.dans.knaw.nl


e-depot for Dutch archaeology (EDNA), http://www.edna.nl


European Desktop Grid Initiative (EDGI), http://edgi-project.eu

European Grid Infrastructure (EGI), https://www.egi.eu

European Middleware Initiative (EMI), http://www.eu-emi.eu

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

Fasti Online, http://www.fastionline.org

German Archaeological Institute (DAI), http://www.dainst.org


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

International Association for Classical Archaeology (AIAC), http://www.aiac.org


Open Context (Alexandria Archive Institute, USA), http://opencontext.org


Strategic Environmental Archaeology Database (SEAD), Umeå University, Sweden, http://www.sead.se

Sustainable Archaeology (Canada), http://sustainablearchaeology.org


Swedish National Data Service (SND), http://snd.gu.se/en/start

tDAR - The Digital Archaeological Record (Digital Antiquity consortium, USA), http://www.tdar.org