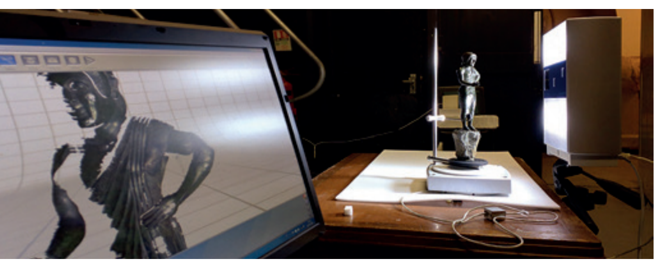
THE ARIADNE IMPACT

EDITED BY JULIAN RICHARDS AND FRANCO NICCOLUCCI








ARIADNE plus

THE
ARIADNE
IMPACT

EDITED BY
JULIAN RICHARDS AND FRANCO NICCOLUCCI

Budapest 2019

The research presented here was funded by the European Commission under the FP7 Programme ARIADNE project, contract no. FP7-INFRA-2012-1.1.3-313193, and under the H2020 Programme ARIADNEplus project, contract no. H2020-INFRAIA-2018-1-823914. Other national or international funding for this research is acknowledged in the relevant sections.

The present publication was funded by the European Commission under the H2020 Programme, as part of the ARIADNEplus project, contract no. H2020-INFRAIA-2018-1-823914.
The European Commission support for the production of this publication does not constitute endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Cover photos:
Front cover (left to right; top to bottom):
Glastonbury Abbey excavations during the 1920s © The Glastonbury Abbey Archaeological Archive Project (https://doi.org/10.5284/1022585); Photograph of a brick found during excavations at Castelporziano. It bears a stamp with Caesar Augustus' name dating it to between 27 BC and 14 AD © Prof Amanda Claridge, Prof Helen Rendell (https://doi.org/10.5284/1000127); The BTC Pipeline Archaeological Excavations in Azerbaijan © David Maynard (https://doi.org/10.5284/1000411); 3D scanning of Pizzidimonte Statuette at the British Museum © VAST-LAB, PIN; Archaeological finds in Pompeii, Italy. Detail © Envato Elements License; Ruins of Apollo temple in Corinth, Greece © Envato Elements License
Back cover (left to right; top to bottom):
Storage of archaeological finds in Pompei, Italy © Envato Elements License; Researcher working on archaeological GIS © VAST-LAB, PIN; Aerial photograph of the Eastern settlement of the ancient Saar complex © Dr Robert C. Killick (https://doi.org/10.5284/1042738); Butser Ancient Farm, Butser Ancient Farm Project Archive 1972–2007 © Roger Hedge, Christine Shaw (https://doi.org/10.5284/1039935)

Copy editor: Kyra Lyublyanovics

Desktop editing and layout: Zsuzsanna Kiss

Cover design: Móni Kaszta

ISBN 978-615-5766-31-2

ARCHAEOLINGUA FOUNDATION
H-1014 Budapest, Úri u.49
kovacsr@archaeolingua.hu
www.archaeolingua.hu
Managing Director: Erzsébet Jerem

Printed in Hungary by Prime Rate Kft.

Contents

ARIADNE and ARIADNEplus

Franco Niccolucci [*] ***and Julian Richards*** [**]

PIN, Prato, Italy
*franco.niccolucci@gmail.com

Archaeology Data Service, University of York, UK
**julian.richards@york.ac.uk

Abstract

Research e-infrastructures, digital archives, and data services have become important pillars of scientific enter prise that in recent decades have become ever more collaborative, distributed, and data intensive. The archaeological research community has been an early adopter of digital tools for data acquisition, organization, analysis, and presentation of research results of individual projects. However, the provision of e-infrastructure and services for data sharing, discovery, access, and (re-)use have lagged behind. This situation is being addressed by ARIADNE and ARIADNEplus. This EU-funded network has developed an e-infrastructure that enables data providers to register and provide access to their resources (datasets, collections) through the ARIADNE data portal, facilitating discovery, access, and other services across the integrated resources. This chapter introduces the ARIADNE programme and outlines it current and potential impact, setting the scene for the individual contributions from ARIADNE partners which make up the rest of the volume.

Keywords: e-infrastuctures; impact; CIDOC CRM; FAIR; EOSC

Introduction

ARIADNE (Archaeological Research Infrastructure for Archaeological Data Networking in Europe) was an infrastructure project funded by the European Commission under the Seventh Framework Programme for the period 2013–2017.[1] ARIADNE's goal was to provide open access to Europe's archaeological heritage and to overcome the fragmentation of digital repositories, placed in different countries and compiled in different languages.

Integration has been achieved using state-of-the-art ICT and open data standards, creating the ARIADNE Catalogue with advanced search functionalities and services to use and re-use data.[2] Innovative vocabulary mapping techniques have brought interoperability to a huge and heterogenous collection of texts, drawings, images, videos, 3D models and maps. ARIADNE succeeded in integrating archaeological

1 http://legacy.ariadne-infrastructure.eu/
2 http://portal.ariadne-infrastructure.eu/

datasets in its Registry, with more than 1,700,000 datasets recorded and managed according to the FAIR principles (Meghini et al. 2017).

In parallel, networking activities raised the awareness of potential users, archaeologists and heritage managers, creating a vibrant transnational community. ARIADNE has been praised by archaeological associations and institutions, including the EAA (European Association of Archaeologists) and it has led to the establishment of data repositories in many countries. ARIADNE succeeded in building a community of use consisting of about 11,000 archaeologists, corresponding to one third of all European archaeologists and probably more than 50% of those using some computer support in their research. Other important results include the development of sector standards: CRMarcheo and CRMba, now officially adopted by the research community; a set of multilingual thesauri; and a number of guides to good practice in digital archaeology. In its 2018 roadmap, ESFRI acknowledged ARIADNE's activity as the response to a 'vital need' (ESFRI 2018, 113).

ARIADNEplus is a new programme of work, funded by the European Commission under Horizon2020.[3] It builds upon the success of ARIADNE, extending its scope and improving the technology to embed the ARIADNE infrastructure in the European Open Science Cloud (EOSC).[4] The overall goal of ARIADNEplus may be summarized as **extending** and **focusing** ARIADNE.

Extending is concerned with the domains served and the users addressed; it has several dimensions including:

- The **geographic coverage**, which in ARIADNE already reached almost all European regions, by integrating in the ARIADNEplus Infrastructure a greater number of archaeological partners and giving particular attention to areas where the coverage was less intensive.
- The **disciplinary coverage**, which in ARIADNE included mainly excavation data and a few specialist topics such as, for example, dendrochronology, by integrating in the new ARIADNEplus Infrastructure data produced by palaeoanthropology, bioarchaeology, environmental archaeology as well as the results of scientific analyses, such as material sciences, dating and so on, and those related to standing structures, be they small remains of ancient constructions or complex and massive monuments as, for example, Hadrian's Wall in the UK or the Magna Graecia temples in Southern Italy.
- The **time-span** considered, pushing back the earliest datasets included, by incorporating palaeoanthropology, and forwards the end-date until recent times, e.g. including industrial archaeology and Cold War archaeology; in practice covering the full time-span of the human presence on Earth.
- The **depth of database integration**, exploiting the potential of well-structured datasets such as databases, for which the interoperability will be extended to item level (in ARIADNE implemented only experimentally), and archaeological

[3] https://ariadne-infrastructure.eu/

[4] https://www.eosc-portal.eu/

Geographic Information Systems (GIS), for which integration will be achieved through the introduction of dedicated services, going beyond mere digital maps and overcoming incompatible reference systems.
- The **integration of text datasets** by extending the use of Text Mining through Natural Language Processing (NLP) and Named Entity Recognition (NER), previously applied only experimentally.
- The **research community** involved. The ARIADNEplus target is to make contact with the majority of all researchers and professionals (particularly important in this domain where research and heritage management often go hand in hand), and address most if not all the needs of computer-aware archaeologists. From existing expressions of interest, including the USA, Japan, Latin America and Australia, it is also already anticipated that ARIADNEplus will attract the international research community.
- The **service portfolio** offered to users, incorporating more advanced tools for digital analysis and interpretation in ARIADNEplus.

Focusing has, instead, only one objective: **innovation**. This 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).

ARIADNEplus is being guided by a suite of principles and strategies, including:

- The Open Science strategy, incorporating Open Access, as fostered by the European Union (European Commission 2019);
- The Open Data EU strategy (European Commission 2019);
- The FAIR data principles to make data Findable, Accessible, Interoperable and Re-usable, consolidating what was already applied in ARIADNE (Wilkinson et al. 2016, 2019);
- Principles for trusted digital repositories and related accreditation, such as the CoreTrustSeal (CoreTrustSeal 2016);
- RDA (Research Data Alliance) recommendations, such as those concerning Data Citation and Permanent Identifiers (PID) (RDA;.

Last, but not least, ARIADNEplus has a close connection with several initiatives related to the European Year of Cultural Heritage. Integrating heritage documentation, and enabling access to it, as planned in ARIADNEplus, is a way to extend the beneficial effects of this important EU initiative in the future. What we expect to achieve at the end of the project, i.e. our mission, is: *to integrate and effectively serve a research community that studies the past to better understand the present with the tools and the methodology of the future, in the service of European culture and society.*

This book arises out of a session on digital infrastructures at the Computer Applications in Archaeology Conference held in Krakow, Poland, in April 2019. The majority of speakers were members of the ARIADNE consortium, including original partners as well as new organisations that have joined as part of ARIADNEplus. All spoke about the importance of ARIADNE and the impact that it had, or was expected to have, on e-infrastructure developments within their own organisations or countries. The work is ongoing but we decided to capture a snapshot of the Impact of ARIADNE at a particular point of time. The result is this volume.

Ambition

The completion of the ARIADNE project fostered a new approach to archaeological research. Data that before were often considered a mere support to documentation started to become the support for new investigations. Important national experiences like those of the Archaeology Data Service (ADS) in the UK had already generated the birth of similar initiatives in other countries, but they still remained fragmented and unable to cross modern administrative borders. With ARIADNE, such experience became the foundations on which Europe-wide integration was built and discovery across boundaries was made possible. This has helped stimulate the growth of a new attitude towards cooperative research, fostering the concepts of data sharing and re-use and a potential new methodology in which data were one of the pillars of archaeological investigations.

With ARIADNEplus this possibility is operational. Extending the integration of data infrastructures to all European regions, all archaeological sectors and all human presence, as ARIADNEplus aims to do, will enable researchers to make use of the powerful tools made available by the project not only to discover what others have already found, but also to aggregate, link and process such results to create new knowledge. It will introduce into archaeological methodology the concept of 'data-based research' that has brought so many advantages in other disciplines.

The networking activities planned in ARIADNEplus will support large scale adoption of this new perspective. The project will carry out activities aimed at increasing the number of 'followers' from the current 10,000 users of the ARIADNE discovery service to a much larger audience. The collaboration with associations and institutional bodies will help facilitate the community penetration of the ARIADNEplus approach. Researchers will be encouraged to pursue data FAIRness for the availability of guidelines and tools provided by ARIADNEplus and tailored to the specific needs of this research community. Researchers willing to participate in the integration process with their data will find an established and safe procedure as most of the issues potentially raising in the integration process will have been tackled, covering almost every possible subdomain or application. In this way ARIADNEplus will achieve not only a critical mass of data, but also a critical mass of supporters. Being a prerequisite for adopting the ARIADNEplus philosophy of data sharing and re-use, the culture of cooperation will be fostered not only among participants but also in the wider research community. The starting point for this cultural leap is Virtual Access to an unprecedented mass of data. Researcher participation will be supported by training activities, with Transnational Access being at the top with the provision of personalized hands-on training to develop integration processes with a user's own content. This novel approach to archaeological research is supported by standardization, state-of-art technology and innovative services, fit for the needs of 21st-century archaeologists.

The standardization promoted by ARIADNEplus covers all the archaeological domains and allows interoperability. The data infrastructure developed by our Joint Research Activities will be supported by cloud technology and will enable the creation of Linked Open Data. The planned services will enable the data processing that archaeological research demands: first of all discovery, of course, within a huge number of datasets provided by the integrated data infrastructures; furthermore, services to manipulate the results and produce visual content and annotation tools. Text mining is then a much-needed functionality in a field where much data consists of textual reports. Furthermore, the multilingual approach, supported by multilingual vocabularies, is of paramount importance. Finally, being able to process data according to space and time and obtain synthetic views of the results is a key factor to stimulate imagination and lead to new knowledge.

The novel approach to archaeological research described above is one of the innovation aspects that ARIADNEplus will introduce to the research community. The innovative services developed by ARIADNEplus and the novel data organization it proposes for integration are other innovation factors. Together they are the starting point for ground-breaking innovation: bringing the digital into a discipline that has made materiality its distinctive character.

The archaeological perspective

ARIADNEplus will extend the integration of archaeological data infrastructures to sub-domains that were not (or only partially) addressed by ARIADNE. The underlying technology will also be improved and updated, in line with the most recent strategies of the EU, enabling a deeper data integration and creating new environments for data-centred research and data reuse, through the standardization, openness and interoperability of existing key infrastructures. Furthermore, the creation of new innovative infrastructures will be fostered where they do not yet exist, ready to be integrated in the overarching ARIADNEplus framework.

Archaeological data infrastructures cover different aspects of archaeological research. Usually they cover a delimited region, or they may address a limited time span. Finally, although archaeological sciences produce a huge amount of digital data, in general only the conclusions of analytical experiments have been used in archaeological research and included in data infrastructures. Such attitudes to disregard the importance of source data are now being phased out by a trend attributing them with a greater importance. Thus ARIADNEplus will extend its integration process to data that received less attention in ARIADNE, extending both its temporal scope and its thematic scope. *Figure 1* illustrates this expansion.

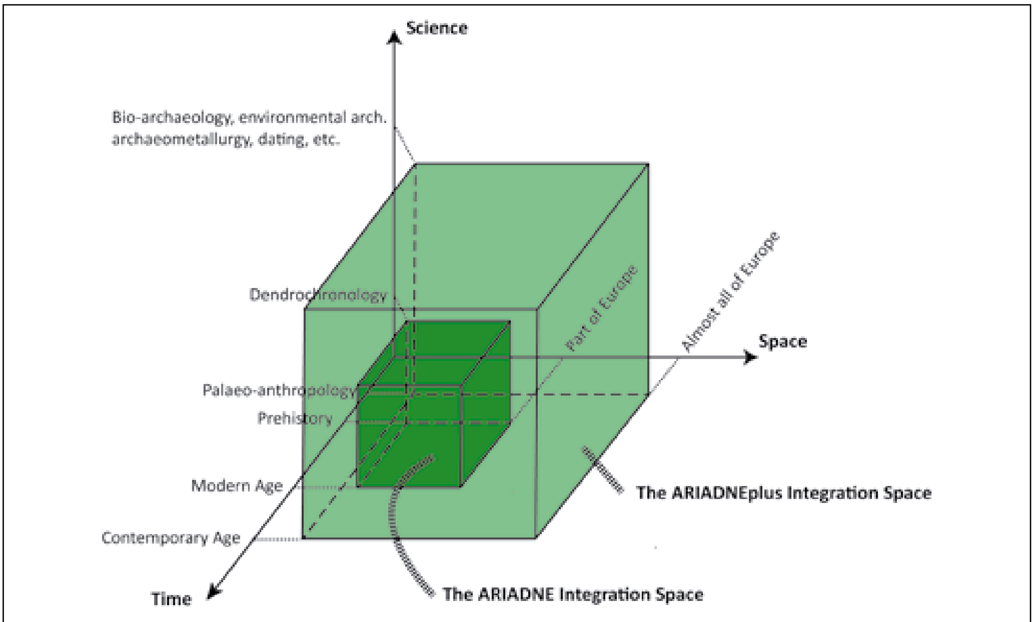


Figure 1: How ARIADNEplus will extend its integration scope in time, space and content

Spatial extension

As a general rule, general-purpose data repositories focus on archaeological work in the region where the repository is placed, or on archaeological assets (e.g. museum collections, monuments and sites) located in the same region. Sometimes, a much smaller part of their data comes from the work of archaeological missions abroad.

This is true also for thematic repositories, e.g. those containing data from scientific analyses, where some integrating initiatives have already been undertaken. In these cases, data refer to contiguous regions, for example Nordic regions, parts of the Mediterranean basin and so on. In sum, archaeological data infrastructures are largely location-based; their content is limited to modern countries, usually without crossing their borders, or their administrative sub-regions, where cultural heritage and archaeology is in charge of federal administrative institutions, like Germany and Spain. Without integration initiatives like ARIADNE, this situation is clearly a serious obstacle to data-based research, as archaeological research questions concern regions of the past that overlap modern countries and cross modern political borders. Easy examples are the Bell-Beaker culture, spanning in various European regions during the Bronze Age; or the Roman provinces forming Gallia, corresponding to present-time France, as well as parts of Spain, Belgium, Switzerland, Germany and the Netherlands; and so on. Due to variability in time, no geographical rearrangement of data may suit all the archaeological needs spanning prehistory to recent times.

If ARIADNE did not exist, any archaeological study concerning such research questions would require access to a number of data repositories located in each of the relevant regions/countries, and possibly others of a thematic nature, with no common perspective nor common user interface and general search tools – let alone research services.

Furthermore, the data schemas and languages used vary among the different data infrastructures, according to the diverse national official languages and approaches, so any integration process requires a labour-demanding process of collaborative work with data owners, concerning the mapping of the metadata descriptors of individual datasets or the mapping of data schemas of their infrastructure. In addition, incorporating their language in multilingual thesauri and other authority files, if not already present, is also necessary. Such collaboration requires a significant effort in time and resources, especially for large data infrastructures: for them voluntary adherence of repository managers to the integrating process is in general feasible only if some funding is added to available local resources.

The approach adopted by ARIADNE took account of this requirement. Its action developed on two levels:

- Convincing data owners that sharing their data was socially important and scientifically useful, as the advantages of a global registry where all the archaeological information was recorded and made available would benefit each researcher in the archaeological community.
- Supporting the effort necessary to start the integration process, providing the resources required by data providers.

When forerunners and early adopters take up this approach and the integration process produces a critical mass, the major problems present in the starting phase are resolved, and the effort required to join is less demanding, but still present. This first objective was achieved by ARIADNE.

ARIADNE also provided the technology necessary for the integration process, with a task force of technological project partners working to set up the structure and functionalities of the ARIADNE Registry. They also collaborated with content-providing partners in the mapping and data ingestion process from their datasets. Different but concurring evidence that the integration process has started to gain traction is the fact that during ARIADNE several other important data infrastructures offered their availability and interest in participating in the integration process, but with a few notable exceptions it was impossible to accomplish such work for lack of resources on the providers' side. In sum, we may conclude that:

- Key archaeological data infrastructures are generally location-based; with a few exceptions for scientific thematic repositories.
- The absence of one or more of such key data infrastructures reduces the effectiveness of all the integration, in proportion to the importance and the extension of the local infrastructure, as they are the sole place where data to be integrated can be found. Large gaps would weaken the integrated infrastructure accordingly.
- Starting the integration process requires resources, which local infrastructure managers do not have available. Maintaining it can instead be supported with much less trouble and with no (or very limited) external funding.

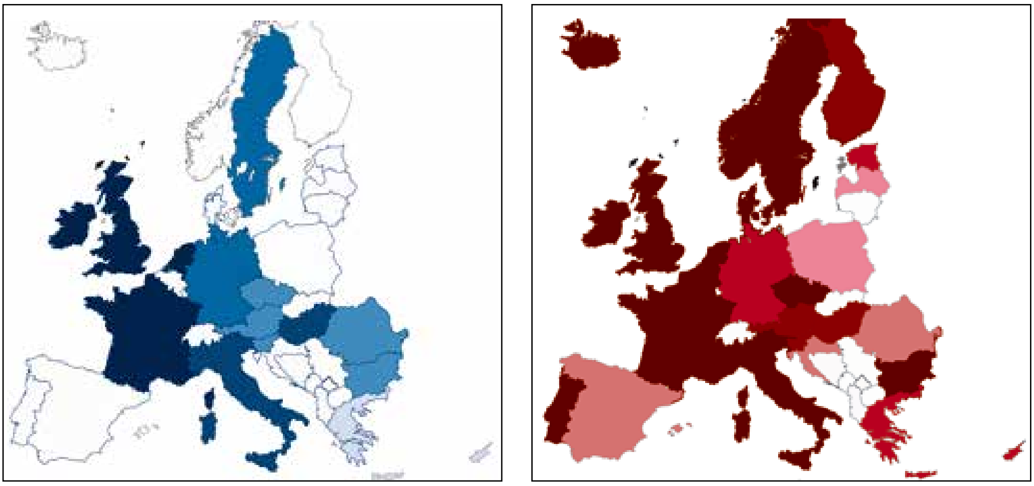


Figure 2: Extension of geographic coverage from ARIADNE to ARIADNEplus

In the former ARIADNE project there were clear gaps in geographical coverage which ARIADNEplus is seeking to address.

In the above maps, the intensity of the fill colour represents the degree of coverage of the online archaeological repositories integrated into ARIADNE. Some countries, especially those where a data infrastructure is missing, are planning to participate thanks to collaboration with the CARARE Europe-wide association. In such countries, ARIADNEplus will endeavour to foster and support a culture of archaeological data publication and sharing, through its networking and TNA activities[5].

ARIADNEplus therefore hopes to create an e-infrastructure of global importance. It already attracts and integrates the most important US initiative tDAR, coordinated by Arizona State University, and the Japanese one created by the national Nara Research Centre, which will contribute their datasets to the ARIADNE Data Infrastructure and promote the ARIADNEplus strategy in their respective countries. In Latin America, Argentina is also participating via its national research council CONICET. Beside the addition of their repositories, the most important aspect of this international network is the adoption of the ARIADNE methodology, which is thus becoming globally accepted.

Thematic extension

At first sight, the present ARIADNE catalogue seems to mainly contain datasets on monuments and sites. This is in part true, and in part a misleading effect of the way in which datasets are recorded in the registry. Regardless of its size, a database of finds, for example, is recorded as a single digital object in the registry. This approach is very

[5] Trans-National Activities (TNA) are a group of activities managed by the infrastructure and taking place at designated partners which provide training to allow researchers to develop – in the case of ARIADNE – skills in specific areas, as for example data organization and repository management.

efficient when the content referred is formed by collections of digital objects, such as the typical archaeological excavation records consisting in texts (the reports), images and drawings (of finds, monuments and sites), and sometimes including also the results of scientific analyses. It is less efficient as regards other databases, where it does not offer any way to directly access individual records or to select choice criteria, as it provides no item level integration. ARIADNEplus is addressing this issue not only by including new artefact datasets, but also enabling item-based searches across them through their federation. Thus an unstructured data repository will be registered in ARIADNEplus as a unique digital object, as in ARIADNE. Structured databases, more frequently used for finds and for scientific databases, will instead be federated and access will be enabled across them.

As regards unstructured datasets with any content, in ARIADNEplus the CRMpe model will be adopted for their metadata organization. This model will be described in greater detail below; here it suffices to state that it has been developed within the PARTHENOS[6] project as a general model for all humanities and heritage catalogues, since it is capable of dealing with activities, actors, procedures, datasets and software. CRMpe is a significant step forward compared with the data model initially used in ARIADNE. The existing ARIADNE catalogue with its 1,700,000 data records has already been converted to this data model and has effectively functioned on this basis since 2017. On the other hand, the semantic glue enabling the federation of the different database schemas will be the common ontology. In such cases a mapping from the schema used in the database to the CIDOC CRM or an appropriate extension of it will be defined. The thematic domains offering databases to be incorporated in the ARIADNEplus Infrastructure include:

- Human Palaeo-biology and Palaeo-environments
  - Palaeoanthropology
  - Bio-archaeology and Ancient DNA
  - Environmental Archaeology
- Analytical Investigations
  - Inorganic Materials Study
  - Dating, including Dendrochronology
- Archaeological Prospection
  - Field Survey
  - Metal Detector Survey (where allowed by the law) and archaeological finds
  - Remote Sensing
- Monuments and Sites
  - Standing Structures
  - Spatio-temporal Data (GIS)
  - Maritime and underwater Archaeology
  - Archaeological Fieldwork
- Inscriptions

[6] http://www.parthenos-project.eu/

The relationship between the ARIADNEplus Infrastructure and its sources

In the first version of the ARIADNE Catalogue, the relationship with the original sources was achieved through linking them to the catalogue item referral. Thus users perform searches on the catalogue using the search tools provided, and at the end of the search refinement they obtain a list of results, which are linked to the original dataset(s). This choice avoided the creation of a monster centralised archive and preserves the content provider's control on the data they produce and own. The same search approach will be supported in ARIADNEplus.

The descriptions collected in the ARIADNE Catalogue were structured according to the ARIADNE Catalogue Data Model (ACDM), which was developed in a consensus-based process involving all stakeholders, who could express their requirements and make sure the ACDM accommodated them. Thanks to this process, the ARIADNE Catalogue was built and used as a basis for the discovery service of the ARIADNE Portal. Later in the ARIADNE project, the ACDM was mapped onto the CIDOC CRM, and the mappings were used, along with an IRI (Internationalized Resource Identifier, extending upon the existing URI - Uniform Resources Identifier) generation policy, to generate the ARIADNE Linked Open Data Cloud. The transformation of the ARIADNE Catalogue into a LOD dataset achieved a twofold effect: from one side, it increased the technical accessibility of the Catalogue, by making it available to the external world via a SPARQL endpoint. From the other, it increased the semantic interoperability of the Catalogue by upgrading it from a domain- and purpose-specific vocabulary (the ACDM) to a general ontology standardized by ISO.

ARIADNEplus will build on this experience, by adopting the ARIADNE LOD cloud as the core of its information base, called the ARIADNEplus data and knowledge Cloud (AC for short). The AC will be born as a LOD dataset, as semantic technologies have reached a mature state of development and, as a consequence, the required resources (ontologies and terminologies) have been made available in one of the semantic web languages. Moreover, the ARIADNEplus AC will extend the ARIADNE Catalogue along several dimensions:

1. **Ontological:** the AC will add representations of the infrastructural entities that participate in the ARIADNEplus network, such as people, institutions and services. Moreover, it will provide a finer-grained characterization of all entities, inspired by recent developments in the modelling of infrastructural resources, such as those of the CRMpe;
2. **Datatypes:** the AC will add to the types of data represented in the ARIADNE Catalogue a whole range of new data types, as described above;
3. **Institutional:** the AC will add to the data of the ARIADNE Catalogue, the descriptions provided by the new partners in the ARIADNEplus consortium;
4. **Granularity:** the AC will include the data resulting from the item-level integration of some datasets collected from the ARIADNEplus network, including overlapping data that can be used to address specific use cases. This extension will push forward the item-level integration experiment successfully conducted by ARIADNE on coin data coming from different institutional repositories and archives.

The richness of the ARIADNEplus AC will be exploited in a more sophisticated semantic discovery service offered via its portal and in a much larger LOD dataset offered to the whole archaeological research community.

The knowledge organization perspective

In ARIADNE the infrastructure was built around a standard that is progressively becoming universally accepted in the archaeological domain: the CIDOC Conceptual Reference Model (CRM).[7] The CRM has been adopted by a large portion of the cultural heritage domain, as witnessed by its influence on the Europeana Data Model. The CIDOC CRM has also created a number of extensions suitable for use within ARIADNEplus.

ARIADNEplus standardization is being defined from the bottom-up, and comprises the following steps:

1. Definition of the necessary application profiles through the collaboration of selected users, i.e. archaeologists with expertise in the different domains providing data for integration, and experts in knowledge engineering. This will produce draft proposals to be checked against perceived user needs and then formalized.
2. Mappings are created (with the mapping service X3ML) and tested on samples from the various involved datasets. Feedback will suggest amendments. This will produce the necessary mapping sets.
3. Vocabularies and gazetteers are selected and improved/adapted.
4. When the mapping is stable, the mapping/conversion/ingestion process starts.

The technological perspective

User operations in the ARIADNEplus framework are enabled by the underlying technological infrastructure and by the services that rely on it. ARIADNE adopted an approach for the integration of key infrastructures that avoided the movement of data around Europe and their concentration into a unique gigantic data deposit. Such a solution would have been hard and expensive to implement, difficult or impossible to maintain and update, and probably ineffective as far as facilitating research is concerned. Instead, ARIADNE created a registry based on the description of datasets stored in the participating infrastructures according to a common data model. The registry can be accessed through the ARIADNE portal and searched by keyword, or according to facets such as place (where), time (when) and content (what). The place and time facets make use of a graphical interface; all three are based on multilingual thesauri.

The goal of ARIADNEplus is not only to extend the ARIADNE registry by incorporating new infrastructures in it, but also to improve the level of integration. This will be achieved by several means:

[7] http://www.cidoc-crm.org/

- Adopting an improved data model for the registry.
- Supporting item-based integration for databases, through database federation.
- Exploiting GIS-based data through item-based integration and services for spatial and temporal analysis.
- Linking text reports to each other and to databases via a common ontology, through NLP and NER applied to texts.
- Enhancing the role of scientific data by linking them functionally to their archaeological context and across investigations for comparisons and syntheses.
- Creating a cloud-based environment where all the above functionalities are implemented, with additional services made available to researchers for data-based investigations.

The ARIADNEplus Cloud and the ARIADNEplus Infrastructure

The ARIADNEplus Infrastructure will be offered on top of the D4Science infrastructure[8] and will consist of a hardware layer and a service layer. The hardware layer will be organized as a dynamic pool of virtual machines, supporting computation and storage, while the services layer will be organized into e-infrastructure middleware, storage, and end-user services. The hardware layer will consist of an OpenStack installation, supporting the deployment of services in the upper layer by provision of computational and storage resources. The service layer (*Figure 3*) will consist of five service frameworks, which can be summarized as follows:

- Enabling Framework, including services required to support the operation of all services and the VREs supported by such services. As such it includes: a resource registry service, to which all e-infrastructure resources (data sources, services, computational nodes, etc.) can be dynamically (de-)registered and discovered by user and other services; Authentication and Authorization services, as well as Accounting Services, capable of both granting and tracking access and usage actions from users; and a VRE manager, capable of deploying in the collaborative framework VREs inclusive of a selected number of 'applications', generally intended as sets of interacting services.
- Storage Framework, including services for efficient, advanced, and on-demand management of digital data, encoded as: files in a distributed file system, collection of metadata records, and time series in spatial databases; such services are used by all other services in the architecture, except the enabling framework; in particular, the ARIADNEplus AC will be implemented on top of the Storage Framework.
- Information Cloud Framework, including all services required to collect, harmonize (transformation), and provide (indexing in different formats and backend) all metadata records describing objects, and links between them, of interest to the ARIADNE community; the ARIADNEplus Infrastructure will be implemented on top of the Information Cloud Framework.

[8] d4science.org

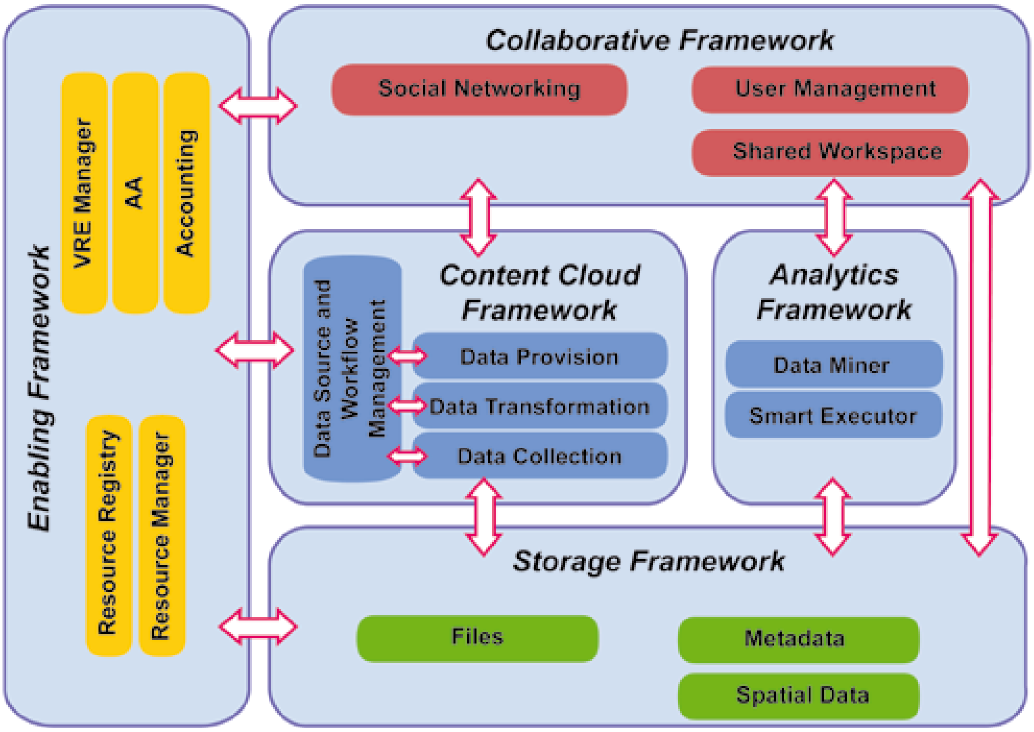


Figure 3: The service layer of the ARIADNEplus Infrastructure

- Analytics Framework, including the services required for running methods provided by scientists taking advantage, in a transparent way, of the power of the underlying computation cloud (e.g. parallel computation) and of a plethora of standard statistics methods, provided out of the box to compute over given input data; the Analytics Framework will be used to account for the Virtual Access to the ARIADNEplus infrastructure offered by the project. Moreover, it will be used in the evaluation of the project impacts, offered to text mining and NLP services and to the ARIADNEplus pilots.
- Collaborative framework, including all Virtual Research Environments (VREs) deployed by researchers, and providing social networking services, user management services, shared workspace services, and WebUI access to the information cloud and to the analytics framework, via analytics laboratory services. ARIADNEplus will use VREs to support the management of the project, as well as the services (including the ARIADNEplus Portal) and the pilots developed by the JRA of the project.

ARIADNEplus services

The front-office services implemented within ARIADNEplus have been chosen as the ones most popular when using data in archaeological research:

- ARIADNEplus visual services
  - Visualization of archaeological imagery, enhancing the visual service already available in ARIADNE and building on the VisualMedia EOSCpilot Science Demonstrator.[9] This will enable the display of visual archaeological information (images, 3D models) in a fast and efficient way. VisualMedia is already developed in a cloud framework, and will be adapted to specific needs of archaeologists, as well as made available in a way coherent with the ARIADNEplus service interface.
  - Visual organization of archaeological data, will build on a tool already developed by CNR to link archaeological documentation to the 3D model of an artefact or monument and visualize it accordingly. Besides a general revision, the existing tool needs porting in the ARIADNEplus cloud environment and adapting it to the service interface.
  - Visual documentation of an archaeological excavation. The service builds on the *Ephemera*[10] service used to visualize in 3D the layers of archaeological excavations and the related documentation. Besides a general revision, it will need porting in the ARIADNEplus cloud environment and adapting it to the service interface.
  - Visualization of scientific data, e.g. the spectra of the numerical results of materials analyses. This builds on tools already available at the Italian National Institute for Nuclear Physics (INFN), but which need embedding in the ARIADNEplus cloud environment.
- ARIADNEplus annotation services
  - Archaeological text annotation. The service consists in an annotation tool for archaeological reports and, in general, texts concerning archaeology. It will build on Open Source annotation tools as, for example, Pundit. The service will be supported by custom archaeological, multilingual vocabularies (also used for NLP and queries).
  - Image annotation. This service is an extension and implementation in the ARIADNEplus cloud framework of the annotation tool DAP developed in the TSS project.[11] The tool permits the annotation of archaeological images in a CIDOC CRM compliant way.
- ARIADNEplus text mining and NLP services
  - The service is based on the previous ARIADNE text mining tool, further developed into TEXTCROWD, a cloud-based NLP tool created as a Science Demonstrator within the EOSCpilot EU project[12]. Work will consist in porting

[9] http://eoscpilot.eu/science-demonstrators
[10] http://ephemera.cyi.ac.cy
[11] http://tss.isti.cnr.it/dap
[12] http://eoscpilot.eu/science-demonstrators

the previous ARIADNE NLP tool in the cloud environment, following what has already been done for TEXTCROWD, and extending the NLP functionality to other languages beyond Italian, English and Dutch. This task will also make extensive use of the vocabularies, gazetteers and time period definitions.

- ARIADNEplus space-time services
  - GIS services comprise the usual services present in GIS systems, for example buffer definition, layer selection, proximity, viewshed analysis and so on, and will be supported by D4Science Cloud geoserver GeoTools, which already has many of the required functionalities built-in. It will also rely on the gazetteers and named time periods vocabularies developed within the project.

The outline of the conceptual approach of ARIADNEplus presented above shows that it is not starting from scratch. It incorporates the results of ARIADNE and builds on them an innovative way of carrying out data-based research in archaeology. It also incorporates the results of other EU-funded projects. Among them, we may quote PARTHENOS, which has contributed to the data model and a number of policy recommendations. Other recent projects providing input are the EOSC-related ones, such as EOSCpilot and EOSChub, with which ARIADNEplus will endeavour to be compatible. Many of the CRM extensions mentioned above were developed within EU projects, for example CRMdig was created in the CASPAR[13] project and refined in 3D-COFORM[14]. D4Science, the backbone of ARIADNEplus processing operations, was originally created with EU funding and further developed in D4Science-II, with services - now to be used in ARIADNEplus - which were implemented for the needs and with the support of Research Infrastructure projects in other domains, such as, for example, the geoserver implementation. Extensive usage of Open Source software will be fostered, including the previously mentioned Geo Tools libraries, the libraries used for NLP and many more. In conclusion, ARIADNEplus will not only integrate data infrastructures. It will also integrate and combine ideas, methods, tools, and results of previous research activity, setting up a new framework aimed at innovating the methodology and the practice of archaeological research.

The ARIADNE community

The ARIADNEplus methodology comprises different components that intersect and are combined to implement the concepts described in the previous section. Teams with different expertise in the ARIADNEplus partnership cooperate in the project activities (*Figure 4*).

We aim to consolidate and extend the existing ARIADNEplus community, paying special attention to Central and Southeastern Europe. The collection of information aimed at defining community needs, actually an update of previous work done within ARIADNE, will also be undertaken. Existing liaisons with major associations

[13] https://cordis.europa.eu/project/rcn/92920/factsheet/en

[14] https://cordis.europa.eu/project/rcn/89256/factsheet/en

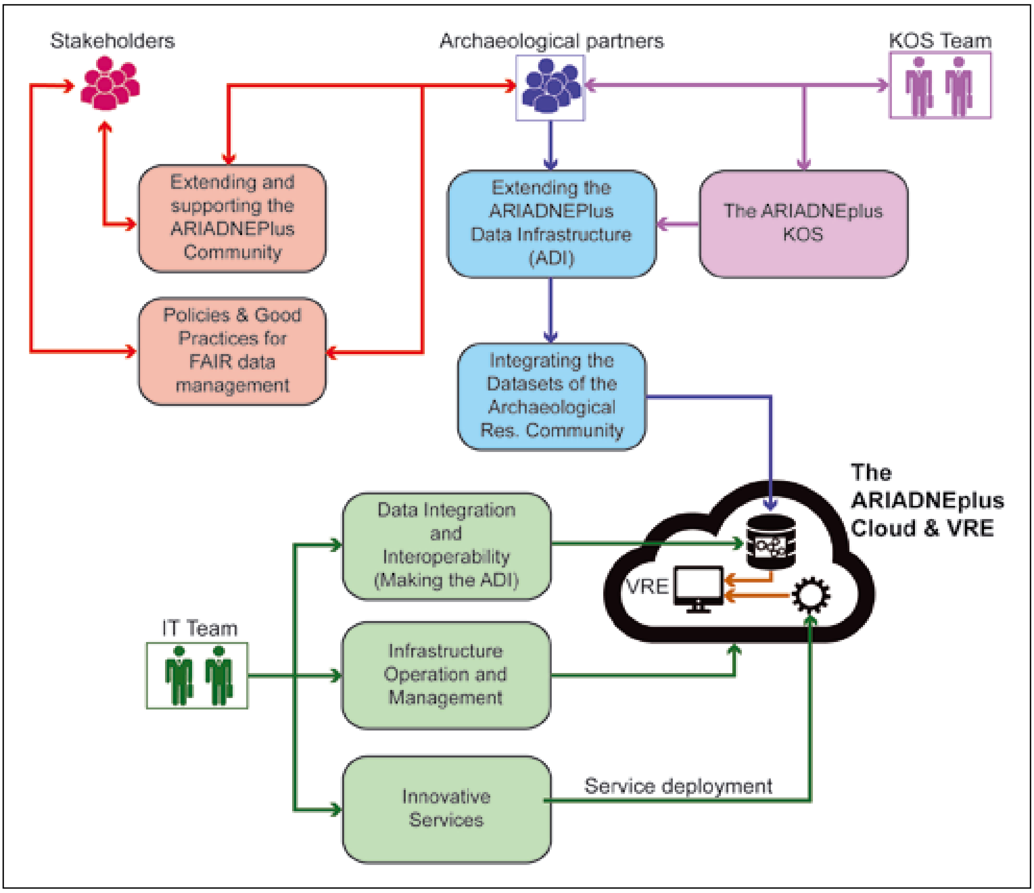


Figure 4: Work methodology in ARIADNEplus

and international bodies such as the EAA and the EAC (European Archaeological Council) will be maintained. Professionals and heritage managers, the 'industry' sector of archaeology will also be embraced. Finally, given that grand archaeological research challenges – such as climate change and migration – are not confined to the continent of Europe, international collaborations will be cultivated and extended.

Developing policies and strategies for FAIR data management targets both archaeological partners and the community at large. The good practices recommendations developed here and the tools, e.g. a Data Management Plan (DMP) template for archaeological data, are provided to users. The ARIADNEplus KOS (Knowledge Organization System) is developed through the collaboration of domain experts, i.e. archaeological partners, and KOS experts, as already mentioned. The application profiles and the vocabularies produced in this activity are then used to create the mappings from each infrastructure to be integrated in ARIADNEplus.

Impact

It is essential, therefore, that ARIADNEplus draws upon its community, but also has an impact on its community. It is anticipated that this may take a number of forms.

(1) Researchers will have wider, simplified, and more efficient access to the best research infrastructures they require to conduct their research, irrespective of location. They benefit from an increased focus on user needs.

Researchers using the ARIADNEplus European infrastructure will have online access to rich knowledge-based data resources from integrated national and domain digital infrastructures, from around Europe and non-European countries (Israel, Japan, Argentina, USA). The ARIADNEplus Infrastructure will allow researchers to search across multiple databases and access data from different archaeological sub-domains.

(2) New or more advanced research infrastructure services, enabling leading-edge or multi-disciplinary research, are made available to a wider user community.

ARIADNEplus will make available a range of advanced and new services for research data providers and users, enabling leading-edge and multidisciplinary research. The project will demonstrate advantages of using ARIADNEplus services and data to the archaeological research community and other user groups.

(3) Operators of related infrastructures develop synergies and complementary capabilities, leading to improved and harmonised services. There is less duplication of services, leading to an improved use of resources across Europe. Economies of scale and saving of resources are also realised due to common development and the optimisation of operations.

ARIADNEplus coordinates its activities with related research infrastructures in the field of humanities and heritage to develop synergies and exploit optimised operations and complementary capabilities. Common development in the field will be continued with regard to harmonised data models (i.e. data catalogues, CIDOC CRM and others). 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.

(4) Innovation is fostered through a reinforced partnership of research organisations with industry.

Private sector actors in archaeology are small businesses of contract archaeologists and consultancies, providing professional services in preventive archaeology. To explore and demonstrate opportunities, project activities address private businesses specifically, including also cultural and creative businesses.

(5) A new generation of researchers is educated that is ready to optimally exploit all the essential tools for their research.

Researchers increasingly need to use advanced tools for data-intensive research paradigms and, at the same time, are requested to acquire skills in data management and sharing based on the FAIR data principles. ARIADNEplus supports these requirements through provision of advanced services, transnational access to the research infrastructures of partners, training offers, and promoting high-quality domain FAIR data and repositories.

(6) Closer interactions between larger number of researchers active in and around a number of infrastructures facilitate cross-disciplinary fertilisations and a wider sharing of information, knowledge and technologies across fields and between academia and industry.

(7) The integration of major scientific equipment or sets of instruments and of knowledge-based resources (collections, archives, structured scientific information, data infrastructures, etc.) leads to a better management of the continuous flow of data collected or produced by these facilities and resources.

Holders of structured data collections will register these in the ARIADNEplus registry, so that they can be discovered and accessed via the e-infrastructure services.

(8) Integrated and harmonised access to resources at European level can facilitate the use beyond research and contribute to evidence-based policy making.

User groups beyond archaeological and cultural heritage researchers will be addressed where appropriate.

Conclusion

It has been argued that archaeologists should excavate less in favour of better exploiting the content already amassed and stored. This is increasingly feasible as the quantity of digital data about new discoveries, or converting/referring to previous non-digital results, is growing. A new Big Data paradigm is appearing in archaeology and in all digital humanities: not the datasets with peta- and perhaps exabytes of numbers, typical of Nuclear Physics, but millions of small files characterize the so-called 'long tail of science', where archaeology belongs with the distinctive characteristics of combining in its knowledge base texts, quantities, and visual content.

Addressing archaeological Grand Challenges such as, for example, early technological advancements, migrations, trade relations, and so on, requires researchers to deal with bits of information dispersed in thousands of datasets hidden among millions of similar (but unrelated) datasets. That is why discovery is the primary concern: creating a sieve retaining only what pertains to a research question. However, it is still necessary to aggregate, combine and further process those filtered data: this is why processing services are paramount. This is how the term *innovation* translates in the archaeological domain: a new way of combining machine processing with human imagination to creatively address Grand Research Challenges about the past and create new knowledge. Discovery as a service was the primary objective in ARIADNE, and is still so - improved and more flexible - in ARIADNEplus. It is complemented by the services described above, which provide a solution to address archaeological research questions. But, above all, ARIADNEplus will create a research ecosystem where data and services are the pillars, openness is the philosophy and collaboration is the attitude.

Acknowledgements

ARIADNE and ARIADNEplus were both funded by the European Commission under the Community's Seventh Framework (Grant no. 313193) and Horizon2020 (Grant no. 823914) programmes.

References

CoreTrustSeal 2016 *An Introduction to the Core Trustworthy Data Repositories Requirements* available at http://www.coretrustseal.org/wp-content/uploads/2017/01/Intro_To_Core_Trustworthy_Data_Repositories_Requirements_2016-11.pdf (accessed 10 August 2019)

ESFRI 2018 *Strategy Report on Research Infrastructures*. http://roadmap2018.esfri.eu/media/1066/esfri-roadmap-2018.pdf (accessed 10 August 2019)

European Commission 2019 'Future of scholarly publishing and scholarly communication. Report of the Expert Group to the European Commission'. doi:10.2777/836532

Meghini, C., Scopigno, R., Richards, J., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F., Felicetti, A., Ronzino, P., Nurra, F., Papatheodorou, C., Gavrilis, D., Theodoridou, M., Doerr, M., Tudhope, D., Binding, C. and Vlachidis, A. 2017 'ARIADNE: A Research Infrastructure for Archaeology'. *Journal on Computing and Cultural Heritage (JOCCH) - Special Issue on Digital Infrastructure for Cultural Heritage, Part 2*. 10.3. doi:10.1145/3064527

RDA. 'All recommendations and outputs'. Available at https://www.rd-alliance.org/recommendations-and-outputs/all-recommendations-and-outputs (accessed 10 August 2019)

Wilkinson, M.D., Dumontier, M., Aalbersberg IJ. J, Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva S., Luiz B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., ,'t Hoen, P.A.C, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. 2016 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data* 3, 160018. doi:10.1038/sdata.2016.18

Wilkinson, M.D., Dumontier, M., Aalbersberg IJ. J, Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva S., Luiz B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., ,'t Hoen, P.A.C, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. 2019 'The Addendum: FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data* 6, 6. doi:10.1038/s41597-019-0009-6

Crossing mountains and digging tunnels from our own backyard to our neighbours

Valentijn Gilissen[*] and Hella Hollander[**]

DANS – Data Archiving and Networked Services
*valentijn.gilissen@dans.knaw.nl
**hella.hollander@dans.knaw.nl

Abstract

Renowned as the Dutch central hub for digital archaeological research data, DANS not only facilitates international research but is also active in enhancing the FAIR qualities of research data as well as enabling archaeological innovation. As a partner of ARIADNE, DANS achieved major contributions to the field of Archaeology including the integration of dendrochronological scientific research data. Additionally, DANS was able to further develop its own services by participation in the ARIADNE project. Research was done by DANS to implement new policies and innovative services into its own infrastructure.

Keywords: implementation of policies; innovation; national hub in international network; project participation

Improvement through project involvement

DANS (Data Archiving and Networked Services) is the Netherlands Institute for permanent access to digital research resources. DANS encourages researchers to make their digital research data and related outputs *Findable*, *Accessible*, *Interoperable* and *Reusable*, according to the FAIR Guiding Principles.

Since 2007, it is stated in the regulations for Dutch archaeologists that all digital excavation project documentation shall be deposited – within 2 years after the fieldwork has been done – in a trusted repository as well as in the governmental depot holding the objects. EASY, the archiving system of DANS is the only qualified repository in the Netherlands. Almost 44,000 archaeological datasets are stored and available (June 2019). The grey literature reports are also kept in the repository of the Dutch Cultural Heritage Agency (RCE), sometimes still in an analogue version.

DANS contributed the archaeological data from EASY as well as the data from the Digital Collaboratory for Cultural Dendrochronology (DCCD) to the ARIADNE infrastructure, making the data better visible internationally via the ARIADNE portal. In close collaboration with Leiden University, pilot work was carried out on data mining and linked data.

The visibility and interoperability of the data greatly benefited from the ARIADNE project. Due to the involvement in the ARIADNE project:

- findability across borders is improved
- successfully mapped vocabularies are enabling concepts-based search
- more research data is made available and accessible
- integrated presentation of datasets is achieved via one central European portal
- archaeological data from grey literature reports are linked with tree-ring data in a pilot on shipwrecks
- the reuse of repository software was promoted by making this available as open source software to stimulate sharing and integration of dendrochronological data.

Thanks to ARIADNE, DANS has also been able to maintain and improve its own preservation services and to set guidelines and standards for automated data deposits at DANS. Giving advice and recommendations to researchers on preferred formats is an important service of DANS.

The EASY and DCCD metadata were aligned to the ARIADNE Dataset Catalogue Model (ADCM) in order to provide valid/correct OAI output/input into the ARIADNE registry. Several EASY Dublin Core metadata fields make use of the archaeological vocabulary ABR+ (Archeologisch Basis Register). The time periods from this vocabulary, used in the Dublin Core field 'Temporal Coverage', were mapped to PeriodO to enable cross-search using absolute date ranges.

All of the conceptual terms from this vocabulary used in the EASY metadata were mapped to SKOS (Simple Knowledge Organization System) and linked to the Art and Architecture Thesaurus (AAT). This enabled concepts-based searching within the ARIADNE portal. This was an invaluable experience from which we obtained knowledge on (meta)data linking. Additionally, the conceptual terms were translated to English, Italian, German, Czech, Spanish and French.

The ARIADNE Map and Timeline features meant that we had to work on the visualisation of metadata. Enabling the display of datasets on the ARIADNE map allowed us to use the methods for working with coordinates within the metadata for our own services. We were able to develop a feature to switch search results from a listing to a map display, showing all results with coordinates in the metadata. Over 3500 archaeological grey literature reports archived in EASY had no coordinates in their metadata. A text mining task was successfully executed in order to identify and extract these coordinates from the PDF files. This highly valuable case of metadata enrichment allowed us to add these records to the display of EASY content on the ARIADNE map.

Additionally, insights from the ARIADNE project helped DANS in formulating the text for the acknowledgement of sources in DOI citation.

Adding 5000 publications and Mediterranean excavation collections

Our work in ARIADNE enabled us to provide the long-term preservation of thousands of archaeological publications and excavation archives for the Dutch Cultural Heritage

Agency and to enable sustainable access to these reports, photos, spreadsheets and maps. This helped us position ourselves as the national hub for archaeological research data; a national hub within an international network.

The 5000 publications obtained from the Cultural Heritage Agency had to be ingested in the DANS EASY archive via a mass import procedure. While bulk-imports had been executed in the past, the means to do so were incompatible with several later improvements of the DANS archive. The necessity and the project means enabled the development of a new, well-structured bulk-import functionality. These methods and tools can easily be used for any deposit of multiple datasets at once. Since the initial import of the 5000 archaeological reports, the bulk-import tools have become part of the standard toolkit of the DANS data managers and have been used multiple times for the import of datasets of various scientific disciplines, including the import of more archaeological datasets which automatically also become findable via the ARIADNE portal.

Figure 1: Zakynthos excavation field photo, Trench D overview[1]

We were able to expand and improve on our existing collection of Mediterranean data. DANS and the UvA (University of Amsterdam) worked to contribute by making archaeological Mediterranean data and corresponding metadata of sites Zakynthos (*Figure 1*), Halos, Boeotia available via DANS in the ARIADNE portal. Data of the Aetolian Studies Project was subsequently added.

Integration of archaeological and dendrochronological data

The DCCD software is an online digital archiving system for dendrochronological data. This software (system) is deployed as 'Digital Collaboratory for Cultural Dendrochronology' (DCCD) at http://dendro.dans.knaw.nl.[2] The vocabularies used within the DCCD Dendrochronology project were cleaned. The controlled vocabularies in SKOS were linked with the Art and Architecture Thesaurus (AAT/Getty) with the help of the University of Glamorgan. This enabled the dissemination of the data within ARIADNE.

1 Wijngaarden 2015: excavation\photos_field\Trench_D\TD_10-07-12\01_Overview_8002_8003_8004_01.JPG

2 More information about the Digital Collaboratory for Cultural Dendrochronology (DCCD) project can be found here: http://vkc.library.uu.nl/vkc/dendrochronology

Working towards dissemination of the DCCD content in ARIADNE also allowed us to improve the DCCD software. A timeline functionality was added to the DCCD services in order to view content and search results on a timeline. GIS functionality and the performance of the interactive map were also improved.

To improve European integration and sharing of dendrochronological data, DANS developed open source repository software from the dendrochronological DCCD-repository of DANS. This enables anyone to use existing components to create their own ARIADNE-compatible dendrochronological archive.[3]

To share expertise within a broader network, DANS and the Dutch Cultural Heritage Agency contributed the Dendrochronology best practice guide. This guide is published as one of the Archaeology Data Service best practice guides (Brewer and Jansma 2016). Another guide on which DANS worked, together with the ADS and IANUS, was on a best practice guide on 3D data (Trognitz, Niven and Gilissen 2016). Building a community of preservation experts is of great value.

(Inter)national networking activities and publications

DANS participated in a number of networking activities and conferences within the context of ARIADNE. First, DARIAH welcomed ARIADNE as an affiliated project within the DARIAH network. In a later stage DARIAH became a full partner of the ARIADNE consortium. As such, ARIADNE was able to benefit from DARIAH, both as a platform for visibility and regarding long-term sustainability. DARIAH developed Virtual Competency Centers (VCC's), with potential use for ARIADNE. As head of the VCC Scholarly content Management DANS participated in the working group 'Thesaurus Maintenance' and shared the spreadsheet with the conceptual terms which were linked to the AAT to incorporate into the Backbone thesaurus to test the new developed tool.

DANS also participated at the launch of the E-RIHS project in Amsterdam. E-RIHS is the European Research Infrastructure for Heritage Science that supports research on heritage interpretation, preservation, documentation and management. DANS joined this Research Infrastructure via ARIADNE and shares its knowledge with this network based on our experience within the DARIAH Research Infrastructure.

In addition, within ARIADNE, DANS was able to build on the deliverables of the successfully completed CARARE project (Connecting Archaeology and Architecture in Europeana). The labelling within the presentation of the use of CARARE data was improved. Hella Hollander and Heiko Tjalsma of DANS contributed to a paper on Archaeology and Intellectual Property Rights (Wright et al. 2016). At the conference Opening the Past – Archaeology of the future in Pisa, 15–16 June 2013, Valentijn Gilissen presented a paper on the ongoing archiving and publication of archaeological data within the broader organisation of DANS. The paper showed how services

3 The open source software is available from the following GitHub repository: https://github.com/DANS-KNAW/dccd-webui

for finding, accessing and re-using data are improved through participation in international projects CARARE and ARIADNE (Gilissen 2013).

DANS also participated in a wide range of dissemination activities under the auspices of ARIADNE. At the CHNT conference on Cultural Heritage and New Technologies in Vienna, 3 November 2013, Hella Hollander presented a paper detailing how the existing infrastructure of the e-depot for Dutch Archaeology allows for sharing of good practices such as long-term preservation, data organisation and data dissemination for accessibility. The paper additionally explains how involvement in projects like ARIADNE helps to develop and expand the archaeological digital infrastructure (Hollander 2013). At the CAA (Computer Applications and Quantitative Methods in Archaeology) conference in Paris, 24 April 2014, Valentijn Gilissen presented a paper on standards and best practices as foundations of services for durable archiving and unlocking of archaeological data (Gilissen 2014).

On March 18th 2016, Hella Hollander presented the importance of a trustworthy archive on the annual meeting of the European Archaeological Council in Brighton. Her article explains the benefits of the existence and use of certified digital repositories saving the cultural wealth of archaeological research data, the impact of national regulations for conducting archaeology, the trend of clustering European infrastructures with a focus on cultural heritage and, finally, gives some future recommendations for shared European archaeological polices to ensure good quality of metadata, data and repositories. This article was subsequently printed in *Internet Archaeology* (Hollander 2017).

At the CHNT conference on Cultural Heritage and New Technologies in Vienna, 21–23 November 2016, Valentijn Gilissen presented a paper on the positioning of the data archive in the research data cycle and the enrichment of data through international collaboration. This paper lead to an article in *Studies in Digital Heritage* (Gilissen and Hollander 2017). DANS also participated as an expert at the forum Digital Archaeology of the future, at the ARIADNE Summer School in Athens.

Training was given to Dutch project leaders of the NWO Odyssee programme, which involved depositing legacy data from decades old archaeological excavations. DANS was present at the Dutch national conference on archaeology called 'Reuvensdagen' in 2017, where a poster showed the improved visibility of the content of EASY through the ARIADNE portal (*Figure 2*).

Figure 2: Data stewards at work, with a view on the Reuvensdagen 2017 poster

All of these dissemination activities serve to share knowledge and best practices in Europe. These activities contribute to the status of DANS as a frontrunner and expert on preservation and enabling reuse of research data. Following the ARIADNE project, DANS became working group leader of other high-profile international projects: SEADDA (Saving European Archaeology from the Digital Dark Age)[4], PARTHENOS (Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies)[5] and ARIADNEplus, in which DANS leads on improving the FAIR qualities of research data.

Advancing the Dutch archaeological data workflow

DANS participated in discussions on the development of an XML metadata quality standard for Dutch archaeological project metadata: SIKB0102. This XML standard enables the exchange of metadata in a uniform manner.

Regarded as an expert on digital preservation, DANS was asked to review a document of the Dutch Heritage Inspection redesigning the national digital infrastructure for archaeology and involved in further negotiations about the flow of the data on a national level. This led to a collaboration between DANS and the provincial and municipal depots wherein DANS provides the certified long-term data repository and the depots are facilitated as a central deposit point for the digital archaeological excavation and survey data. The Provincial depots are enabled to retrieve information at file level of archaeological data collections stored at DANS. Datasets need to be submitted with the aforementioned SIKB0102 XML metadata in order to be processed within this workflow, thereby ensuring standardized and interoperable metadata.

A key identifier within Dutch archaeology is called the 'Zaakidentificatie' (case identification) given to every archaeological project registered with the Cultural Heritage Agency. This identifier was newly introduced in 2016 as an improvement of a preceding identifier from the Cultural Heritage Agency. DANS made a separate field within the dataset metadata for this identifier in order to remove ambiguity in content filtering and therefore improve searching and linking of the content of the archive. The usage of this key identifier also allowed us to further integrate the data at DANS at a national level: to the Cultural Heritage Agency and within the workflow of the Provincial Depots.

Looking forward

Apart from continuous work on Interoperability, our ARIADNEplus focus is on Reusability.

Scientific research data of specialists are fully incorporated in final archaeological reports. To get an indication of the reuse of this information, DANS produced

[4] https://www.cost.eu/actions/CA18128/

[5] http://www.parthenos-project.eu/

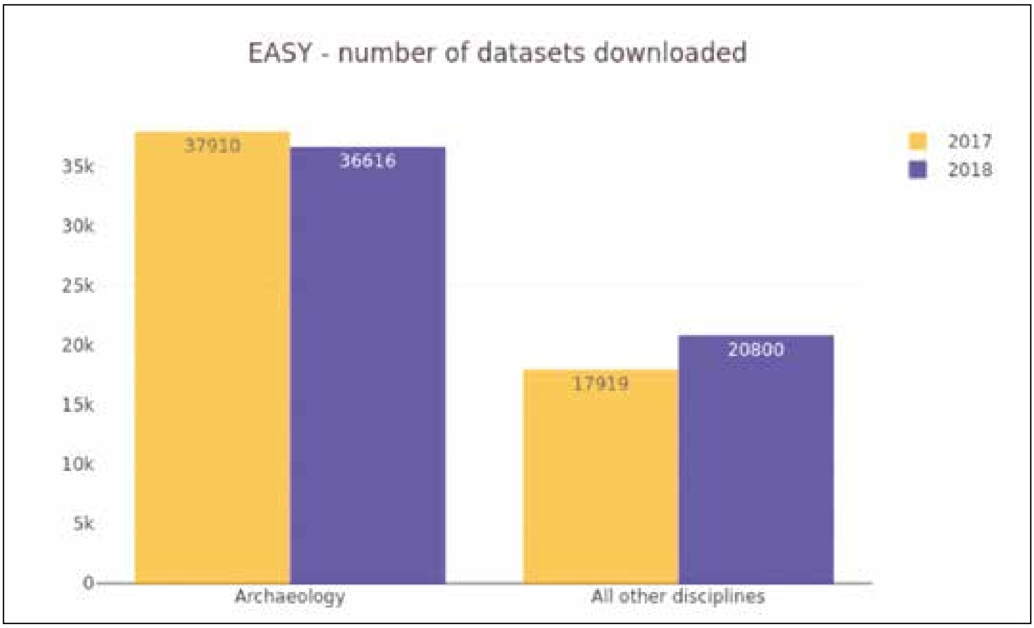


Figure 3: Total number of datasets downloaded from DANS EASY over the years 2017–2018

download statistics in April 2019 for the years 2017 and 2018. This showed that in those years over 74,000 archaeological datasets were accessed. The huge amount of downloads and the fact that, as shown from combining account data with statistics logs, 70% of the downloads is done by accounts belonging to professional archaeologists, means that reuse of archaeological data stored in the trusted repository of DANS is a natural part of the research process (*Figure 3*, with special thanks to Henk van den Berg).

The establishment of DANS as a national hub for digital archaeology means that large volumes of data need to be processed: ingested, validated, preserved and disseminated. There is a growing need for (semi-)automated workflows, wherein the role of the data manager shifts from curating the data after submission, to guiding and monitoring data providers in their commitment to creating FAIR data. Experiences, guidelines and tools from ARIADNE serve as foundations for these new workflows.

References

Brewer, P., and Jansma, E. 2016 'Dendrochronological Data in Archaeology: A Guide to Good Practice', Archaeology Data Service. Available http://guides.archaeologydataservice.ac.uk/g2gp/Dendro_Toc (Last accessed 28 June 2019)

Gilissen, V. 2013 'Past the Opening: Building Towards the Present, On-Going Dissemination of Dutch Archaeological Data as Part of the DANS Archive', *Opening the Past 2013: Archaeology of the Future*, MapPapers 1-III, 42–3. doi:10.4456/MAPPA.2013.17. Available http://www.mappaproject.org/wp-content/uploads/2011/08/OP2013_pre_atti_rid.pdf (Last accessed 28 June 2019)

Gilissen, V. 2014 'Dykes of standards supporting polders of data/ The practices used in the Netherlands for making archaeological data available and accessible'. http://hdl.handle.net/20.500.11755/fcea71b9-8e44-4b0a-af31-1b912398d64e

Gilissen, V. and Hollander, H. 2017 'Archiving the Past While Keeping up with the Times', *Studies in Digital Heritage* 1(2), 194–205. doi:10.14434/sdh.v1i2.23238

Hollander, H. 2013 'The e-Depot for Dutch Archaeology – Archiving and publication of archaeological data'. *CHNT 18, 2013 – Proceedings*, ISBN 978-3-200-03676-5. Available https://www.chnt.at/wp-content/uploads/Hollander_2014.pdf (Last accessed 27 June 2019)

Hollander, H. 2017 'Saving Treasures of the World Heritage at the Digital Archive DANS', *Internet Archaeology* 43. https://doi.org/10.11141/ia.43.9

Trognitz, M., Niven, K. and Gilissen, V. 2016 '3D Models in Archaeology: A Guide to Good Practice', Archaeology Data Service. Available http://guides.archaeologydataservice.ac.uk/g2gp/3d_Toc (Last accessed 28 June 2019)

Wright, H., Green, K., Schäfer, F, Hollander, H.S.H. and Tjalsma, H.D. 2016 'Archaeology and Intellectual Property Rights: CARARE – Project Deliverable'. http://hdl.handle.net/20.500.11755/b420f7b0-c2e1-4517-86f4-a1e7c08983c0

Wijngaarden, dr. G.J.M. van (University of Amsterdam, The Netherlands) 2015 *The Zakynthos Archaeology Project*. DANS. https://doi.org/10.17026/dans-27f-zd2u

The advantage of participating in an EU-project

Ulf Jakobsson

Swedish National Data Service (SND), Gothenburg, Sweden
ulf.jakobsson@snd.gu.se

Abstract

When the first archaeological data was deposited to the Swedish National Data Service (SND)[1] *in late 2011, we were not prepared at all. Neither our management system nor our data catalogue, built for social science survey data, could handle the new types of metadata that were required to make the archaeological data searchable, findable, and visible in our data catalogue. Two years later when the ARIADNE project was launched, SND became one of the technical partners. Developing and implementing services for the ARIADNE portal meant new technical knowledge that could be used to further develop the search system and the catalogue at SND. A new portal based on the same technology as ARIADNE's portal, has been operative since early 2017.*

Keywords: data catalogue; metadata; research infrastructure

Background

Svensk Nationell Datatjänst (SND), or in English: Swedish National Data Service, is a national research infrastructure for making research data accessible both nationally and internationally. We support open access to research data through education, assistance and technical facilities. The basis for this is a web-based research data catalogue with detailed data descriptions and the options to directly download or to order data.

SND was founded in 2008 after a call from the Swedish Research Council for an organization that could make research data accessible for secondary use. Several universities answered the call but the University of Gothenburg ended up as the 'winner' and became the host organization. SND has a background in the organization *Svensk Samhällsvetenskaplig Datatjänst* (SSD) that had been active for more than 25 years, managing and making research data available within the area of social science. SSD had been cooperating with, among others, the Inter-university Consortium for Political and Social Research (ICPSR) as well as the Consortium of European Social Science Data Archives (CESSDA). With the new organization (SND), the scope had to expand from the social sciences to include also the humanities as well as health medicine.

From 1 January 2018, SND became a consortium of seven universities with its main office at the University of Gothenburg. The consortium members provide expertise in research data and data management from areas in which they excel. With

[1] https://www.snd.gu.se/en

the consortium, our scope increased further so that we now are to handle all types of research data from all research areas. Due to this, we are building a distributed uniform system to make data FAIR together with a group of national higher education institutions (universities and university colleges) and research institutes (the SND-network that has more than 30 members). Members of the network have committed to create a data access unit (DAU or equivalent). The function of each DAU is to manage, store and make research data accessible from their own university with the support of the SND-consortium, with the option to make the data accessible via SND's data catalogue.

The beginning

Since SND originated from a social science data organization, and the majority of the staff had that background as well, it is no surprise that knowledge about archaeological data was 'limited'. When I, as the first archaeologist, started working at SND in December 2009, data was primarily defined as structured, rectangular, and most often survey-based. File formats that were used were based on software like SPSS, STATA, and EXCEL.

In August 2011 we were contacted by the researcher Daniel Löwenborg, at the University of Uppsala, who wondered if we could make archaeological data accessible via our data catalogue. The data he wanted to deposit were GIS-data (shape-files) and Access databases. From here, a massive work took off.

At this time, the organization had almost no knowledge of how archaeological data was structured, what kind of data, metadata, and documentation that was used, what file formats to expect, prefer, or even recommend. Suddenly this had to be learnt and implemented. We had an internal management system adapted to social science data, specializing in questionnaire-based surveys. We needed software to view and further uniform the structure of the data and metadata and we had to learn new types of research data: How was the material documented? What information within the files could be used in our data catalogue to document and describe the data and what information should be used as search criteria in our systems? Could we extract that information from the files and if so how, or would we have to do the extraction manually via copy/paste? We also had to make changes and additions to the internal management system so that it would allow better descriptions of digital archaeological materials.

Since I, at that time, was the only one who had any sort of experience and knowledge about archaeological data it naturally fell on me to define what changes that had to be made to our management system, our data catalogue as well as search system. I was quite often questioned about why this and that type of information was needed and who would use it (like geographical information, time periods, map search…).

First, we had to make adaptations, both in our management system as well as in our data catalogue, to the concept 'geographical information'. Here we looked into how to incorporate and later present information like coordinates (bounding box, polygons), new administrative areas in hierarchies[2], and possibly a map search. We also had to look at new types of vocabularies and keywords with archaeological terminologies (until that time we only used ELLST[3]), time periods and the possibility to directly download the data from the web catalogue without registration. Another thing we had to consider was the different administrative identifiers that we might be able to use to connect the data stored at SND with data at the Swedish National Heritage Board as well as social infrastructures on municipal level.

During spring 2012, the data for 230 rescue/contract archaeological surveys covering 269,659 square meters were delivered. After discussions with the researcher, we sent the data back for restructuring of some of the information as well as enhancement of metadata.[4] After some adjustments made by the researcher and his team, and a re-delivery of the data, we published the first 20 archaeological surveys in May/June 2012. This was the first time data was directly accessible via a link. It now became possible to directly download Shapefiles and Access databases without contacting SND.

At SND, we had up to that moment made some changes to our management system. It was now possible to add hierarchical geographical information (based on administrative areas), new types of keywords (based on keywords in the data, including time periods), and it was possible to directly download the data from the data catalogue, but it was not yet possible to add coordinates… Questions I got during these adaptations were like 'Why do you want coordinates?', 'Who's using map search?', 'Aren't keywords enough?, 'How many variables are there in a GIS-file? Not necessary? Why not? It's a survey, isn't it? In social science it's important to mention how many variables… ', 'Time periods? We need a start date and an end date… Does not always apply? Why? You say it's a Time Period! It has to be years then! Not? Why? Why call it a time period if it is not years?' and so it went on. We still wanted to be able to display the surveys via a map interface and for that we needed coordinates.

During the autumn of 2012, we added coordinate fields (Xmax, Xmin, Ymax, Ymin) to our management system so that we at least could present Bounding Box[5] information (manual input though). At the same time, we were investigating whether

2 In Sweden, we have two parallel systems for geographical areas. Simply put, there is a modern system used on a daily basis by local authorities etc. that has an administrative function (civil parish, municipalities and provinces). SND were using this type of information for social science-based data. There is also one that has a more historical tradition and that is used by genealogists, historians, and archaeologists but also authorities like the Swedish National Heritage Board and the Swedish Museum of Natural History. This one has no administrative function today (church parish, counties and provinces). One of the benefits of using this system is that the borders of these areas never change.

3 The European Language Social Science Thesaurus (ELSST) is a broad-based, multilingual thesaurus for the social sciences. https://elsst.ukdataservice.ac.uk/

4 For further reading about this work see: 18th International Conference on Cultural Heritage and New Technologies (CHNT 18), Wien, 2013-11-13, "The OAIS reference model and archaeological data", http://www.ariadne-infrastructure.eu/Resources/Presentations/CHNT-2013-Workshop

5 https://wiki.openstreetmap.org/wiki/Bounding_Box

or not it was possible to extract polygon-information from the shape files directly into our management system and use that in a future map system in our data catalogue. In other words, there was still a lot that needed to be done. One problem was that SND, like most organizations, had a limited staff and limited funding in comparison to what we wanted to do. Developing these kinds of services and systems is expensive.

ARIADNE

When the ARIADNE project started in February 2013 SND had the fortune to become both technical partner and content provider. As technical partner we now had, together with other partners, to build the portal with all its functions and services. We also had, as a content provider, to adapt the metadata and other information we had in our internal management system and data catalogue so that we could share that information into the ARIADNE portal.

When it came to the metadata and the information we had in our systems we had to map and translate the vocabularies we used to the AAT[6], we also had to better define the time periods we used so that it could be shared via PeriodO[7] and later used for metadata forwarded to the ARIADNE portal.

For the ARIADNE project, several services were incorporated into the portal[8]. In addition, metadata from the content providers was used to populate the catalogue. For the search engine, it was decided to use Elasticsearch[9]. One of the benefits of using this system was that it is a full-text search. It has also capabilities to handle structured geographical information to enable exploration and visualization. Geographical metadata could then be explored and visualized using OpenStreetMap[10]. The results of a search in the map interface would then be shown with markers and corresponding information in a table next to the map including a link to the original resource. The resources could also be filtered based on subject and/or time period among others. On the landing page Elasticsearch was also used to find related resources based on geographical and topical similarities.

As mentioned before, participating in the project also meant that we as content provider had to adapt our metadata model so that its structure would fit the requirements by the project. At SND we realized that the techniques and software used and developed during the project could be useful for our own data portal which was based on SQL and was lacking geographical capabilities and fast full-text search.[11]

6 https://www.getty.edu/research/tools/vocabularies/aat/
7 https://perio.do/en/ To find the one used by the project choose: "ARIADNE Consortium. ARIADNE Data Collection. 2015."
8 http://portal.ariadne-infrastructure.eu/services
9 https://www.elastic.co/products/elasticsearch
10 https://www.openstreetmap.org
11 https://snd.gu.se/en/catalogue

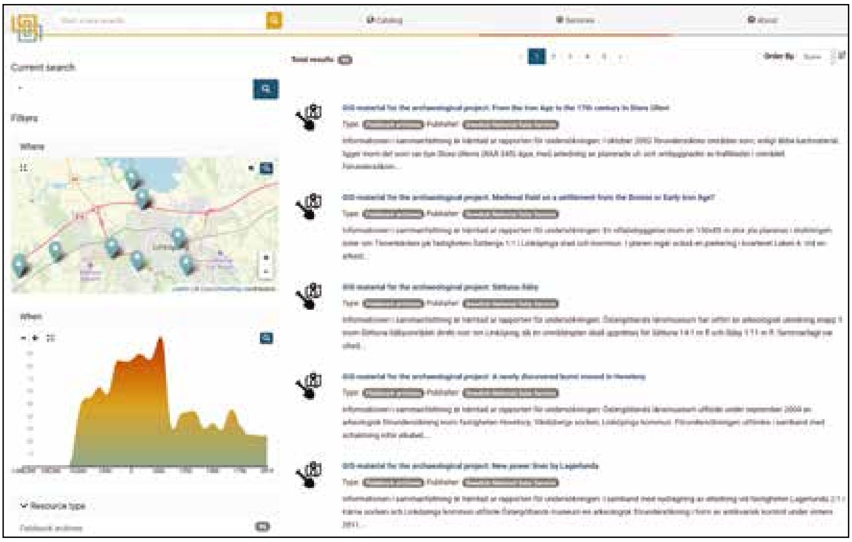


Figure 1: The ARIADNE catalogue

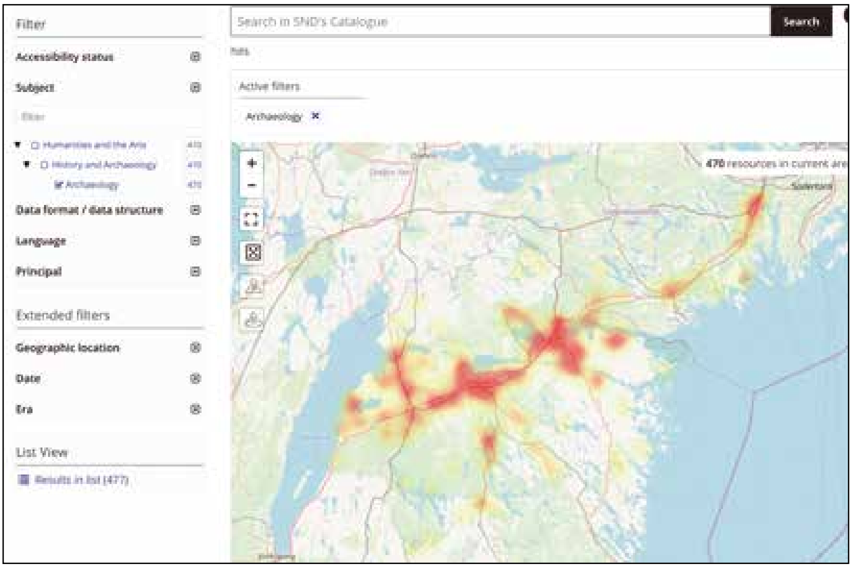


Figure 2: SND's map search with filter search set to archaeology

A quick time jump forward… The project went on and we learned new techniques that we used to adapt our catalogue so that our metadata would better fit the requirements from ARIADNE. In early 2017, SND launched a new portal with the same search engine and techniques as the ARIADNE portal, with Elasticsearch, Map search, time periods, etc.

The new catalogue was better adapted to the scope, types and presentation of data archaeologists and historians are looking for. There are still a lot of things that

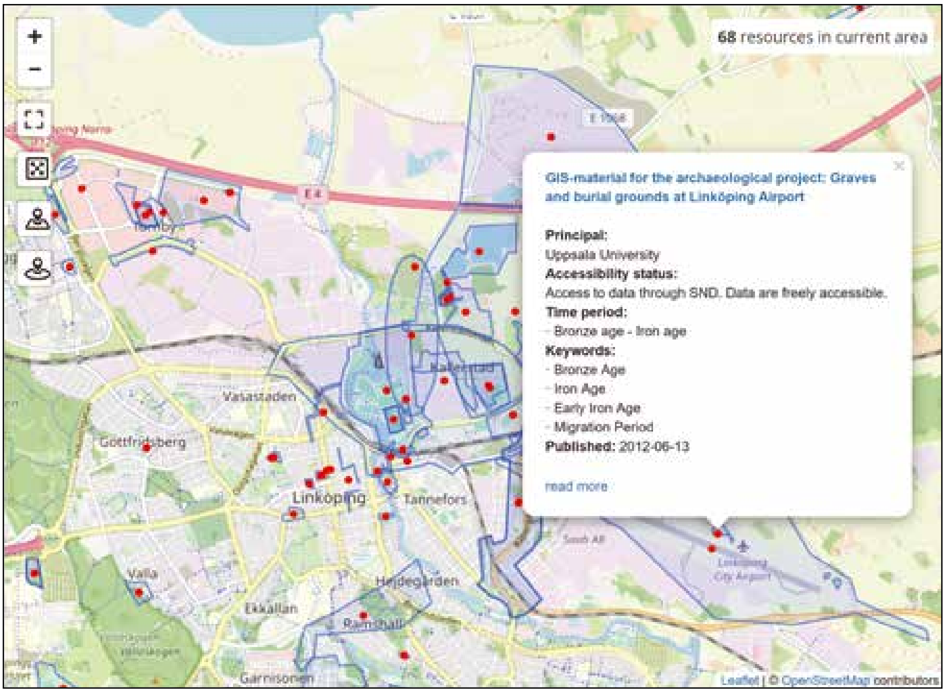


Figure 3: Zoomed map search with the alternatives Marker view and Polygon view both enabled

can be done to better fulfil the requirements by the research community, including improved filtering, better map interface, more services and so on.

The ARIADNEplus project started in early 2019 and it will provide for further services, tools and a better catalogue than the one we developed for the ARIADNE project. As both technical partner and content provider, we are looking forward to implement some of these as well…

On a more personal note, this was the first time I participated in an EU-project, and doing so had many positive aspects, as well as more challenging ones. Deadlines, requirements at short notice, and sometimes miscommunications due to inexperience, as well as long distance travel and language problems can be very stressful. It is imperative that one has institutional support and colleagues with whom to exchange ideas and who sometimes can take over tasks when there is simply too much to do. On the positive side, meeting new people, learning new things, cooperating with people from organizations with other types of knowledge and experience is very rewarding. In addition, having a project that has requirements and needs can sometimes be very useful to get things done in one's own organization. To be able to say that the project requires this can be very beneficial as it can sometimes lead to faster outcomes than individual requests, even though it is the same thing being asked for. If ARIADNEplus can provide further stimulus for internal change then its impact in Sweden will be significant.

The impact of European cooperation on national archaeological research organisations: The example of ARIADNE and Inrap

Kai Salas Rossenbach*, Amala Marx** and Emmanuelle Bryas**

Inrap, Institute for Preventive Archaeological Research, Paris, France
*kai.salas-rossenbach@inrap.fr
**amala.marx@inrap.fr
***emmanuelle.bryas@inrap.fr

Abstract
In 2012, the French National Institute for Preventive Archaeological Research (Inrap) was invited to join the ARIADNE consortium to build a European infrastructure for archaeological digital data. This chapter provides a qualitative assessment of the impact of ARIADNE on the Institute. It demonstrates how this impact had two natures, on the one hand very technical and on the other hand very 'cultural' in fostering an internal culture of data sharing. These impacts have to be understood as the consequence of a formal project agenda and objectives but also as very positive side-effect of the cooperation itself.

Keywords: impact; digital; data; Europe; preventive archaeology

Introduction

Being the largest European research institute working in the field of preventive (development-led) archaeology, the French National Institute for Preventive Archaeological Research (Inrap) is a massive archaeological data producer. With more than 2000 field research operations each year, Inrap generates an immense quantity of digital reports, field records and image data. Data management has therefore become a key strategic issue in recent years. When in 2012 the Institute was invited to join the ARIADNE consortium, the main objective of building an infrastructure for digital data sharing was timely in terms of Inrap's own digital challenges. Shortly after the end of the first phase of the project, it is useful to take a look back and to evaluate the impact of ARIADNE on our own institution. A qualitative assessment allows us to define two broad types of impact. In fact, as we will see, ARIADNE's impact is not only due to Inrap's assigned project objectives but also to significant positive side effects. Defining such impacts requires us to explain the context of the cooperation, the life cycle of the project, to characterise its impacts, and finally to describe the possible futures for our institution within this promising European ecosystem.

Figure 1: The distribution of Inrap archaeological centres

Context

As in everyday life, the digital environment has changed our way of working and our tools but also modified our processes, from the administrative context to the research perspective; the digital development of our world is influencing our way of doing research and possibly our way of thinking. These changes have had a great effect on archaeology with impacts throughout the whole research process, from the field to the library, from data production to data re-use. Inevitably, Inrap has to face this last decade of digital evolution of our discipline and practices.

Created in 2002 by a national law on preventive archaeology (Décret n° 2002-90) under the Malta convention principles (Council of Europe 1992), beyond its mission of carrying out archaeological evaluation and excavation on French territory, Inrap is legally charged with ensuring: 'the scientific exploitation of preventive archaeological operations and the dissemination of their results. It contributes to the teaching, cultural dissemination and enhancement of archaeology'. Our institution is therefore responsible for the dissemination of the results of its operations to archaeologists, to scientific communities at large, but also to citizens.

Inrap carries out about 50% of preventive archaeological excavations on French territory, which represents around 2000 archaeological operations each year. To fulfill its mission, the institute is structured around its headquarters in Paris and fifty archaeological research centres spread across France and the French territories (*Figure 1*).

A large part of the research data is born digital or digitized and every year this process is getting more and more important. Here is an example of the digital data production for one excavation:

- Administrative documentation:
  - Administrative and financial documents prepared before the start of the excavation
  - Information linked to the excavation implementation (mail, minutes…)
  - Scientific documentation gathered before the excavation
  - Documents relating to excavation logistics
  - Documents relating to the security of the site
- Archaeological material documentation:
  - All documentation linked to excavated artefacts, labelled samples…
  - Raw data produced during the fieldwork
  - Excavation recording: field database…
  - Images: georeferenced pictures, video, metadata files…
  - Scanned field diaries
- Spatial data:
  - Georeferenced images
  - Photogrammetry used in field recording
  - Topographic data
  - Georeferenced vector data
- Post excavation studies:
  - Stratigraphic analysis, documentation studies, laboratory analysis, archeological material studies (seeds, ceramics…)
- Excavation report:
  - Report, inventories, figures…
- Dissemination documentation:
  - Video, poster, articles, dissemination support…

Excavation data is the most important part of the institute's data production but data from research projects, digital outreach programmes or digital educational support can be added. The amount of data produced is enormous, even if we are not currently able to quantify its volume. One of the reasons is the dispersion of the data throughout the research centres, each of them managing their own data (*Figure 1*). A lack of an institutional repository to gather and better organize digital documentation, and of a standardized data management policy are important obstacles to data preservation, access and re-use.

However, despite a sense of urgency and of lagging behind the rapid changes facing our discipline, a number of initiatives were undertaken and tools developed prior to the beginning of ARIADNE. The excavation report, simultaneously an administrative document, an activity report and a scientific document was naturally (as a compulsory requirement of preventive archaeology), first digitized and managed using a unique tool: a digital documentary resources catalogue called DOLIA created in 2009 (Bryas et al. 2013). Following the development of the use of GIS at Inrap, an information system called CAVIAR was deployed in 2014 to gather all GIS data collected in the field and allow internal reuse. Finally, and in view of the use of many recording systems by Inrap's archaeologists, from 2014 a field recording system EDArc was developed to address issues related to harmonization, interoperability and re-use of recording data (Tufféry and Augry 2019). But all of these tools and initiatives had to be connected and developed following the rapid evolution of the digital environment, in order to allow data to be more findable, interoperable and reusable according to international norms.

At this stage, in 2012 when Inrap was first invited to become part of ARIADNE, DOLIA was already a fully operational tool, even if evolution was needed (Bryas et al. 2013, 80), whilst CAVIAR and EDArc were still under development. Discussion of the organization, harmonization and conservation of digital data was already underway but many issues remained about regulations and national policies concerning the dissemination and re-use of data.

ARIADNE offered participation in the creation of a digital platform enabling access to data generated across Europe, but also the creation of a dynamic community of users, fostering sharing, use and re-use of digital data. It opened a European space for the discussion and sharing of concerns about harmonization, interoperability, conservation and re-use of archaeological datasets and for the development of skills and expertise within several key areas of archaeological data collection, management and integration.

Project life cycle

Inception

The decision of the ARIADNE coordination team to invite Inrap to become an ARIADNE partner was due to two main factors. Firstly the Institute had, at the time, already been involved in European cooperation for many years, specifically coordinating multiannual projects. This early connexion with some ARIADNE partners and the confidence created through these European projects was a strong argument for integrating Inrap in the consortium. However, the main reason for involving Inrap in the ARIADNE consortium was its position in the data creation ecosystem. Preventive archaeology, whatever the national regulation system, has become the main producer of archaeological data in Europe. The overwhelming amount of research activities generated by development in the last thirty years has completely transformed the archaeological sector in terms of jobs, practices, knowledge and more specifically data management. Including Inrap in ARIADNE was a way to represent this strategic sector in the consortium, where the Institute became a full partner and the sole representative of France. Inviting Inrap to participate was also a way to cover this key country for archaeological research in Europe. The fact that the Institute is a public organisation also reflected the French context, as France is strongly dominated by the public sector. Furthermore, in 2013 the Institute was at this digital turning point, with many digital systems being set up but a global digital framework and strategy still to be established. In this particular aspect, ARIADNE offered Inrap a strong networking channel to partners who represented the international state-of-the-art of digital humanities.

Figure 2: Workshop to test the ARIADNE infrastructure at Inrap headquarters (Paris) 31st May 2016, with the participation of the CNRS, Traces, Citeres and Bibracte

The role of Inrap

In the overall division of the consortium between technical and archaeological partners, Inrap was clearly included in the latter. Its main role was to provide input from its archaeological perspective, i.e. confronting project development with the everyday life and regulation of the archaeological ecosystem in France. As such, the added value of the institute was around archaeological data management in the field of preventive archaeology where topics such as grey literature (reports) or field database management are key issues. Inrap also connected the project with the French community not only by disseminating its surveys and results but also in proposing that French institutions join the consortium as associate partners. This led to the integration of the PACTOLS thesauri as a core reference for ARIADNE vocabularies in the French language. Inrap took charge of the mapping of the French thesauri to the Getty's Art and Architecture Thesaurus, allowing full access to French data.

Inrap also played the key role of being one of the strongest and earliest data-providers of the project. Making available to the infrastructure many datasets, principally a huge set of archaeological reports, Inrap is one of the major contributors to the infrastructure with c.28,000 items. Furthermore, the Institute also led a task with the objective of testing the services provided by the infrastructure. Bringing together all the archaeological partners of the project this task had two components. Firstly, it consisted of coordinating the involvement of European partners in testing each service from an archaeological perspective. Secondly, Inrap organized several workshops with French archaeologists employing a methodology of testing using a bottom-up perspective (*Figure 2*).

Project impact

European state of the art

The main impact of ARIADNE on Inrap has been increased knowledge of digital data management. The experience of working inside the consortium clearly benefited Inrap professionals in term of skills as well as an appreciation of the importance of data management within the institution more widely. The project provided the opportunity on the one hand to exchange ideas with other European archaeological institutions and benefit from their experience in facing their own data challenges; on the other hand it provided an opportunity to work with more technical partners specialised in humanities IT, allowing Inrap to gain knowledge and to experiment with state-of-the-art technologies. These included the use of the CIDOC CRM; the exchange of experience with the top-level Archaeology Data Service; the use of SKOS; or work on grey literature within CNR. This exchange of knowledge took the form of informal training in the way of 'learning by doing' but also in a more formalised way when Inrap professionals participated in summer schools and training throughout the project. Bringing this European knowledge into France and our Institution was a major result of the project.

ARIADNE was also a means by which to highlight the quality and quantity of data production in Inrap. Sharing our datasets and reports within the ARIADNE infrastructure drastically improved the visibility of our research work in France and Europe. Thus, the project took Inrap further in the European Research Area and helped the institution to take an important reputational step forward.

Cultural impact towards data sharing

The impact of ARIADNE on Inrap should not be seen as only technical, however, and is also 'cultural'. When the project began, Open Science policies were just being generalised at a European level. The project represented a clear opportunity to be in the forefront of the data sharing movement. Inrap was not fully prepared to work in such an ecosystem (as is still the case with the vast majority of archaeological institutions). The culture of data sharing was not as widespread as it is now and archaeological partners had to adjust their own practices to the project's objectives and requirements. In this sense, ARIADNE was very innovative in terms of its impact on institutional culture change. From this point Inrap initiated a change in its strategy towards data sharing practices. Even if this transformation is not yet completed, its primary origin can be found in ARIADNE. But how did this work? We can say that it was due to two mechanisms. First, Inrap had a contractual obligation to provide the infrastructure with data. Complying with this objective led Inrap to launch reflexions on its own practices and solutions to make its data FAIR. This began in the preparation of the project and continued until the end, four years later. At the same time, participating in this international consortium led to a cultural assimilation with European colleagues and institutions that had already advanced in this way. It is worth mentioning here the digital repositories (ADS, DANS, SND) which directly influenced our institution by being good examples of concrete implementations of data sharing. With their specificities, successes and failures these top-level organisations have a very strong influence on many institutions and specifically ours. The contact with other archaeological partners trying to deal with the same issues in the project also gave us some comparison elements and some very good examples of ways to follow. This cultural shift for our organisation and the confrontation with European partners showed Inrap that sometimes we were doing very well (indeed better than we thought) and sometimes that the road was still very long.

In the course of the project, it became clear that our own disciplinary ecosystem cannot do everything by itself and that is very dependent upon the evolution of national and European policies and regulations. In this sense, the legal panorama in 2012 is completely different from today. Our institute has new national legislation fostering data sharing in the public and research sector (Law for a digital republic - LOI n° 2016-1321, etc.) and at European level, the Open Science policy has become a key aspect of the European Research Area. In summary, the impact of ARIADNE on Inrap is very powerful in the way it has influenced our strategies towards data sharing and in this regard, it is a great success for our Institute, whatever the outcome.

Limitations and prospects

Thus, the impact of ARIADNE on Inrap was significant but we cannot hide the difficulties we faced, particularly during the first phase of the project. The large size of our institute, its national dimension, and difficulties in understanding what a European project could bring to our discussion and organization about digital data management impeded the development of the project internally. The project got off to quite a slow start requiring time-consuming internal communications. Numerous presentations and explanations of the project helped to build trust in ARIADNE and helped Inrap professionals understand that this project could provide active support without imposing anything.

One of the root misunderstandings came from the English language and led to difficulties in gaining a clear understanding of the core concept of the project, generating fear of uncontrolled data dissemination. Much discussion focused on the legal right of Inrap to provide access, even restricted, to the grey literature and raw data, as they were considered both as state-owned administrative documentation and subject to copyright. Despite the open data and open access movement launched 10 years earlier in the French research domain, our institute, as mentioned above, had to await a change in the national regulatory framework with the promulgation of the Law for a Digital Republic in 2016 (LOI n° 2016-1321). With a few exceptions, this makes it mandatory to publish digital administrative documents online. Produced as part of a public service mission, the data resulting from excavations may be affected by this law.

One of the other aspects that initially slowed down the project was the lack of human resources to pursue and develop project activities both within the institute and outside. Following the recruitment of a dedicated person integrated within the Inrap project management team, everything became easier and more efficient.

During the two-and-a-half-year interregnum between ARIADNE and ARIADNEplus, Inrap launched many institutional initiatives to identify, structure and preserve its archaeological data in order to prepare for their sharing and reuse. Actions were also undertaken to raise the quality of data production in the field in accordance with the FAIR principles.

ARIADNEplus will give Inrap the opportunity to provide access to more data by referencing new datasets in the ARIADNE infrastructure, including new excavation reports but also research databases, absolute dating documentation, and so forth. We will rely on this European project to push forward data structuration and data openness. We will take the opportunity provided by ARIADNE to disseminate the FAIR principles and data management good practice through our colleagues and processes. We will continue to develop the French user community in full collaboration with the other French partner, the CNRS and the MASA consortium. Our new challenge is to continue to promote the ARIADNE portal within the French archaeological community but also to raise awareness about ARIADNEplus at the political level (namely ministries) and amongst the heritage management community.

As leaders of a work package, Inrap will coordinate and develop pilots to test the infrastructure on real research cases. It is fundamental that archaeological partners test the infrastructure and services under real conditions of use from an end-user perspective. We will work on seven case studies to define the innovative methods enabled by ARIADNEplus services for archaeological research communities and demonstrate the advantages of using ARIADNEplus to archaeologists. The ARIADNEplus data and services will also be showcased to build applications for professionals and heritage managers, but also for the wider public.

Conclusion

Shortly after the end of the first phase of the ARIADNE European digital infrastructure project it has been very useful to have a look back and to try to analyze its impact on our Institute. Having explained the context of the project, as well as the position and role of Inrap, we qualitatively described the means and effect of its impact. When we look closely at these four years of cooperation we see that ARIADNE's impact followed two channels. Firstly, the predefined project objectives guided the institute towards new reflections and decisions and secondly, there was a complex process of assimilation and example-based learning. Highlighting this double phenomenon leads us to propose, for future cooperation, the reinforcement, inside such consortia, of bilateral associations of institutions susceptible to share common objectives and experiences. Our post evaluation shows also that the concrete impact has two forms. ARIADNE clearly allowed us to take a technical and technological step forward towards the European state-of-the-art in digital humanities. But, maybe more importantly, ARIADNE has had a strong 'cultural' influence on our institution by promoting a culture of data sharing amongst our professionals and internal decision-makers. In this sense the innovation potential of ARIADNE has to be seen technically speaking, but also in the social innovation it has created in the archaeological digital ecosystem.

References

Bryas, E. Bellan, G. Speller, A. and Carpentier, C. 2013 'Dolia, le système documentaire de l'Inrap. État des lieux et perspective', *Archéopages* 37, 72–80. doi:10.4000/archeopages.357

Council of Europe 1992 *European Convention on the Protection of the Archaeological Heritage (Revised), Valetta, European Treaty Series n°143.* Available https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/090000168007bd25 (Last accessed 17 July 2019)

Décret n° 2002-90, *Décret n° 2002-90 du 16 janvier 2002 portant statut de l'Institut national de recherches archéologiques préventives.* Available https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000770535&dateTexte=&categorieLien=id (Last accessed 17 July 2019)

LOI n° 2016-1321, *LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique.* Available https://www.legifrance.gouv.fr/affichLoiPubliee.do?idDocument=JORFDOLE000031589829&type=general&legislature=14 (Last accessed 19 July 2019)

Tufféry, C. and Augry, S. 2019 'Harmonisation de l'acquisition des données d'opérations d'archéologie préventive. Retours d'expériences et perspectives à partir de l'application EDArc', *Actes de l'atelier DAHLIA dans le cadre de la conférence EGC 2019,* 21–27. Available http://dahlia.egc.asso.fr/atelier_DAHLIA2019_actes.pdf (Last accessed 17 July 2019)

ARIADNE and ARIADNEplus in Austria

Edeltraud Aspöck

Austrian Centre for Digital Humanities, Vienna
edeltraud.aspoeck@ÖAW.ac.at

Abstract

In Austria, the ARIADNE project was vital in creating an awareness of how fragile our digital heritage is and how necessary standards and data management plans are to protect archaeological data from destruction. The project accompanied the creation of a data archive for the humanities at its host institution, the Austrian Academy of Sciences and it facilitated the creation of two new datasets applying standards and Open Science principles: the Digitizing Early Farming Cultures *and the* A Puzzle in 4D *projects. The ÖAW contributed to the development and testing of the CIDOC CRM ontology extension CRMarchaeo for archaeological data.*

Keywords: archive; CIDOC CRM; collaboration; data management; open science

Introduction

In 2013, at the beginning of the ARIADNE project, many topics of ARIADNE were new to Austrian archaeologists. Austria, like most of the ARIADNE partner countries, did not yet have data repositories in place. Hence the idea of large-scale integration of archaeological datasets held in repositories to create a data infrastructure was far from the reality of Austrian archaeologists (compare Corns et al. 2014). Nevertheless, the project received a lot of attention from its outset, both in the media (there was a number of online press articles and even a radio broadcast) and from the archaeological community. During its four years the ARIADNE project played an important role in creating an awareness of how fragile our digital heritage is and how necessary therefore standards and data management plans are to protect archaeological data from destruction.

In Austria, the Austrian Academy of Sciences (Österreichische Akademie der Wissenschaften) is the archaeological partner of ARIADNE and ARIADNEplus. The Austrian Academy of Sciences (ÖAW) is the national body for science and research in Austria and conducts research over a wide range of disciplines. The ARIADNE project was based at the Institute of Oriental and European Archaeology (OREA), where research activities cover archaeological fieldwork and material studies, together with interdisciplinary work in areas such as archaeozoology, anthropology, geoarchaeology and landscape modelling. The ARIADNEplus project is hosted by the ÖAW Austrian Centre for Digital Humanities, an institute which facilitates digital research in the humanities and which runs the new ÖAW digital repository ARCHE.[1]

[1] https://arche.acdh.ÖAW.ac.at/browser/

The second Austrian ARIADNE and ARIADNEplus partner is Salzburg Research, an independent research institute with a focus on information technologies. In ARIADNE, Salzburg Research was responsible for the assessment of user needs, community building and development of an innovation agenda for the project. This article will focus on the work carried out at the ÖAW and the impact these activities had for the archaeological community in Austria. ÖAW activities in ARIADNE included the organisation of training events to foster data management in archaeology, which provides the basis for standardized datasets which can be related to other data and integrated into an infrastructure. Furthermore, OREA datasets were prepared for the integration into the ARIADNE infrastructure, and one important focus of activities has been the work with the CIDOC CRM standard, the ontology that formed the basis for the ARIADNE data model and for deep data integration. Finally, ARIADNE paved the way for the development of two new ÖAW datasets, the *Digitizing Early Farming Cultures* and *A Puzzle in 4D* projects, both adhering to data standards and implementing the ARIADNE approach to data integration and Open Science principles.

The community: caring about preservation of archaeological data

In the beginning of ARIADNE, many of the topics of the project, such as data integration and data infrastructures were new to archaeologists. As a consequence, a discussion started about the state of the art of data standards, data management and digital long-term preservation in archaeology in the partner countries. It turned out that there are big differences across Europe and hence across the ARIADNE partner countries and institutions. Basically, there are those countries who already have data repositories and hence standards in place (a few) and those who have not (most) (Corns et al. 2014; Wright and Richards 2018). A need to address this issue and to bring all partners into the boat was identified early. Training events in data management and digital long-term preservation as well as support for the development of new datasets that adhere to standards were seen as necessary.

An inventory of digital resources that was carried out in 2013 at the ÖAW Institute for Oriental and European Archaeology (OREA) confirmed this view (Aspöck and Masur 2014). The survey was undertaken to provide an overview of digital resources held at OREA as a basis for activities in ARIADNE. OREA was a newly founded Institute at the ÖAW, unifying three former independent Commissions of the Academy. However, part of the survey has also been an assessment of data standards, data management practices and needs. In this survey, archaeologists expressed a strong need for guidance, training and support in data management in mainly two areas: database design and data management during the lifetime of a project (how to curate digital data, updates, storage, naming conventions). At that time there was little concern about issues of long-term preservation – the print publication was still seen as the end point of the life of any digital object. The only role of digital data was seen to provide the basis and building blocks of a final publication. There was no awareness of possible data re-use and hence no need to prepare data for long-term preservation.

Conference sessions and workshops eventually led to an increased awareness in the archaeological community that our digital cultural heritage is a fragile property and that something needs to be done to preserve it. Knowledge and tools to achieve this have been disseminated. The ÖAW has been organising and contributing to a series of such events as part of ARIADNE activities. Awareness was first raised at a session at the 'Cultural Heritage and New Technologies' (CHNT) conference in Vienna in November 2013, where the Austrian ARIADNE partners Guntram Geser (Salzburg Research) and Edeltraud Aspöck (ÖAW), organised a session titled 'Infrastructures and services for sharing of archaeological documentation', introducing the ARIADNE project (Aspöck and Geser 2014). A year later, an ÖAW workshop on repositories was organised by the ÖAW institutes OREA and the Austrian Centre for Digital Humanities (ACDH). This workshop, entitled 'Save the data' invited representatives from data archives and infrastructure initiatives in the humanities and archaeology to present their answers to a set of questions about repositories raised by the organisers. Delegates from the British Archaeology Data Service (ADS) and the Dutch Data Archiving and Networked Services (DANS), as well as from the ARIADNE project presented their responses. Finally, in January 2016, a very successful ARIADNE data management training event was held at the Academy of Sciences in Vienna (*Figure 1;* Wright and Richards 2018). Holly Wright (ADS) and Kate Fernie (PIN and 2Culture Associates) discussed why special processes are needed for data archiving, which data to archive and what to discard, what sort of equipment is needed and good practices and case studies. They also spoke about the current state of affairs in Europe and Edeltraud Aspöck presented the perspective of archaeological data policies and standards in Austria. In 2016, a workshop titled 'Old excavation data – what can we do?' was held at the 10th International Congress on the Archaeology of the Ancient Near East (ICAANE) in Vienna. The idea for the workshop came about as part of our work on the 'A puzzle in 4D' project (see below), which aims to digitise the resources from the Austrian long-term excavation project at Tell el-Daba, Egypt to make them available open access online. At the beginning the project was facing many questions, such as how to organise the digitisation process of the vast number of analogue resources, including some complicated and deteriorating materials, such as colour film negatives? Hence we invited similar projects to introduce their approaches to our questions. The workshop is now being published as a book (Aspöck et al. forthcoming), introducing projects that publish excavation data online (such as the Ur online project) and software solutions (e.g. OCHRE data service, the ARCS project). Finally, the end of the ARIADNE project was again marked with an event at the CHNT conference in Vienna in November 2016. At a round table, the Austrian ARIADNE partners invited archaeologists to discuss 'Long-term preservation and access: Where is an archive for my data?' – the facilities for long-term preservation were still not been there.

The two most important outcomes of these dissemination activities were that first, a larger number of archaeologists in Austria – and in Europe – became aware that digital data is fragile and that it is imperative to do something to preserve our digital cultural heritage. The second outcome, which ARIADNE activities contributed to, was that in December 2017 the ÖAW institutional repository ARCHE (A Resource Centre

Figure 1: An ARIADNE data management training event was held at the Academy of Sciences in Vienna in January 2016

for the HumanitiEs) was launched, and in 2018 ingest of archaeological datasets from ÖAW institutes started. Furthermore, the University of Vienna launched its repository Phaidra,[2] which has also already been tested for its suitability for archaeological data (Hagman 2018).

Contributing to the CIDOC CRM ontology and extensions

The ontology CIDOC CRM (ISO 21127:2006) enables information exchange and integration of heterogeneous sources from the cultural heritage domain. The ARIADNE Reference Data Model is based on the CIDOC CRM and extensions to allow deep integration of scientific and cultural information. The ÖAW contributed to the development of the CIDOC CRM extension CRMarchaeo, which supports integrating archaeological excavation data (Masur et al. 2014). There are a variety of ways to document excavations and excavation activities. In some countries there are standardized recording sheets, and in many countries particular documentation methodologies have become a de-facto standard, e.g. the Museum of London 'single context recording system' (MoLAS 1994). Guidelines and standards for archaeological practice have been published in many European countries. For example, in Austria guidelines for archaeological excavations were first released in 2012 by the Austrian Bundesdenkmalamt (Federal Monuments Authority Austria) (BDA 2018). Hence, excavation documentation usually varies according to country. There are also special recording forms for the requirements of specific types of sites, such as underwater and Palaeolithic sites. Published and unpublished excavation sheets from different

2 https://phaidra.univie.ac.at/

countries representing different excavation methods and guidelines were collected and compared. Using case studies it was shown how CIDOC CRM can support their integration (Masur et al 2014).

ÖAW datasets were used as case studies for deep integration of scientific and cultural-historical evidence (Doerr et al. 2015). This included the *dFMRÖ* (Digitale Fundmünzen der Römischen Zeit in Österreich/ Digital Coin-finds of the Roman Period in Austria) an online MySQL database of the Numismatic Research Group of the Austrian Academy of Sciences.[3] Since the 1990s it documents coin-finds from the Celtic and Roman periods that have been published in various printed volumes of the FMRÖ (Fundmünzen der Römischen Zeit in Österreich / Coin-finds of the Roman Period in Austria) from the 1970s up to 2007. The dFMRÖ database currently hosts about 76,000 finds from Austria and Romania. The second test case was the cemetery database 'Franzhausen-Kokoron', holding data from 400 cremation graves from the late Bronze Age Urnfield Culture (1050–800BC; Lochner and Hellerschmid 2010). A Microsoft Access database was created in 2006 to catalogue and analyse the graves. In 2010 the database was published online with an interactive cemetery plan interface for viewing the records of each grave. The data were all mapped to the CIDOC CRM and it could be demonstrated that these databases about quite specific archaeological content can successfully be integrated with the rather generic ARIADNE Reference Data Model without loss of specificity of meaning (Doerr et al. 2015). This was a major step towards large-scale, high-quality information integration to create resources for the researcher. Finally, all the ÖAW datasets we worked on were of course integrated into the ARIADNE catalogue and can be queried and accessed via the ARIADNE user interface.

Following the ARIADNE approach: the *Digitising Early Farming Cultures* and *A Puzzle in 4D* projects

ARIADNE supported the development of two new datasets at the ÖAW. The projects follow the ARIADNE approach in usage of data standards and aim to provide open data according to the FAIR principles.

The *Digitising Early Farming Cultures* (DEFC) project[4] integrated information from resources pertaining to the Neolithic periods of Greece and Anatolia, including databases, publications and a pottery collection (Aspöck and Masur 2015; Andorfer et al. 2016; Štuhec et al. 2016). The project was supported by ARIADNE and received additional funding from the National Foundation for Research, Technology and Development (ACDH 2014/22). The resulting 'DEFC-App' is structured like a site database but, unlike most site databases, its very granular data model enables integration of detailed information on finds. Hence, the application allows the user to query for information about typical finds from that period and region. This includes 3D-models of pottery from the Schachermeyr pottery collection, a teaching collection for professionals and students (Schachermeyr 1991; *Figure 2*).

[3] https://www.ÖAW.ac.at/antike/fmroe/content/suche.de.php

[4] https://defc.acdh.ÖAW.ac.at/

Figure 2: 3D-models of pottery from the Schachermeyr pottery teaching collection (Schachermeyr 1991) typical for Neolithic periods in Greece and Anatolia are available via DEFC app

In the DEFC project, 'open' pertains to nearly all aspects of the project. The DEFC app provides linked open data and employs several standards, including mappings to the CIDOC CRM. Access to the data is open – data can be queried and downloaded via the application. The vocabulary of the thesaurus is also openly available for reuse in other projects. Information about the creation of the application can be accessed via the project homepage in the 'building the defc app' blog. The posts provide information about the methodology, workflows and tools that were used, and, most importantly, also about things that did not work out and where other ways had to be sought. Furthermore, the source code of the DEFC app is available on github.[5] The code has already been reused for a second project, the CBAB (Cremation Bronze Age Burials) application.[6] The DEFC app can be used by scholars and students internationally for resource discovery (to get basic information about an archaeological site) and to familiarise themselves with typical finds from that period and region. Currently, the value of the application as an actual research tool is tested in a case study of temporal-spatial distribution of finds.

5 https://github.com/acdh-ÖAW/defc-app
6 https://cbab.acdh.ÖAW.ac.at/

The *A Puzzle in 4D* project provides digital long-term preservation for the resources from the Austrian excavations in Tell el-Daba, in the eastern Nile delta regions of Egypt (Bietak 1996; Aspöck et al. 2015). Austrian excavations at Tell el-Daba took place since 1966, revealing evidence for a wealthy society which had contacts with many parts of the eastern Mediterranean dating to the 12th to 18th dynasties (early second millennium BC). After 50 years of fieldwork, huge and heterogeneous amounts of digital and analogue resources such as photographs, plans, drawings and written documentation have been created and are held at the archive of the Institute for Oriental and European Archaeology (OREA) at the Austrian Academy of Sciences. The overall aim of the 'A Puzzle in 4D' project (2015–2020) is to provide digital long-term preservation of the knowledge and information from the over 50 years of excavations by preparing them for long-term archiving based on national and international standards in the ÖAW institutional repository ARCHE. Parts of the Tell el-Daba archive will be made available open access online for researchers and the interested public. The project has been supported by ARIADNE and received funding from the National Foundation for Research, Technology and Development (DH 2014/12).

For the digital archiving of the resources, we had to create a data model that represents the relationships between the information in the analogue and digital documentation, the actual archaeological evidence, the process of excavating this evidence and, of course, of digitizing and processing the resources. We have used the CIDOC CRM ontology to represent this complex network of information. Using the CIDOC CRM for cultural heritage documentation did not only allow us to create the complex relationships that we needed for our data model, but it also enabled the data to be encoded in a machine-readable format. This is important if other computers want to access the information from the Tell el-Daba archive.

To prepare the analogue resources for digital long-term preservation paper records such as field drawings had to be scanned and metadata for the scans had to be created. Metadata includes information about the archaeological content of the resource as well as characteristics of the analogue resource and information about the digitization process (*Figure 3*). The organization of the metadata follows the standards set by the CIDOC CRM ontology. Typical metadata records in *A Puzzle in 4D* for a digitized photo of a find are for example the inventory number of the find, the date when the photo was made and information about the scanning. The metadata about the finds include information about the type of find, the dating of the find and the find spot. The relationships that we have created between the different types of metadata allow complex querying of the archive. In the case of digital resources, the original files have to be converted into durable file formats suitable for long-term archiving and relevant documentation has to be added. Best-practice guides from the repository ARCHE and IANUS are being used.[7] Finally, the files will be imported into the new ÖAW data hosting service ARCHE. The repository builds on the well-established open-source repository software Fedora Commons version 4 which provides a sound technological basis for implementing the OAIS (Open Archival Information

[7] https://www.ianus-fdz.de/it-empfehlungen/inhalt. https://arche.acdh.ÖAW.ac.at/browser/

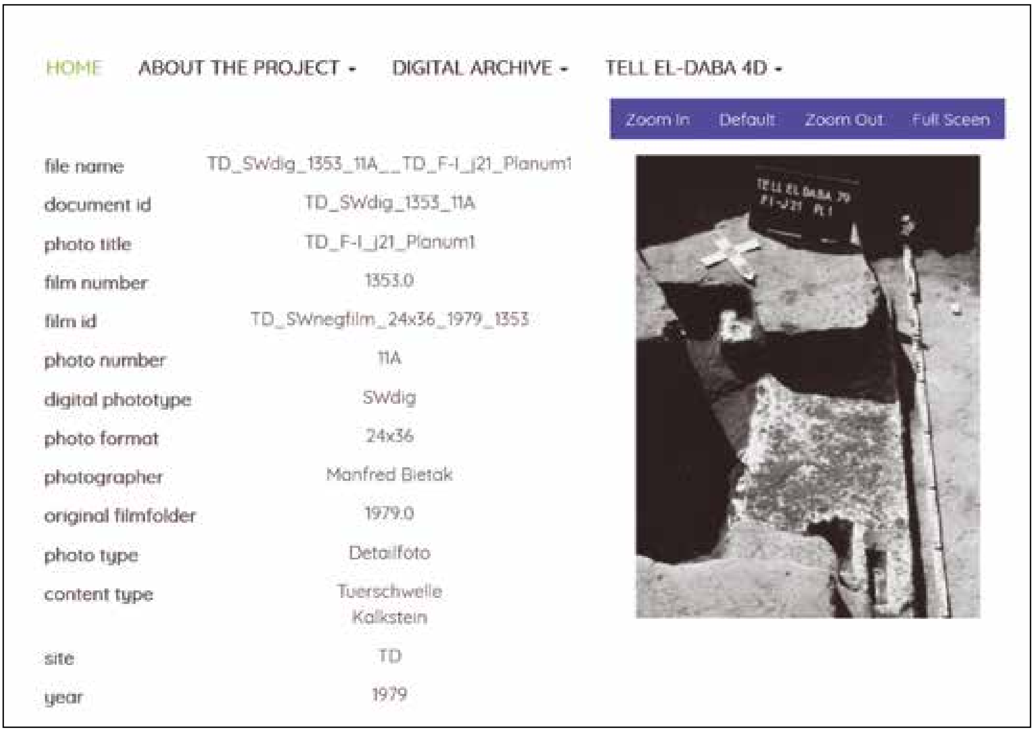


Figure 3: Metadata of the resources from the Tell el-Daba excavations includes information about the archaeological content, the analogue resource and the digitization process

System) reference model. ARCHE has a custom metadata schema, distinguishing project-, collections-, and file-level metadata.

The workflow and technology stack of the project were described in the ÖAW contribution to the ARIADNE session at the 2019 Computer Applications in Archaeology conference in Kraków (Aspöck and Hiebel 2019). The innovative part of the project is the semantic workflow that has been developed for the integration of data and metadata (Aspöck et al. forthcoming). We are using Karma (ISI 2016), a semantic web tool to map metadata and vocabularies to the CIDOC CRM data model and SKOS. Karma creates a knowledge graph to represent the information and exports it in RDF (Resource Description Framework). Then the RDF structure is ingested in a triple store, where the resources are linked through the unique identifiers. Resources are either linked on a class level (because they belong to the same CIDOC CRM class, e.g. 'document' or 'physical thing'), on the SKOS concept level (because the same thesaurus term was attributed to them, e.g. 'field drawing') or on an instance level (because they describe the same excavation area or archaeological feature/find, e.g. 'Site TD, Area F/1, SQUARE j/21, Planum 3').

To create, manage and query metadata and digital documents of the Tell el-Daba excavation documentation we identified three main components within the system architecture: 1. Data Creation & Curation, 2. Data Integration, Storage & Archiving and 3. Data Presentation & Publication. The goal was to develop a system with open and well-defined interfaces between the components, as the leading idea was that the data are the most important asset within the project and it should be possible to choose different software products for each system component – and, if necessary, replace them individually if a better one comes up for the specific purpose.

Data creation takes place at OREA by a team of 4-5 students. We chose MS Excel for the metadata entry and management of the controlled vocabularies. The advantages of Excel in comparison to databases are the flexibility offered by MS Excel in comparison to other systems which would need development of a user interface or customisation of an existing interface. MS Excel allowed us to start the metadata entry process quickly and it also allows us to enter values quickly (e.g. entering the same value to many cells at the same time, whereas a database only allows the user to enter one value at a time). However, data modelling and possibilities for data validation are limited (no referential integrity, check for allowed values, concurrent user access etc.) and this method is more prone to errors, as identifier handling and management is performed by humans and requires constant monitoring and regular quality assessment. However, as only few students carry out the data entry, they have become experts in TD documentation in the meantime and mistakes have become less frequent.

The *A Puzzle in 4D* project also employs Open Science principles, however, there are several restrictions in their implementation, such as that a large amount of the documentation in the archive has not yet been published and is not made available openly. Information about the workflows is available via the project homepage, where access to the digital resources will also be provided via a web application.[8] The code of the web application will be made available via github.

Outlook: ARIADNEplus

The new home of ARIADNEplus at the ÖAW is the Austrian Centre for Digital Humanities (ACDH). The ACDH is a research institute that was set up to foster the humanities by applying digital methods and tools. The institute cooperates closely with all three archaeological institutes at the ÖAW (Institut für Kulturgeschichte der Antike/Institute for the Study of Ancient Culture (IKAnt), Österreichisches Archäologisches Institut/Austrian Archaeological Institute (ÖAI) and OREA, which had hosted ARIADNE). Cooperations include a series of digital archaeology projects, such as the two projects described in the previous section. Most importantly however, ACDH is the home of ARCHE ('A Resource Centre for Humanities Related Research in Austria'), the new repository for the humanities at the ÖAW which has been launched in December 2017 (ARCHE). Hence, here an activity most relevant for ARIADNEplus has

8 https://4dpuzzle.orea.ÖAW.ac.at/

been started, the archiving of archaeological datasets at the ÖAW. Additionally, ACDH is the Austrian host of fellow infrastructure projects DARIAH-EU and CLARIN-ERIC.[9]

In ARIADNEplus, the ÖAW is responsible for the testing of the new ARIADNEplus project ontology. This task builds on the activities of mapping ÖAW datasets to the CIDOC CRM in ARIADNE, which have been described above. The ÖAW is also part of the ARIADNEplus dissemination team and has the task of planning dissemination at archaeological events, which will be mainly including activities at the major archaeological conferences. Furthermore, the ÖAW is leading a work group on the integration of datasets on cemeteries and graves. This will be an interesting new case study on item-level data integration, which will hopefully result in a dataset fully functioning for researching questions in mortuary archaeology. And, last but not least, the archaeological datasets which are being ingested into the institutional repository ARCHE since last year will be integrated into the ARIADNEplus data infrastructure.

For the Austrian – and European – archaeological community ARIADNEplus will continue to be representing standards in data management through project activities. For example, a round table on ARIADNEplus and SEADDA will be held at the Cultural Heritage and New Technologies (CHNT) conference in Vienna in 2019, not only introducing the two projects but also the FAIR data principles.

Conclusions

In Austria, ARIADNE has played a vital part in changing archaeological data management culture. Archaeologists are now aware of the importance of data management strategies to preserve their digital assets not only during the lifetime of the projects but also beyond, to allow reuse of their research data. Training events have provided important guidance and tools to achieve this. Most importantly, Austria has now facilities that care for the long-term preservation of (archaeological) research data. ARIADNE has also contributed to the development of new datasets according to the FAIR principles at the ÖAW and projects that embrace Open Science principles. Open Science is not yet part of mainstream archaeology and hence these projects represent important case studies for future research. The ÖAW also contributed to the development of the CIDOC CRM extensions for archaeological data and carried out mappings of their data for testing the suitability of CIDOC CRM and extension for data integration. This contribution on the development of standards has prepared the way for relating archaeological data from dispersed resources and is relevant to the archaeological community internationally.

ARIADNEplus will continue the work that was started in ARIADNE. There will be further events, promoting the infrastructure and what it has to offer, but also to support archaeologists in FAIR data management. ÖAW will again be working on the development of standards in archaeology and run an innovative test case for item-level data integration.

9 https://www.ÖAW.ac.at/acdh/projects/clarin/. https://www.ÖAW.ac.at/acdh/projects/dariah-eu/

Acknowledgements

I want to thank those who contributed to the successful delivery of ARIADNE at the ÖAW: Anja Masur, Seta Štuhec, Irene Petschko and Elisabeth Pichler. Further thanks goes to my (new) colleagues from the ÖAW-ACDH for their continuing cooperation and support and to my fellow ARIADNE partners for the community and friendship that has developed over the four years of the project. And, finally, I want to say thank you to ARIADNE and ARIADNEplus coordinators Franco Niccolucci and Julian Richards for their support and for doing a great job handling this large partnership.

References

Andorfer, P., Aspöck, E., Ďurčo, M., Masur, A., and Zaytseva, K. 2016 'The DEFC-App: A Web-based Archaeological Data Management System for 'Digitizing Early Farming Cultures" in *Digital Humanities 2016: Conference Abstracts.* Kraków: Jagiellonian University & Pedagogical University, 726–728.

Aspöck and Hiebel 2019 'A puzzle in 4D': using semantic technologies for the integration of resources from a long-term excavation project', Digital Infrastructures for Archaeology: Past, Present and Future directions, Computer Applications in Archaeology Conference Kraków, 24 April 2019. Available https://2019.caaconference.org/wp-content/uploads/sites/25/2019/04/CAA2019_programabstracts_v20190423.pdf (Last accessed: 04 July 2019)

Aspöck, E., K. Kopetzky, B. Horejs, M. Bietak, M. Kucera and W. Neubauer 2015, 'A puzzle in 4D digital preservation and reconstruction of an Egyptian palace' in *2015 Digital Heritage,* Granada, 675–678. doi:10.1109/DigitalHeritage.2015.7419596

Aspöck, E. and Masur, A. 2015 'Digitizing Early Farming Cultures customizing the Arches Heritage Inventory & Management System' in *2015 Digital Heritage,* Granada, 463–464. doi:10.1109/DigitalHeritage.2015.7419549

Aspöck, E. and A. Masur 2014. Data Inventory OREA. Unpublished Report, ARIADNE WP 3.

Aspöck, E. and Geser, G. 2014 'What is an archaeological research infrastructure and why do we need it? Aims and challenges of ARIADNE' in Börner, W. and Uhlirz, S. (eds) *Proceedings of 18th Conference on Cultural Heritage and New Technologies (CHNT)* 2013. Vienna: Stadtarchäologie, Museen der Stadt Wien, 1–10. Available http://www.chnt.at/wp-content/uploads/Aspoeck_Geser_2014.pdf (Last accessed: 03 July 2019)

Aspöck, E., Štuhec, S., Kopetzky, K. and Kucera, M. (eds) forthcoming 'Old excavation data – what can we do?' *Proccedings from workshop at 10th ICAANE, Vienna April 2015.* OREA Series, Vienna: Austrian Academy of Sciences Press.

Aspöck, E., Hiebel, G., Kopetzky, K., and Durco, M. forthcoming 'A puzzle in 4D: Archiving digital and analogue resources of the Austrian excavations at Tell el Daba, Egypt' in Aspöck, E., S. Štuhec, K. Kopetzky and M. Kucera (eds), *Old excavation data – what can we do? Proccedings from workshop at 10th ICAANE, Vienna April 2015.* OREA Series, Vienna: Austrian Academy of Sciences Press.

Bietak, M. 1996 *Avaris. The Capital of the Hyksos,* London: British Museum Press.

Bundesdenkmalamt 2018 Richtlinien für archäologische Maßnahmen 5. Fassung, 1.Jänner 2018. Available https://bda.gv.at/publikationen/standards-leitfaeden-richtlinien/richtlinien-fuer-archaeologische-massnahmen/ (Last accessed: 03 July 2019)

Corns, A. Kennedy, L. Štular, B. 2014 'Impressions from the ARIADNE community: Slovenia and Ireland, Two Perspectives on Archaeology Data' in *ARIADNE. The Way forward to digital archaeology in Europe,* 83–89.

Doerr, M., Theodoridou, M., Aspöck, E. and Masur, A. 2015 'Mapping archaeological databases to CIDOC-CRM' in Campana, S., Scopigno, R. Carpentiero, G. and Cirillo, M. (eds) *CAA 2015, Keep The Revolution Going, Proceedings of the 43rd Conference on Computer Applications and Quantitative Methods in Archaeology.* Oxford: Archaeopress, 443–452.

Masur, A., K. May, G. Hiebel and E. Aspöck 2014 'Comparing and mapping archaeological excavation data from different recording systems for integration using ontologies' in Börner, W. and Uhlirz, S. (eds) *Proceedings of 18th Conference on Cultural Heritage and New Technologies (CHNT)* 2013. Vienna: Stadtarchäologie, Museen der Stadt Wien, 1–11. Available http://www.chnt.at/wp-content/uploads/Masur_etal_2014.pdf (Last accessed: 03 July 2019)

Hagman, D. 2018 'Überlegungen zur Nutzung von PHAIDRA als Repositorium für digitale archäologische Daten' *Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare Mitteilungen* 71, 1, 53–69.

ISI 2016. Karma: A Data Integration Tool. http://www.isi.edu/integration/karma/ (Last accessed: 09 August 2019).

Lochner, M. und Hellerschmid, I. 2010 *Ein Gräberfeld der jüngeren Urnenfelderkultur aus Franzhausen-Kokoron (UFK).* Katalog und Abbildungen. Available http://hw.ÖAW.ac.at/franzhausen-kokoron2/ (Last accessed: 03 July 2019)

Museum of London Archaeology Service (MOLAS) (1994) Archaeological Site Manual, Hampshire, RAS Printers Limited, 7. Available https://achill-fieldschool.com/wp-content/uploads/2016/01/molasmanual942.pdf (Last accessed: 03 July 2019)

Niccolucci, F. and Richards, J.D. 2013 'ARIADNE: Advanced Research Infrastructure For Archaeological Dataset Networking in Europe', *International Journal of Humanities and Arts Computing* 7.1-2, 70–88. Edinburgh University Press, doi:10.3366/ijhac.2013.0082

Schachermeyr, F. 1991 *Sammlung Fritz Schachermeyr: Die neolithische Keramik Thessaliens. Aus dem Nachlaß bearbeitet von Eva Alram-Stern,* Veröffentlichungen der Mykenischen Kommission 13, Vienna: Academy of Sciences.

Štuhec, S., Aspöck, E., Masur, A., Andorfer, P., and Zaytseva, K. 2016 'Putting 3D models into context. Schachermeyr pottery collection and the DEFC app' in *Proceedings of the 8th International Congress on Archaeology, Computer Graphics, Cultural Heritage and Innovation. Advanced 3D Documentation, Modelling and Reconstruction of Cultural Heritage, Monuments and Sites,* Valencia, Spain.

Wright, H. and Richards, J.D. 2018 'Reflections on Collaborative Archaeology and Large-Scale Online Research Infrastructures' *Journal of Field Archaeology* 43, S60-S67, doi:10.1080/00934690.2018.1511960

The Hungarian archaeology database in the light of ARIADNE

Attila Kreiter

Laboratory for Applied Research, Hungarian National Museum, Budapest
attila.kreiter@gmail.com

Abstract

In Hungary there has been an increasing need to provide archaeologists with an accessible online catalogue of archaeological sites including site metadata and documentation. In this respect the ARIADNE project had a great impact on how Hungarian archaeologists consider access to archaeological data and documentation. Open access to archaeological data is still in its infancy in Hungary, but ARIADNE and ARIADNEplus respectively made it possible for the Hungarian National Museum to develop an archaeology database (https://archeodatabase.hnm.hu/en) and continuously mobilize more and more archaeological data. Without the ARIADNE and ARIADNEplus programmes this could not have been possible.

Keywords: archaeology database of Hungary; archiving; data catalogue

Access to archaeological data and so-called grey literature (unpublished site reports, maps, inventories, drawings, photographs etc.) is particularly important in Hungary since in the last three decades development-led excavations provided an immense amount of information on tens of thousands of sites. These excavations clearly shed a completely new light on previous results, and it is our professional responsibility to make as much data available for the scientific community as possible. Through the archaeology database of the National Museum some of these data can now be accessed by the research community.

During the development of the archaeology database we had different perspectives in mind. Firstly, to systematically collect as much metadata on sites as possible to provide information to the scientific community and archaeology students. For the first time in Hungary, the data model, descriptive concepts, terms and temporal coverage of archaeological sites became standardised (ACDM native XMLs/CIDOC-CRM, Getty AAT, WGS84, PeriodO). Secondly, to collect site documentation and make these documents accessible online through different access levels. This process includes all types of documentation, from short reports (called 30-day reports in Hungary) through databases of finds to the results of interdisciplinary analyses (anthropology, archaeozoology, ceramic/stone petrography, macro/micro archaeobotany, dendrochronology, leather/textile analysis, ^{14}C, XRD, XRF, LA-ICP, FTIR etc).

In Hungary the submission of primary documentation (30-day report, 1-year report) is regulated. By law these documents have to be submitted to the District Heritage Office, Heritage Registry Office of the Prime Minister's Office, Hungarian National Museum, county museum and to a museum where the finds are stored. In some cases an excavation is not carried out by the local museum, but if the finds will be stored there eventually the local museum also receives a copy of the abovementioned documentation. (27/4§ of the 68/2018 (IV.9) Government Decree) However, further documentation (inventories, databases, scientific assessment, interdisciplinary analysis, photos, drawings etc.), which are mainly held in the museum that carried out the excavation or in the museum where the material is finally stored, may not end up in the District Heritage Office, Heritage Registry Office of the Prime Minister's Office, Hungarian National Museum and county museum. Thus, these documents are stored in the local museums that carried out the excavations and/or where the finds are kept. In other words, there is no officially appointed centralized digital repository in Hungary where all the documentation is stored and made accessible. Thus, there are millions of files stored in museums all over the country without being used or re-used, and without the archaeological community, or even central institutions, being aware of them. In this system only some of the documents (30-day report, 1-year report) are available and accessible in central institutions to which primary documentation has to be submitted. In this way the site documents live completely separate lives and local museums are the only places where all documents are kept together. Nevertheless, digital archiving of archaeological documentation is not regulated and therefore digital data are often archived as objects (CDs, DVDs receive inventory numbers in an archive) rather than computerised information. This is a major drawback not only in terms of the FAIR principles but in terms of archiving and long-term preservation of data as well. The archaeology database of the National Museum provides a solution in this respect and it can serve as a national repository where documentation can be stored and accessed online in one place (through various access levels).

Another important aspect of providing access to data and documentation for professionals is to support research and the usability/re-usability of documentation. For example, all archaeologists have probably found that when they wanted to work with an assemblage they had to travel to different museums to look at the material and the documentation of a site and to decide if the quality is suitable for answering their archaeological questions. This is a time-consuming and costly endeavour and there are cases when the material is not as suitable for research as expected. The archaeology database also provides help in this respect. Researchers can access information on the spatial distribution of sites within certain periods (or whatever combination of search terms they chose), they can see where the material is stored, who they should contact to access those materials, and if documentation/inventories etc. are available online; they can even assess the quality and quantity of the assemblage. In other words, the suitability of the assemblage for the planned research can be checked beforehand to save time and resources. In the light of planning a national archaeological strategy the ministry (Ministry of Human Capacities), heritage managers or other decision makers can also have information on the type of sites and their spatial distribution within their field of interest, the type of available documentation, where the assemblage is kept, who is scientifically responsible for the documentation and so on. In this way professionals can make a well-informed decision about the sites and assemblages to work with.

Another aspect we had in mind during the development of the database is that Hungarian archaeologists work on several, even dozens of excavations from all sorts of periods during their careers. It is easy to see that it is often impossible to complete the analysis of so many sites and assemblages. There are thousands, or rather tens of thousands, of sites in Hungary that are still awaiting detailed scientific assessment. Considering how difficult it is for archaeology students to access materials for MSc, MA and PhD theses, these sites could be perfect starting points for them. Moreover, despite the amount of money invested in these excavations and post-excavation works, their results are hardly known, or not known at all, even by the archaeological community. Why not make them accessible to others?

The archaeology database opens completely new avenues for assessing the extent and nature of our national heritage, but comprehensive and systematic records are imperative in this respect. The scientific potential in the data of registered archaeological sites in Hungary (almost 60,000) is immense and the database would be perfectly suitable to help plan local or even regional scientific projects. The sites associated with research projects can even be showcased in the database and it can host all documentation and metadata that otherwise could not be published. The emphasis here is that an unpublishable amount of data can be stored, preserved, and linked to a scientific project. This is a professional way to show the results of research projects, not just to the archaeological community but to national or international funding bodies (*Figure 1*).

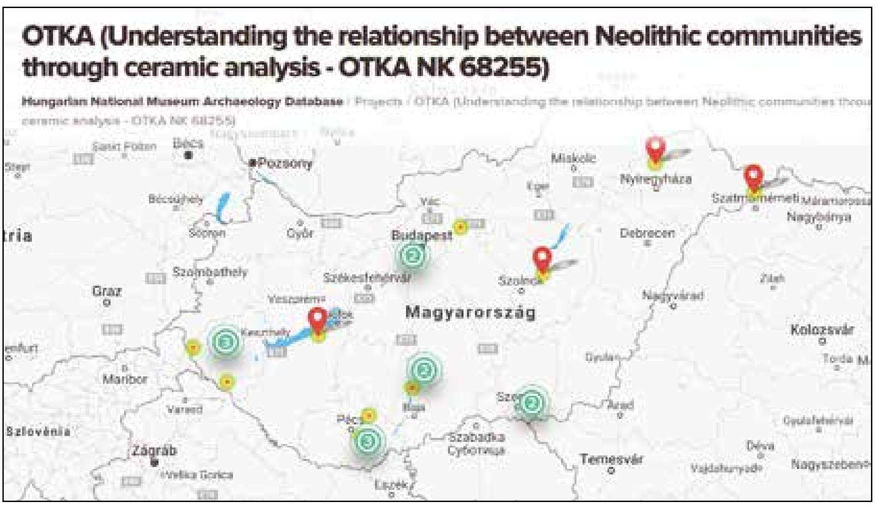


Figure 1: Showcased sites associated with a research project. By clicking on the sites, all data and scientific documentation can be accessed

Since there were no existing publicly accessible online archaeology databases in Hungary, we planned to develop a portal that could be used by professionals, students of all ages, and the general public as well. This posed a challenge because at the same time archaeological sites also had to be protected and access to sensitive information had to be restricted. Therefore, different access levels were introduced.

Non-professionals (public) can access short reports of sites published in the volumes of *Archaeological Investigations in Hungary*. Not just the volumes of this series are accessible through the database, but all the short reports within the books are linked to archaeological sites and the contents of *Archaeological Investigations in Hungary* are also included in the search function of the database. In this way the public can also access valuable information on archaeological sites and could easily find out about the prehistory of their area, and the data can be used as a teaching aid in schools.

Figure 2: Archaeological Investigations in Hungary are available for public access

Professionals may have full access to the database (see 'how it works' menu in the database) including access to all uploaded documentation. The purpose of the archaeology database of the National Museum is to support research, teaching and learning and also to acknowledge and protect the intellectual property and copyright of the authors of the documentation. Therefore, uploading site documentation to the database is voluntary. Short reports (called 30-day reports) are exceptions; these are uploaded without permission but access to these is still restricted to professionals. For all other documentation types permission is required from all contributors in order for their documentation to be uploaded to the database. Having known the difficulties in acquiring permissions for archaeological documentation in Hungary it was a great surprise to us that the overwhelming majority of contacted people gladly provided permission for uploading their documentation to the database, and more and more colleagues are willing to join. This was a very important sign of a paradigm change in the Hungarian archaeology community. The majority of colleagues responded very positively and were also clearly aware of the fact that they can only work with the minority of sites and materials that they had excavated so why not give access to them and make them available to the research community.

Developing a national archaeology database and making site data and documentation accessible to professionals (under restricted access) is a sound way of assessing our national heritage and becoming familiar with what our heritage comprises and with what we need to protect and preserve. The increasing content of the database is a shared success; it could not have been possible without a joint venture of all archaeologists who so far have showed their support and valuable contribution.

In order to involve archaeologists and museums, and to facilitate an increase in the quality and quantity of metadata of sites, there is a function in the database through which professionals can comment on or add metadata to any site in the database. In this way the quality of data continuously improves, and since the launch of the database (May 2016) thousands of items of metadata have been corrected or added to sites. Moreover, there is a feature through which archaeologists can upload sites to the database and send documentation to be uploaded.

A recognized benefit of the database is that there is virtually no size limit for documents to be uploaded and an unpublishable amount of data can be stored in one place, and accessibility (through access levels), archiving and long-term preservation is also assured. The database was developed with a long-term plan in mind that it can serve as a national repository for archaeological data and documentation. The archaeology database indeed provides a complex solution for keeping data and documentation in one system and can serve as a repository for other Hungarian museums as well.

The content of the database was not accessible previously, but is now a primary research tool for archaeologists in Hungary. In particular, since access is restricted and in this way sites are still protected, this encouraged archaeologists to make their documentation accessible online. The success of ARIADNE in Hungary can clearly be viewed through numbers. ARIADNE has been promoted in Hungary for years among professionals, which finally made it possible to mobilize data on all registered Hungarian sites. We originally planned to include the data of 600 sites because in 2012 these were available for us, but as promotion has continued this number has increased to 1500 sites. The excavations of these sites were coordinated by the former Field Service for Cultural Heritage and National Heritage Protection Centre. By the time the first phase of the ARIADNE programme had finished we had almost 60,000 sites in the catalogue, and it now hosts all registered Hungarian archaeological sites, their metadata and, where available, documentation – a total of almost 1,200,000 files. A major leap in the increase of database content (to almost 60,000 sites) could not have been possible without the support of the former Forster Centre and its governing body the Prime Minister's Office.

The main benefits of the Hungarian archaeology database can be summarized as follows:

- Archaeological sites are protected since sensitive data can only be accessed by professionals.
- Where available, all documents are presented together for a site (from 30-day report to interdisciplinary results). Each site has a unique URL for referencing.
- Unpublishable amounts of documentation can be uploaded for each site, there is virtually no size limit. Archiving and long-term preservation is also assured.
- The database is updated regularly with new and controlled data. Professionals can also modify site data (controlled process); in this way the quality of site data continuously increases.
- Detailed search functions allow assessment of available data and/or the planning of regional/national scientific projects.
- Sites involved in a research project can be linked together and the results can be visualized/demonstrated in the database and all the results and documentation can be accessed in a controlled manner. This is a great way to show funding bodies the results of a research project.
- Online available information on the quantity and quality of the assemblages speeds up research, saves time and resources, and also speeds up the process of working with legacy sites, mobilising their data and increasing scientific exploitation.
- Hungarian professionals and museums can join the database and upload sites and documentation using a controlled process.

In summary, this is our success story with ARIADNE and ARIADNEplus. ARIADNE not only made it possible to develop a Hungarian archaeology database, but through the ARIADNE Portal our data are linked to those of other European data providers. Nevertheless, there is still an immense amount of work to be carried out – but the benefits of online access are being recognized in Hungary, just as ingrained attitudes towards the documentation of sites are changing. These achievements could not have been possible without the ARIADNE and ARIADNEplus programmes. As part of the ARIADNEplus programme the National Museum, among others, continues to increase the content of the archaeology database, in particular including data and documentation on interdisciplinary analyses. We encourage Hungarian archaeologists and museums to join this unprecedented endeavour to assess our national heritage and make it accessible to the research community in the archaeology database of the National Museum. I have no doubt that the archaeology database of the Hungarian National Museum and access to an increasing amount of data have given a great push to our profession, and will make it possible to re-evaluate and re-position our national heritage and the results of Hungarian archaeology not only in Hungary but also in Europe.

The Zbiva Web Application: a tool for Early Medieval archaeology of the Eastern Alps

Benjamin Štular

ZRC SAZU: Research Centre of the Slovenian Academy of Sciences and Arts
bstular@zrc-sazu.si

Abstract

'Zbiva' is a research database for the archaeology of the eastern Alps and its surrounding regions in the Early Middle Ages. Its inception in the early 1980s was deeply rooted in the scientific research context of the time. In 2016 the database front-end was migrated to the 'Zbiva web application' based on an open source Arches 3.0 platform. Zbiva is a GIS-enabled web application focused on catering to the needs of highly invested researchers. Some of the most important design ideas for the application were informed either by the ARIADNE User Needs report or by discussion with many collaborators within the ARIADNE consortium. The design maxim for Zbiva was to focus on highly motivated and invested users. This was only possible because Zbiva's data top level data search is 'outsourced' to the ARIADNE portal.

Keywords: Early Medieval archaeology; Eastern Alps; web application; database

Introduction

Zbiva (Pleterski and Belak 1995; Kastelic et al. 2016) is a research database for the archaeology of the eastern Alps and its surrounding regions in the Early Middle Ages. Its current front end is the Zbiva web application,[1] a GIS-enabled web application focused on catering for the needs of highly invested researchers. The aim of this chapter is to present the development of the Zbiva web application with an emphasis on the role the ARIADNE project played in the application's design.

Scientific background

Zbiva was designed as a tool to study the so-called Carantanian-Köttlach archaeological culture. This means that – for historic reasons – its chronological focus was on the time from the settlement of the Slavs in the area in the 6th century AD until the end of habitual deposition of grave goods in the 11th century AD. It primarily holds data from the archaeologically relevant region that includes nowadays Slovenia, Austria, NW Croatia, and NE Italy. For comparative purposes it also includes selected relevant sites from neighbouring regions and from the preceding period.

[1] http://zbiva.zrc-sazu.si

The history of Zbiva dates back to 1980 and its inception is deeply rooted in the scientific context of Early Medieval archaeology in the Southeastern Alpine area in particular, and in the Central European tradition of archaeology in general. In order to understand the circumstances a brief history of research is needed (*cf.* Štular and Pleterski 2018).

Research into the Early Medieval archaeology of the Southeastern Alpine area began with the publication of 'unusual enamelled jewellery' found in 1853 by workers digging for gravel at Köttlach in Lower Austria (Franck 1854). When the author of this publication asked himself a series of questions, including who were the people to whom these finds belonged, and when did they live, he set the research agenda for more than a century. The number of similar finds rapidly increased thus revealing the full extent of the phenomena.

The main research agenda until 1980s was to define the archaeological culture that the material belonged to, as well as its chronological and ethnic definition. Dating these artefacts to the early medieval period was soon clear. However, in accordance with the cultural-historical understanding of archaeological artefacts (*cf.* Jones 2003), a controversy arose regarding the ethnicity of the people to whom this enamelled jewellery belonged. Some scholars assumed that it belonged to the Slavs, others disputed that it was exclusively Slavic, and others saw the same artefacts as evidence for the early medieval presence of Germans in the Eastern Alps. (For the period until the First World War see Pleterski 2001; for distinctively different understanding during and immediately after the Second World War see Dinklage 1941a; 1941b; 1941c; 1943 and Korošec 1947). Over time terms Carantanian – after the connotations with the early medieval Duchy of Carantania (Schmid 1913) – and Köttlach – after the site of first discovery – were coined into *terminus technicus* Carantanian-Köttlach (archaeological) culture. In historiography this population was termed Alpine Slavs (Grafenauer 1954; *cf.* Kahl 2002). The Early Medieval archaeology of the region, its sites (242 at that time) and the artefacts were presented by P. Korošec (1979), who indicated chronological and cultural differences within the Carantanian-Köttlach culture. Almost simultaneously Jochen Giesler's chronological essay on the same material proposed a very different chronological interpretation (Giesler 1980).

Subsequent discussion was only driving the protagonists further apart, which suggested that the entire scientific discourse needed to be built anew on fresh foundations. And it was to this end that the concept of the archaeological database Zbiva was envisaged at ZRC SAZU[2] in 1980 under the lead of A. Pleterski and M. Belak. The database took its first digital shape in 1987 (Zbiva v1) as a closed system based on a single PC. An early demonstration of the database's potential was an analysis of Early Medieval church organisation (Pleterski and Belak 1995). The web-based version (Zbiva v2) was deployed in 2000. Due to the technical limitations only sites and the

[2] Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti (Research Centre of Slovenian Academy of Sciences and Arts). The work on both Zbiva and ARIADNE took place within the Inštitut za arheologijo (Institute of Archaeology).

bibliography were accessible. In 2016 a full-blown GIS-enabled web application featuring the entire data set (Zbiva v3) was released.

A key strength of this data set is that since its launch Zbiva is regularly (monthly) updated by scouring the relevant literature. Access to the latter is based on the systematically built and sustained literature exchange network between ZRC SAZU and all of the major institutions contributing to the topic. At the time of writing Zbiva holds data on 2944 sites from the core region and a further 435 comparative sites.

This is a significant departure from the 242 sites discussed by Korošec and Giesler in 1979 and 1980. The difference is in part due to the increase in known sites published after 1979 and in part the result of the long-term systematic approach to data collection. To be precise: 50.6% of sites in the current Zbiva database have first been published in or prior to 1979. This means that a very diligent and capable researcher using a classical approach in 1979 was able to gather data on 14.1% of sites known at the time. By extension, it can be conjectured that a diligent researcher using the same approach nowadays would be working with less than 500 sites, whereas Zbiva at the time of writing enables an analysis of 3379 sites.

Zbiva v3

The trilingual (Slovenian, English, German) Zbiva database is an aggregation of:

- a sites database
- a graves database, and
- an artefacts database.

The *Sites database* (*Figure 1*) includes spatial information and site type, chronology, and bibliography. This being a research database the bibliography is the most important and also most diligently curated data set. At a first glance a bibliographic collection may seem obsolete due to the abundance of online resources devoted to scientific literature. However, this is far from true since (i) most of the relevant books and periodicals are still only published in print or are behind a paywall (ii) a significant portion of the bibliography stems from old publications that are not likely to be digitised in the foreseeable future and, most importantly, (iii) the whole bibliography has been enhanced with tagging using a controlled vocabulary (e.g. flat graves, hoard, church; monograph publication, article, report; head-circlet, necklace, ring).

The *Graves database* (*Figure 2*) comprises data on individual graves from selected cemeteries. Each grave is described by selected criteria (e.g. grave features level 1 and 2, body features level 1 and 2, dimensions, orientation, stratigraphic relations), free text and images. Currently only six cemeteries out of 1354 are included. However, by far the largest relevant cemetery in the region (Župna cerkev in Kranj, Slovenia) is included in its entirety and the second largest (Ptujski grad, Ptuj, Slovenia) is to be added in 2019. The graves database was not envisaged to provide comprehensive coverage of all published cemeteries but rather as a tool to be used for cemeteries under investigation. As such it is open for contributions and can be used as a research tool by all researchers with suitable dataset.

The *Artefact database* is similar to the graves. Free text description and images are supplemented by typological determinations (only for pottery, jewellery and knives). This portion of the database currently includes data on over 10,000 individual artefacts from sixteen sites although it was also not designed to provide comprehensive coverage.

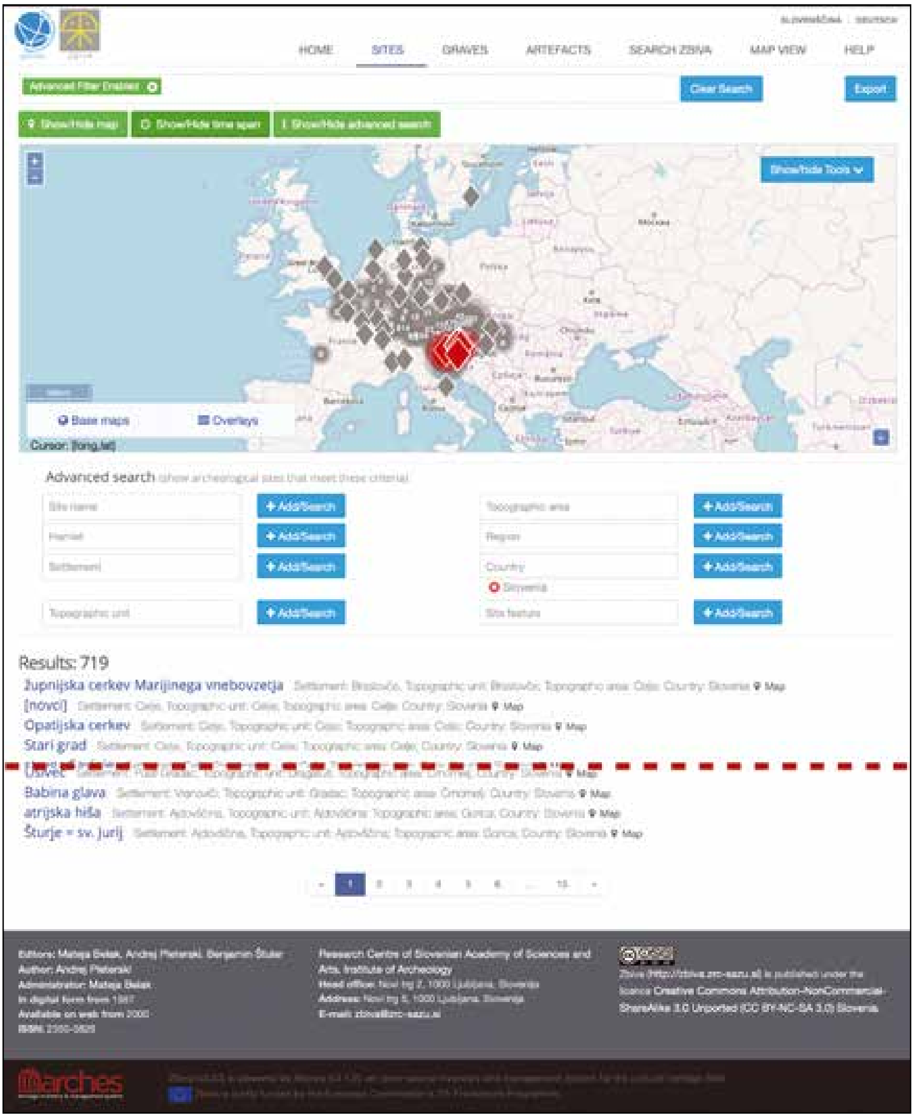


Figure 1: The Zbiva web application (English interface), Sites; results for advanced search term 'Country – Slovenia' are shown in red

The Zbiva web application is based on the Arches 3.0[3] open source inventory and data management platform. The Arches project originated in 2004 when the Getty Conservation Institute and World Monuments Fund formed the Iraq Cultural Heritage Conservation Initiative. Due to the security uncertainty the project was moved to Jordan and in June 2010 development of MEGA-Jordan was completed. This is a web-based, bilingual geospatial information system built with open-source tools designed to serve as an archaeological site inventory and management system. In April 2011 a prototype of MEGA-Iraq was made available but never deployed. During the development of the MEGA system many heritage organizations around the world stated their interested in using the system. This led the Getty Conservation Institute and World Monuments Fund to develop a user-friendly, low-cost, web-based geospatial information system designed to help inventory and manage all types of immovable heritage, including archaeological sites, buildings, structures, landscapes, and heritage assemblages or districts. In June 2011 the development of Arches as an open-source project began and Arches 1.0 was deployed in 2013. Zbiva is based on the third version (3.0) and the current version since February 2019 is 4.1.4 (Kastelic 2015; Kastelic et al. 2016; cf. Lee Enriquez, Myers and Dalgity 2018).

Based on experience with the MEGA project and extensive research on best practices and standards the following guiding principles have been set for Arches (Kastelic 2015):

1. *Standards*. The system must be based on internationally adopted standards for information technology, heritage inventory, and heritage data management (e.g. the CIDOC CRM). The incorporation of such standards is necessary for the creation of a generic system for heritage inventory and management anywhere in the world and promotes sharing and longevity of data regardless of inevitable technological advances.
2. *Accessibility*. To allow for maximum accessibility the system must be web-based and as end-user friendly as possible.
3. *Efficiency*. As an open-source system it has to be provided free-of-charge and at the same time it must provide support for long-term sustainability.
4. *Upgradability*. The system must be modular so that it can be easily tailored and upgraded. One of the key features in this regard is multilingual support.
5. *Security*. The system must allow for different levels of access, e.g. open, closed or any combination in between.

The existing features of Arches obviously set the frame for the Zbiva web application but upgrades enabled us to tweak the application according to our needs. The design maxim envisaged for Zbiva was that the application is to be used by highly motivated and invested users. Since Zbiva's data on sites is included in the ARIADNE portal the accessibility for top level – and hence less invested – research was 'outsourced' and therefore not our first priority. However, once assured that Zbiva offers data relevant for her/his research the researcher becomes committed to drill deeper into the data. This means that Zbiva's design priorities need not be simplicity,

[3] http://archesproject.org

speed and visual allure. Rather, we were able to focus on a fully commited user solving specific research questions. In other words, we were designing the application based on our own experience and use case scenarios.

A practical example of this design approach is the depiction of search results (*Figure 1*). By default the Arches platform is set to return five hits per page and as a consequence the operation only takes a fraction of a second. In addition, the results are presented in an expanded font that is pleasing to look at. However, in our use cases a typical search would easily return a hundred or even hundreds of relevant results, e.g. female graves with a finger ring. The aim of the researcher is not necessarily to narrow the search scope but to inspect all of the results. In order to better facilitate such scenarios Zbiva now returns fifty results per page in a more condensed font. Thus some of the visual appeal and speed has been sacrificed for usability in a realistic use case.

The following are mayor upgrades developed for Zbiva (Kastelic et al. 2016):

- additional multilingual support for data import
- support for new document types (sites, graves, artefacts)
- advanced search capabilities tailored for structured search (separate search for each document type, multi-level search, search according to the dimension using limiter from … to)
- automated import from the central MS Access database into the Arches environment
- several map extensions (export, import, enlarge window, search within polygon/radius/distance)

The most important feature of Zbiva from the user's perspective is its twin search engine combining elasticsearch[4] and SQL since Zbiva is a relational database. The strength of a relational database structure, from the perspective of the intended Zbiva user, is the ability to efficiently drill into the data by multi-level search queries enabled by an upgrade, e.g.

- body feature – level 1: grave goods
- body feature – level 2: ring

Another huge advantage of this data structure is that after the initial investment into the translation of controlled vocabulary the database is able to operate in three languages.

However, data input in this relational database with controlled vocabularies is time-consuming since the data must be interpreted prior to the input. This is, at least in the case of Zbiva, an assiduous scientific process rather than a mechanical data input. In addition, the search in relational database is most efficient if the user is familiarized with the database structure.

To mitigate these downsides most of the entries in the Zbiva database also include descriptive fields, the upkeep of which is more time-efficient. This free text can be

4 https://www.elastic.co

searched using the elasticsearch engine. The results are less predictable and will often return a larger number of results with a lower degree of relevance. This search is also limited by the language of the free text that is currently mostly Slovenian. Such usage can be described as unstructured search, e.g. 'prstan' (*en.* finger ring).

Perhaps the biggest improvement of the Zbiva web application over the preceding version of is that it is GIS-enabled, i.e. for the first time it is equipped with map tools. Any search can be spatially tailored. For example, any search can be limited within a user-defined polygon or within a set distance from a designated point or line. In combination with a time span slider Zbiva has become a fiercely effective tool. Some of the most common types of search in archaeology are a combination of spatial and chronological attributes, e.g. finding settlements within a 50km radius dated to the 9th century. Without the access to the data held in Zbiva such search takes weeks. But more importantly, even using the Zbiva database (but not the Zbiva web application) in combination with desktop GIS tools such a search takes considerably longer than in the Zbiva web application where it can be done in under a minute.

Use cases

It is not the intention of this chapter to delve into the particularities of Early Medieval archaeology. However, two use cases will be briefly presented to demonstrate the capabilities of the Zbiva web application.

The first example is an analysis of the graves of the Župna cerkev in Kranj cemetery. This is a large cemetery with 2945 medieval and post-medieval graves excavated. Close to 1000 of those are Early Medieval graves (overview in Štular and Štuhec 2015), which is as many as the next three largest cemeteries in the region combined (*cf.* Štular and Pleterski 2018 with bibliography). The analysis of the site has been hindered so far by two factors. Firstly, it was excavated in numerous expeditions between 1953 and 2013 which makes the data sets very heterogeneous (*cf.* Štular and Belak 2012a; 2012b; 2013; Belak 2013; 2014; Sagadin and Belak 2014). Secondly, the sheer quantity of data and density of burials (at one point reaching 19 burials per square metre) demands the use advanced spatial analysis (*cf.* Achino et al. 2019). Obviously, an in-depth analysis of such a site can only take place within a long-term dedicated research project. Such a project is underway (Pleterski, Štular and Belak 2016; 2017; Pleterski, Štular, Belak and Bešter 2019) and as a part of this project data on individual graves has been added to Zbiva.

Earlier researchers, whose engagement preceded the advent of digital tools, struggled to answer a seemingly simple research question: was the spatial distribution of Medieval burials static or dynamic?

To answer this question the general search and mapping capabilities of the Zbiva web application have been employed. First the distribution of a typical Early Medieval head-jewellery (head circlets) was mapped (*Figure 2a*). To demonstrate the extent of the High Medieval and Early post-Medieval cemeteries several typical artefacts made of brass have been mapped (*Figure 2b*). The comparison of the two distribution maps clearly demonstrates the much larger extent of the Early Medieval cemetery. In further refinement a selection of typical Early Medieval artefacts have been mapped. The first is a type of head circlet typical for the first half of the 9th century AD (*Figure 2c*) and the second group are two types typical for the late 10th and 11th centuries (*Figure. 2d;* cf. Pleterski 2013). While both the earliest and the latest types of Early Medieval female jewellery are distributed evenly throughout the cemetery there is a discernible difference between the distribution of the two latest types. This clearly demonstrates the dynamics in the process of selection of burial plots within the cemetery other than purely chronological factors.

Such quick insights into the cemetery's dynamics are by no means sufficient for any sort of final interpretation and can only be seen as a starting point for in-depth research. This is exactly how the project team is using the Zbiva web application: as a tool for hypothesis testing. However, the depth of analysis enabled by the Zbiva web application exceeded our expectations. It is especially noteworthy that the use case above has been performed on non-structured data, i.e. on free text descriptions of individual burials taken from digitised archives (Štular, Belak 2012a; 2012b; 2013; Belak 2013; 2014; Belak, Sagadin 2014).

The second use case demonstrates the use of the Zbiva web application as a planning tool. As mentioned, the backbone of Zbiva is data on more than 3000 sites. One of the most exciting expectations from the Zbiva web tool, when it was designed, was the promise of an insight into the chronology of settlement dynamics. Once the application was operational it was a simple task of using the combination of time span and map to compare spatial distributions. An example of this is a comparison between the spatial distribution of cemeteries in the 8th and 10th centuries. In this period it is known that burial shifted from cemeteries without a church to cemeteries with a church (Pleterski and Belak 1995). Alas, the result was inconclusive since no discernible pattern is noticeable (*Figure 3*). Further examination confirmed that the underlying chronological data is simply too coarse to answer this question.

This was, however, not a failure of the application. Rather, the Zbiva web application enabled us to pinpoint the weakness in our dataset. This revelation was turned into a convincing research question and ultimately in the successful bid for a € 600,000 research project. In other words, the Zbiva web application proved itself to be an invaluable tool for assessing the data quality of the Zbiva database.

Conclusion: The impact of ARIADNE

Just as the scientific background was important for the inception of the Zbiva database in the 1980s so was ARIADNE important for the creation of the Zbiva web application: not necessarily the only reason but surely a key influence. When ZRC SAZU joined the ARIADNE project in 2011 it had neither a vast experience in big European projects nor an archaeological IT department to speak of. It was therefore our primary intention to keep our heads down and eyes open. However, within the

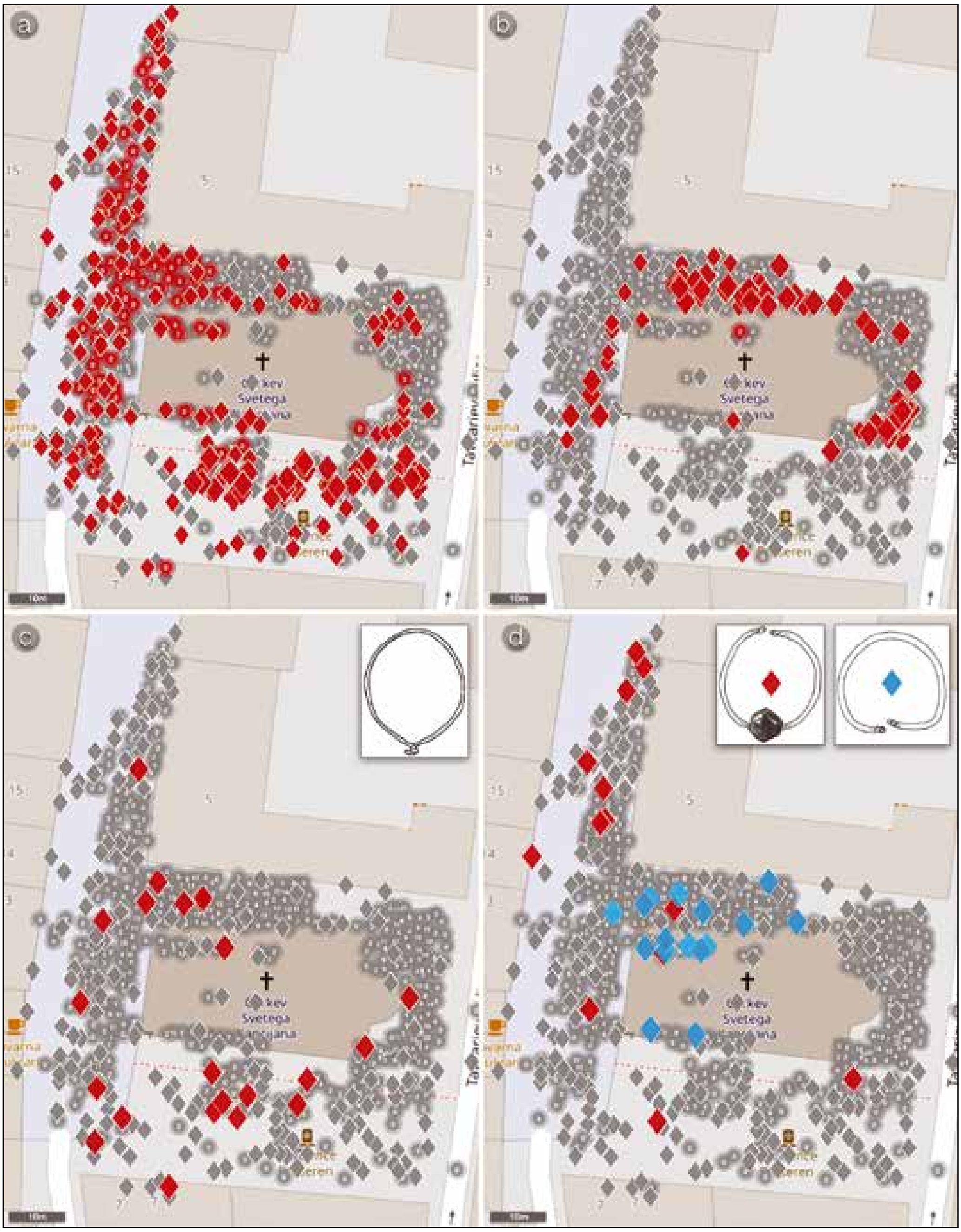


Figure 2: Zbiva web application search results on a Župna cerkev in Kranj cemetery: a – head circlets typical for Early Medieval jewellery; b – brass jewellery typical for late High Medieval and early post-Medieval period; c – specific type of head circlet with two hooks (see inset images) typical for 8th/9th century AD; d – two specific types of head circlet with double thickenings (see inset images) typical for 10th/11th century AD

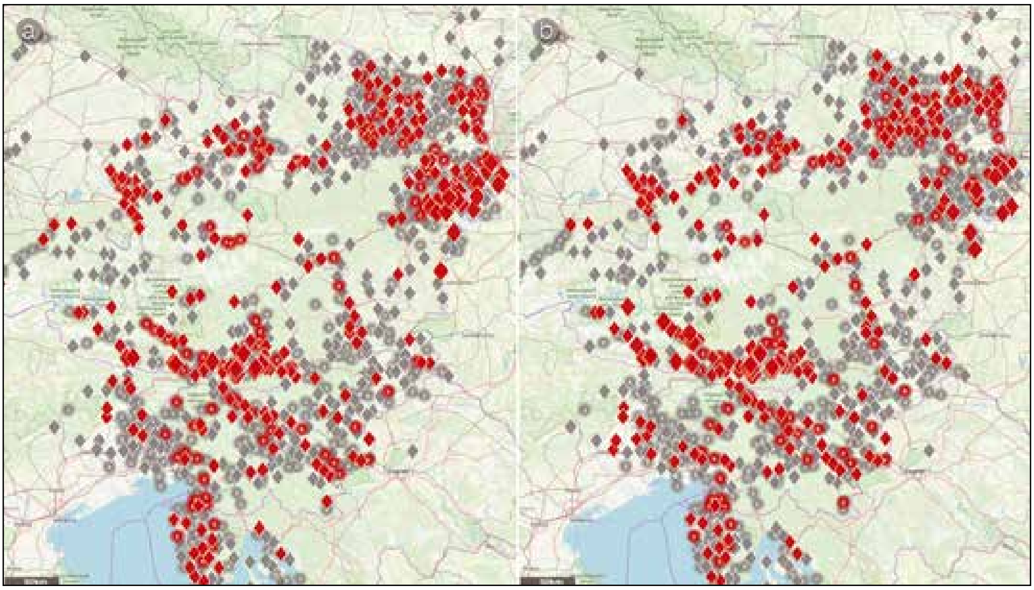


Figure 3: Zbiva web application search results on sites: a – 8th century AD cemeteries; b – 10th century AD cemeteries

first day of the project's kick-off meeting it became clear that we are not alone in this preconception. The consortium has been built from two types of partners that, for the purposes of clarity, the project's principal investigator Franco Niccolucci divided into the 'technical partners' and the 'content providers'. The former are some of the biggest European institutions in the field of digital archaeology and the latter are a selection of the most prominent archaeological institutions in Europe. It was as if a huge burden has been taken off our shoulders when we realized that we were not alone in the situation that could be best described as technologically challenged. As a matter of fact, ZRC SAZU at the time was probably a good example of an average institution as far as digital infrastructure in the humanities was concerned. And this, indeed, was the vision behind ARIADNE in the first place: to bring together the best and brightest in the field of digital infrastructure in archaeology and the best and brightest in the field of archaeology.

For the remainder of ARIADNE, work on the project was a pleasant experience, but more importantly, a hugely rewarding one. The main stated task of ZRC SAZU in the ARIADNE project was to make all the necessary preparations in order to include ARKAS and Zbiva in the ARIADNE portal[5] (*cf.* Štular, Niccolucci and Richards 2016, 160). During each stage we were forced to learn and adapt which greatly improved our understanding of the role of digital infrastructure in 'everyday' archaeology. With this we began to better understand not only our weaknesses but also our strengths.

5 http://portal.ariadne-infrastructure.eu/

The biggest strength of ZRC SAZU in this regard is systematic gathering of selected scientific data for the past four decades. As with most archaeological research institutions in a Central European milieu ZRC SAZU's Institute of Archaeology is internally organised according to archaeological periods and the focus of our research is to publish research in scientific journals and monographs of the highest quality. However, since the outset in 1947 there was always an additional task of systematic collection of data (Pleterski 1997). The work of the first generation of post Second World War archaeologists in Slovenia culminated in an atlas of archaeological sites comprising almost 5000 entries (ANSL 1975). The work of the next generation, apart from continuously upgrading the data set, was to migrate this data first into a digital and then to the web-accessible format. This was implemented at ZRC SAZU as ARKAS (Archaeological Cadastre of Slovenia).[6] The other archaeological database of note at ZRC SAZU is Zbiva. Both these data collections are perhaps best described as 'slow data', as opposed to 'big data', emphasizing 'thoughtful digital curation' instead of 'quantity trumps quality' (Kansa and Whitcher Kansa 2016; Huggett 2019; *cf.* Huggett 2015).

Throughout the duration of ARIADNE the same project team at ZRC SAZU has also been deeply involved in a project analyzing an Early Medieval cemetery. The necessities of the latter coupled with an exchange of ideas and project results taking place within ARIADNE gave birth to the idea to elevate an antiquated database Zbiva (v2) into a modern web application. The third piece of the puzzle was a computer scientist, M. Kastelic, who introduced the Arches platform to Slovenian archaeology (Kastelic et al. 2016).

Most of the design ideas behind Zbiva, which elevate Zbiva above a simple implementation of the Arches platform, were informed either by the ARIADNE User Needs report (Selhofer and Geser 2015) or by discussion with our many collaborators within ARIADNE. The most important contribution of ARIADNE to Zbiva, perhaps, was that it enabled us to acquire an overview of the digital archaeology landscape in Europe. This led us to realistically position ourselves in that landscape and, as mentioned above, to realize our weaknesses and our strengths. While we will probably never spearhead technological developments we can employ the ever emerging new digital tools to create added value that on occasions can rise to be second to none.

References

Achino, K.F., Štular, B., Rihter, Je. and Rihter, Ja. 2019 'Assessing the intentionality of spatial organization. Cemetery of Župna Cerkev (Kranj, Slovenia) case study', *Arheološki vestnik* 70, 297–313. Available http://av.zrc-sazu.si/pdf/70/AV_70_10_Achino_et_al.pdf (Last accessed: 28 June 2019)

ANSL = Gabrovec, S. et al. (eds) 1975 *Arheološka najdišča Slovenije,* Ljubljana: Državna založba Slovenije.

6 http://arkas.zrc-sazu.si

Belak, M. (eds) 2013 *Grobišče Župna cerkev v Kranju : dnevniki izkopavanj 1969 do 1973,* Ljubljana: Založba ZRC. Available http://iza.zrc-sazu.si/pdf/e_monografije/Kranj_4.pdf (Last accessed: 28 June 2019)

Belak, M. (ed) 2014 *Grobišče Župna cerkev v Kranju. Grobni zapisniki,* Ljubljana: Založba ZRC. Available http://iza.zrc-sazu.si/pdf/e_monografije/Kranj_6.pdf (Last accessed: 28 June 2019)

Dinklage, K. 1941a 'Die frühdeutschen Bodenfunde aus Krain und Untersteiermark', *Germanen-Erbe* 6, 69–80.

Dinklage, K. 1941b 'Frühdeutsche Volkskultur der Ostmark im Spiegel der Bodenfunde von Untersteiermark und Krain', *Mitteilungen der Anthropologischen Gesellschaft in Wien* 71, 235–259.

Dinklage, K. 1941c 'Oberkrains Deutschtum im Spiegel der karolin- gischen Bodenfunde', *Carinthia* I/131, 360–391.

Dinklage, K. 1943 *Frühdeutsche Volkskultur in Kärnten und seinen Marken,* Ljubljana: Ljudska tiskarna.

Franck, A. 1854 'Bericht über die Auffindung eines uralten Leichtenfeldes bei Kettlach unweit Gloggnitz und über einige andere bemerkenswerte Fundstücke', *Archiv für Kunde österreichischer Geschichtsquellen* 12, 235–246.

Giesler, J. 1980 'Zur Archäologie des Ostalpenraumes von 8. bis 11. Jahrhundert', *Archäologisches Korrespondenzblatt* 10/1, 85–98.

Grafenauer, B. 1954 *Zgodovina slovenskega naroda I. Od naselitve do uveljavljenja frankovskega fevdalnega reda*, Ljubljana: Kmečka knjiga.

Huggett, J. 2015 'A Manifesto for an Introspective Digital Archaeology', *Open Archaeology* 1 (1), 86–95.

Huggett, J. 2019 'Is Less More? Slow Data and Datafication in Archaeology' in: K. Garstky (ed) *Critical Archaeology in the Digital Age. 12th IEMA Visiting Scholar Conference, 6–7 April 2019, Greiner Hall, North Campus University at Buffalo, Abstract book,* Buffalo: IEMA, 13–14. Available http://iema.buffalo.edu/wp-content/uploads/sites/18/2019/03/2019-IEMA-Conference-Program_LowRes.pdf (Last accessed: 28 June 2019)

Jones, S. 2003 *The Archaeology of Ethnicity. Constructing identities in the past and present,* New York: Routledge.

Kahl, H.-D. 2002 *Der Staat der Karantanen. Država Karantancev. Fakten, Thesen und Fragen zu einer frühen slawischen Machtbildung im Ostalpenraum (7.–9. Jh.),* Ljubljana: Narodni muzej Slovenije.

Kansa, E. and S. Whitcher Kansa 2016 'Toward Slow Data in Archaeology' *Presented at The 81st Annual Meeting of the Society for American Archaeology,* Orlando, Florida doi:10.6067/XCV8930W5N

Kastelic, B. 2015 *Predstavitev izuma in razvoja zgodnjih plovil s platformo Arches.* MA thesis, University of Ljubljana, Ljubljana.

Kastelic, B., M. Belak, A. Pleterski, B. Štular and M. Erič 2016 'Zbiva in EWD, spletni orodji za arheološke raziskave' in T. Erjavec and D. Fišer (eds) *Zbornik konference Jezikovne tehnologije in digitalna humanistika (Proceedings Language Technologies & Digital Humanities),* Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani, 198–199. Available http://www.sdjt.si/wp/wp-content/uploads/2016/09/JTDH-2016_Kastelic-et-al_Zbiva-in-EWD.pdf (Last accessed: 28 June 2019)

Korošec, J. 1947 *Staroslovenska grobišča v severni Sloveniji*, Celje: Družba sv. Mohorja.

Korošec, P. 1979 *Zgodnjesrednjeveška arheološka slika karantanskih Slovanov,* Dela 1. razreda SAZU 22: 1, 2, Ljubljana: Slovenska akademija znanosti in umetnosti.

Lee Enriquez, A., Myers, D. and Dalgity, A. 2018 'The Arches Heritage Inventory and Management System for the Protection of Cultural Resources' *Forum Journal* 32 (1), 30–38.

Pleterski, A. 1997 *Inštitut za arheologijo – polstoletnik (Fiftieth Anniversary of the Institute of Archaeology),* Ljubljana: Založba ZRC.

Pleterski, A. 2001 'Staroslovansko obdobje na vzhodnoalpskem ozemlju. Zgodovina raziskav do prve svetovne vojne', *Arheo* 21, 73–7.

Pleterski, A. 2013 'Korak v kronologijo zgodnjesrednjeveškega naglavnega nakita vzhodnih Alp = A step towards the chronology of early medieval head ornaments in the Eastern Alps, *Arheološki vestnik* 64, 299–334. Available http://av.zrc-sazu.si/pdf/64/AV_64_2013_Pleterski_obrocki.pdf (Last accessed: 28 June 2019)

Pleterski, A. and M. Belak 1995 'Zbiva. Cerkve v vzhodnih Alpah od 8. do 10. stoletja (Zbiva. Archäologische Datenbank für den Ostalpenbereich. Die Kirchen in den Ostalpen vom 8. bis 10. Jahrhundert)', *Zgodovinski časopis* 49 (1), 19–43.

Pleterski, A., Štular, B. and Belak, M. (eds) 2016 *Izkopavanja srednjeveškega in zgodnjeno-voveškega grobišča pri Župni cerkvi v Kranju v letih 1964–1970 (Excavations of Medieval and Post-Medieval cemetery at the Župna cerkev in Kranj from 1964 to 1970), Ljubljana:* Založba ZRC. Available https://iza2.zrc-sazu.si/sites/default/files/ISBN9789612549732.pdf (Last accessed: 28 June 2019)

Pleterski, A., Štular, B. and Belak, M. (eds) 2017 *Arheološka raziskovanja grobov najdišča Župna cerkev v Kranju med letoma 1972 in 2010 (Archaeological investigation of graves from Župna cerkev site in Kranj between 1972 and 2010), Ljubljana:* Založba ZRC. Available https://iza2.zrc-sazu.si/sites/default/files/kranj_2_oias_37_2017_m_0.pdf (Last accessed: 28 June 2019)

Pleterski, A., Štular, B., Belak, M. and Bešter, H. (eds) 2019 *Začetek in konec raziskovanj grobišča pri Župni cerkvi v Kranju (1953 in 2011–2013) (The beginnning and end of exploration at the Župna cerkev cemetery in Kranj (1953 and 2011–2013)),* Ljubljana: Založba ZRC. Available https://iza2.zrc-sazu.si/sites/default/files/kranj_3_oias_38_2019_m.pdf (Last accessed: 28 June 2019)

Sagadin, M. and Belak, M. (eds) 2014 *Grobišče Župna cerkev v Kranju. Dnevniki izkopavanj 1984, Ljubljana:* Založba ZRC. Available http://iza.zrc-sazu.si/pdf/e_monografije/Kranj_5.pdf (Last accessed: 28 June 2019)

Schmid, W. 1913, 'Römische Forschung in Österreich 1907–1911' *Bericht der Römisch-Germanischen Kommission* 6 (1910–1911), 79–113.

Selhofer, H. and Geser, G. 2015 *D2.2 Second Report on Users' Needs.* Salzburg: ARIADNE. Available http://legacy.ariadne-infrastructure.eu/wp-content/uploads/2019/01/ARIADNE_D2.2_Second_report_on_users_needs.pdf (Last accessed: 28 June 2019)

Štular, B. and M. Belak (eds) 2012a *Grobišče Župna cerkev v Kranju. Dokumentacija o izkopavanjih v letu 1953,* Ljubljana: Založba ZRC. Available http://iza.zrc-sazu.si/pdf/e_monografije/Kranj_1.pdf (Last accessed: 28 June 2019)

Štular, B. and M. Belak (eds) 2012b *Grobišče Župna cerkev v Kranju. Kartoteka najdb iz leta 1953,* Ljubljana: Založba ZRC. Available http://iza.zrc-sazu.si/pdf/e_monografije/Kranj_2.pdf (Last accessed: 28 June 2019)

Štular, B. and M. Belak (eds) 2013 *Grobišče Župna cerkev v Kranju: dokumentacija o izkopavanjih v letih 1964, 1965 in 1966*, Ljubljana: Založba ZRC. Available http://iza.zrc-sazu.si/pdf/e_monografije/Kranj_3.pdf (Last accessed: 28 June 2019)

Štular, B., F. Niccolucci and J. Richards 2016 'ARIADNE: povezani odprti podatki (LOD) v praksi (ARIADNE: Linked Open Data (LOD) in practice)' in T. Erjavec and D. Fišer (eds), *Zbornik konference Jezikovne tehnologije in digitalna humanistika (Proceedings Language Technologies & Digital Humanities),* Ljubljana, 158–163. Available http://www.sdjt.si/wp/wp-content/uploads/2016/09/JTDH-2016_Stular-et-al_ARIADNE-povezani-odprti-podatki-LOD-v-praksi.pdf (Last accessed: 28 June 2019)

Štular, B. and A. Pleterski 2018 'Prolog. Zgodnjesrednjeveška arheologija jugovzhodnoalpskega prostora: nekoč, danes, jutri [Prologue. Early Medieval archaeology in the South eastern Alpine area: past, present, future]' in J. Lux and B. Štular and K. Zanier (eds) *Slovani, naša dediščina [Our heritage: the Slavs],* Ljubljana: ZVKD.

Štular, B. and S. Štuhec 2015 *3D Archaeology. Early Medieval Earrings from Kranj,* Ljubljana: Založba ZRC. Available https://books.apple.com/si/book/3d-archaeology/id972355479?ign-mpt=uo%3D4 (Last accessed: 28 June 2019)

The Italian Geoportal for Archaeology: a new tool for archaeological data integration and reuse

Valeria Acconcia*, Valeria Boi**, Sara Di Giorgio, Annalisa Falcone**** and Paola Ronzino********

Service II – Protection of Archaeological Heritage, MiBAC, Rome, Italy
*valeria.acconcia@beniculturali.it
****annalisa.falcone@beniculturali.it

Central Institute for Archaeology, MiBAC, Rome, Italy
**valeria.boi@beniculturali.it

ICCU, MiBAC, Rome, Italy
***sara.digiorgio@beniculturali.it

PIN, Prato, Italy
*****p.ronzino@gmail.com

Abstract

The Italian Central Institute for Archaeology (ICA), in collaboration with the Directorate-General Archaeology Service Ii - Protection of Archaeological Heritage (DG ABAP), has begun the development of a National Geoportal for Archaeology (GNA), with the technical support of a team of researchers from the VAST-LAB of PIN. The GNA is intended as an online platform that will act as a single point of access and exchange, through which it will be possible to archive, search and learn about archaeological datasets. The Geoportal is the result of an ongoing discussion about the management of archaeological information in Italy, begun in 2007 and continued by universities and members of the Italian Ministry of Cultural Heritage and Activities (MiBAC). The GNA will provide archaeological researchers and professionals with open and accredited access to resources and data already held by MiBAC. Moreover, it will encourage universities to contribute existing and newly created geodatabases that are the result of archaeological fieldwork. The GNA will be developed following the paradigm of the ARIADNE portal, allowing data visualization and browsing by applying filters based on type, chronology and location. Furthermore, it will ensure item-level integration of geographical information and enable integration with other public geodatabases.

Keywords: geoportal; INSPIRE; spatial data; ontology; preventive archaeology

Introduction

As stated in the Italian Code of Cultural Heritage, 'Protection consists in the exercise of functions and in the discipline of activities aimed at identifying the assets constituting cultural heritage and guaranteeing their protection and conservation for purposes of public use, on the basis of an adequate cognitive activity' (Article 3 of Legislative Decree 42/2004).[1]

In law, therefore, knowledge is at the core of cultural heritage protection. There is no protection without knowledge, and the ultimate goal is public benefit. This brief statement of principles outlines the role of MiBAC in promoting knowledge of heritage assets, in guaranteeing their preservation, and in making them available to public. All actors that contribute to the creation, collection, management and storage of data in Italy (the Ministry, universities, Italian and foreign research institutes, and professionals) are called upon to guarantee, within the framework of their respective rights and duties, that activities translate into shared knowledge.

In the case of archaeological heritage, the process of knowledge creation generally involves a destructive activity, i.e. excavation. The documentation produced during fieldwork is the only record of stratigraphy and structures removed. This documentation, delivered to the ministerial offices at the end of an excavation, whether carried out with an excavation permit or other planned research activity (e.g. preventive or emergency investigations) plays a crucial role for future studies, conservation, and data re-use.

For technological reasons, access to such documentation was once only possible in the archives that stored them, with all the resulting limitations. One of the most controversial points in the management of archaeological data was the impossibility of bringing together, within the same digital spatial platform, the information resulting from research conducted by universities and research institutions, and data held by MiBAC. A conspicuous amount of archaeological data, at times completely or partially unpublished, are currently held in different archives, sometimes only in paper copies.

In recent years, thanks to the ever-increasing availability of digital archiving tools and the sharing of consistent information on the web, many projects have launched the publication of archaeological data online. In particular, for datasets with spatial information, the diffusion of GIS applications and increased awareness regarding open data has led to the proliferation of WebGIS portals with the consequent production, at the European level, of a vast amount of spatial data and the adoption of different standards and vocabularies, preventing the creation of interoperable systems (McKeague et al. 2012; Azzena et al. 2016; McKeague et al. 2017; Oniszczuk and Makowska 2017). Meanwhile, in cartography, spatial data homogenisation has been achieved at the European level through the INSPIRE Directive of the

[1] Codice dei beni culturali e del paesaggio, ai sensi dell'articolo 10 della legge 6 luglio 2002, n. 137, art.3 D.L. del 22 gennaio 2004, n. 42 (G.U. n. 45, suppl. 28, 24 febbraio 2004).

European Commission (2007/2/EC INSPIRE).[2] In Italy, however, the current situation is represented by a lack of homogeneity, including highly advanced systems (both at the ministerial and university levels) with portals that offer data through web services compliant with the INSPIRE legislation, together with various WebGIS and geoportals born as a result of individual projects, and therefore not compliant with common standards (Ronzino et al. 2018).

Within archaeology, a high degree of interoperability has been achieved by ARIADNE, with an infrastructure that integrates archaeological datasets scattered across Europe, making them available through its portal.[3] Driven by the need to take advantage of a unique platform for discovering and accessing archaeological information, the GNA will be configured as a hub that will index resources, making them easier to find and guaranteeing their online publication, while at the same time preserving their individuality and ensuring that the ownership of data remains with the data producer. Following the success of ARIADNE, the GNA will be designed to allow data visualization and browsing across type, chronology and location of the archaeological resources (Aloia et al. 2017; Meghini et al. 2017).

The management of archaeological datasets in Italy

The GNA project originated from the observation that currently existing web-platforms do not allow integrated management and examination of known archaeological information located within Italy. For this reason, the project is committed to developing a unique platform for accessing the whole documentation, promoting the adoption of open licenses (*Figure1*). The reasons leading to the creation of an infrastructure for national data able to manage this type of information are well-known to the stakeholders: first of all, the need to reorganize and digitize archaeological documentation stored in the archives of MiBAC's peripheral institutes. The results of archaeological excavations conducted by universities and research institutes (which sometimes focus on limited areas and thus provide answers to very specific scientific queries), are often made available only through traditional printed editions, whilst the bulk of the data collected on site is not accessible online because of problems related to the sustainability of projects beyond their lifetime. The possibility of interacting with other databases produced by urban and regional management and planning bodies in Italy is another aspect that is strictly connected to landscape protection. Within MiBAC, ICA acts as the collection centre for archaeological data, unifying the requirements concerning protection/research/dissemination, proposing shared guidelines, and organizing and disseminating such data in a consistent way.

A recent assessment carried out by the authors revealed the presence of a large number of online databases, WebGIS and geoportals that were born as a result of

[2] European Commission 2007 Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) http://eur-lex.europa.eu/legalcontent/ EN/ALL/?uri=CELEX:32007L0002.

[3] ARIADNE portal, http://portal.ariadne-infrastructure.eu/

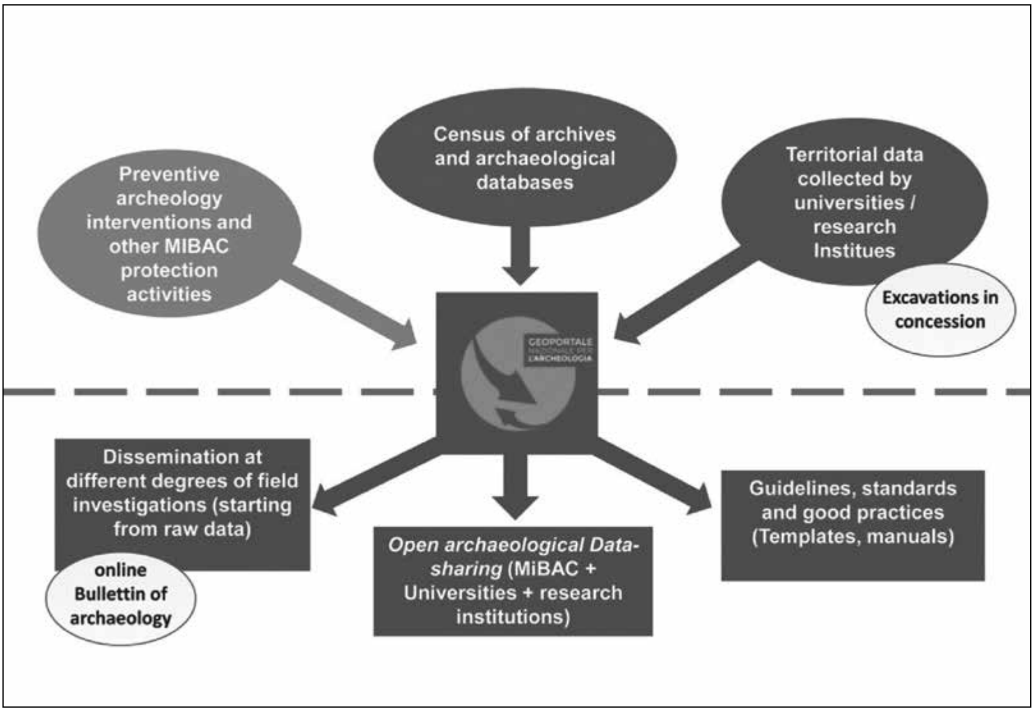


Figure 1: This image shows documentation that will be managed by the GNA and the expected outcomes and services that will be provided to users

different activities involving the public, private or academic sectors (Ronzino et al. 2018). Each database, whether the result of a specific research project or an institutional remit, is characterized by diverse structures and contents, according to the aims of data collection and reuse. The diverse IT solutions adopted and the dispersion of data across platforms makes it difficult to understand the state of knowledge, resulting in a tangible obstacle to its preservation, and to the study of a specific landscape or chronological context. The information held in the many archives scattered throughout the country is enormous but unfortunately, its actual availability is limited by a series of factors that contribute to greatly limiting knowledge and reuse. The authors of the archives find it difficult to guarantee medium-long term sustainability and technological updates. As a result, the data collected is rarely published online at the end of the project, and even when they are digitized and shared online the databases are not easily findable. On top of this, the extreme heterogeneity in the methods of collection, organization and publication of information makes it extremely difficult for the user to compare different sources of information, especially if quantitative or statistical data are sought.

Real interoperability between such different databases is certainly not a simple goal to achieve. However, it is essential to avoid what can be defined as a 'waste of data', or rather the lack of disclosure of information, in particular that deriving from either fully or partially unpublished research, which runs the risk of obsolescence before the scientific community has ever had access to it. This fragmented situation and the availability of many archaeological institutes to contribute their data for consultation and reuse has mobilized ICA to develop a national spatial infrastructure for the integration of archaeological data. Furthermore, the adoption of widely accepted standards, such as those being developed by ARIADNEplus (Niccolucci 2018) and its compliance with the INSPIRE legislation, represents a first step towards the integration of spatial data at the national and European level.

The stages of the GNA project development

The design of the geoportal was preceded by the assessment and analysis of the existing online resources, which revealed the heterogeneity of standards and forms used to structure the related databases (including projection systems, accuracy, format and content). The complex task of surveying and collecting data to publish on the GNA proceeds in parallel with the organization and design of the portal. Collecting the datasets developed by contributors of the geoportal will make the results of national research projects accessible online (from bibliographical and archival research, to field survey and actual excavation). One fundamental aspect of the new portal's structure will be its ability to accept heterogeneous datasets, organized in a such a way as to fulfil the requirements of diverse yet complimentary objectives, for example those related to research and protection. The geoportal will adopt the technologies and tools developed within ARIADNE – particularly the mapping and aggregation tools – to allow discovery and access of data contained in several datasets: not only the data stored directly in the geoportal's dedicated servers, but also the information that has already been made available online by its authors through WMS-WFS exchange protocols. In both cases, of course, the digital resources remain the property of the rights-holders, either the author or the repository where it is stored, and any limits to accessibility or consultation foreseen by the rights-holder for each single dataset are taken into account.

Within the remit of the protective measures carried out by MiBAC, the publication of data on the geoportal fulfils a double role: on the one hand, it allows internet access to documentation that is currently held in the archives of its peripheral offices; on the other, it sets up tools for the submission of records originating from local heritage protection. The tools currently available to the public administration allow not only its own staff but also external users to digitally manage the flow of information from the very start, facilitating conservation, consultation and reuse through different access procedures if these are necessary.

Another activity carried out to support the design of the GNA was the identification of user needs, collected through a short survey based on the simplified Cockburn method (Cockburn 2000). The survey was submitted to a group of officers of MiBAC, experts within the domain, who provided feedback on the features and tools required to carry out their routine work (Ronzino 2018). Almost all experts agreed upon the need to make data open and easily accessible to everyone, making it easier to protect the cultural heritage. The groups of experts who participated in the user requirements survey ranked preventive archaeology first in importance (60% of answers). This analysis contributed to identify preventive archaeology as a priority area for the National Geoportal's further strategic work. Other activities envisaged by users concern the analysis of the archaeological landscape and activities related to public archaeology and academic research.

The preventive archaeology procedure

At the beginning of 2018, as part of the preliminary activities for the design of the geoportal, ICA, in collaboration with the Central Institute for Cataloguing and Documentation (ICCD) and Service II of DG ABAP, began to survey and digitize the data collected in the context of preventive archaeology carried out between 2006 and 2016. Among these, impact assessments are of great importance because data is collected as the result of activities of public interest financed with public investments. Data collection and processing, carried out by professionals or universities, includes bibliographic and archival research, photo interpretation, archaeological survey and, in a second phase, excavation. Without access to data, there is the risk of repeating studies in the same areas, thus multiplying costs not only for the contractors but also for the general public. At present, tests have been conducted with the collaboration of the Superintendencies of three pilot regions: Piedmont, Tuscany and Apulia. These regions have started surveying the documentation related to preventive archaeology held in the archives of its peripheral offices, proceeding to the digitization of the data by filling in templates prepared according to ICCD documentation standards, the information template - MODI 4.00.[4] (*Figure 2*)

The need to describe both the administrative data concerning the assets subject to preventive archaeology and the large number of archaeological sites/areas has led to the preparation of two separate forms: the MOPR - Project Module and the MOSI - Sites/Areas Module. The first form (MOPR) collects information about public works, the second (MOSI) collects information on individual sites or archaeological areas that fall within the affected area. In the development of both forms it was decided to make the compilation of some information mandatory: type of work and client, type of investigation and sources collected, dating, site dimensions, geographical data and research impact. Collecting spatial information is crucial for future integration with other geographical data held by MiBAC and future publication on the GNA.

Some of the peripheral offices completed the task of entering data, making it possible in January 2019 to start verifying the completed forms and preparing the tools for publication. In particular, the collaboration established between ICCD and the Higher Institute for Conservation and Restoration (ISCR) to harmonize the standards for the description and representation of information between the GNA, the General

[4] MODI http://www.iccd.beniculturali.it/index.php?it/473/standardcatalografici/Standard/105

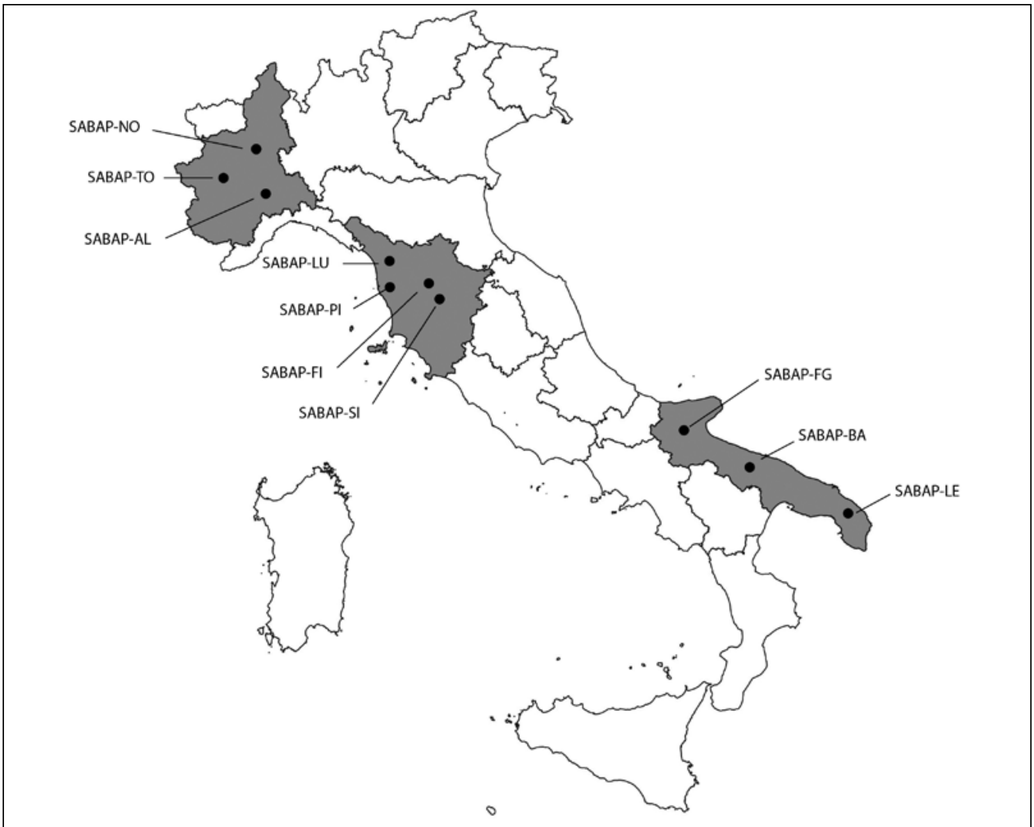


Figure 2: Superintendencies participating to the test of the new standards (MODI 4.00)

Information System of the Catalogue (SiGeCweb) and Vincoli In Rete (VIR)[5] platform was of great importance. The working group's priorities are the definition of file formats and the enhancement of data visualization functions, as well as the definition of download and reuse policies for data from other institutions. Once testing is over, the system used to collect information from legacy documentation will be adapted to the procedure for submitting new archaeological impact assessments. Suitable tools will be made available for drafting project documentation and for digitally organizing scientific information arising from the various phases of the archaeological study (archival and bibliographic research, non-invasive investigations, excavation) in order to make them immediately available and accessible online through the GNA.

The positive impact of this initiative can be seen in the systematization and sustainability of data storage, which will be accessible both to the officers of the Superintendences, and to scholars and professionals. It is hoped that this method of data publication will stimulate the production of more accurate predictive studies and archaeological risk assessments, making heritage protection more effective.

5 Vincoli in Rete http://vincoliinrete.beniculturali.it/VincoliInRete/vir/utente/login#

The GNA and the addition of new data

Besides giving full visibility and access to existing archaeological datasets, the GNA will allow archiving and management of newly produced information. In the coming months, ICA in collaboration with Service II of the Directorate-General Archaeology will promote the use of standard formats promoted by ICCU to professionals and archaeological companies for the submission of scientific excavation data to the offices of the Superintendence. The related metadata will be drafted based on AO-Cat, the formal ontology currently under definition by ARIADNEplus. Consequently, the documentation archived in the offices of the Superintendence will eventually be digitally surveyed and published on the GNA, guaranteeing access to a minimum amount of information through the online platform.

Thus, archaeological information finally comes full circle: the data collected by the Superintendences and professional archaeologists will be made available to archaeologists evaluating archaeological impact, and to researchers and governmental bodies managing the landscape (*Figure 3*). Improved accessibility to known data will contribute to the creation of more accurate studies, and to shared landscape planning. By overcoming the dichotomy between domain experts and the general public, full accessibility of online data, as promoted by the geoportal, will be offered to all those who are interested to obtain archaeological information at various levels of detail, therefore meeting the constantly increasing demand for publicly accessible and shared archaeology. The needs of landscape protection, archaeological research and data accessibility will thus be combined with citizen participation.

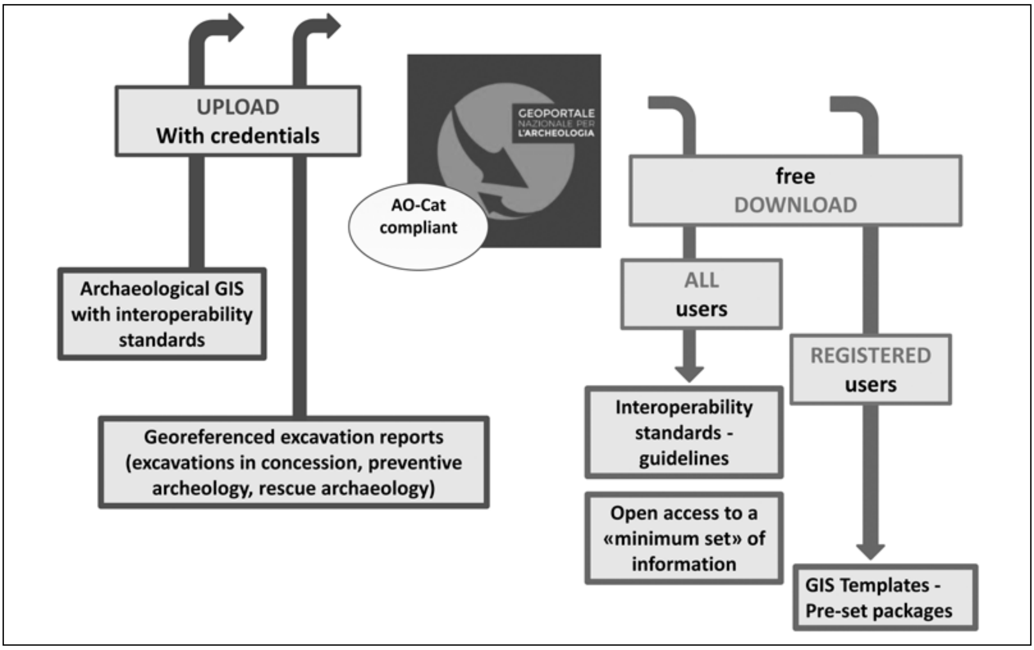


Figure 3: The full circle of archaeological information

Conclusions

This chapter has described the activities underway at the Central Institute for Archaeology (ICA), together with the Directorate-General Archaeology's Service II - Protection of Archaeological Heritage, in collaboration with the VAST-LAB team, for the creation of a geoportal (GNA) that acts as a national aggregator for archaeological datasets, in particular those with a geospatial component.

The portal will not only hold data about preventive archaeology and emergency excavations enacted by MiBAC. In the spirit of data sharing and collaboration between MiBAC, the academic world and other actors working in the sector, it will also include the results of research carried out by universities and research institutes, thus contributing to greater awareness and protection of the archaeological heritage. This initiative has already had a positive impact on the activities carried out by the Ministry. The simplified retrieval of information and a deeper knowledge of the landscape has assisted the complex task of managing daily activities; alleviating the situation where archives are located in different offices, or where management suffers from a turnover of local officials.

The positive repercussions also extend to the construction of a more profitable dialogue with local authorities, who can count on a platform that is easy to access and use, thus enabling a first assessment of archaeological evidence in the context of landscape planning and large infrastructure developments. For individual researchers and professionals collaborating with the Ministry in the field of preventive and emergency excavations, the opportunity to find information easily will result in more detailed and reliable assessments.

Lastly, we believe that the circulation of information can generate a virtuous loop, in which awareness of future publication can stimulate the production of quality excavation documentation, with greater attention paid to understanding and reusing information, thereby contributing to the dissemination of good practices and sharing and reuse of data.

References

Aloia, N. et al. 2017 'Enabling European Archaeological Research: The ARIADNE E-Infrastructure', *Internet Archaeology* 43. doi:10.11141/ia.43.11

Azzena G., Busonera R., Nurra F. and Petruzzi E. 2016 'Homogenization of the Archaeological Cartographic Data on a National Scale in Italy' in S. Campana, R. Scopigno, G. Carpentiero and M. Cirillo (eds) *Proceedings of the 43rd CAA Conference, 2016,* Oxford: Archaeopress, 283–292.

Cockburn, A. 2000 *Writing Effective Use Cases.* Boston, USA: Addison-Wesley Professional.

McKeague P, Corns, A. and Shaw, R. 2012 'Developing a spatial data infrastructure for archaeological and built heritage', *International Journal of Spatial Data Infrastructures Research* 7, 16–37. http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/239

McKeague, P., Corns, A. and Posluschny A. 2017 'Why the Historic Environment needs a Spatial Data Infrastructure', *Internet Archaeology* 43. doi:10.11141/ia.43.7

Meghini, C.et al. 2017 'ARIADNE: A Research Infrastructure for Archaeology', Journal on Computing and Cultural Heritage 10.3. doi:10.1145/3064527

Niccolucci F. 2018 'Integrating the digital dimension into archaeological research: the ARIADNE project', *Post-Classical Archaeologies* 8: 63–74.

Oniszczuk, A. and Makowska, A. 2017 'Archaeological Data in the GIS Portal of the National Heritage Board of Poland', *Internet Archaeology* 43. doi:10.11141/ia.43.5

Ronzino, P., Acconcia, V. and Falcone A. 2018 'Towards the integration of spatial data through the Italian Geoportal for Archaeological Resources', *Digital Heritage International Congress (DigitalHeritage), San Francisco.*

From ARIADNE to ARIADNEplus: Developing the concept of archaeological digital data and increasing its 'FAIRness'

Sorin Hermon[*] and Valentina Vassallo[**]

The Cyprus Institute, Cyprus
*sorin.hermon@gmail.com
**v.vassallo@cyi.ac.cy

Abstract

The Cyprus Institute, a non-profit private research and academic institution, represented Cyprus in ARIADNE and ARIADNEplus, with multiple roles reflecting its research and social mission in the MENA (Middle East and North Africa) region at the southern edge of Europe. As such, its objectives were, amongst others, to provide a bridge between MENA regional players and European partners, to raise local and regional awareness of the importance of archaeological data sharing and to enable the integration of multi-disciplinary archaeological data and its scientific visualisation through web-based solutions. Direct outcomes of the experiences gained throughout ARIADNE were the participation of our researchers in related EC-funded initiatives, development of various archaeological data-related services for the regional communities and establishing the role of our institution as a regional leader in archaeological data management.

Keywords: heritage at risk; scientific visualisation; FAIR data in archaeology; partonomy

Introduction

The Cyprus Institute was founded more than a decade ago with the mission to create an international research and academic institution, with strong ties to European partners, benefitting from support of key US-based research institutions and with a wide regional network, in order to have a positive impact on the regional scientific agenda, improve the societal impact of research, and create a suitable environment for innovation in research. Since its establishment, it has tackled research topics of global importance and regional relevance through the application of advanced technologies and sciences, focusing on a rather narrow range of subjects, while investing in their in-depth investigation and by developing unique state-of-the-art laboratory infrastructures. The three initial research subjects were climate change, cultural heritage and computational science, recognizing the potential of cultural heritage as a trigger for societal, economic and cultural benefit in the region. The Science and Technology in Archaeology and Culture Center has been advancing, since

Figure 1: The data repository: home page, browsing options and visualisation of single assets

its launch in 2009, new digital methods, tools and methodologies to promote research in Cultural Heritage (CH), amongst which data management has a fundamental role (Niccolucci et al. 2009). Soon after its establishment a data repository (Damnjanovic et al. 2012; Vassallo et al. 2013) was created (*Figure 1*), hosting various types of archaeological data, such as images, texts, 3D models, video and audio files and architectural drawings, created through a wide range of research collaborations with local and regional institutions, as well as hosting data from local entities, such as museums and art galleries, or foreign archaeological excavations in Cyprus. Its main guiding concepts were the organisation of data by collections and types of data, with searchable and browsing options and visualisation of single items through dedicated windows, allowing visualization of the associated metadata and visual interaction with the digital object. The repository was instrumental in the participation of the institute in various European initiatives on CH data repositories and related portals (such as Europeana) while at the same time it posed research questions and challenges on how to gather, describe and manage such data.

The ARIADNE experience in Cyprus

Several fundamental questions guided the development of our archaeological data repository:

- which kind of data it should contain,
- how this data should be accessed and by whom,
- data management in a broader, national, regional and European perspective.

Archaeology is an interdisciplinary subject, combining research in the humanities, social, natural and exact sciences. As such, any repository for archaeological data should prioritise the concept of interoperability as a key factor for data gathering as well as data use and re-use. The CIDOC CRM is an optimal domain ontology that allows the flexibility of semantic descriptions required for such inter-disciplinary domains as archaeology (Sugimoto et al. 2007), while at the same time it is built in such a way that it allows the development of specific extensions aimed at grouping and representing information around a focused topic, such as the CRMarchaeo for describing the archaeological excavation process (Doerr et al. 2016) and its digital data provenance outcomes (CRMdig) (Pitzalis et al. 2010) or data describing building

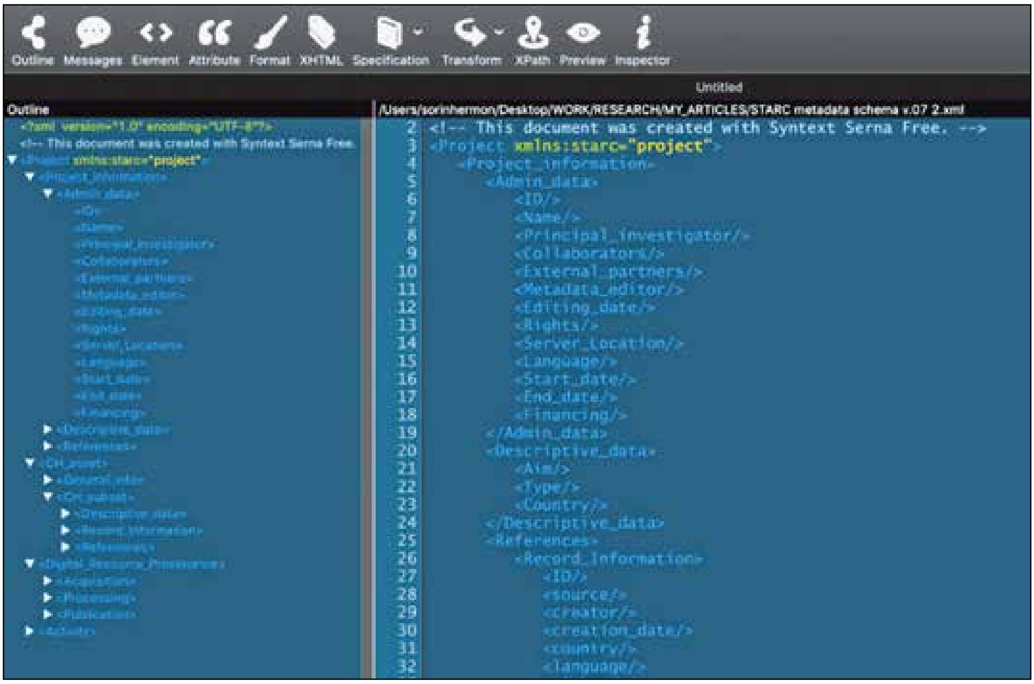


Figure 2: The Cyprus Institute metadata schema for Digital Cultural Heritage

information (CRMba) (Ronzino et al. 2016). Based on the above concepts and having the CARARE EU-funded project schema in the background, we have developed a schema (*Figure 2*), describing archaeological datasets (Ronzino et al. 2012).

The schema was built to provide a holistic description of a physical CH asset and its digital surrogate, along with its provenance and qualitative properties, as well as information about the overall framework of activities which led to the creation of the specific digital content. The repository described in the previous section and the metadata schema detailed above, along with strategic partnerships with major CH administrative public and private bodies, such as The Cyprus Department of Antiquities (DoA) and the Leventis museum, served as the basis for the participation of our institute in ARIADNE, as a content and archaeological knowledge provider, and it enabled our researchers to participate in various thematic workgroups of the project.

The main impacts of ARIADNE on our institute can be summarised as follows:

- *Generation of a new approach to data repositories* – working within a large community pushed us to adopt strategies for strengthening local and regional cooperation among academic, research and administrative institutions.
- *New knowledge for scientific community* – following the ARIADNE initiative, we have started to consider integrating archaeological data along its development pipeline, for data acquisition to processing, post-processing, analysis and interpretation. Such an action enabled a more transparent approach to collaborative

research, in particular for the assessment of data quality and provision of insights into the interpretative process.

- *Tools for online data analysis* – providing access to data is only a first step towards collaborative research; we have developed a set of tools (see below) that enabled a richer interaction with the digital content and consequently a new approach to data investigation.
- *Knowledge repositories as archaeological research environments* – a major achievement of ARIADNE was the recognition that archaeological research can be conducted much more efficiently in a collaborative environment, with data which can easily be quality assessed, and which originates from multiple sources. Moreover, integrating data from various disciplines allows a better evaluation of the impact and contribution of each set of data on the overall archaeological interpretation.
- *Involvement of public institutions* – working closely with governmental bodies, such as The Department of Antiquities, helped us prioritise our research and devote a substantial effort to solve real-life challenges, while from the other end the DoA recognized the importance and the beneficial aspects of initiatives such as ARIADNE for their daily needs.
- *Stimulated further research to fill identified gaps* – the first stage of ARIADNE clearly delineated the subjects for future research, such as the need for more focused CRM extensions, enrichment of browser-based interfaces for interaction with digital content, development of tools for further exploration of archaeological data, such as scientific visualization tools, or new protocols for data ingestion and retrieval.
- *Participation in further research initiatives* – a major outcome of ARIADNE was the establishment of our research group as experts in archaeological digital data management, amongst other things, which lead to the creation of new research partnerships for various EU funding initiatives and projects.
- *Participation in E-RIHS (ESFRI infrastructure)* – following the development of our digital data repository and consequent scientific expertise, we were nominated as national representatives of Cyprus within E-RIHS: the European Research Infrastructure for Heritage Science.

From ARIADNE to ARIADNEplus

Once the foundation for a collaborative archaeological research environment, based on data repositories was laid, the next targets for our research group are well aligned with the priorities of ARIADNEplus. We aim to include more heterogeneous datasets and to create a variety of browser-based tools to interact with such content, while at the same time increasing data transparency and reliability (Niccolucci and Hermon 2015, 2016), better defining archaeological concepts (Hermon and Niccolucci 2017) and sharpening the semantic descriptions of archaeological information (Niccolucci et al. 2015). Moreover, scientific effort has been directed into the development of dedicated services that would facilitate the inclusion of new datasets, enhance the semantics-based data description and facilitate the integration of analytical data with

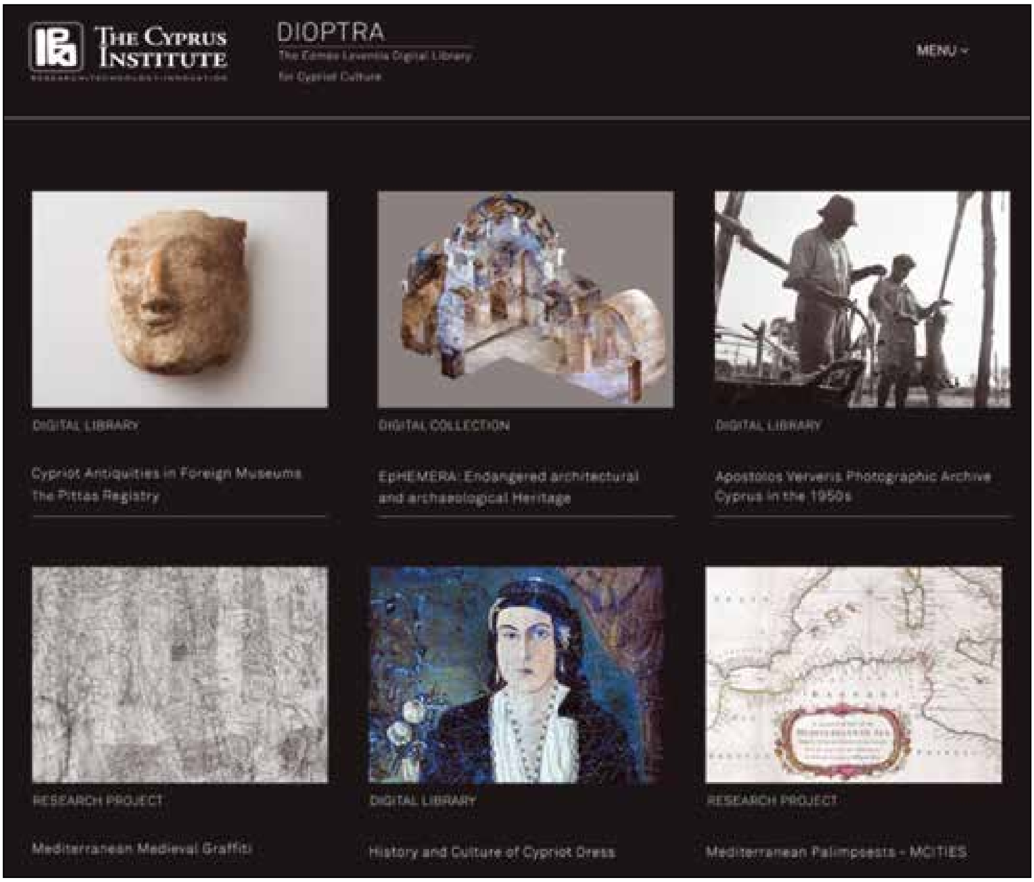


Figure 3: DIOPTRA – the digital library on Cyprus Heritage

3D visualisation datasets. Thus, DIOPTRA, the Digital Library for Cypriot Heritage was created (*Figure 3*) in collaboration with various CH bodies in Cyprus and the region, offering access to descriptions of technologies, conceptual and methodological details, as well as a set of collections with their respective interaction tools.

DIOPTRA builds on our initial repository and extends its functionality and scope, by providing not only access to data, but also by expanding the interaction with such data through dedicated browser-based tools (Damnjanovic et al. 2017), detailed descriptions of how the digital library was built, and concrete examples of the methodological steps and the decision-making process for how it was conceived, planned and implemented. Below are a few examples describing some of the datasets hosted by DIOPTRA. The Cypriot medieval coins history and culture collection (*Figure 4*) has been built in collaboration with the Cultural Foundation of the Bank of Cyprus in order to present the history of medieval coinage in Cyprus to the public. Users can browse or search the dataset (left image), and receive information on the selected coin through descriptive metadata (middle image) while having the

Figure 4: CMC – The Medieval coins history and culture collection

possibility to use browser-embedded Reflectance Transformation Imaging (RTI) tools to change the light direction and its intensity and thus discover details on the surface of the coin (right image).

Another example is EPHEMERA (Abate et al. 2017), the digital collection dedicated to the 3D scientific visualisation of endangered heritage (Hermon et al. 2014), rescue archaeological excavation sites and physically inaccessible heritage (Hermon et al. 2013), developed in collaboration with the Cyprus Department of Antiquities (*Figure 5*).

Its main functionalities enable a word-based search of the collection, 3D visualisation of selected data and related metadata and a set of tools allowing a range of measurements of the 3D model, such as calculation of distances between points, areas and volumes, as well as the extraction of cross-sections (*Figure 6*). EPHEMERA provides a browser-based virtual environment for exploring 3D models of built heritage assets and is fully compliant with the principles for the visualisation of Cultural Heritage, as set by The London Charter (Niccolucci et al. 2010).

Another consequence of ARIADNE was that it raised awareness of the potential of such collaborative environments based upon archaeological digital data, as exemplified by the EC-funded GRAVITATE project (Hermon 2018; Mortara et al. 2017; Phillips et al. 2017; Biasotti et al. 2015). The project dealt with 3D digital models of fragmented statues physically held in various museums across the UK and in Cyprus, thus highlighting another important aspect of archaeological data digital collection, namely remote access to digital surrogates of physical objects geograph-

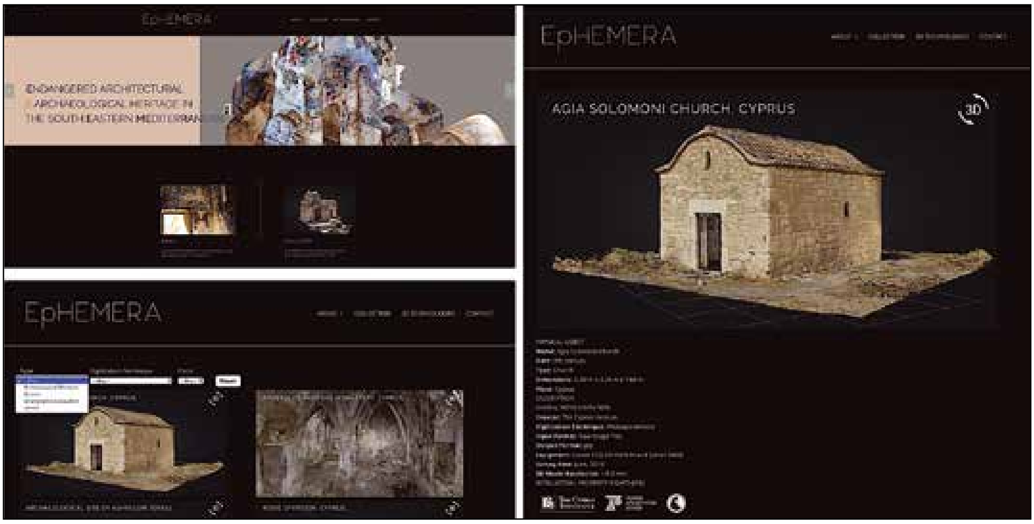


Figure 5: EPHEMERA, the digital library of 3D models of Heritage at Risk

Figure 6: Functionalities of EPHEMERA: browse/search the collection (left) and scientific visualisation interaction with the 3D models (right)

ically distributed over a large area. More than 250 fragments of heads, torsos, limbs, etc., probably belonging to c.30 statues, have been 3D digitized and semantically described (*Figure 7*), in order that they can be further explored by digital tools for virtual restoration, re-association and re-unification of the fragments. As such, tools to describe the rugosity of the surface, measure complex geometries (Scalas et al. 2018) and delineate features were created and assembled within a dashboard that allows visual inspection and precise measurements of selected fragments (*Figure 8*).

Figure 7: The digital collection of fragments (left) and one of the more complete fragments (right – almost a life-size terracotta statue)

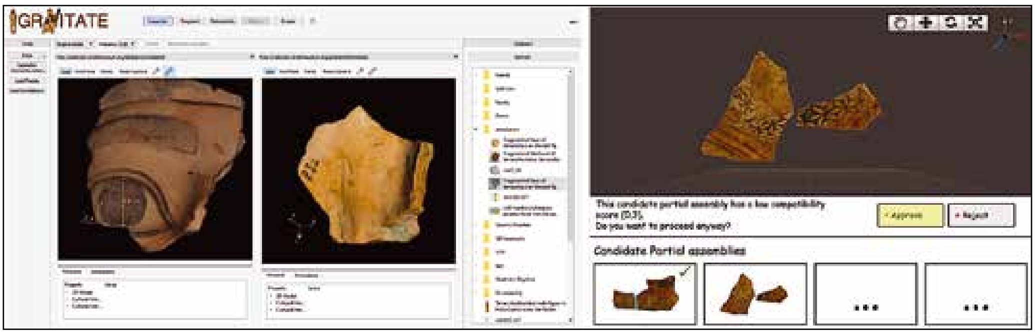


Figure 8: Functionalities of the GRAVITATE dashboard

Figure 8 presents some snapshots exemplifying how 3D geometry analysis is used to enrich semantic descriptions and how both are used to select from the collection fragments which have the potential for restoration.

A further development within GRAVITATE (Vassallo 2016) which was a direct result of the discussions previously held in ARIADNE regards the semantic description of components and relationships between them and the whole of a CH asset. Such a partonomy (Catalano et al. 2019, Vassallo et al. 2018), illustrated in *Figure 9* and based on semantic descriptions, is fundamental for an accurate description of an asset by its components, with implications for virtual restoration and re-association of fragments.

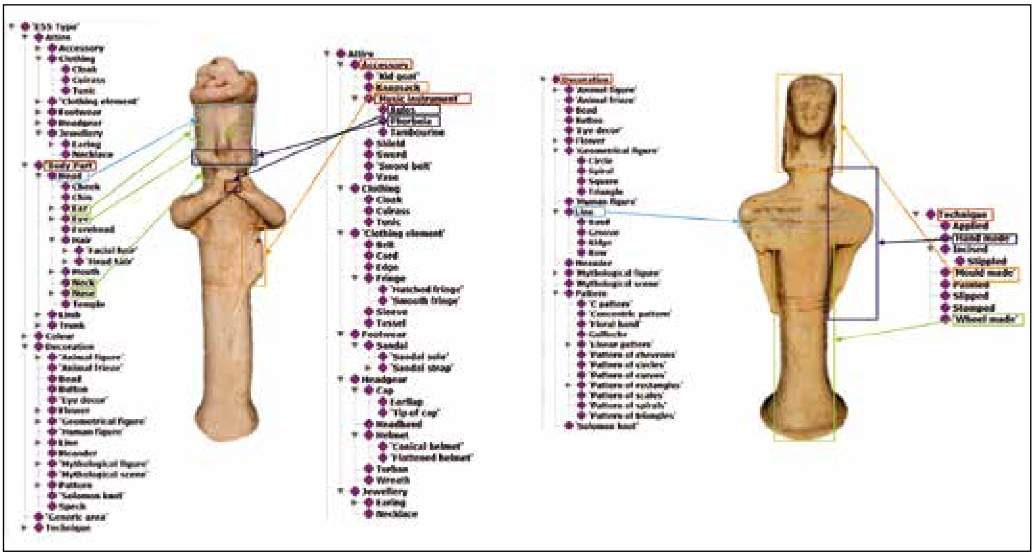


Figure 9: Partonomy description of terracotta figurines

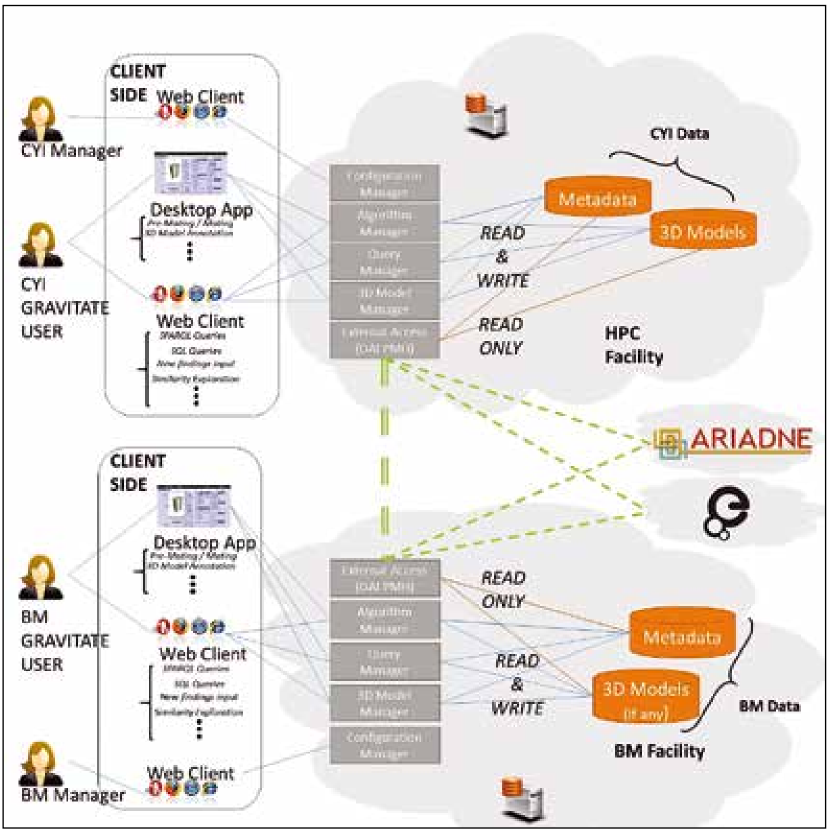


Figure 10: Diagram exemplifying the GRAVITATE deployment within ARIADNE

Figure 11: New datasets to be provided by CyI

A final example of a return benefit for ARIADNE is the attempt to deploy the GRAVITATE platform into the ARIADNE infrastructure, as a service for archaeologists working on fragmented items. *Figure 10* exemplifies the model envisioned for such an exercise, where ARIADNE service platform integrates GRAVITATE for the benefit of the wider archaeological community.

Towards ARIADNEplus

The cross-domain search of multi-disciplinary datasets may have a huge impact on future CH research, unlocking the immense potential of data to advance the discipline. As such, we are looking to expand the range of datasets to be provided by the Cyprus Institute to ARIADNEplus and to add information on medieval graffiti, their provenance and interpretation (Demesticha et al. 2017), analytical and art historical analysis of masterpieces (Gasanova et al. 2017), 3D models of archaeological artefacts with Cypro-Minoan inscriptions and 3D models (a PhD topic in our research group), and 3D models of medieval churches and their removed fresco fragments (Abate et al. 2016). *Figure 11* presents some of these datasets; a fundamental characteristic that unites them is their multidisciplinary component, integrating analytical measurements with technical imaging, 3D models and textual information, thereby posing further challenges for the ARIADNEplus platform, as well as exemplifying the richness of information available for study.

Lesson learnt and future work

Perhaps surprisingly, one of the major challenges that the ARIADNE approach had to overcome in Cyprus was the reluctance of researchers to share data or to shift the focus of their research from data processed (by others) to primary (raw) data, as well as the low awareness of FAIR issues related to data transparency and reliability. While there is still a lot to do, we can state with confidence that we have overcome these initial reservations and we are now moving into the second phase of work. This will prepare the ground for a paradigm shift in archaeological research, where data have to be gathered with regard to the FAIR principles and assuring transparency and further re-use, and where tools and methods are developed for the advancement of related collaborative digital environments.

References

Abate, D., Avgousti, A., Faka, M., Hermon, S., Bakirtzis, N. and Christofi, P. 2017 'An online 3D database system for endangered architectural and archaeological heritage in the south-eastern Mediterranean', *ISPRS International Archive of Photogrammetry, Remote Sensing and Spatial Information Science Conference Proceedings,* Volume XLII-3/W3, 1–5.

Abate , D., Hermon , S. and Eliades , I. 2016 'Virtual and physical re-composition of fragmented ecclesiastical frescoes using a photogrammetric approach' *Proceedings of the International Society for Photogrammetry and Remote Sensing, Commission VI,* Prague, 153–160. doi:10.5194/isprsarchives-XLI-B5-153-2016

Biasotti, S., Thompson, E.M., Barthe, L., Berretti, S., Giachetti, A., Lejemble, T., Mellado, N., Moustakas, K., Manolas, I., Dimou, D., Tortorici, C. and Hermon, S. 2018 'Recognition of geometric patterns over 3D models' *Proceedings of the 11th Eurographics Workshop on 3D Object Retrieval* Eurographics Association, 71–77.

Catalano C.E., Vassallo V., Hermon S. and Spagnuolo M. 2019 'A Cultural Heritage partonomy for the documentation of 3D digital artefacts of Cypriot coroplastic art' *Proceedings of the CIDOC 2018 conference,* Heraklion, Crete. doi:10.5281/zenodo.2536814

Damnjanovic, U., Vassallo, V. and Hermon, S. 2017 'Integration of Multimedia Collections and Tools for Interaction with Digital Content. The case study of the Archaia Kypriaki Grammateia *Digital Corpus' in S. Orlandi, R. Santucci, F. Mambrini, and P.M. Liuzzo (eds) Digital and Traditional Epigraphy in Context Proceedings of the EAGLE 2016 International Conference,* 247–262.

Damnjanovic, U. and Hermon, S. 2012 'Interactive Workspace for Exploring Heterogeneous Data' in G. Earl, T. Sly, A. Chrysanthi, P. Murrieta-Flores, C. Papadopoulos, I. Romanowska and D. Wheatley (eds) *Archaeology in the Digital Era* Volume II, e-Papers from the 40th CAA Conference, Amsterdam University Press, 110–119.

Demesticha, S., Delouca, K., Trentin, M.G., Bakirtzis, N. and Neophytou, A. 2017 'Seamen on Land? A Preliminary Analysis of Medieval Ship Graffiti on Cyprus', *International Journal of Nautical Archaeology,* 46(2), 346–381.

Doerr, M. et al. 2016 Definition of the CRMarchaeo - An Extension of CIDOC CRM to support the archaeological excavation process, version 1.4. Available at http://www.ics.forth.gr/isl/CRMext/CRMarchaeo/docs/CRMarchaeo_v1.4.pdf

Gasanova, S., Bakirtzis, N. and Hermon, S. 2017 'Non-invasive sub-surface analysis of the male portrait underlying the Titian's Studio Ecce Homo', *Heritage Science* 5:33. doi:10.1186/s40494-017-0145-2

Hermon, S. 2018 'Integrated digital approaches to the virtual restoration, re-unification and re-association of fragmented cultural heritage artifacts, the EU-funded GRAVITATE project' in D. Hookk (ed) Виртуальная археология с воздуха, на земле, под водой и в музее, Proceedings of the 3rd conference on Virtual Archaeology, Hermitage State Museum, St. Petersburg, 265–270.

Hermon, S., Iannone, G. and Amico, N. 2014 'Three-Dimensional Approach to the Documentation and Analysis of Heritage Sites – a Case Study from the Cypriot Cultural Heritage Landscape' *Proceedings of the 1st International Conference on Best Practices in World Heritage: Archaeology, Menorca,* Spain, 27–37.

Hermon, S. and Niccolucci, F. 2017 'Formally defining the time - space – archaeological culture relation in archaeology - problems and prospects', *Archeologia e Calcolatori* 28(1), 93–108.

Hermon, S., Niccolucci, F., Yakoupi, K., Kolosova, A., Iannone, G., Faka, M., Kyriacou P. and Niccolucci, V. 2013 'Documenting Architectonic Heritage in Conflict Areas. The case of Agia Marina Church, Derynia, Cyprus,' in R. Gabaglio and D. Gulotta (eds) *Heritage 2013 - Monitoring Conservation and Management,* Milano, 800–808.

Mortara, M., Pizzi, C. and Spagnuolo, M., 2017 'Streamlining the preparation of scanned 3D artifacts to support digital analysis and processing: the GRAVITATE case study' *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage,* Eurographics Association, 165–169.

Niccolucci, F., Hermon, S. and Doerr, M. 2015 'The Formal Logical Foundations of Archaeological Ontologies' in J.A. Barceló and I. Bogdanovic (eds) *Mathematics and Archaeology,* CRC Press, 86–99.

Niccolucci, F., Beacham, R., Hermon, S. and Denard, H. 2010 'Five Years After: The London Charter Revisited' in A. Artusi, M. Joly, G. Lucet, A. Ribes and D. Pitzalis (eds), VAST 2010 The 11th International Symposium on Virtual Reality, Archaeology, and Cultural Heritage, Eurographics Association, 101–104.

Niccolucci, F., Felicetti, A., Hermon, S. and Nys, K. 2009 'Managing Full-text Excavation Data with Semantic Tools' in K. Debattista et al. (eds) *VAST 2009 The 10th International Symposium on Virtual Reality, Archaeology, and Cultural Heritage,* Eurographics Association, 125–132.

Niccolucci, F. and Hermon, S. 2016 'Expressing Reliability with CIDOC CRM', *International Journal on Digital Libraries, Special Issue on Networked Knowledge Organization Systems (NKOS)* 18(4), 281–287. doi:10.1007/s00799-016-0195-1

Niccolucci, F. and Hermon, S. 2015 'Representing Gazetteers and Period Thesauri in four-dimensional space-time', *International Journal on Digital Libraries, Special Issue on Networked Knowledge Organization Systems (NKOS)* 17(1), 63–69

Pitzalis, D., Niccolucci, F., Theodoridou, M. and Doerr, M. 2010 'LIDO and CRMdig from a 3D cultural heritage documentation perspective' in A. Artusi, M. Joly, G. Lucet, D. Pitzalis and A. Ribés (eds) *Proceedings of the 11th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST 2010.* Eurographics Association, 87–95.

Phillips, S.C., Walland, P.W., Modafferi, S., Dorst, L., Spagnuolo, M., Oldman, D., Tal, A., Shimshoni, I. and Hermon, S. 2017 'GRAVITATE: Geometric and Semantic Matching for Cultural Heritage Artefacts' in C.E. Catalano and L. De Luca (eds) *Proceedings of the 14th Eurographics workshop on graphics and cultural heritage.* Eurographics Association, 199–202. doi:10.2312/gch.20161407

Scalas, A., Vassallo, V., Mortara, M., Spagnuolo, M. and Hermon, S. 2018 'Shape analysis techniques for the Ayia Irini case study' in R. Sablatnig and M. Wimmer (eds) *The 16th EUROGRAPHICS Workshop on Graphics and Cultural Heritage.* Eurographics Association, 255–8.

Ronzino, P., Hermon, S. and Niccolucci, F. 2012 'A Metadata Schema for Cultural Heritage Documentation' in V. Capellini (ed) *Electronic Imaging & the Visual Arts: EVA 2012 Florence,* Firenze University Press, 36–41.

Ronzino, P., Niccolucci, F., Felicetti, A. and Doerr, M. 2016 'CRMba a CRM extension for the documentation of standing buildings', *International Journal on Digital Libraries*, 17(1), 71–78.

Sugimoto, G., Felicetti, A., Perlingieri, C., and Hermon, S. 2007 'CIDOC CRM Spider: Stonehenge as an Example of Semantic Data Integration' in D. Arnold, S. Chalmers and F. Niccolucci (eds) *Future Technologies to Empower Heritage Professionals, VAST 2007,* Eurographics Association, 47–54.

Thompson, E. M., Biasotti, S., Sorrentino, G., Polig, M., and Hermon, S. 2018 'Towards an Automatic 3D Patterns Classification: the GRAVITATE Use Case' in R. Sablatnig and M. Wimmer (eds) *GCH 2018 - Eurographics Workshop on Graphics and Cultural Heritage,* Eurographics Association. https://doi.org/10.2312/gch.20181372

Vassallo, V., Athanasiou, E., Hermon, S., and Eliades, I. 2013 'Publishing Cultural Heritage content for Digital Libraries: the case of the collections of the Byzantine Museum and Art Gallery of the Archbishop Makarios III Foundation' in A. Addison, G. Guidi, L. De Luca, and S. Pescarin (eds) *Digital Heritage International Congress,* Marseille, IEEE, II, 479–484.

Vassallo, V., Sorrentino, G., Gasanova, S., and Hermon, S. 2018 'Integrating Analytical with Digital Data in Archaeology: Towards a Multidisciplinary Ontological Solution - The Salamis Terracotta Statues Case-Study' in M. Matsumoto and E. Uleberg (eds) *CAA2016: Oceans of Data.* Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology. Oxford: Archaeopress, 109–118.

Portuguese cultural heritage data access tools and policies – anticipated impacts of ARIADNEplus

Maria João Correia*, António Santos Silva**, Filipa Neto and Miguel Martins*****

LNEC: National Laboratory of Civil Engineering, Lisbon, Portugal
*mjmcorreia@lnec.pt
**ssilva@lnec.pt

DGPC: Directorate-General for Cultural Heritage
***fneto@dgpc.pt
****adolfo.miguel.martins@gmail.com

Abstract: *The management, protection and enhancement of cultural heritage in Portugal, according to current policies, is ensured by DGPC which is also responsible for managing various information systems (Ulysses and Endovelico for archaeological assets). Another system, DB-HERITAGE managed by LNEC, records the history, properties and performance of construction materials, thus integrating fundamental information for the management of heritage assets. The anticipated impacts of ARIADNEplus include greater awareness of useful strategies and best practice for data management, and further opportunities for consolidating a comprehensive, active and informed data-sharing community.*

Keywords: cultural heritage protection; Portuguese information systems; heritage assets management

The Portuguese law nº 107/2001 of 8 September 2001 established the current policy for the protection and enhancement of cultural heritage in Portugal. In addition, it established systematic and up-to-date compliance with inventory, inspection and prevention, avoiding the degradation or loss of cultural assets, as well as promoting systematic data collection and providing access to citizens and interested bodies. The Directorate-General for Cultural Heritage (*Direção Geral do Património Cultural – DGPC*) ensures the management, safeguarding, enhancing, conserving and restoring of assets integrating immovable, movable and intangible cultural heritage.

According to specific policies, different departments of the DGPC are responsible for organising and updating various information systems for cultural heritage, ensuring that information is made available under the rights enshrined in the Portuguese Constitution and compliant with the provisions for the protection of personal data.

Generally, the legal protection of cultural heritage is based on its classification and inventory. In particular, archaeological assets, from archaeological works or random finds including nautical and underwater assets, are inventoried and listed through

Figure 1: Illustration of the Archaeologist's Portal

the creation of national databases. These are *Ulysses* for listed monuments and sites, and *Endovelico* for archaeological heritage, both linked to a GIS.

The Portuguese archaeological database – *Endovelico* – was specially designed for the upload of data collected during archaeological activity. This database has constantly been optimised and updated by the DGPC. The entries in *Endovelico* have their provenance from archaeological field reports. In this way, research projects, commercial archaeology activity and bibliographic information contribute to wider knowledge of Portuguese heritage. This data is stored under specific protocols on the DGPC servers and can be partially accessed online via *Portal do Arqueologo*.[1]

A few known limitations have been acknowledged, including a lack of data for archaeological sites located in the islands of Azores and Madeira which have a regional and autonomous government.

Despite only 80% of archaeological sites being geo-referenced, Endovelico is an efficient and integrated digital platform providing an up-to-date tool for the protection of the archaeological heritage. The data is then linked to GIS software and used for strategic planning of land usage, implementation of mitigating measures, research, public awareness and outreach, conservation and environmental impact assessment purposes.

Included in Endovelico is a section dedicated to intertidal and underwater archaeological sites which is being updated and improved. The reason for dividing the database into terrestrial and underwater archaeology is related to the need for dedicated database fields to obtain a comprehensive end-product. In addition, this section also includes shipwrecks and artefacts located in international waters.

1 http://arqueologia.patrimoniocultural.pt/?sid=sitios

Figure 2: Geoportal search on the Archaeologist's Portal

To improve the database a set of fields are being added for bioarchaeological data from ancient cemeteries and burial grounds found through archaeological fieldwork. Moreover, it is expected to include other areas of archaeology such as dendrochronology, zooarchaeology, as well as artefact storage location and more. This will widen knowledge about surviving artefacts and their archaeological context and therefore increase efficiency for archaeological heritage management.

Reports and bibliographic information are also included in Endovelico as references so as to provide relevant literature and supporting information. Access to immovable heritage and associated research data is provided through web geoportal platforms, namely the Archaeologist's Portal for archaeological data and the Listed Property Heritage Atlas for monuments which are listed or in the process of being listed. Public disclosure of information about works and interventions on cultural assets is mandatory by Portuguese Law (Law 140/2009 of 15 June and decree–Law 164/2014 of 4 November). However, its systematic public dissemination is dependent on the human and financial resources of the central administration body (DGPC), which with some effort maintains up-to-date inventories with the latest information gathered from reports and technical sources.

The databases run by DGPC are useful tools:

- to support policies, strategies and intervention actions in the territorial management, urban qualification and rehabilitation, and the safeguarding and enhancement of the architectural and archaeological heritage;
- to promote the production and collection, processing and conservation, access and use of more and better information on architecture and on architectural, urban and landscape heritage.

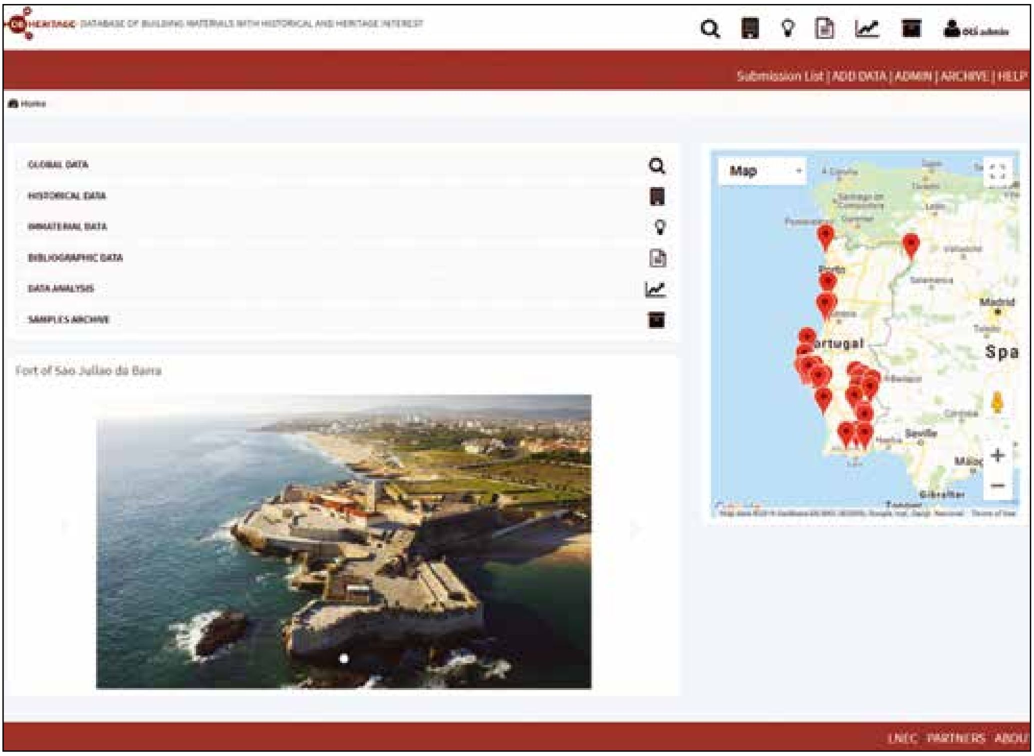


Figure 3: DB-HERITAGE database frontend

Another tool available in Portugal is the DB-HERITAGE database, which provides systematic recording of data on the history, properties and performance of building materials.[2] Material characterisation and condition assessment are fundamental for intervention planning, and for choosing materials and techniques for the conservation and restoration of heritage assets. In practice, there is a large amount of scattered information about performance of materials, currently at risk of being lost, but which is essential for modelling and cost/risk/safety analysis and for supporting informed decision-making. Documenting and preserving data about materials is important not only for improving research and heritage management but also for providing social benefits related to educational and historical value.

The DB-HERITAGE system, managed by the National Laboratory for Civil Engineering (*Laboratório Nacional de Engenharia Civil* – LNEC), archives data about the source, use, dates of interventions, material properties, environmental exposure conditions, and assessment details, and integrates a management tool for the physical repository of samples of materials. However, data is uploaded and owned by different Portuguese bodies. All authorised contents, as defined by the data owners, are freely available to the general public, but editing is only available for registered users.

[2] http://dbheritage.lnec.pt/

Despite the various information systems related to cultural heritage currently available in Portugal, there are noticeable problems in getting the most out of FAIR data. In order to improve scientific research and the management of heritage assets (both for establishing policies and strategies and planning interventions) it is essential to increase awareness of the benefits of data-sharing.

Portuguese participation in ARIADNEplus should provide greater awareness of useful strategies for data, such as the application of the FAIR principles, and help share best practice regarding the management and sustainability of repositories. Contributing to the searchable ARIADNE catalogue by adding Portuguese terms, and improving multilingual vocabularies, should also enable queries in Portuguese. It is also envisaged as an opportunity to consolidate a comprehensive, active and informed data-sharing community, dedicated to improving the effectiveness of research, re-using the research results, and producing savings.

CENIEH: A relevant source of digital paleoanthropological datasets for ARIADNEplus

Mohamed Sahnouni*, M. Isabel Sarró Moreno**, Cecilia Calvo Simal*

National Research Centre on Human Evolution (CENIEH), Burgos, Spain
*mohamed.sahnouni@cenieh.es
**misabel.sarro@cenieh.es
***cecilia.calvo@cenieh.es

Abstract

CENIEH is a Spanish acronym for Centro Nacional de Investigación sobre la Evolución Humana *(in English: National Research Centre on Human Evolution), based in Burgos, Spain. Its mission is twofold: (1) Conducting world-class multidisciplinary research in Paleoanthropology and Quaternary Sciences, focusing on human biological and cultural evolution worldwide; and (2) Providing services to the field of Social and Quaternary Sciences, as well as to the technological and Industrial communities at large. The research is carried out through three main programs: Geology and Geochronology; Hominid Paleobiology, and Archaeology. To fulfill its mission, the Centre is equipped with state-of-the-art laboratories, including: Digital Mapping and remote sensing, Geology and Geochronology (Sediment analysis, Archaeomagnetism, Cosmogenic, U-Series, Luminescence, Electron Spin Resonance (ESR)), Archaeometry, Microscopy, Micro-Computed Tomography, Experimental Archaeology, Use-wear, Referential and Archaeological lithic collections, and hominid and animal fossil bone collections.*

As part of ARIADNEplus, CENIEH has the potential to provide a full range of datasets in Paleoanthropology generated through its research. These include, for example, in situ *Gamma dose spectrometry measurement pertaining to ESR and Luminescence dating, data generated from automated analysis of sediment and rock cores such as magnetic susceptibility and quantified colour of sediments, micro-computed tomography data, and 3-D scan images and datasets of lithic and fossil bone collections. As an ARIADNEplus partner, CENIEH will therefore enhance the European digital archaeological infrastructure with valuable paleoanthropological and geochronological datasets covering early prehistory, as well as disciplines and regions that were previously unrepresented.*

Keywords: paleoanthropology; heritage; datasets; ARIADNEplus; CENIEH

Introduction

CENIEH, Centro Nacional de Investigación sobre la Evolución Humana (in English: National Research Centre on Human Evolution), is based in Burgos, Spain. It is open to scientific and technological use by the international scientific and technological community. Its research activities are mainly focused on human evolution during the Late Neogene and Quaternary, and include collaborative projects at excavations and geological deposits worldwide. Special attention is paid to reaching society through awareness raising and knowledge transfer. In addition, CENIEH is responsible for the conservation, restoration, management and recording of archaeological and paleontological collections, in particular from the Atapuerca sites (Burgos).

CENIEH was included in ARIADNEplus as part of the overall goal of the new project summarized as *'Extending and Focusing ARIADNE'.* Extending includes chronological and geographical data covering the full timespan of human presence on Earth around the world. Focusing includes multidisciplinary coverage as well as the results of scientific analyses, such as material sciences, dating, and so on. Specifically, the Centre reinforces knowledge and information with reference to Quaternary Science, Geochronology, Archaeology, Paleoanthropology and Human Evolution.

The success of ARIADNE is summarized in a statement from the ESFRI (2016) Roadmap.[1]

> *The enthusiastic reviews of these initiatives testify the success of their action to advance knowledge and to establish a research community, acknowledged as "advanced" in official EU documents concerning conservation, or quickly growing in the field of archaeology as shown by the performance indicators of the relevant project ARIADNE. [...] In the archaeological sciences the ARIADNE network developed out of the vital need to develop infrastructures for the management and integration of archaeological data at a European level. As a digital infrastructure for archaeological research ARIADNE brings together and integrates existing archaeological research data infrastructures so that researchers can use the various distributed datasets and technologies.*

One of the characteristics of Archaeology and Paleoanthropology is that they are transdisciplinary; for that reason, sometimes information is difficult to find or is distributed across different sources. The participation of CENIEH in ARIADNEplus will facilitate the integration of a large amount of data and information covering Archaeology, Geochronology, Archaeometry and Cultural Heritage Conservation, thanks to the information obtained through our research projects and the data obtained in our laboratories.

CENIEH Research

Research constitutes the core activity of CENIEH. Research carried out at the Centre is of a markedly interdisciplinary nature and is organized in three programs: Archaeology, Geochronology and Geology, and Paleobiology, combining different scientific fields and approaches. Our research focus is on human evolution during the Late Neogene and Quaternary, but the Centre also promotes awareness and knowledge transfer for society and encourages and supports collaboration in conducting excavations of deposits of these periods worldwide. These sites give us information about the first

1 https://www.esfri.eu/esfri_roadmap2016/roadmap-2016.php

traces of early hominids and their culture, as well as the most ancient tools or rock art. This evidence represents the link to our ancient cultural past and they probably represent the link between our species (*Homo sapiens*) and other human species. For this reason some of these sites are known as exceptional and universal Cultural Heritage sites, and are included in the UNESCO World List.

Due to transdisciplinarity in cultural heritage research, CENIEH represents an extraordinary instrument for knowledge transfer and research applied to conservation, as the objectives of the National Research Plan state the following:

- To increase historical, archaeological, artistic or anthropological knowledge applied to conservation, mostly obtained through international programs and interdisciplinary research projects.
- The improvement of storage conditions by identifying materials and diagnosis of the conservation status and factors influencing degradation; the proposal of new materials and techniques for restoration and protection; and the design of suitable systems related to storage, transport, packaging and exhibition.

The limited number of resources devoted to research in heritage conservation is insufficient to address the many aspects required, but CENIEH tries to combine efforts in research on human evolution with conservation and cultural heritage. The Centre´s excellent facilities give us the opportunity to combine all these aspects, integrating artistic-historical-scientific studies with archaeological diagnosis.

Singular Scientific and Technical Infrastructures

CENIEH has been awarded the ICTS label (Singular Scientific and Technical Infrastructures) by the *Ministerio de Economía y Competitividad (MINECO)* of Spain. The aim is to test and develop technologies related to scientific domains. That is why CENIEH-ICTS hosts the most modern technologies for the development of high quality research, and its laboratories are open to the entire scientific community. Technological experiments, support services and working spaces have been equipped to achieve the highest standards of research and services for industry and society.

CENIEH is also a member of ESFRI (the European Strategy Forum on Research Infrastructures), a network devoted to develop the scientific integration of Europe and to strengthen its international outreach, and it is partner in E-RIHS (the European Research Infrastructure for Heritage Science)[2], an infrastructure that supports research on heritage interpretation, preservation, documentation and management. The competitive and open access to our infrastructure supports and benchmarks the quality of the activities of European scientists, and aims to attract the best researchers from around the world. Of course, some of the CENIEH-ICTS activities include cultural heritage research and its equipment offers numerous applications to this field. ARIADNEplus is another way to give added value to the ICTS, especially to provide services for the Social and Quaternary Sciences and Palaeoanthropology communities.

[2] http://www.e-rihs.eu

CENIEH's laboratories and its datasets

CENIEH includes infrastructures in the field of geological and material characterization, geochronology, an area devoted to fossil collection, restoration, and preservation and other support laboratories.

Digital mapping and 3D analysis

The activities of this laboratory are focused on cartography, reconstruction, and spatial analysis of geological and archaeological elements. Among its products, the following are particularly noteworthy: geological and geomorphological cartography, paleogeographic reconstructions, analysis of spatial patterns, 3D reconstructions, and the morphometric analysis of archaeological sites and objects for research projects in Europe and Africa.

These activities, which take place both in the field and inside caves, can cover temporal sequences. Thus, the digitization, reconstruction and analysis of natural and cultural heritage can also be approached from a 4D perspective.

The laboratory has high-precision topographic equipment such as total stations, GPS/GNSS systems, photogrammetry systems, drones and spherical photography, and laser scanners for use in the field and in the laboratory. This equipment is complemented by the necessary hardware and software for data processing and analysis, such as Geographic Information Systems (GIS), tele-detection and 3D applications.

The types of data generated in this laboratory comprise photos (jpg; tiff), shape files (shp), topographic and cartographic outputs (xls; shp; txt), vector outputs (shp; CAD), raster (tiff; geotiff), photogrammetry and 3D scanner outputs (stl; avi; tiff; raw), and high precision GPS files (txt; xls; CAD).

Material Characterization Laboratories

These laboratories are equipped for archaeometry, microscopy, and micro-computed tomography. The main objective of the *Archaeometry Laboratory* is the characterization of materials, based on the study of their chemical and mineralogical composition, and their vibrational and thermal analysis. It seeks to support research in the areas of geology, archaeology, heritage, chemistry, and pharmacy, as well as to provide services in the fields of construction, civil engineering and nanomaterials. It is fitted with equipment to conduct non-invasive analyses, without the need for the prior physical and chemical preparation of the sample, and without altering the surface on which the analysis is carried out. There is also portable equipment for field use.

This laboratory has various analytical techniques for the characterization of solids, such as: i) x-ray crystalline powder diffraction, ii) x-ray fluorescence spectrometry, iii) Raman spectroscopy, iv) infrared spectroscopy and v) thermal analysis.

The Microscopy and Micro-Computed Tomography Laboratory boasts a wide range of equipment enabling the microstructural and elemental characterization of all types of materials (for example, biological, organic and inorganic materials), with practice-oriented analysis, quality control and basic and applied research.

Microscopy area: This laboratory has different types of microscopes, making it possible to distinguish four main areas within the field of Microscopy: i) optical, fluorescence and metallographic microscopy; ii) particle analysis; iii) confocal laser microscopy; and iv) scanning electron microscopy. This wide range of scientific equipment makes it possible to study multiple parameters for different types of materials: from the analysis and automated classification of particles to quality control of electronic components.

Micro-computed tomography (microCT) area: a microCT (model V|Tome|X s 240 by GE Sensing & Inspections Technologies) provides high-resolution assessments of the density, geometry and microarchitecture of mineralized tissues, such as bones and teeth, calcification as a result of pathology, or soft tissues and biomaterials stained with radiographic contrast media and allows us to obtain 3D models. The image analysis software used to process the images obtained with the MicroCT provides a powerful tool with which to study different parameters in the materials.

The type of data generated in these laboratories are photos in optical, confocal and electron scanning microscopes (jpg; tiff), analytical information (SEM/EDX-WDS) (elemental composition, etc. in pdf, xls, tiff files), and 3D models (jpg; tiff; stl; avi; txt; xls CAD).

Geochronology Laboratories

The skilled technical staff combined with the most advanced technology available allow CENIEH's scientists to determine the age of rocks and other materials, dating important events in geological and biological history. The research interests of the Geochronology Program include the establishment of the time framework for human evolution and its interaction with the environment. A great deal of effort is put into methodological advancements. Together, the techniques are capable of providing absolute or relative chronologies spanning the entire Quaternary period (0 – 2.6 Ma) and beyond.

Through understanding such information in its geological context, our research provides key insights into such processes as continental drift, volcanism, mountain building, mass extinctions, climate change, and the evolution of humankind itself, cultural heritage authentication and conservation. Geochronology is critical not only for establishing the time axis along which these various phenomena occur, but also for determining the relationships among them.

These laboratories are equipped for archaeomagnetism, electron spin resonance, luminiscence, and uranium series dating methods. The coordination and collaboration between the laboratories and the development of new methodologies has been an essential tool in cultural heritage investigations with applications that cover the span of structural geology, tectonics, stratigraphy, paleontology, petrology, paleoclimatic and geochemistry studies.

The *Archaeomagnetism* facility boasts all of the instruments necessary for the study of the magnetic properties of rocks in order to develop magneto-stratigraphies in sediments. This enables the determination of the magnetic remanence, demagnetization, IRM and ARM in sediments, rocks and archaeological artefacts and also makes it possible to obtain hysteresis loops, FORC diagrams and susceptibility and anisotropy. The 2G magnetometer, the laboratory's main piece of equipment, enables the automatic measurement of standard samples and 'u-channel' probes of up to 150 cm in length, as the basis for chronostratigraphic studies.

Electron Spin Resonance dating is mainly devoted to dating geological and/or archaeological materials such as tooth enamel or optically bleached quartz grains extracted from sediment. Occasionally, it also carries out work in dosimetry and characterization of modern or old geo-materials.

Luminiscence dating is based on the ability of certain minerals (quartz and feldspar) to accumulate electrical charges within their mineral structure and to release such light energy charges when they are submitted to an external stimulus. According to the type of external stimulus applied, there are different types of luminescence, TL, OSL, IRSL, depending on whether the external source is heat, a visible light source or infrared.

Finally, the *Uranium Series* facility conducts research work and implement services involving elemental and isotopic analyses of different types of material.

The type of data generated in these laboratories are photos in optical from fieldwork collected as photos, Gamma and beta spectrometry (cnf; tka) (xls; txt; dat), Software Genie 2000, or elemental and isotopic information of sediments, and dating ages for sediments, speleothems, teeth and bones.

Collections, Conservation and Restoration Area

One of the roles of CENIEH is the curation of collections stored as the result of excavations at archaeological and paleontological sites. It is hoped that such collections will steadily expand with the entry of more and more archaeological and paleontological remains from different excavations, particularly those linked to projects carried out by the centre's teams.

It therefore boasts facilities that are suitable for the storage of a large volume of remains, ensuring optimal conditions for their long-term curation. It also has a modern database system which has been specifically designed to catalogue its collections and for the management of the associated contextual information, including the identification of pieces through RFID labels and barcodes. The collection management system has been designed to facilitate the study of the collections by researchers and to provide them with contextual information. CENIEH's collections are:

The *Comparative Anatomy Osteological Collection* (COAC) includes specimens belonging to living species and also casts of Quaternary fossils, particularly casts of human fossils from all geographical regions and time periods. Osteological collections are essential reference material for palaeontological and zooarchaeological studies. The collection is constantly growing. At the present time 592 specimens are included; 272 of which are complete skeletons and the remaining specimens are skulls, incomplete skeletons and single skeletal parts. Natural bone specimens from modern species represent 50% of the collection while the other half corresponds to casts of fossil specimens and high-quality replicas of crania from recent species.

The non-human mammals section, currently consisting of 106 items, includes specimens from recent species and casts of Pleistocene fossils. The aim of the collection is to include at least one adult male and one adult female skeleton for each recent species included, in addition to one or more immature individuals of different ages. The ornithological section aims to include skeletons of an adult male, an adult female and an immature individual of every species currently present in the Iberian Peninsula. Presently, the ornithological section includes 162 skeletons.

The Anthropological section, with 234 specimens, includes casts of skeletons of modern male and female humans of various ages, and casts of fossil hominin specimens from Europe, Africa and Asia. It is actually composed of 81 skull casts (complete or fragmentary), 9 casts of complete skeletons and 144 casts of isolated skeletal elements.

Another reference collection has been created to answer the growing necessity of identifying and classifying the lithic raw materials used in European and African archaeological sites that are currently under research by CENIEH. The *Mineral Collection (LITHO),* incorporating physical samples of a wide array of raw materials (flint, quartzite, basalt, rhyolite, trachyte, etc.), is under development with the purpose of accommodating petrological samples of materials that have been used or that may have been used to create stone tools during the Palaeolithic and in recent prehistory. The objective of this mineral collection is to become a reference collection, including thin sections and mineralogical an elemental composition information, to serve as support to research into stone artefacts. These materials have been classified and documented, and each one incorporates basic data (source coordinates, geological provenance, etc.), photos and thin-sections. Some of them have also associated geochemistry data and 3D models.

The *Experimental Traceology Collection (CET)* is also under development, bringing together tools made from different raw materials (mainly stone and bone) which have been used on an experimental basis in different activities related to prehistoric ways of life. The objective of the CET is to provide experimental models in order to identify signs that the stone materials, analyzed from a traceological perspective, have been used.

The *'Ratón Perez' Collection* is based on one of the most abundant and well preserved remains: the teeth of ancient and recently deceased individuals. This is often the most important element from which to recover useful information (e.g. species, age, sex, stress events, diet) that may help to identify the individual they belong to. Each individual possesses unique dental characteristics (shape, size, morphology, pathologies) that may be used in different fields for different purposes. In forensic odontology the identification of unknown human beings has been generally achieved by comparing the dental evidence of the deceased with ante-mortem dental records (e.g. medical history, x-rays, dental biometrics).

Since most of the time the deciduous teeth are discarded, our purpose was to collect them to form a large reference collection that could be used by scientists from several disciplines. Thus the main aim was to gather the maximum amount of data, not only from the collected teeth, but also important information about the donor and their relatives. Although the sample is still being increased, from the first collection campaign in 2014 to date, we have gathered more than 1200 teeth of children whose ages of tooth loss are between 4 and 12 years. Each tooth is associated with basic information about the individuals and their ancestors (sex, date and place of birth, dwelling place), as well as with important data about early life history (pregnancy duration, type of weaning) and other relevant information that the volunteers could provide. All these details and pictures will be shown in the database.

As not many deciduous teeth samples are available, the 'Ratón Pérez' collection represents a unique collection for the study of variables useful in a wide range of disciplines such as forensic, dental and anthropological fields. We want to make the research community aware of the existence of this collection.

Finally, the *Sediments Collection* includes screen-washed sediments from the Sierra de Atapuerca sites. The sites have been systematically excavated since the 1980's. All the sediment is screen-washed and sieved for the recovery of microfossil remains. The product of the wet-sieving process is a concentrate consisting of calcareous fragments from the cave walls, fossil remains of small vertebrates, and fragments of bones of large vertebrates. These concentrates are packed in plastic bags labeled to indicate their stratigraphic origin. Roughly, 12-15 tons of sediments are processed in a single campaign, giving rise to several hundred bags of concentrate (Cuenca-Bescós et al. 2015). Only a small number of the thousands of bags produced in the last decades have been surveyed by the Atapuerca research team, to obtain a representative sample of the small fossil remains from each stratigraphic unit. CENIEH's Sediments Collection comprises the non-processed bags of concentrate from 1988 to 2010. It currently includes 7,688 bags of concentrate, mainly from the Gran Dolina site, although materials from Galería, Covacha de los Zarpazos, Sima del Elefante, Sima de los Huesos, Trinchera Penal, Portalón de Cueva Mayor and El Mirador are also present.

Planned objectives

As ARIADNEplus partner, CENIEH will share its new Database (DB) System to the ARIADNEplus infrastructure. This System has been funded by ERDF (European Regional Development Fund) and is being designed to provide an integrated solution for the management of the palaeontological and archaeological collections housed at the Centre. It is intended to provide easy access to all the information for authorized users, whilst allowing the departments and laboratories of the Collections Area to keep control of the information generated by themselves. As many user profiles as desired may be created, assigning functions and privileges according to role (restoration staff, collection technician, curator, scientific researcher, etc.).

The DB will record all the information regarding collection items: catalogue data, technical description, precise location in the stores, conservation state etc. In addition, the DB will allow the staff of the Collections Area to track the entire 'history' of each item from the moment it was unearthed to the present, and it stores electronic copies of all related documents (from facility reports to scientific publications). The Restoration and Conservation Area staff uses the DB to record and document all their interventions and to manage and organize their work. Furthermore, the DB incorporates powerful tools to manage researcher access as well as any movement of the items either inside (to the Restoration and Conservation Area, to the studio of photography, etc.) or outside the CENIEH facilities (temporary exhibitions, etc.).

CENIEH's collections (COAC, LITHO, CET, Ratón Pérez, and Sediments) together with research results, laboratory data, and scanned items from reference collections (MicroCT and Laser scanner) will become findable via ARIADNEplus (*Figure 1*). In addition, the DB is complemented by other laboratory data that can be shared, such as the results of elemental or mineralogical analysis of different materials, 3D scans, radiographies, RAMAN spectrometry databases for pigments, gamma spectrometry measurements in relation to GPS information, etc. These are currently in development to be shared as soon as the information is organized in compatible datasets.

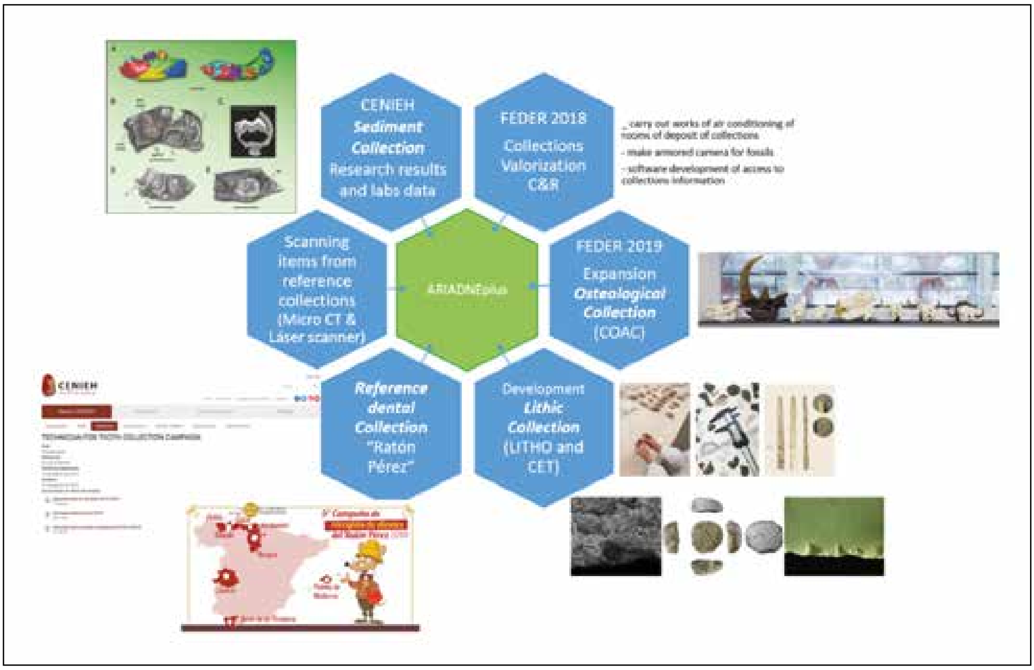


Figure 1: Types of data generated at CENIEH that will be shared with ARIADNEplus infrastructure

Conclusion

As an ARIADNEplus partner, CENIEH will enhance the European digital archaeological infrastructure with valuable paleoanthropological and geochronological datasets, covering early prehistory, as well as disciplines and regions that were unrepresented in the previous ARIADNE infrastructure.

Currently CENIEH, together with a research group of the University of Burgos, is working on the design and deployment of the infrastructure, and on the guidelines which will describe the way to share the digital data and its limits. The elements being considered include the data access policy; the data re-use policy; open data: its organization and management; the identification of data and the classification by typology, origin and format; and the use of a common language. These guidelines are based on the FAIR (Findable, Accessible, Interoperable and Reusable) principles.

The value of the research data is not only scientific, they are key for economic and innovation progress too. To improve knowledge transfer, our Centre publishes multiple papers each year, takes part in numerous national and international projects and is present in multiple excavations around the world. ARIADNEplus provides us with a new way for organizing and sharing our data and information.

Acknowledgments

For the CENIEH team, special thanks to Jesús Rodríguez and Joseba Rios of the Collections area, and to Marina Martinez de Pinillos of the CENIEH's Palaeoanthropology research line.

References

Barros, M. A. and Sarró, I. 2013 'Laboratorios de vanguardia para los estudios geológicos, arqueológicos y geocronológicos en el Centro Nacional de Investigación sobre la Evolución Humana', *Tierra y Tecnología* 43, 85–8.

Sarró, I. and Barros, M.A. 2013 'CENIEH: a research centre for human evolution. Innovation and technology for cultural heritage conservation and the advancement of our understanding of human origins', *Coalition* 24, 2–10.

Cuenca-Bescós, G., Blain, H.-A., Rofes, J., Lozano-Fernández, I., López-García, J.M., Duval, M., Galán, J. and Núñez-Lahuerta, C. 2015 'Comparing two different Early Pleistocene microfaunal sequences from the caves of Atapuerca, Sima del Elefante and Gran Dolina (Spain): Biochronological implications and significance of the Jaramillo subchron', *Quaternary International* 389, 148–158.

The ADED project – a Norwegian infrastructure for excavation data

Christian-Emil Smith Ore [*] ***and Espen Uleberg*** [**]

Unit for Digital Documentation, Department of Linguistics and Nordic Studies, Humanities Faculty, University of Oslo.
*c.e.s.ore@iln.uio.no

Museum of Cultural History, University of Oslo
**espen.uleberg@khm.uio.no

Abstract

A major issue of the last 10-15 years has been to rescue, preserve and provide access to datasets from archeological excavations. The EU infrastructure projects ARIADNE and now ARIADNEplus are a driving force in this work. Still there is a huge number of datasets, which have been definitely lost or are not accessible or reusable. For the available datasets there are only weak links between the excavation and the data sets on the one hand and the museum collections (find repositories), site and monument registries and publications on the other. To strengthen the FAIRness of the datasets such links have to be strengthened or at least established.

In Norway a new infrastructure project, ADED (Archaeological Digital Excavation Documentation) was launched in 2018 with the objective of creating a repository for data sets and establishing the aforementioned links. The outcome will be an integrated part of the MUSIT system, a collaboration between the University Museums. In this infrastructure, the CIDOC CRM suite will be applied as semantic glue to facilitate cross-project queries based on geography and metadata as well as linking Norwegian excavation data with archaeological information from other countries in ARIADNE.

So far, most archaeological excavations in Norway have been conducted by the university museums, but in the coming years the counties will conduct more surveys and small excavations. Therefore, ADED will include the data flow from fieldwork to permanent repositories at the museums, Directorate for Cultural Heritage (the sites and monuments register Askeladden*) and the counties.*

Keywords: Norway; ADED; excavation documentation; CIDOC CRM; digitization

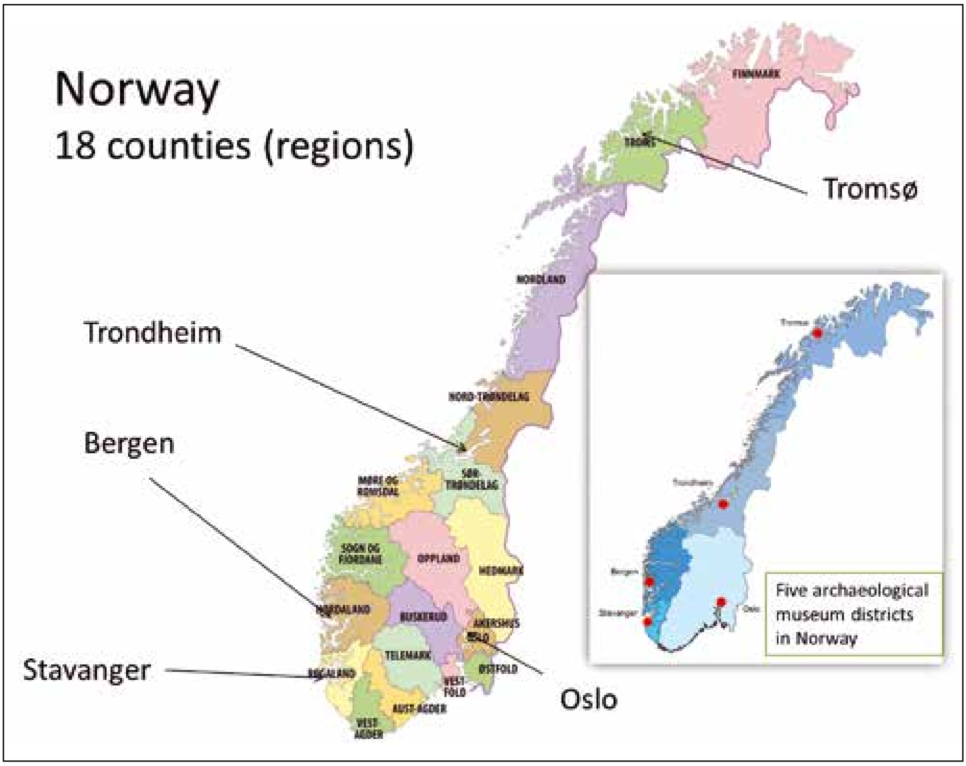


Figure 1: Archaeological museums, museum districts and counties in Norway.
Nord- and Sør-Trøndelag were merged into one county in 2019

The Norwegian archaeological institutions

In Norway there are five archaeological museums: those established in Trondheim[1] in 1775 (connected to the Royal Norwegian Society)[2], Oslo[3] in 1811/1829 (as part of the new university), Bergen[4] in 1825, Tromsø[5] in 1872, and Stavanger in 1877. The museums of Bergen, Trondheim and Tromsø formed the nucleus of the universities founded in the 20th century. At the end of the 19th century, the five museums became responsible for archaeological surveys and excavations in their respective parts of Norway. Today each archaeological museum is a part of a university

Since 1905, the archaeological university museums have had the responsibility to preserve all archaeological objects found during land-based excavations in Norway. The university museums also conduct the archaeological excavations within their museum districts. There are two exceptions; the semi-private NIKU (Norwegian

1 https://www.ntnu.edu/museum
2 http://www.dknvs.no/?lang=en
3 https://www.khm.uio.no/english/
4 https://www.uib.no/en/universitymuseum
5 https://en.uit.no/startsida

Institute for Cultural Heritage Research)[6], excavates in medieval cities and churches, and the county archaeologists undertake excavations related to surveying. In addition, there are two maritime museums in Oslo and Bergen, which are responsible for marine archaeology in their regions.

From its establishment in 1844 until 1909, the National Trust of Norway (*Fortidsminneforeningen*)[7] was the de facto cultural heritage authority and closely collaborated with the museums. In 1909 its responsibilities were taken over by a new governmental Cultural Heritage Authority (*Riksantikvaren*)[8], which developed into the current Directorate for Cultural Heritage.

Until 1990 the five archaeological museums and the Directorate for Cultural Heritage had the responsibility for all land-based excavations and surveys in Norway. This structure implied that the Norwegian university museums together functioned as the distributed repository for archaeological documentation. For example, the Norwegian university museums have topographically ordered archives (TopArk) containing complete and detailed information about sites, monuments, archaeological surveys, excavations, finds and instances of destruction for the full period from the museums' establishment in the early 19th century to 1990. In 1990, the responsibility for the administration of the sites and monuments was delegated to the 18 counties.

As a consequence of the 1990 decentralization, the detailed information on sites, monuments, surveys and excavations is kept in at least 25 different administrative archival systems which also contain information about all kinds of everyday administrative issues. It is difficult to extract all of the information about a given site, survey or protected building, since the object identifier from the sites and monuments registry is not necessarily used as one of the archival keys. One conclusion that can be drawn from this situation is that decentralization of responsibility requires centralization of information maintenance. The specific obligatory rules for local and governmental archives make this a complicated task. The situation in Norway is in many ways comparable to that observed in Sweden (Ore 2017): cultural heritage information is kept but 'lost' in a sea of general administrative information. This problem is rarely discussed in the literature. Still, we believe it to be one that exists in many countries.

Digitization and the museums

The work on digitizing the museum catalogues, archival material and photos started with the Documentation project in 1991 (Aukrust and Hodne 1998; Ore and Kristiansen 1998). It was realized that several institutes and museum collections at the universities had large amounts of digital data that ought to be better available for research and the public. The result was an extensive project at the faculties of humanities at the Norwegian universities in the period 1991 until 1999. The university museums with archaeological and numismatic collections took part in the project, and the first online

6 https://www.niku.no/en/

7 https://www.fortidsminneforeningen.no/about-us

8 https://www.riksantikvaren.no/en/

versions of museum catalogues were published in 1995. The work at the museums continued in the Museum project from 1999, a project that included ethnographic and natural historic collections at the university museums. Towards the end of the project period, it was clear that this work should not be financed through projects, and a permanent organization, MUSIT (MUSIT 2019), was established from 2007.

Over the years, this continuous cooperation has resulted in common database solutions for the university museums, which are used to catalogue all new acquisitions. The system contains authoritative lists for artefact names, finds categories, and materials that the museums in MUSIT have agreed on. The documentation of the artefacts is published online through a common web portal,[9] and as of June 2019 more than 1.4 million entries have been published.

The topographically ordered archives (TopArK) were digitized for the museums in Bergen and Trondheim and the far largest archive, in Oslo, remains to be digitized although the work has been initialized. Buskerud County and the Museum of Cultural History in Oslo have launched a project where they will digitize and publish the archival material pertaining to Buskerud County. This will be available together with archival material pertaining to a few selected areas in Akershus County. The ultimate objective is to digitize all the pre-1990 archives and thereby make the documentation of the activities accessible. As mentioned, the documentation after 1990 is spread over many archival systems and is not always easily accessible. Digitization alone does not imply easy access to the complete documentation. The work process has to be conducted by persons with deep insight into the subject matter, a good overview of the totality and advanced archival competence.

Digitally-born documentation of excavations

In the early 1990s digital documentation of excavations was introduced at a few, mainly larger, excavations. PenMap (from Trimble Navigation) was used mainly as a drawing tool in the 1990s, and there is generally a great variety in attribute registration in the early projects. From 2000 digital documentation has been increasingly used on excavations. The degree of digitizing in the interventions done by the counties is very variable.

The university museums use the common MUSIT system for their collections in general and for artefact cataloguing especially. Until 2011, there was no commonly used system for field documentation. In 2011 the museums decided to use the same program for field documentation, and the Swedish GIS-based system Intrasis (Intrasis 2019) was chosen after an evaluation process. At that time, Intrasis had been used for almost 10 years in Norway starting with the large excavations at Kaupang from 2000–2003 (Pilø 2007), and at some excavations in the following years.

Intrasis is a very flexible framework and can be adapted to most excavation practices. However, the flexibility can create more variation, which makes it more

9 http://unimus.no/

difficult to integrate separate projects in a common database structure. Therefore the museums specified a common set-up and interface to the program, in order to make it possible to combine results from several separate excavations at a later stage.

Even though the Swedish Intrasis system has been the standard tool for documenting excavations since 2011, the backlog of digital excavation data from 1990 and later is a problem in Norway. There is no common repository for archaeological datasets, archival practices vary, and the state of preservation of datasets from 1990 to at least 2005 is unclear.

There are around 150 archaeological excavations performed by the university museums each year in Norway. In addition, there are numerous archaeological surveys conducted by the county archaeologists. The counties also perform minor excavations related to the surveys. Positive results from these interventions, that is, detection of remains of human activities before 1537, are registered in the national registry of protected buildings, archaeological sites and monuments (SMR), *Askeladden.* The registry is maintained by the Directorate for Cultural Heritage. It is used by the county administrations as an important basis for their administration of land development and building proposals, and can only be updated by specially authorized persons in the Directorate, the 18 counties and the university museums. The general public has access through special web-portals.

The SMR *Askeladden* contains information about all *known* protected sites. It is used in early stages of a planning process to assess the potential for unknown sites. The assessment would be better when based on *all* interventions in an area. A future development of *Askeladden* will therefore include registration of negative results, i.e. the extent of every archaeological survey will be recorded. This will not only give a better knowledge base for planning and cultural heritage management, but it can also be useful for landscape archaeology and predictive modelling.

Access to archaeological excavation reports and other grey literature is handled separately by the institutions doing archaeological fieldwork. More than 90% of the fieldwork comprises rescue excavations, and the general rule is that the report shall be ready and published within 18 months of the fieldwork. The Museum of Cultural History in Oslo publishes all its excavation reports at DUO, the Research Archive at the University of Oslo[10]. Excavation projects are registered in a database that is used to create a map-based online interface to the reports. At present, more than 1100 reports have been published, which is almost all excavations from the period 2000–2018 and also a few older excavation reports. The database contains geographic information, coordinates, site-ID, a link to the report in DUO, as well as the ID-number used for images.

Digital images are uploaded to the MUSIT media database, and published at the MUSIT photo portal[11]. The images are published with a Creative Commons license, most of them as CC 4.0 BY-SA, that is, they can be freely used as long as they are

[10] https://www.duo.uio.no/

[11] http://www.unimus.no/foto

supplied with information about the source and they must be open for further sharing (Creative Commons 2019). Geographical information and site-ID are among the recorded metadata. This makes it possible to link the images to other digital information and in this way give the public access to more information.

Public access and dissemination of archaeological data

In Norway, there is a tradition of making archaeological investigations available for researchers and the public. From 1834 to 1842 the museum catalogues and surveys were published in *Urda,* a periodical published by Bergen museum. From 1866 until the beginning of the 20th century, the archaeological museums published their acquisitions in the yearbook of the National Trust of Norway. Larger excavations where published as articles with the catalogue text as part of the article in the same volume as the other acquisitions of that year. This tradition of published museum catalogues continued in the museums' yearbooks, and later in dedicated publications. The archaeological collection at the University of Oslo, *Oldsaksamlingen,* continued this analogue tradition until the publication of the acquisitions of 1999. It has continued as an online publication with weekly updates from the collection management system (Matsumoto and Uleberg 2015; Uleberg and Matsumoto 2019).

The ADED project. Archaeological Digital Excavation Documentation

There is vast information potential in digital assets that have been published in various ways. What is lacking is the structure that will provide well-designed access to this data for researchers and the public. It might be somewhat presumptuous to call this Linked Data, but it can certainly be described as Linkable Data (Ore 2017). The ADED project will contribute to better access to the wide range of digital archaeological information in Norway.

The project was launched in 2018 with the objective to develop a common repository for all digitally-born excavation data in Norway. ADED will be integrated in the national MUSIT system, and will provide a link between the documentation of artefacts and their excavations. ADED will convert existing and future digital documentation of archaeological excavations into a common format compliant with the CIDOC CRM conceptual model and its specialization for archaeology.[12] In addition to the necessary APIs, the repository will be supplied with a web interface for excavation documentation and in this way make it accessible for research, cultural heritage planning and the public.

The Museum of Cultural History in Oslo is the leader of the project. Projects partners are the Directorate for Cultural Heritage, MUSIT and the university museums in Stavanger, Bergen, and Tromsø. The university museum in Trondheim is an associated partner and participates in the development in the project. The project builds on

[12] http://www.CIDOC-CRM.org

the well-established cooperation among the Norwegian university museums. The fact that all museums have used the same documentation system for fieldwork since 2012 gives a good starting point for the ADED project. NIKU (Norwegian Institute for Cultural Heritage Research), which is responsible for excavation in the medieval cities, is not a partner, but is involved in working groups. NIKU also uses the Intrasis documentation system, and delivers the field documentation to the museums together with the finds.

Up to now the counties send their written reports to the museums, but detailed documentation, drawings, images, GIS-data and results from analyses are, as mentioned, archived in in the local systems of the counties. ADED will include this detailed documentation in the common repository. In this way, all results from archaeological activity will have one point of access. To achieve this, the counties will get access to a web interface to upload their results.

It has been argued elsewhere (Ore 2018) that archaeological excavation repositories can be characterized by three levels:

- A repository is a device for safe storage such that one can extract in an unspoilt condition what was originally inserted. The availability of safe data silos for long term, say 100 years, preservation of digital excavation data must be a basic requirement.
- A data set from an excavation corresponds to a book in a digital library or a box of documents in a traditional archive. To find the relevant material, users of libraries and archives are dependent upon a good catalogue with detailed metadata about each archival unit and books. For an excavation archive, this will be detailed information about the excavation, for example: where (coordinates), when, how, what was excavated, and who was responsible. In addition to being a finding aid in a given archive, the metadata from all archives should be accessible via APIs and as linked (open) data.
- In the spirit of the open-the-silos slogan, the elements in the data sets should be made available as linked (open) FAIR data. That is, the data sets have to be opened such that each element of the set, be it a profile, a posthole, a photo or a LIDAR point cloud is searchable and accessible.

Level 1 is a prerequisite for levels 2 and 3 and requires stable maintenance organisations. It corresponds to the focus of the ARIADNE project (ARIADNE 2014). Levels 2 and 3 correspond to the FAIR data principles (FAIR 2019) and are two of the objectives in the follow-up project ARIADNEplus (ARIADNEplus 2019).

The challenge is that a meaningful linking of data (and data sets) requires sufficiently advanced and compatible metadata schemas, (see Oldman et al. (2016) for a detailed discussion and Hyvönen et al. (2016) for a very elegant example of how to use CIDOC CRM as semantic glue). The ARIADNEplus project has chosen the CIDOC CRM conceptual model and its specialization for archaeology, CRMarcheo, as the basic semantic glue. This requires that the participants' data schemas can be mapped to CRMarcheo.

In the ADED project we follow the same strategy. The Intrasis datasets documenting the excavations have to be mapped to a CIDOC CRM/ CRMarcheo compatible format. Tests have showed that this can be done without any problems by a small group consisting of archaeologists, an ontologist and a database expert.

In the ADED project the level 1 repository will consist of xml/json encoded database exports from the Intrasis instances (currently postgreSql/postGis database schemas). Each database dump will be mapped to a dataset in a CIDOC CRM/ CRMarcheo compatible form. These data as well as the datasets as a whole will be imported into a common index system with a GIS component, whether a graph-database or a traditional database. This will constitute levels 2 and 3.

An important objective of the ADED project is to link the excavation data to the relevant data in the MUSIT museum databases (artefact cataloguing data, photo, media and document repositories), the national SMR, *Askeladden,* and maybe other relevant databases. Fortunately, almost 20 years ago it was decided to design the conceptual schema for the MUSIT databases according to the event-centric principles behind the CIDOC CRM (Jordal et al. 2010). Closer studies of the design of the conceptual schema of the *Askeladden* system a few years ago indicate that it is not difficult to map the necessary data to a CIDOC CRM compatible form. One can either design CIDOC CRM/CRMarcheo compatible APIs for federated search or import the necessary data from these sources into the common index system. The latter solution will make a faster system

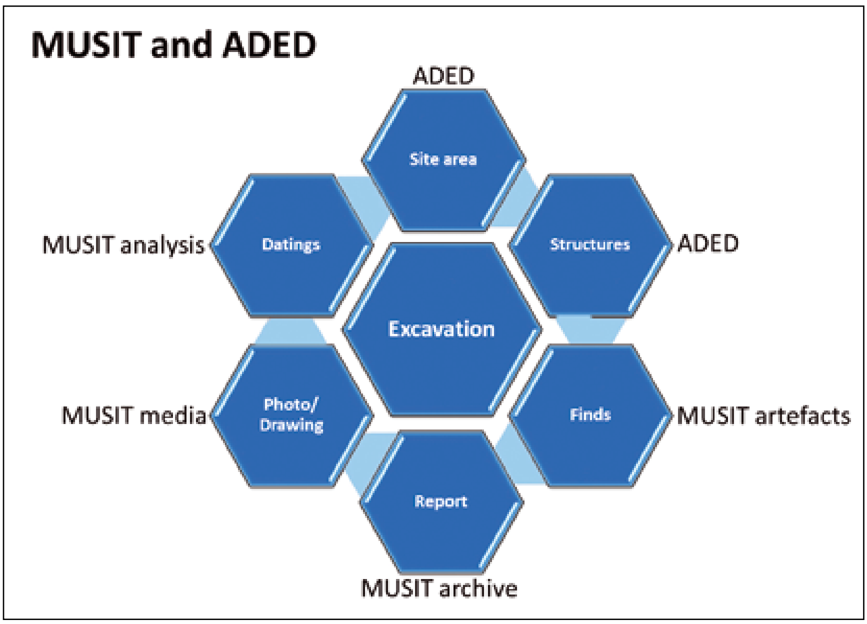


Figure 2: Modules in the MUSIT / ADED system

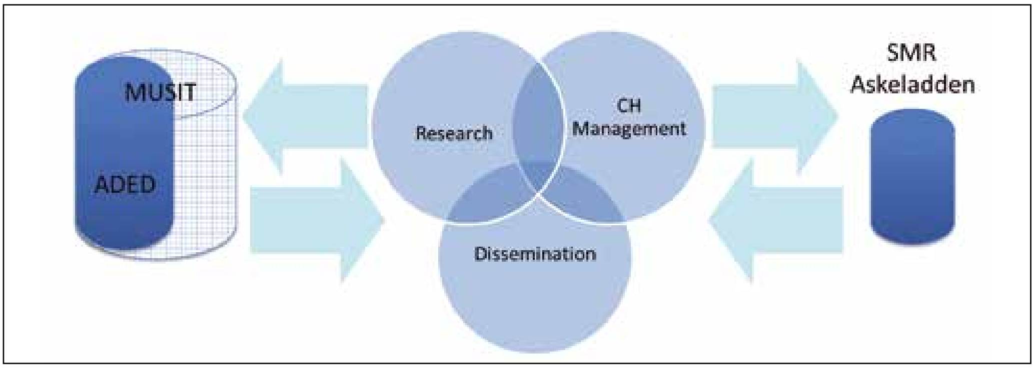


Figure 3: Dataflow in MUSIT/ADED and the SMR data systems. Both will deliver and collect information from the work being done within the fields of Research, Dissemination and Cultural Heritage Management

The results of three decades of digitizing and more and more documentation being born digital are a vast amount of information that can be available as linked data. ADED will facilitate facetted search across single archaeological interventions. It will be possible to list all Iron Age houses in Norway, or all structures dated to the Bronze Age. Some of the basic requirements to achieve this are already in place. One of this is that the MUSIT system allows queries across all the museum collections, e.g. to query for all objects of a specific type from a specified archaeological period. The artefacts are linked to images in the image database. Samples are registered with the artefacts, and the results of analyses are also registered in the MUSIT system. Excavation reports, artefacts as well as images are tagged with the site-ID, and this makes it possible to use this together with the national SMR as Linked Open Data. ADED will provide aggregated detailed excavation documentation, and further develop the linking between the different parts of the MUSIT system. In this way, ADED will create a hub for archaeological information in Norway, and an entry point for international data exchange

It must however be acknowledged that the site-ID is only a weak link between excavation and data sets; the site-ID is not sufficiently precise. One archaeological intervention can cover more than one registered site, and there can be more than one intervention at each site. To achieve this, it will be necessary to use UUIDs in a more precise way than at present. Not only each intervention but also each documented structure needs unique IDs so that the relationship between structure, artefact, documentation and later analyses can be preserved.

Conclusion

The systematic digitalization work and cooperation among the university museums has created a unique opportunity to create interfaces that interlink vast amounts of archaeological information. Relatively few institutions are allowed to do archaeological fieldwork in Norway. The university museums have a long history of cooperating to create database systems, and results of this work are found in web pages that allow nationwide queries.

All artefact descriptions made in the museums as part of the collection management are documented and entered as events in the artefact database. This includes updated terminology and new knowledge on what materials the objects consist of. However, there is a need to have better routines to incorporate or at least link to research results. The databases should be a good tool for research and not just a collection management system.

The ADED project will create a repository for excavation data sets and establish links between these and datasets covering museum collections, sites and monuments registries and publications. This will strengthen the FAIRness of the datasets.

References

ARIADNE 2014. *ARIADNE. The way forward to digital archaeology in Europe.* Available http://www.ariadne-infrastructure.eu/About (Accessed 3 July 2019)

ARIADNEplus 2019. Available https://ariadne-infrastructure.eu/ (Accessed 3 July 2019)

Aukrust, K. and Hodne, B. (eds) 1998 *Fra skuff til skjerm*. Oslo: Universitetsforlaget.

Creative Commons 2019. Available https://creativecommons.org/ (Accessed 3 July 2019)

Engevik jr., A., J. Holmen, S. Innselset and J. Stabell 2004. 'Digital Archaeological Resources at the University of Bergen: an Efficient Tool in Research and Heritage Management?' in Fennema, K. and Kamermans, H. (eds) *Making the connection to the Past. CAA99. Computer Applications and Quantitative Methods in Archaeology: Proceedings of the 27th Conference, Dublin, April 1999*. Faculty of Archaeology, Leiden University, Leiden, 43–47. Available https://proceedings.caaconference.org/files/1999/06_Engevik_et_al_CAA_1999.pdf (Accessed 3 July 2019)

FAIR 2017. https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf (Accessed 3 July 2019)

Holmen, J. and Uleberg, E. 1996 'The National Documentation Project of Norway—the archaeological sub-project' in H. Kamermans and K. Fennema (eds) *Interfacing the past. Computer Applications and Quantitative Methods in Archaeology CAA95 Vol. I. Leiden: University of Leiden, 43–46. Analecta Praehistorica Leidensia* 28. Available https://proceedings.caaconference.org/files/1995/06_Holmen_Uleberg_CAA_1995.pdf (Accessed 3 July 2019)

Hyvönen, E., Heino, E. Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., and Mäkelä, E. 2016 'Publishing Second World War history as Linked Data Events on the Semantic Web' in *Digital Humanities 2016, Conference abstracts*. Kraków: Jagiellonian University and the Pedagogical University of Kraków. Kraków, 571–573. Available http://dh2016.adho.org/abstracts/222 (Accessed 6 July 2019)

Intrasis 2019. Available http://www.intrasis.com/ (Accessed 3 July 2019)

Jordal, E., J. Holmen, S.A. Olsen and C.-E. Ore. 2010 'From XML-tagged Acquisition Catalogues to an Event-based Relational Database' in F. Niccolucci and S. Hermon (eds) *Beyond the Artifact. Digital Interpretation of the Past. Proceedings of CAA2004, Prato 13–17 April 2004.* Archaeolingua, Budapest, 81–85. Available https://proceedings.caaconference.org/files/2004/12_Jordal_et_al_CAA_2004.pdf (Accessed 3 July 2019)

Matsumoto, M. and Uleberg, E. 2015 'Digitizing archaeology in Norway', in K. Piotrowska, and P. Konieczny (eds) *CONDITION.2015 Conservation and Digitalization.* Gdańsk: National Maritime Museum in Gdańsk, 159–64.

MUSIT 2019 Available at http://www.musit.uio.no/musit/musitweb/html/english.html

Oldman, D., Doerr, M., and Gradmann, S. 2016 'Zen and the art of Linked Data: new strategies for a Semantic Web of humanist knowledge' in S. Schreibman, R. Siemens and J. Unsworth (eds) *A new companion to digital humanities.* Chichester, UK: John Willey & Sons, 251–273. doi:10.1002/9781118680605.ch18

Ore, C.-E. 2017 'Digitization and Preservation of Cultural Heritage Information – A Nordic focus' in B. Franz and G. Vinken (eds) *Das Digitale und die Denkmalpflege: Bestandserfassung – Denkmalvermittlung – Datenarchivierung – Rekonstruktion verlorener Objekte,* Heidelberg: arthistoricum.net. Veröffentlichungen des Arbeitskreises Theorie und Lehre der Denkmalpflege e.V. 26, 140–149. doi:10.11588/arthistoricum.263.348

Ore, C.-E. 2018. Oceans of Data: Creating a Safe Haven for Information, in M. Matsumoto and E. Uleberg (eds) *Oceans of Data Proceedings of the 44th Conference on Computer Applications and Quantitative Methods in Archaeology.* Archaeopress, London 2018, 3–9.

Ore, C.-E. and Kristiansen, N. (eds) 1998 *Sluttrapport 1992–1997.* Oslo: Dokumentasjonsprosjektet. Available http://www.dokpro.uio.no/sluttrapp.pdf (Accessed 6 July 2019)

Pilø, L. 2007 'The Fieldwork 1998–2003: Overview and Methods' in D. Skre, *Kaupang in Skiringssal,* Kaupang Excavation Project Publication Series I, Norske Oldfunn XXII, 143–60.

Uleberg, E. and Matsumoto, M. 2019 'Georeferencing museum collections' in M. Doerr et al. (eds) *Humanists and the digital toolbox. In honour of Christian-Emil Smith Ore.* Oslo. Novus forlag, 111–132.

DataArc: a case study using CIDOC CRM as an ontology for transdisciplinary research

***Gísli Pálsson*[*] and Rachel Opitz*[**]**

Fornleifastofnun Íslands (Institute of Archaeology, Iceland)
*gisli.palsson@gmail.com

Department of Archaeology, University of Glasgow, UK
**rachel.opitz@glasgow.ac.uk

Abstract

The DataArc Project aims to create digital resources and tools intended to encourage integrative collaborative research on long-term human ecodynamics of the North Atlantic. The research community engaged in the project works across this transdisciplinary domain, and engages specialists from environmental geography to tephrochronology to zooarchaeology to saga studies. Defining a shared data model and conceptual framework to build links across the data created and studied by each of these disciplines is a key part of DataArc's digital cyberinfrastructure[1] development. DataArc aims to create a coherent and useful set of links across diverse data sources and data models belonging to each discipline, and to provide the potential for compatibility with external projects. To accomplish this a tiered structure has been developed, with data model items, or combinations of these items phrased as queries, mapped to a shared conceptual framework through pointers referred to as 'combinators' which serve to provide expert commentary on the mappings. The agreed community conceptual framework, formulated as a topic map, is then mapped to the International Committee for Documentation Conceptual Reference Model (CIDOC CRM). This paper reflects on the arrangement of the information and infrastructure developed in the project.

Keywords: CIDOC CRM; ontology; North Atlantic archaeology

Introduction

Archaeological research in the North Atlantic islands,[2] and critically in Iceland, has had a long focus on long-term human-environment interactions (e.g. Harrison and Maher 2014; McGovern et al. 2007), exemplified by research efforts led by the North Atlantic Biocultural Organization (NABO).[3] NABO is an international group of researchers working in transdisciplinary domains, with disciplinary specialisms ranging from

[1] The term 'cyberinfrastructure', widely used in the USA, is equivalent to the term 'e-infrastructure' commonly used in Europe.

[2] Taken here to include Iceland, Greenland, the Faeroe Islands and the Outer Hebrides.

[3] www.nabohome.org

environmental geography to tephrochronology to zooarchaeology to saga studies. The data generated through this kind of interdisciplinary research illustrates the challenges of contemporary archaeological data in that it draws on a multiplicity of specialists' expertise across traditional disciplinary divides, generates a wide variety of types of data, and embeds itself in equally diverse conceptual frameworks and assumptions. The situation of archaeological research in the North Atlantic islands, and in Iceland in particular, likewise exemplifies the role of archaeological data and the knowledge generated from it in the context of contemporary political and ecological debates, a situation which highlights the need for data that is not only openly accessible, but is legible to a wide variety of potential audiences. In order to develop an infrastructure that addresses these needs, the community must address hard socio-technological challenges surrounding the integration of interdisciplinary data and conceptual frameworks.

The first stage of ARIADNE explicitly addressed these problems at a high level through the adoption of the CIDOC CRM and the development of several extensions to it. The official remit of the CIDOC CRM is defined as connecting data from diverse sources through a heritage domain ontology. The CRM has been expanded several times, reflecting the broadening definition of domains connected to heritage that can be linked through the ontological model. This paper reflects on the impact of the CRM and the approach it promotes on the work of the DataArc Project to integrate data and concepts from North Atlantic island archaeological research, centered on NABO affiliated projects.

The NSF-funded project *DataArc: Linking Data from Archaeology, History, Sagas and Climate* began developing an ontology and cyberinfrastructure to enable data-driven research by NABO and its allied researchers in 2014. The cyberinfrastructure will link together key databases and support the creation of analytical research tools with the initial phase focusing on Iceland, Greenland, and the Scottish Northern Isles. The aim is to augment research surrounding the history of the North Atlantic by linking and presenting diverse datasets in new ways, leveraging the improved technological capacity created by this cyberinfrastructure. In its first phase, the project is focused on datasets which include the Strategic Environmental Archaeology Database (SEAD, paleoecological data)[4], NABone (zooarchaeological data stored in tDAR), the Icelandic Saga Map (ISM; geolocated saga references)[5], Storied Lines (a historical geodatabase of land use in Iceland)[6], paleoclimate proxies, Grefill (Icelandic excavation data), and Tephrabase (tephra stratigraphy and chronology)[7], with plans to expand to hold further data sources in future phases of work.

The DataArc Project serves as a potential model for the integration of a broader set of Icelandic data into the ARIADNE infrastructure, and for other situations where interdisciplinary and transnational research agendas lead to data management and synthesis

4 http://qsead.sead.se/
5 http://sagamap.hi.is/is/
6 http://jardabok.com
7 http://www.tephrabase.org/

challenges that are difficult to address using existing infrastructure because many extant data repositories are either specific to national data or knowledge domains, leading to unconnected data silos. As noted above, this is essential for archaeological data to speak to contemporary social issues such as observing and explaining long-term human environment interactions, which has been identified as one of the 'grand challenges' of archaeology, key to understanding global social trajectories, and to assessing human vulnerability to climate change (Kintigh et al. 2014, 2).

The CIDOC CRM as an ontology for interdisciplinary research

DataArc is one of several projects in the domain of human environment interactions that is addressing broad and, its proponents would argue, critical questions in an explicitly interdisciplinary framework. One challenge of interdisciplinary, data-embedded research is the clear impossibility of being sufficiently expert in all the domains involved to engage with the detail of their data and conceptual models. The combination of deep expertise in one domain and shallow expertise in allied domains introduces problems at each stage of the research process. At the question development phase it becomes difficult for research groups to understand where they are likely to be able to successfully address a complex question. At the data discovery stage, the tasks of efficiently finding data from other domains, understanding if that data is relevant and, if so, how it can be used, present real challenges, particularly to individual researchers wishing to engage seriously with data from allied domains.

Consider as an example a researcher investigating the role of driftwood in Medieval Iceland from c. 900–1700 AD in a human ecodynamics framework. This kind of topic clearly benefits from an interdisciplinary approach. An effective data infrastructure should facilitate integrative and multiproxy approaches by drawing together information contained in multiple data sources, for instance: the presence and provenance of driftwood in inland archaeological sites, place names and references in the Icelandic Sagas regarding the use and control of driftwood, documents detailing the ownership of driftwood in conjunction with other data modelling the high status sites in control of the resource. These combined data, together with environmental proxies and indicators of economic stability, wealth, or hardship, can in turn shed light on the socioeconomic and political circumstances surrounding the role of driftwood, providing new perspectives on a subject traditionally studied primarily through a historical lens. By drawing together these data sources it becomes possible to investigate driftwood through complementary lenses, for example as a process of social power by considering the societal influence of the farmsteads that controlled the resource in Medieval Iceland.

To respond to research questions such as the example given above, we argue that when developing a system designed to give broad access to complex data, it is not enough simply to link data with minimal or domain-specific metadata, because this does not provide sufficient explanation to effectively cut across disciplinary boundaries. In an interdisciplinary framework, it is necessary to use computational ontologies to map shared concepts to domain specific concepts in which data are implicitly embedded. DataArc adopted and adapted the CIDOC CRM for this purpose. The CRM was selected for its nuanced and flexible ontology, and on the basis of its growing influence in archaeoinformatics (e.g. Doerr et al. 2004; Binding et al. 2008; Niccolucci and Richards 2013).

The CIDOC CRM fulfills its primary aim 'to enable information exchange and integration between heterogeneous sources of cultural heritage information' (Crofts et al. 2011, 9) by explicitly describing the underlying semantics of database schemata and document structures which enables connectivity across datasets to be defined and mapped. These descriptions draw on a large vocabulary of so-called classes and properties. A class may be treated as a node in a graph structure, and is defined as one or more items sharing common traits used as criteria for defining a class. These traits are called the intensions of the class (Crofts et al. 2011, 9). Properties, which may be treated as links in a graph structure, define specific relationships between classes.

The relative abstraction and high level of these classes and properties in themselves facilitate interdisciplinary work because they are general enough to accommodate data and their attendant concepts from a variety of domains allied to heritage.

Recently, CIDOC members, along with collaborators from the ARIADNE network have developed a number of extensions to CIDOC CRM, aimed at expanding its capability to handle diverse archaeological information. These include CRMarchaeo (Hiebel et al. 2014), providing a semantic map for excavation data; CRMgeo for spatiotemporal data (Hiebel et al. 2017), CRMsci for scientific observations (Doerr et al. 2014), and CRMepi for epigraphs (Felicetti et al. 2015). The process of extending the CRM to various specific domains linked to the overarching heritage domain illustrates the utility of the CRM for connecting across domain specific conceptual models and ontologies, and highlights the need within the community served by ARIADNE for an interdisciplinary framework that connects data and concepts between allied domains connected to heritage broadly defined. The success of these efforts is illustrated by projects that have used the CIDOC CRM to bring together interdisciplinary and multifaceted datasets e.g. for Stonehenge (Sugimoto et al. 2007), in the global rock art database (Haubt 2015), and for the COSCH Project on Color and Space in Cultural Heritage (Bentkowska-Kafel et al. 2015), all of which provide useful models for DataArc's work.

DataArc: applying CIDOC CRM from the concepts to the data

Conventionally, the CRM is intended for mapping from data, where the ontology is used to describe the data structure, and the data structure defines the scope of the parts of the ontology used. This approach is well suited to working within single, well defined disciplinary contexts, where implicit meanings of key concepts are broadly agreed and readily understood. Within the context of interdisciplinary research, as is the case for DataArc, it is useful to add a second level of ontological structure by using the CRM to explicitly map the connections between domain-specific concepts used by different disciplines and broader shared concepts. This approach allows each discipline to maintain well defined, meaningful concepts, which can be described within the infrastructure, and shows how they connect to shared concepts, many of which may be viewed by domain experts as 'further from their data' or as a necessary compromise between what their data actually says and the high-level questions asked by a broader research project.

The process of mapping the conceptual network that spans multiple allied disciplines is essential to the development of useful data infrastructure, and has subtle but important socio-technical implications. To take an example from the DataArc context, a disciplinary expert in zooarchaeology may be justifiably hesitant to explicitly and directly map their data on butchery marks on sheep bones to a high-level concept like 'land use' because there are several intellectual leaps between the data and the high-level concept. Further, the disciplinary expert may argue, their data is necessary but insufficient to address the high-level concept. This expert may further point out that data and mid-level concepts from other disciplines are necessary to map their data to the high-level concept. Capturing the complexity of this conceptual scaffolding is, we argue, necessary both to reassure experts that their data are not inherently being misrepresented by how they are mapped into a data infrastructure and to illustrate to an end user where greater conceptual complexities are at play in relating data to the question that motivated their search.

Motivated by the need for explicit conceptual scaffolding in data infrastructure, and taking advantage of a network of researchers which included many key domain knowledge experts working in the North Atlantic, DataArc undertook the exercise of characterizing the research that had taken place in the region with an explicit aim of identifying common ground between disciplines. Rather than starting from mapping each and every data element in the various participant's databases to CIDOC classes and properties, we began by mapping major research themes in the study of North Atlantic human ecodynamics as a network of concepts. We then classified the nodes and edges in the network using CIDOC CRM classes and properties.

This exercise, beyond producing a shared ontology that is currently implemented in the project's prototype cyberinfrastructure, encouraged thinking about concepts that have, in graph theory terms, high betweenness-centrality, that is concepts that are key connectors between data and conceptual frameworks across the disciplines represented in the shared ontology. These concepts may be considered primary targets for future research because they sit at the intersection of many datasets and domains. Conversely, assessment of the ontology's graph structure facilitates the identification of concepts that are poorly connected, and therefore not well addressed by the collective data and expertise of the domains represented. In short, identifying key gaps and opportunities for research is facilitated by the ontology and data mapping exercises.

Explaining connections

In a transdisciplinary context, more of each domain's expertise and more explanation need to be embedded into each mapping from data to concept, and arguably also

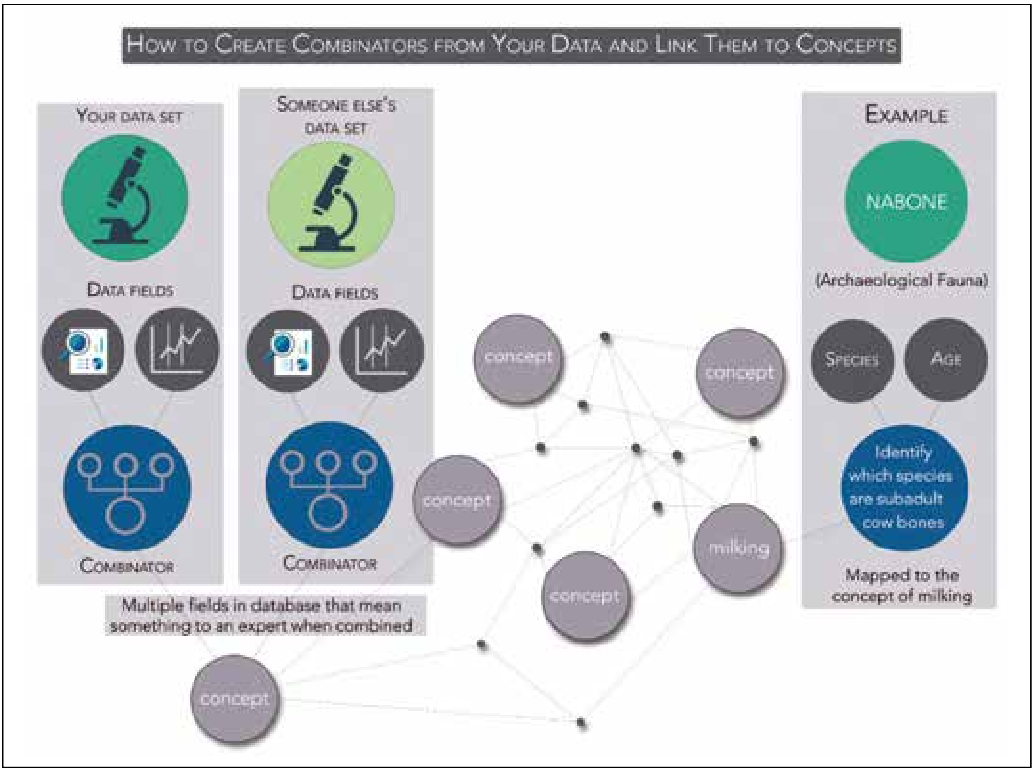


Figure 1: A schematic overview of the DataArc approach to interdisciplinary data and concept mapping to connect domain specific databases to a shared ontology via 'combinators' that provide expert commentary and bibliography

between concepts, because implicit domain-specific background knowledge cannot reasonably be assumed, a problem well illustrated by Kansa and Kansa (2013). The individual data types or combinations of data types or other aggregations that are relevant to a given concept are only really legible outside a community of experts if their mappings to the concept map include intelligent commentary on their relevance and a bibliography that supports the expert commentary, effectively data-to-concept metadata explaining not what is in the data, but why it is relevant to the concept at hand.

This kind of interdisciplinary integration effort goes beyond basic synthesis in that it is explicitly made at the data level, and therefore involves the concepts and standards implicit in how each discipline models their data, as well as the structure of the data sources used by various research communities. Bringing these data together requires a formal framework like the CRM to clearly express the links between shared ideas and questions and the various data lenses that address them.

In this sense, the DataArc system goes beyond the scope of LOD, which describes the data but leaves it to the user to connect each data type to the problem space. The definition of 'discoverable' given in the context of research as part of the International

Polar Year (Parsons et al. 2011) implies what we are driving towards – data must be readily generally assessed for its suitability, and achieving this is one of the project's socio-technical problems. A key question becomes how to create this contextualizing descriptive metadata. The desired expanded interconnection between data sources from different fields relies, at a certain level of usefulness, on increasingly nuanced semantic connections. The problem has also been well identified elsewhere: 'There are great challenges to sharing sufficient context and provenance information across disciplinary boundaries for users to effectively evaluate and correctly apply data' (Parsons et al. 2011). Along similar lines, Scheider et al. (2017) note, 'Very often, scientists need to know more about foreign data than can be discovered from its surface, causing what one might call *science friction*. This becomes even more important if we want to exploit the advantages of citizen science, i.e. local and traditional knowledge from the citizens, and increased participation for the citizens.' In sum, while the challenge is known, practical effective solutions remain elusive. Scheider et al. (2017) describe what we have called the data-to-concept metadata as 'synthetic context' which 'consists of observation, data and model selection, data derivation and interpretation'.

Towards driving explicitly interdisciplinary search

Using the data and conceptual mappings and explanatory 'combinators' described above, DataArc is implementing a search tool that intentionally presents related results from other disciplines in addition to the results that respond to a domain specific query, and clearly explains why the results are relevant. To design this tool, we begin by considering how a researcher with a question approaches searching a database with which he or she is familiar. An archaeologist might have a question about when a farm began to decline and become less productive. To address this question, the researcher may want to draw upon a constellation of evidence, searching for collapsed buildings, counts of animal bones summarized as counts of animals, and counts of human burials. Each of these are in themselves concepts that feed into the higher-order concept of farm decline. Each one can be linked to a CRM 'E' property, man-made features, biological objects, and sites are all likely candidates. To drive broader, more diverse search results, DataArc is applying two tactics. First, when a researcher searches for a specific data category (a certain field or a limited set of related fields), DataArc returns results that have been mapped to the same concept as that initial field, or results mapped to concepts one or two degrees away from the original concept within the network of connected concepts. Second, DataArc is able to limit or describe the elements of this broader result set and their relationship to the concept indicated by the original search through their CRM properties. In our approach to search and data discovery, the CRM plays an essential role in explaining high level connections between concepts and data categories attached to them.

Consider how we might connect the ideas expressed in two passages taken from *Jarðabók Árna Magnússonar & Páls Vídalín* (1913–1943), an early 18th-century land use census to evidence from archaeological survey and paleoenvironmental models using CRM properties. The first passage reads 'Litlunúpar is the name of an ancient farm mound within the lands of this farm, a little way from the farmstead ... clear signs of buildings are here, both structures and boundary walls. It is rumoured that this place became desolate due to a haunting' (Magnússon and Vídalín 1913–1943, XI, 220; translation by the authors).

The second reads '[the farm Folafótur has good peat on its land in] Faxadalur, but is never used, as it is believed that bad things happen to those who cut peat there' (*ibid.,* VII, 184). These are both abandoned or actively disused areas of the landscape which are out of use because they are associated with something supernatural, bad, or dangerous. Within our concept model we have a concept of 'cursed spot', which is a general place type under which both of the places described in these passages might fall.

The implementation of the concept-to-concept mappings have important implications for the result sets delivered by the search tool. For example, we might handle these places by creating new sub-concepts of the 'cursed spot' specific to 'haunted peat' and 'abandoned farm'. Alternatively, we might handle the inclusion of the second of these places by creating two peatland concepts: unmanaged peatland and managed peatland, which will be more generally applicable across our data. Then we can make an explicit subcategory of 'haunted peatland' under unmanaged peatland which is connected to 'cursed spot' or 'ghost' with the relationship 'may be created by' P94 (here and elsewhere we refer to specific E- element and P- property codes in the CIDOC CRM or its extensions) and a haunting event or activity. Similarly, we treat the bad things that happen to people who use haunted peat as the creation of a conceptual object (an imagined consequence) created by the presence of a ghost or haunted place. The concepts of unmanaged peatland and managed peatland, both types of peatland, would connect to evidence drawn from the paleoecological data, represented by pollen and/or insect assemblages. These concepts would also connect to the archaeological evidence, differentiated by the presence of evidence for peat cutting. We use the same logic to connect 'cursed spot' or 'ghost' and a haunting activity or event to the concept of a 'farm' to define the concept of a 'farm abandoned due to haunting'. This set of relationships puts the archaeological evidence for built structures two degrees away from eco-code concepts. A search for built structures in the archaeological data, a common search for an archaeological specialist, widened to include results one conceptual degree out, would pull in the idea of haunted or abandoned places, and widened two degrees out would connect to specific ecological signatures. This is a chain of reasoning we would not be surprised to find in an article or book chapter, codified into one of the mechanisms driving search results. That said, there is an important difference in the implementation of this kind of chain of reasoning in a search tool in that different choices in how the conceptual connections are organized will deliver different sets of search results based on the degrees of conceptual proximity defined.

The mapping of another high-level concept, ecclesiastical power, to the various data sources available draws on similar chains of reasoning, and provides another

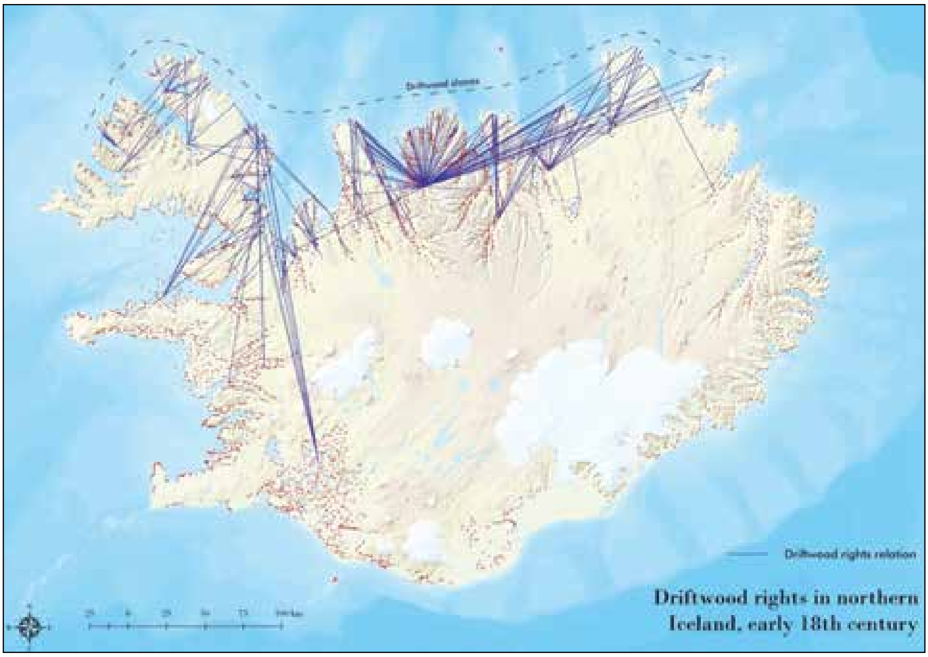


Figure 2: The control of driftwood shores in early 18th century Iceland. The dataset covers only the northern shores, indicated by the dotted line

useful example for consideration. Driftwood was a crucial resource in Medieval Iceland as it constituted the main source of wood for construction, as well as being the main source of firewood for several regions along its coast (Kristjánssons 1980). The 16th-century text *de Mirabilis Islandiae* clearly indicates late medieval attitudes toward the resource (Oddsson 1917). Driftwood shores were prized, and for the most part controlled by wealthy church farms often located far away from the driftwood-rich shores (*Figure 2*).

The control of driftwood has an impact on settlement structure, requiring that we make links between textual and archaeological data sources, between legal objects E72 and physical things E18. The medieval legal codex *Jónsbók* has several clauses relating to driftwood rights that give a clearer picture of the work involved in harvesting the resource. Landholders were allowed to wade into the sea and mark driftwood that drifted close to the shore. In other words, the control of driftwood necessitated a persistent human presence at the coastline. Bountiful driftwood beaches are frequently found in places with very poor conditions for traditional Icelandic subsistence agriculture, a conclusion drawing on links to models of environmental conditions, treated here as E28 conceptual objects. However, and for that reason the settlement structure facilitating the control of driftwood appears to have been maintained, at least in part, through Iceland's medieval tenure system (see Júlíusson 2013). In other words, driftwood does not only have an agency on the coastal settlement structure itself, but so do the high-status church farms ultimately

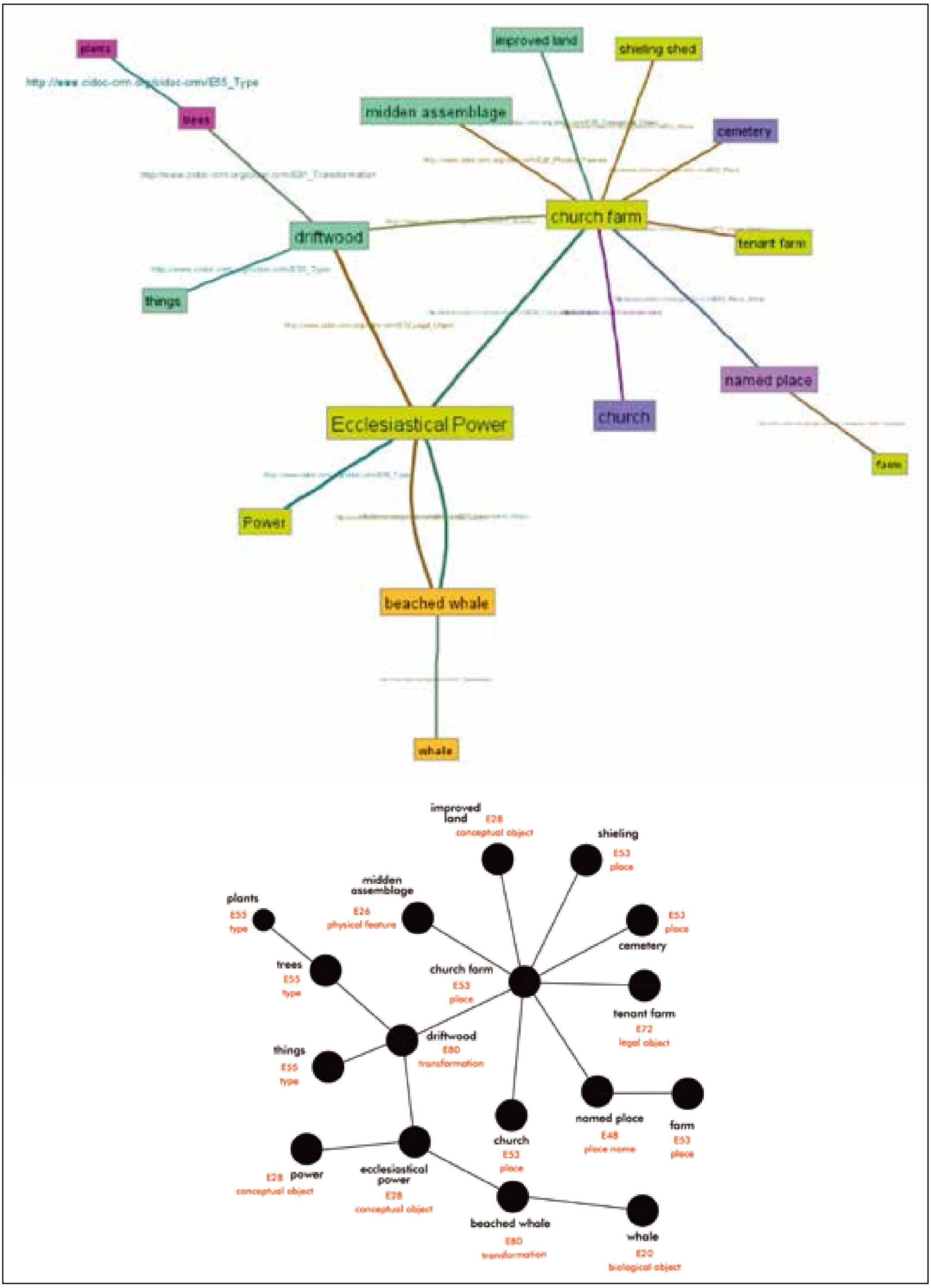


Figure 3: The topic map surrounding ecclesiastical power, where edges are codified as CIDOC 'P' properties and nodes are mapped to CIDOC 'E' elements (top); CIDOC 'P' Properties define edges on the topic map, and are bidirectional, including the outgoing and inverse properties (bottom)

in control of the resource. Further, it would be difficult to understand the settlement pattern when combined with the environmental data without an understanding of the legal framework and ecclesiastical landscape.

This example highlights the challenges created through the potential for multiple equally valid mappings, which for the topics we have just discussed seem necessary. For example, within the various DataArc sources driftwood is simultaneously construed as a legal object E72, a biological object E20, and a physical thing E18. As a legal object rights may be held (P105) by an actor. As a biological object, it sits one degree away from plants, having been transformed (P124), and two degrees away from ecological types. Similarly, a church farm is both a conceptual object E28, and a site E27. Church Farms in their site E27 mode are composed of (P46) physical features E26 that include built structures or structural elements commonly identified in the archaeological data. As conceptual objects, they collect together legal objects and connect to actors through P105 rights. The multiple mappings given here codify our understanding of this complex set of relationships between the textual, archaeological and environmental factors. If we choose a single mapping, we are likely both to misconstrue how one domain on another understands a concept, and to cut off links to relevant data, which would be entirely counterproductive to our goal of driving interdisciplinary search results. However, creating large numbers of parallel mappings introduces complications in the practical implementation of the search mechanism, as one of the modes in which the object in question is construed must be used to define the search terms.

Open questions: Bending, deviations, and incompleteness

As evident in the brief discussion above, while the CIDOC CRM can be used as the ontological basis for an interdisciplinary search tool, the process of mapping the concepts of diverse domains together through the CRM has raised questions on several fronts. One set of questions surrounds the problem of multiple parallel mappings. We have given the example of driftwood as a physical object, biological object and conceptual object in our own model. The same problem has been expressed as a problem in other projects. Kansa (2014) gives the example of mapping potsherd colors, 'For example, we recently had a discussion with a librarian trying to use the CIDOC CRM to organize some archaeological data from a survey for publication in Open Context. The librarian used the CIDOC CRM property "P3 has_note" as a predicate for use with Munsell color readings of potsherds. This raised some interesting issues. It is probably debatable if a Munsell color reading is simply a descriptive "note" or if a Munsell color reading is more of a measurement. If the latter, then the CIDOC CRM property "P43F has_dimension" would probably be a more appropriate predicate. In theory, Munsell can be seen as an objective measurement. In practice, many researchers take Munsell readings because they vaguely think they should, and then they do not adequately control for all sorts of issues (lighting conditions, dampness, color blindness, etc.) that may impact a Munsell reading. The example above illustrates how difficult the

CIDOC CRM can be to use in practice.' The possibilities of multiple ways of codifying the relationships between any two given topics and the presence of ambiguities is a problem recognized by the CIDOC community, as well as groups attempting to implement the standard. As noted by Tudhope et al. (2011), 'many of the entities in the ontology are fairly abstract; understanding the conceptual complexity of the CIDOC CRM poses a challenge to some non-specialists. It can also be possible for different people to make alternative valid mappings to the ontology for the same situation, raising difficulties for semantic interoperability'.

From our perspective, the participation of each concept in multiple CRM elements is to be expected, and the role switching as each idea participates in different relationships is important. Limiting the number of assignments and relationships is, as mentioned above, essential to the coherence of the mapping, and to effectively driving search. If we 'map the world' and connect everything, as is clearly possible if not productive, then the search results suggested by pulling in connected concepts will return all results every time, which is not the intended result. Rather, we attempt to create sets of mappings sufficient to capture the 'why' of the primary relationships between concepts, as defined by the DataArc community.

A second important set of questions relates to the defined CRM terms not quite fitting the concepts used by the DataArc community. This is not new, and individual research communities have extended or amended the CRM, as noted in the Introduction. The work by ARIADNE resulting in CRMarchaeo and CRMba specific to the archaeological community, and the development of other extensions such as CRMsci, CRMgeo, or CRM-EH are cases in point (Binding, May and Tudhope 2008; Doerr et al. 2014; Doerr et al. 2015; Hiebel, Doerr and Eide 2016; Ronzino 2016). When the existing ontology was found to be not quite satisfactory, revisions or extensions were developed (e.g. Vassilakaki et al. 2015). However, in these cases, larger organizations were involved, as was the case with English Heritage, or the extensions and revisions were deemed applicable for a large group, as was the case with CRMsci. The proposal of new extensions or revisions to be incorporated into the model is clearly an accepted process within the active CRM community, as evident from discussions on their issue tracker page[8] and the development of new extensions addressing limitations in the current framework (e.g. CRMtex; Felicetti and Murano 2016).

Given the relatively small and specialized research community within DataArc, formally extending the CRM does not seem an entirely sensible solution, as such an extension may not be useful to a broad set of researchers. Perhaps a compromise is needed, to declare some 'creative use' of the existing CRM and its extensions, without going so far as to define a new extension to the standard.

The DataArc approach to actors is illustrative of a 'creative use' of the CIDOC CRM that might not be useful as a formal modification or extension of the standard. The CIDOC CRM defines the concepts of an actor E39 and actor appellation E82. The CRM conception of an actor refers explicitly to a person, 'This class comprises people,

[8] http://cidoc-crm.org/Issue

either individually or in groups, who have the potential to perform intentional actions for which they can be held responsible'. Some archaeologists working within symmetrical archaeology (e.g. Hamilakis and Overton 2013; Preucel 2016; Lindstrom 2015) have promoted a theoretical framework in which non-human actors have this active potential. While we would not go so far as to ascribe intentional action in the way intended by the CRM to sheep, in the context of our questions and data, they play a more active role than simply a biological object, and can certainly be held partly responsible for certain changes in land cover.

Without entering into an extended theoretical debate on the agency of non-humans, we wish to ascribe more active roles to the animals, plants, and things within our framework. The primary motivation for allowing insects and sheep to be actors in our conceptual model is to provide a better balance and more equivalencies between the disciplines represented and their data. In a conceptual universe that includes paleoecology and paleoentomology, where there are many pieces of data specifically about insects and very few explicitly about humans, it would seem odd to only allow humans to be active and responsible players.

Creative use of the CIDOC CRM concepts highlights a tension between the needs of individual projects or contributors to a broader infrastructure like ARIADNE, and that of the infrastructure's governing community. From an individual project perspective, a note explaining DataArc's expanded definition of actors may be deemed sufficient and developing a full formal extension of the CRM may seem excessive. From the perspective of the broader ARIADNE community, however, accepting and integrating data mappings and ontologies that creatively use the CIDOC CRM may be viewed skeptically. A process of negotiation between individual projects and the broader ARIADNE community seems necessary to develop solutions that will meet the needs of both project-specific readings of the ontology's classes and properties and the broader infrastructure's conceptual coherence.

Discussion and conclusions

The utility of the CIDOC CRM as a mechanism for interdisciplinary data and conceptual framework integration constitutes an important, if unplanned, contribution of the first phase of ARIADNE. This chapter has illustrated this by showcasing its implementation by DataArc to facilitate research on human ecodynamics in the context of the North Atlantic islands. Crucially, this mechanism brings the process of interdisciplinary negotiation and synthesis from later stages of research, enacted through informal discussions, into the early question formation and data gathering stages of research, and formalizes it within a data search tool.

The process of developing the DataArc data and conceptual model, while resulting in a working prototype, highlighted key socio-technological obstacles to formalizing the process of interdisciplinary synthesis. In interdisciplinary research, publication phase integration is not uncommonly the result of extended discussion between specialists working on a shared question, and the experience of concept modeling and mapping has paralleled this, involving wide-ranging and repeated conversations between the community's specialists, as the contribution of each data source is highlighted and reservations about the reliability or need for contextualization of individual data types are expressed. This closely parallels the experience described by Doerr (2002, 15) on working with another interdisciplinary group mapping to the CRM in the relatively early years of the standard's development, 'Philosophical considerations and long discussions were necessary to clarify the role of the modelled knowledge with respect to the working concepts of the domain experts. Without such clarifications, no consensus on the relevant concepts could be achieved.' It is also echoed in Kansa's discussion of the difficulty of mapping to the CIDOC CRM (Kansa 2014).

The experience of the DataArc project is illustrative of the situation faced by projects and organizations attempting to integrate data from the diverse domains that fall under the heritage umbrella, and particularly those of groups aiming to address current societal challenges that require drawing on interdisciplinary data and concepts. The need to address diverse data and concepts in an integrated, coherent framework is increasingly prevalent in the archaeological and heritage communities, as we are collectively asked to engage with contemporary social and environmental agendas and questions. To meet these needs at a broader scale will require continued capacity building to enable explicitly interdisciplinary and transdisciplinary research. As it continues to grow, the ARIADNE infrastructure is well positioned to meet this challenge.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Numbers 1439389 and 16370706. The authors gratefully acknowledge the support and contributions of the DataArc community and collaborators, who have enthusiastically participated in discussions and shared data.

The current prototype data discovery tool is available at: https://beta.data-arc.org/. The code including the topic map is available on github at: https://github.com/digital-antiquity/DataARC.

References

Bentkowska-Kafel, A., del Hoyo Melendez, J.M., MacDonald, L.W., Mathys, A. and de Almeida, V.M. 2016 'Colour and Space in Cultural Heritage in 6Ds: the Interdisciplinary Connections', in S. Campana, R. Scopigno, G. Carpentiero and M. Cirillo (eds) *Proceedings of the 43 Annual Conference on Computer Applications and Quantitative Methods in Archaeology CAA2015,* 953–962.

Crofts, N., Doerr, M., Gill, T., Stead, S. and Stiff, M. 2011 'Definition of the CIDOC conceptual reference model' *ICOM/CIDOC Documentation Standards Group. CIDOC CRM Special Interest Group* 5.

Binding, C., May, K. and Tudhope, D. 2008 'Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC CRM' *Research and Advanced Technology for Digital Libraries,* 280–290.

CRMarchaeo 2016 *CRMarchaeo: The Excavation Model.* http://www.cidoc-crm.org/crmarchaeo/ (Accessed July 2019)

CRMgeo 2016 *CRMgeo: Spatiotemporal Model.* http://www.cidoc-crm.org/crmgeo/ (Accessed July 2019)

CRMsci 2016 *CRMsci: Scientific Observation Model.* http://www.cidoc-crm.org/crmsci/ (Accessed July 2019)

Doerr, M. 2002 'The CIDOC CRM: An Ontological Approach to Semantic Interoperability of Metadata', *AI Magazine, Special Issue on Ontologies.*

Doerr, M., Kritsotaki, A., Rousakis, Y., Hiebel, G. and Theodoridou, M. 2014 *CRMsci: the Scientific Observation Model.*

Doerr, M., Theodoridou, M., Aspöck, E. and Masur, A. 2015 'Mapping archaeological databases to CIDOC-CRM', in S. Campana, R. Scopigno, G. Carpentiero, G. and M. Cirillo (eds) *Proceedings of the 43 Annual Conference on Computer Applications and Quantitative Methods in Archaeology CAA2015,* 443.

Duşa, A., Nelle, D., Stock, G. and Wagner, G.G. 2014 *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences.* Berlin: SCIVERO Verlag.

Felicetti, A. and Murano, F. 2016 'Scripta manent: a CIDOC CRM semiotic reading of ancient texts', *International Journal on Digital Libraries,* 1–8.

Hamilakis, Y. and Overton, N.J. 2013 'A multi-species archaeology', *Archaeological Dialogues,* 20(02), 159–173.

Harrison, R. and Maher, R. (eds) 2014 *Long-Term Human Ecodynamics in the North Atlantic: An Archaeological Study.* Lanham: Lexington Publishers.

Haubt, R.A. 2015 'The global rock art database: developing a rock art reference model for the RADB system using the CIDOC CRM and Australian heritage examples' *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(5), 89.

Hiebel, G., Dörr, M. and Eide, Ø. 2017 'CRMgeo: A spatiotemporal extension of CIDOC-CRM', *International Journal on Digital Libraries,* 18(4), 271–279.

Júlíusson, Á.D. and Jónsson, J. 2013 *Landbúnaðarsaga Íslands.* Reykjavík: Skrudda.

Kansa, E.C. and Kansa, S.W. 2013 'Open archaeology: we all know that a 14 is a sheep: data publication and professionalism in archaeological communication' *Journal of Eastern Mediterranean Archaeology and Heritage Studies,* 1(1), 88–97.

Kansa, E. 2014 'Open Context and Linked Data' *ISAW Papers* 7.10.

Kintigh, K.W., Altschul, J.H., Beaudry, M.C, Drennan, R.D., Kinzig, A.P., Kohler, T.A., Limp, W.F., Maschner, H.D.G., Michener, W.K., Pauketat, T.R., Peregrine, P., Sabloff, J.A, Wilkinson, T.J., Wright, H.T. and Zeder, M.A. 2014 'Grand challenges for archaeology' *Proceedings of the National Academy of Sciences,* 111(3), 879–880.

Kristjánsson, L. 1980 *Íslenzkir sjávarhættir.* Reykjavík: Bókmenntaútgáfa menningarsjóðs.

Lárusson, Ó. 1934 *Lögbók íslendinga, Jónsbók, 1578.* Copenhagen: Levin and Munksgaard.

Le Boeuf, P., Doerr, M., Ore, Ch.E. and Stead, S. 2015 *Definition of the CIDOC Conceptual Reference Model.* http://www.cidoc-crm.org/official_release_cidoc.html

Lindstrøm, T.C. 2015 'Agency "in itself". A discussion of inanimate, animal and human agency', *Archaeological dialogues,* 22(2), 207–238.

Magnússon, Á. and Vídalín, P. 1913–1943 *Jarðabók Árna Magnússonar og Páls Vídalín.* Copenhagen: Hið íslenzka fræðafjelag.

McGovern, T.H., Vésteinsson, O., Friðriksson, A., Church, M., Lawson, I., Simpson, I.A., Einarsson, Á., Dugmore, A.J., Cook, G., Perdikaris, S., Edwards, K.J., Thomson, A.M., Adderley, W.P., Newton, A., Lucas, G., Edvardsson, R., Aldred, O. and Dunbar, E. 2007 'Landscapes of settlement in Northern Iceland: Historical ecology of human impact and climate fluctuation on the millennial scale', *American Anthropologist,* 109(1), 27–51.

Niccolucci, F. and Richards, J. D. 2013 'ARIADNE: advanced research infrastructures for archaeological dataset networking in Europe', *International Journal of Humanities and Arts Computing,* 7(1-2), 70–88.

Oddsson, G. 1917 [1638] *Annalium in Islandia farrago; and, De mirabilibus Islandiae.* Ithaca: Cornell University Library.

Parsons, M.A., de Bruin, T., Tomlinson, S., Campbell, H., Godoy, O., LeClert, J. and the IPY Data Policy and Management Subcommittee 2010 *The State of Polar Data—the IPY Experience.* Unpublished Report, March 2010.

Preucel, R.W. 2016 'Pragmatic Archaeology and Semiotic Mediation', *Semiotic Review* 4. Available at https://www.semioticreview.com/ojs/index.php/sr/article/view/11 (Date accessed: 24 July 2019)

Ronzino, P. 2016 'Harmonizing the CRMba and CRMarchaeo models', *International Journal on Digital Libraries*, 1–9.

Scheider, S., Degbelo, A., Lemmens, R., van Elzakker, C., Zimmerhof, P., Kostic, N. and Banhatti, G. 2017 'Exploratory querying of SPARQL endpoints in space and time', *Semantic web,* 8(1), 65–86.

Sugimoto, G., Felicetti, A., Perlingieri, C. and Hermon, S. 2007 'CIDOC-CRM spider: Stonehenge as an example of semantic data integration', *Proceedings of the 8th International conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage,* 47–54.

Tudhope, D., May, K., Binding, C. and Vlachidis, A. 2011 'Connecting archaeological data and grey literature via semantic cross search', *Internet Archaeology* 30. doi:10.11141/ia.30.5

Vassilakaki, E., Zervos, S. and Giannakopoulos, G. 2015 'CIDOC-CRM extensions for conservation processes: A methodological approach' *AIP Conference Proceedings* 1644(1), 185–192.

MASA Digital ecosystem for the French archaeological community

Olivier Marlet [*] ***and Xavier Rodier*** [**]

CNRS, Université de Tours
*olivier.marlet@univ-tours.fr
**xavier.rodier@univ-tours.fr

Abstract

In France, the MASA Consortium strives to disseminate good practices in digital data within the archaeological community, and more particularly in terms of the interoperability of this data. Inspired by the choices of the European ARIADNE program, the MASA Consortium disseminates the FAIR principles and applies them as a proof of concept. Among its achievements, OpenTheso is an application designed to manage thesauri and in particular the multilingual PACTOLS thesauri; OpenTermAlign is an application to align an unstructured vocabulary with a standardized thesaurus; OpenArchaeo is an intuitive query platform using triplestores whose data are mapped with the CIDOC CRM.

Keywords: interoperability; CIDOC CRM; thesaurus alignment; semantic web; FAIR principles

Introduction

Created in 2012, the *Mémoires des Archéologues et des Sites Archéologiques* (MASA) Consortium has been certified by the Very Large Research Infrastructure Huma-Num. MASA was born from the experience acquired by and within several contributors to French archaeology, in particular *Maisons des Sciences de l'Homme,* Inrap and the Frantiq Network, in processing the documentation produced by archaeologists. MASA's partners have pooled their skills to meet the needs of the archaeological community. The main issues raised concern the use of digital technology in archaeology and therefore the good practices to be deployed in order to ensure that digital archaeological data can be used under the best conditions. Of course, this data processing involves online publishing and feeding the semantic web with heritage data and more particularly archaeological data.

The issue of data interoperability was raised from the beginning of the MASA Consortium in 2014 and experiences in this area had recently emerged in the community. The work carried out by European programmes such as Humanistica for the human and social sciences or Europeana for cultural heritage has obviously served as a reference to guide the work of the MASA Consortium. More pragmatically, French archaeologists felt more interested in the work carried out by the ARIADNE programme, in which Inrap has been involved since the beginning. The ARIADNE platform has shown great interest in making archaeological data interoperable by aligning them with recognized standards (Art and Architecture Thesaurus – AAT - of the Getty Museum, PeriodO and GeoNames) and by mapping the data with the ontology of CIDOC CRM, a conceptual reference model perfectly adapted to cultural heritage data. For these reasons, the MASA and 3D-SHS Consortia, as representatives of the CNRS, have joined the new ARIADNEplus program.

Digital archaeological data

Among archaeological data, archaeological archives have a special status that makes them valuable, related to the nature of the discipline itself: by digging, the archaeologist irreparably destroys his own object of study. Even with a rigorous protocol for recording data during their production, the experiment is not reproducible. This places a particular responsibility on the archaeologist and gives his records the status of primary data. The return to this data is often necessary for comparative purposes and reinterpretation.

With the emergence of digital technologies, databases have supplemented, or sometimes even replaced, our paper records; all documentation (field notebooks, photos, drawings, field recordings, etc.) have been digitized to make it easier to consult: indexing makes it easier to search and digital consultation spares sometimes fragile archives. Excavation archives are made up of various recordings and artifacts collected, which represent a considerable mass of material elements. The use of information systems enables us to rationalize the information and make data more easily usable. However, digital has not replaced paper and digital archives are doubling, sometimes complementing, paper archives. As the two are inseparable, information systems must ensure the preservation of the link between them.

The increasing use of information technology tools has also caused the proliferation of archaeological databases with a wide variety of formats and content. Archaeologist often design these heterogeneous databases without a methodological and technical choice protocol. Prior work is therefore necessary to make these databases standardised and interoperable (RDF, SKOS, DC), to use common gazetteers that make it possible to consider linking these data on the semantic web, to document them with new metadata where appropriate, to enhance them and to facilitate access while ensuring their sustainability.

FAIR principles and reliability

In the framework of the MASA consortium, we assessed the status of our datasets at our laboratory, according to the Five Stars LOD (Berners Lee 2016) and the FAIR principles (FORCE 11 2016). This assessment demonstrated the obvious: the older the datasets are, the less they comply with FAIR principles. The FAIR evaluation helps us to identify gaps to be filled in some of our datasets. For each dataset, it is possible to show its state of progress in the stars of the LOD and in the FAIR principles.

We have therefore been able to evaluate the quality of our data according to the criteria of these two evaluation systems: sustainable identification of resources, metadata, accessibility, standard format, standardized vocabulary, data linked to gazetteers, licenses, compliance with community standards. Although Aalto University (Finland) has been working to extend the five stars to seven (Living Laboratory Data Service for the Semantic Web 2014) and even if the two evaluation methods are similar, the FAIR principles method seemed more detailed and complete than the Five Stars Linked Open Data method, so we focused mainly on the FAIR principles. Using a very subjective evaluation grid, we were able to evaluate all the remaining work to improve the different datasets we publish on the web.

Within the online publication of databases, the MASA consortium encourages digitization of the paper recording sheets, to put them online, combined with their computerized version in DBMS. This entire digitized archive group must be accurately described in an EAD (Encoded Archival Description) file, an XML-based encoding standard for archival finding aids. In addition to the physical description and description of the archival units constituting the collection, this metadata file specifies the collection organization, the classification method and elements on the recording methodology. By defining the conditions for data acquisition and the way they are structured, the user takes a critical look at the quality and reliability of the data.

The MASA digital ecosystem

To meet these objectives, the MASA consortium proposes to the archaeological community a process of data manipulation from acquisition to publication according to a systemic approach that respects the FAIR principles. The MASA ecosystem is composed of bricks for archiving and sharing archaeological datasets. Once processed, documented and standardized, the archaeological datasets are put online according to the standards in force (XML, TEI, EAD...). Standardised gazetteers are used for spatial (GeoNames), temporal (PeriodO) and descriptive (PACTOLS via the OpenTheso thesaurus manager) information. The online application OpenTermAlign manages the alignment of dataset vocabulary with standardized thesauri and generates a SKOS file of the aligned vocabulary. The OpenArchaeo platform ensures their interoperability in a MASA triplestore and allows their interrogation via a simplified HMI that translates requests into SPARQL according to a generic model for mapping archaeological data with the CIDOC CRM ontology. To document each step of this process, MASA developed OpenGuide, a platform to publish good practices guides. In response to a call for projects from the European programme DARIAH, a series of workshops were organised around the use of the Standardization Survival Kit set up by PARTHENOS (H2020). One of these workshops was organized in collaboration with the MASA Consortium to work on the transfer of MASA good practice guides to the SSK, whose compatibility is simplified by the use of TEI by OpenGuide and SSK. The LogicistWriter tool is helpful to write in a logicist way and offers the matching of inferences with the CIDOC CRMinf extension on reasoning. We will not go into more detail here on OpenGuide because it is only a tool to formalize in a simple way the dissemination of good practices, nor LogiscistWriter, which, beyond the interoperability of data sets, is part of the scientific publication policy. The publication of the archaeological site at Rigny is a demonstration of this approach (Marlet et al. 2019a).

Standard vocabularies alignment

The alignment of our vocabularies with standards thesauri seemed essential to us and we explored several possibilities. If the Getty Museum's AAT seemed to us to be an attractive option, it had the flaw of not offering a French translation and dealt relatively little with the vocabulary of field archaeology. The thesaurus that interested us the most was the thesaurus "Sujets" from the PACTOLS thesauri developed by the Frantiq network, with the advantage of being multilingual. Initially, these thesauri were dedicated to the indexing of historical and archaeological publications, but their scope has grown quickly over time and PACTOLS are now widely used to index any data in the field of cultural heritage. Like the AAT, PACTOLS do not necessarily have all the required vocabulary to index field archaeology data. However, thanks to the PACTOLS enhancement policy and also thanks to the involvement of the Frantiq network within the MASA Consortium, it is possible for us to enrich the PACTOLS with the vocabulary that may be lacking.

The Frantiq network has set up a very useful online tool: OpenTheso. It is a thesaurus manager that enables us to navigate through the various PACTOLS thesauri and see all their aspects: hierarchical position in the thesaurus, links with Wikidata and Wikipedia, permanent identifier and several translations. In addition to this first function of thesaurus exploration, OpenTheso offers several web services that enable PACTOLS to be interfaced with local applications or CMS such as OMEKA-S.

The matching of dataset vocabulary with standardized thesauri can be achieved through the online application OpenTermAlign, which generates a SKOS file of the aligned vocabulary (*Figure1*).

Following the aim of providing tools for archaeologists in order to help them to improve their data and bring their datasets to the semantic web, the MASA Consortium has developed OpenTermAlign, a web application that enables the alignment of poorly structured or non-structured vocabularies with standardised thesauri. Marion Lamé (CITERES-LAT, Tours), in collaboration with Federico Ponchio (PIN, Pisa), developed this tool during a post-doctoral position. OpenTermAlign has been tested with PACTOLS and various archaeological vocabularies. It offers the possibility to store the entire alignment process in a SKOS file, which can then directly interface with the concerned applications, thus simplifying both alignment updates and the enhancement of the reference thesauri with new vocabulary.

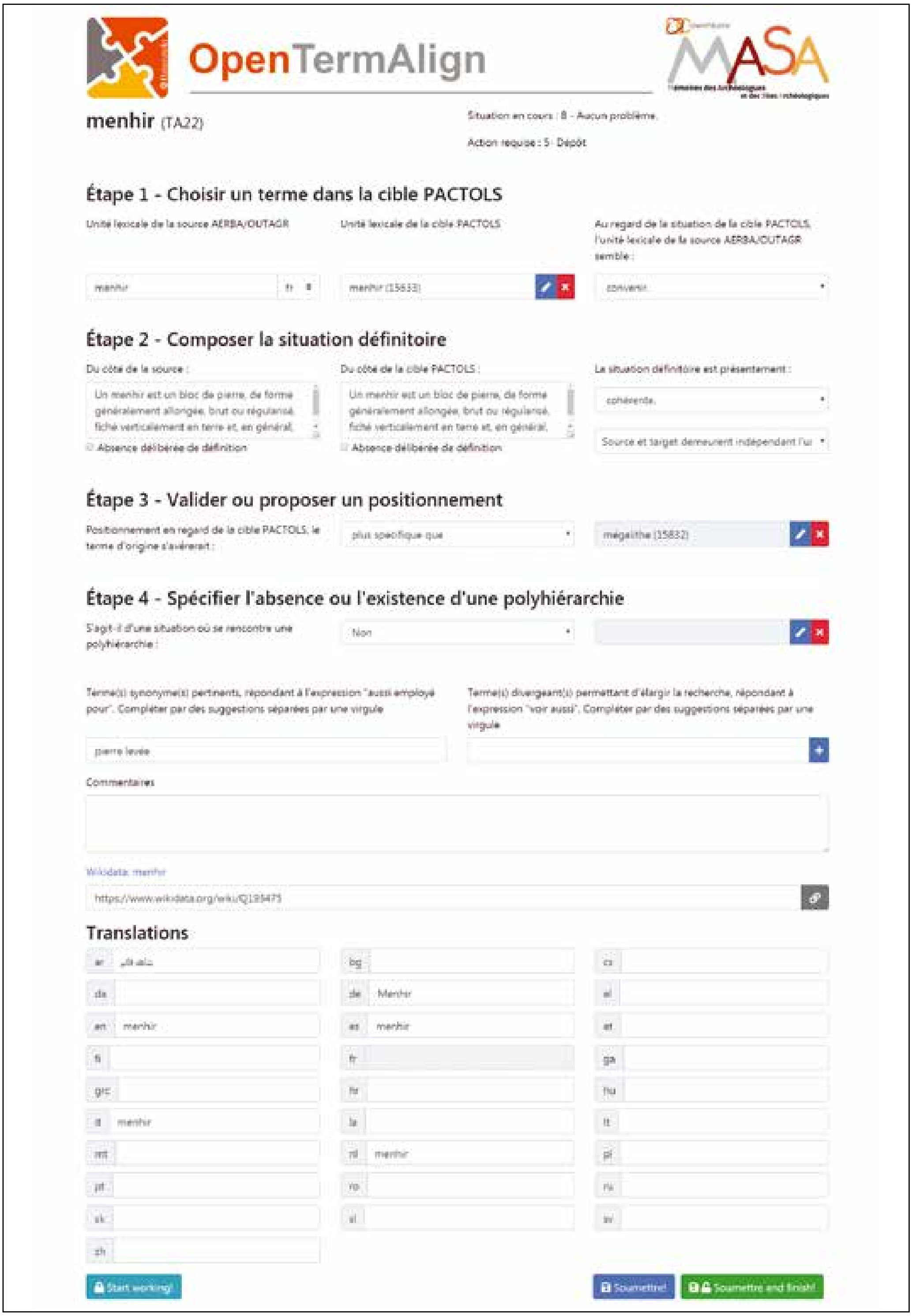


Figure 1: The OpenTermAlign application (prototype)

OpenArchaeo

In the context of ARIADNEplus, the MASA Consortium has agreed to share its datasets, together with OpenArchaeo as an experimental tool which has been devised for exploiting them. More and more projects have mapped older data sets to the CIDOC CRM for interoperability purposes. It is now necessary to have intuitive search tools to explore these interconnected datasets. The MASA Consortium proposes that OpenArchaeo can be this tool for the archaeological community. It provides an online interface for querying several archaeological datasets, meeting the needs of a user-friendly query interface, the use of external thesauri, and API for web services with a SPARQL endpoint. Thus, OpenArchaeo is a way for exploring data from distributed autonomous data providers, without duplicating all data by locally loading it. OpenArchaeo was developed by the LAT with the Sparna company for the MASA consortium (Marlet et al. 2019b).

We devised interfaces for applications, for administrators, and for end-users. OpenArchaeo's end-users are archaeologists, whether researchers or amateurs, that is people who know what an excavation is. For such people, we soon realized that the same generic model we built as a guide for mapping excavation data to the CIDOC CRM should be the right conceptual model for guiding the querying of these data.

The mapping of archaeological databases to the CIDOC CRM is a prerequisite for their interoperability. Many tools are currently available to perform this operation depending on the existing database format. The first dataset to integrate OpenArchaeo is ArSol, the *Laboratoire Archéologie et Territoires* relational database for excavation data management and research, created in 1990 and available online since 2014. For ArSol, we used Ontop, an Ontology-Based Data Access tool designed at the University of Bolzano, which enables us to define mappings with a Protégé plug-in. Those mappings can be used either to directly query the CRM ontology and get results from the connected relational database, or to export the relational database into an RDF graph of CIDOC CRM instances (Marlet et al. 2016). The second dataset is an XML dataset, AERBA, the Atlas of rural settlements in ancient Beauce (France). For this second dataset, we used the Mapping Memory Manager (3M), the online visual application provided by ICS-FORTH team in Heraklion to map an XML dataset to the CIDOC CRM.

In both cases, a generic data model extracted from the CIDOC CRM and its extensions has been implemented, establishing the minimum elements that can be found in most archaeological datasets, even on different scales (*Figure 2*). The same generic model is currently used by the MOM (*Maison de l'Orient et de la Méditerranée,* Lyon, France) for mapping their data on excavations of the Kition-Pervolia site in Cyprus to the CIDOC CRM, in order to add them to OpenArchaeo.

This model is a selection of a few CIDOC CRM, CRMsci, CRMarchaeo and CRMba entities and properties that are necessary and sufficient for representing the core of excavation data. It is important to notice that our objective is to federate several autonomous datasets and the OpenArchaeo platform is not intended to provide

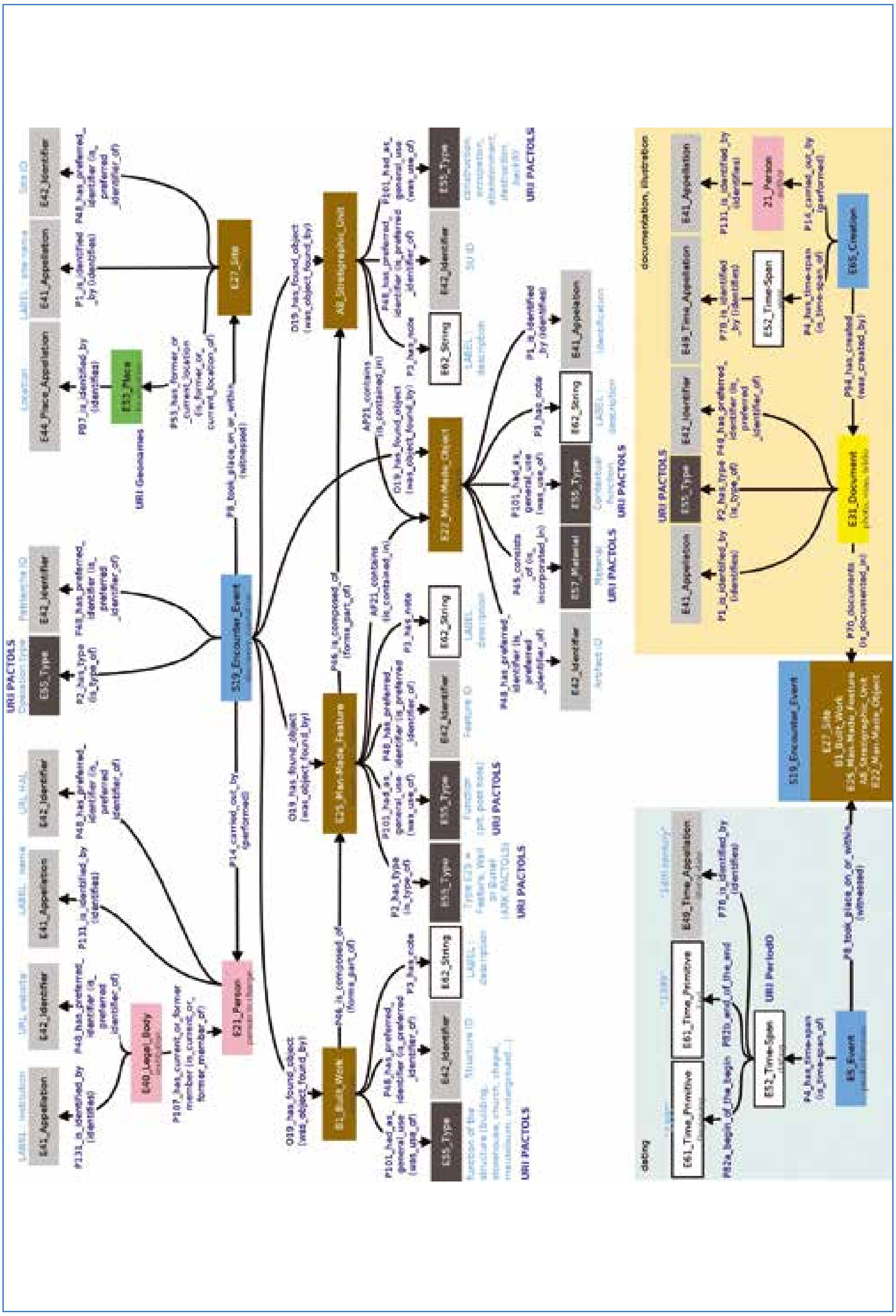


Figure 2: Archaeological generic mapping model for OpenArchaeo with CRMbase 6.2.3, CRMsci 1.2.5, CRMarchaeo 1.4.8, CRMba 1.4

access to all specific elements of each corpus. It is therefore not necessary to perform a completely detailed mapping of each dataset to the CIDOC CRM. Its purpose is to answer fairly simple queries and to provide, in the answers, the URLs to access the detailed records in each data source. If a source handles specific issues, it is by switching from the results given by OpenArchaeo to the online database of this source that the researcher is able to query these specificities more accurately. Thus, the queries concern a general level that is common to most archaeological datasets. This level concerns the site, its location, the person in charge of the operation, the structures, features, walls, burials, stratigraphic units and artifacts.

For each of these items, attention has been paid to standard descriptions, possible dates and related documentation. Following the spirit of CIDOC CRM, the central entity is the event of encountering/excavating a site and this is done under the responsibility of a person. As usual in the CIDOC CRM, each entity is associated with an identifier, a type (preferably from a standard repository) and possibly an appellation. The identifier is generally the inventory number assigned during recording and provides a unique identifier in the dataset. In the case where a permanent identifier has not yet been deployed to identify an online resource (ARK or Handle), it is this record identifier that can be used to access the resource.

The user-friendly interface provides a list of available data sources to start with. In our prototype, we currently have only two data sources but others are in preparation (such as Kition) and we are open to any submissions to increase the resources and improve OpenArchaeo. We devised the intuitive visual query interface of OpenArchaeo based on the generic model and by following the main visual guidelines chosen by ResearchSpace, to the best of our knowledge, the only existing comparable visual querying tool. ResearchSpace was developed at the British Museum based on the CIDOC CRM with the aim of connecting researchers, data and practices. Inspired by this model, a system of icons has been set up to identify the main components of the archaeological data: the site, the operation manager, the archaeological structure, the archaeological feature, the wall, the burial, the stratigraphic unit and the archaeological artifacts. The choice of this thematic entry level in the CIDOC CRM corresponds on the one hand to the desire to address the archaeological community directly, and on the other hand not to go to the highest level of each data set in order to maintain a level of generality necessary and sufficient for interoperability. This is the main difference with ResearchSpace, whose entry into ontology is at a much higher level like 'Think' for example.

We start by selecting the entity for which we want to obtain an answer. Then, OpenArchaeo suggests either questioning directly the criteria of these burials as its dating, or investigating entities directly related to the 'burial' entity. When users wish to connect two elements (burial and site for example), the interface automatically suggests the available relationships between these two entities. This enables users to formulate their request in a simple way without having to know either the entities and properties of CIDOC CRM, or the structure of the system.

The SPARQL queries that correspond to the sentences visually built by users are automatically computed. Thus, the user does not need to know the underlying model or the SPARQL query language. OpenArchaeo relies on a MASA triplestore used as a repository for several datasets, but OpenArchaeo also queries other external triplestores (such as Kition). Thanks to the queries federation, OpenArchaeo returns all the results in a homogeneous way. The standard visualisation is a table of results. If the results are geolocalized (the sites), they can be shown on a map. By clicking on the URI of a result, the user opens the online page of the resource and can consult its details and specificities (*Figure 3*).

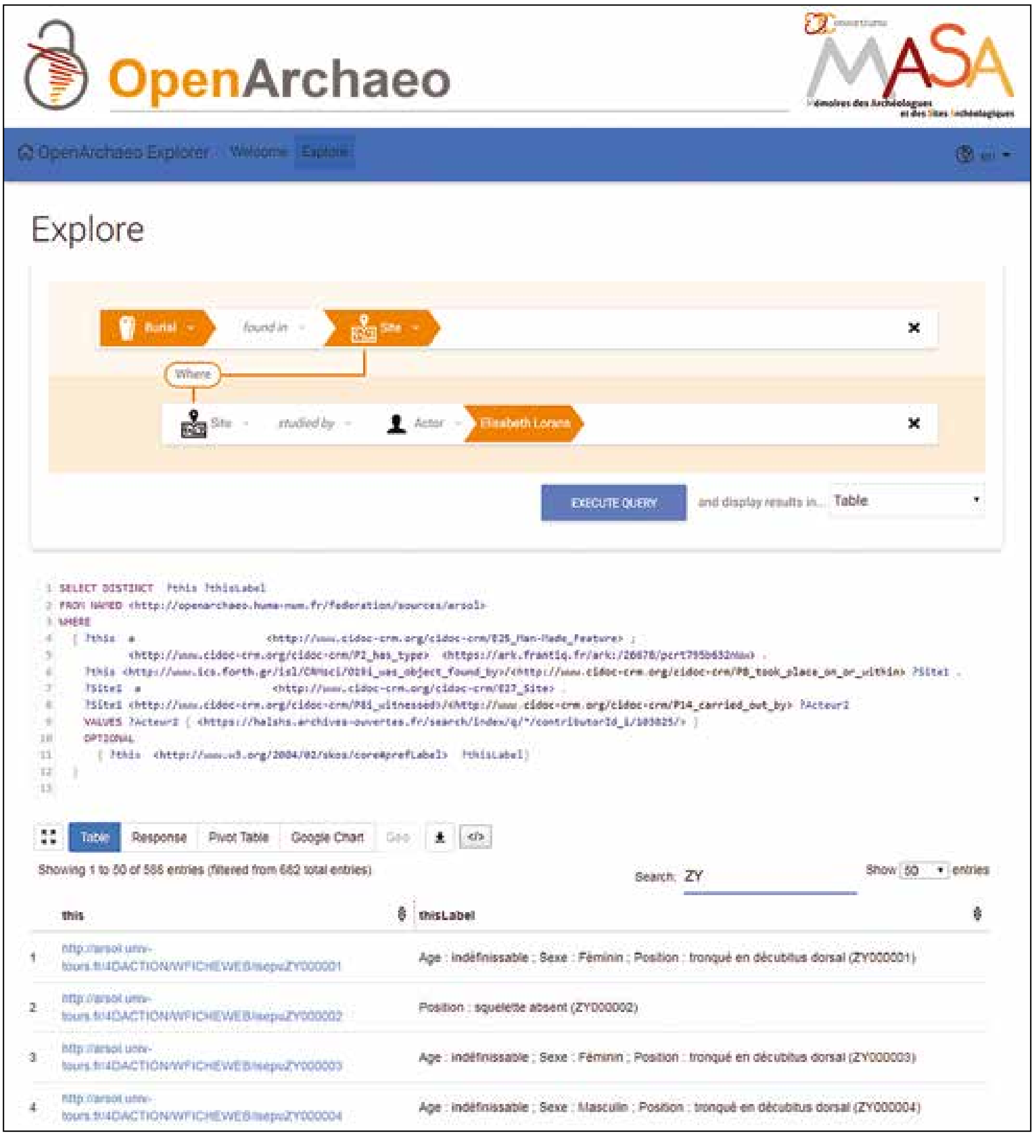


Figure 3: The OpenArchaeo platform (prototype)

OpenArchaeo is able to integrate several external thesauri that are useful for querying excavation datasets. Currently, it enables users to formulate their queries with the vocabulary collected in the PACTOLS thesauri. It is also possible to use GeoNames for spatial searches and PeriodO for temporal searches, as shown in these two examples.

Enabling end-users to explore datasets integrated with OpenArchaeo is of primary importance. To go further, the Linked Open Data cloud currently exists for developers to build innovative applications based on the published linked datasets. The full potential of the CIDOC CRM integrating capabilities is not necessarily limited to the visual querying provided by OpenArchaeo. To make it usable in other ways by everyone, it also provides tools for applications to query the datasets via the integrated access point implemented by OpenArchaeo. In this way, programs in Java, PHP, Python and so on, can use OpenArchaeo's SPARQL Endpoint to automatically compute statistics on the participating datasets, or verify periodically the updates in datasets and, more generally, perform any query that could be built in the user interface. So, OpenArchaeo enables experienced users to interactively explore the datasets with their own SPARQL queries.

OpenArchaeo is designed to be extendable and reusable in other contexts. This can be done at a lower cost, starting from a new mapping model with CIDOC CRM and a new translation of the requests for the interface. It has the advantage of being flexible, in particular it is easy to extend the system with new data providers.

Conclusion

With this digital ecosystem, the MASA consortium relies on the data culture of archaeologists and their long experience in computerization to encourage the community to respect the FAIR principles and to open these corpora as Linked Open Data. The ARIADNE program has set the same objectives and produced enough resources to guide the MASA Consortium's choices when it was launched. It was therefore natural for the MASA Consortium to join the other European partners when the ARIADNEplus programme was renewed.

Joining the ARIADNEplus program is a great opportunity for the MASA consortium because we have the same objectives at the international level. The comparison of the choices made by MASA and the tools proposed with those of the ARIADNE infrastructure enables us to consolidate or redirect them if necessary to respect international standards and FAIR principles as well as possible. The datasets processed by MASA will feed the ARIADNE platform and thus offer a much better visibility to French archaeological data. In addition, the original developments proposed by MASA, such as OpenTermAlign and OpenArchaeo, will be integrated into the ARIADNE platform for the benefit of the entire community. For MASA and these teams, ARIADNEplus offers both a recognition of the work carried out in the consortium and above all the necessary support to boost the transmission of the FAIR principles within the French archaeological community. This community remains to a large extent to be convinced and the services offered by the ARIADNE platform are fundamental to demonstrate the relevance of the approach at European level.

References

Berners-Lee, T. 2006 *Linked Data.* https://www.w3.org/DesignIssues/LinkedData (Last accessed: 25 January 2019)

FORCE 11 2016 *FAIR principles.* https://www.force11.org/group/fairgroup/fairprinciples (Last accessed: 25 January 2019)

Living Laboratory Data Service for the Semantic Web 2014 *Linked Data Finland.* http://www.ldf.fi/ (Last accessed: 17 June 2019)

Marlet O., Curet S., Rodier X. and Bouchou-Markhoff B. 2016 'Using CIDOC CRM for dynamically querying ArSol, a relational database, from the semantic web' in S. Campana, R. Scopigno, G. Carpentiero and M. Cirillo (eds) *CAA2015 Keep the revolution going, proceedings of the 43rd Annual CAA Conference* (Sienna, Italy, 30 March-3 April 2015), Archaeopress Archaeology, 241–250.

Marlet O., Zadora-Rio, E., Buard, P.-Y., Markhoff, B. and Rodier, X. 2019a 'The Archaeological Excavation Report of Rigny: An Example of an Interoperable Logicist Publication', *Heritage* 2019, 2(1), 761–773. doi:10.3390/heritage2010049

Marlet, O., Francart, T., Markhoff, B. and Rodier X. 2019b 'OpenArchaeo for Usable Semantic Interoperability' in Poggi A. (ed) *Proceedings of the First International Workshop on Open Data and Ontologies for Cultural Heritage* (Rome, Italy, 3 June 2019), Central Europe Workshop Proceedings, vol. 2375, 5–14. http://ceur-ws.org/Vol-2375/paper1.pdf (Last accessed: 17 June 2019)

Archaeological digital repositories: Fostering networks from the Global South

Andrés Darío Izeta[*] and Roxana Cattáneo[**]

CONICET, Instituto de Antropología de Córdoba (IDACOR); CONICET,Universidad Nacional de Córdoba, Argentina

Universidad Nacional de Córdoba, Departamento de Antropología y Museo de Antropología, Argentina

*aizeta@ffyh.unc.edu.ar

**roxanacattaneo@gmail.com

Abstract

This chapter considers the role of computing and digital media in archaeology in Argentina. It describes the current state of the art regarding the use of digital methods (for creating, analysing, storing, retaining and reusing data, and as a means of communication amongst various interest groups). PAD, the Digital Archaeology Program of the Museum of Anthropology (IDACOR) is presented as an example of actions taken by public state agencies, including universities, research centres and scientific bodies. Finally, we discuss how PAD has become a game changer in digital archaeology, fostering national and international networks.

Keywords: digitization; thematic repositories; Argentina; Programa de Arqueología Digital; Digital Public Archaeology

Introduction

Archaeology is about studying the past: a past that may seem so far away that it may resemble, as Lowenthal (1995) suggested, a foreign country, or a very recent past, moments before the present. Although the past and how people interact with each other and their environment has always been one of our main interests as archaeologists, the way we approach it becomes everyday more sophisticated and hi-tech. As Zubrow (2006, 11) has said 'technology is a tool to control nature, and archaeologists – similar to other workers – use their new tools to find and control the past'. In other words, archaeologists need modern tools to build new narratives of the past. However, what are these modern tools? Computers may be the answer, but they have been around for more than 40 years now (e.g. Costopoulos 2015; Dallas 2015; Evans and Daly 2006). Therefore, it seems that it is not the hardware itself that will improve archaeological research but the long-lasting growth in processing capacity added to network connectivity.

As soon as computers became available, archaeologists incorporated them into their professional practice. Digital catalogues and databases built from registers of diverse material culture attributes allowed more accurate seriations, organization of items and finally the classification of object similarities (and therefore cultural affinities). As the capacity of computers grew, just as their cost went down, the use of these technologies became more widespread (Evans and Daly 2006). Many tasks that had required much effort became easier when mediated by the use of computing. Spatial analysis, various types of quantification (zooarchaeology, lithic studies, etc.) and the use of statistics led to the creation of various specialisms within the discipline. Quantitative archaeology, spatial archaeology, and virtual archaeology benefited and even originated from the use of computers, different types of software, and especially the connectivity and mobility achieved in recent years.

These technologies also allowed faster communication of archaeological results. In addition, information and communication technologies (ICT) were used to communicate with different publics (e.g. Richardson 2013). In recent years applying for grants, publishing papers or presenting technical reports are only possible in electronic format. The means used for the dissemination, teaching or mere transmission of archaeology were also subject to this digital revolution (e.g. Evans and Daly 2006; Izeta and Cattáneo 2018). The greater ease of access to digital technologies increased the content related to archaeological subjects. Video games based on archaeological topics, films and documentaries made by research teams, independent producers or state agencies began to proliferate, and there has been an increasing presence of web pages, Facebook pages, Twitter or Instagram accounts referring to research teams scattered throughout the country (see Izeta and Cattáneo 2018; Morgan 2016).

In this instance, where we see an accelerated use of computing and digital media we believe it becomes necessary to consider their role in the modelling of our practice as archaeologists and how they connect us with a public interested (or not) in archaeology. Here we need to ask ourselves if we need to be reflexive about 'the digital'. In this sense, what is the role that 'the digital' plays in mediating between archaeologists and non-archaeologists? This question can be answered if we think of the digital as synonymous with Internet-mediated (Boellstorff 2012) and we can paraphrase and say that all archaeology is digital archaeology in one way or another. In fact, we must consider whether there is an ontological transformation or change as proposed by Kockelman (2013) in the sense of the software algorithms (hidden and unknown to the vast majority of users) that silently guide our professional practice. In this context, Richardson (2013) asked whether it is possible to speak of a Digital Public Archaeology, understanding it as the way professional archaeologists adopt to communicate the results of archaeological knowledge to those considered non-archaeologists, in this case through digital media. This question becomes significant when considering that this is a relatively new dimension of contemporary practice and has not yet had much theoretical analysis (Richardson 2013).

To answer it, we must recognize that very early in the disciplinary development of archaeology different ways of communicating the results of explorations, excavations and the analysis of archaeological material were considered. In fact, one of the first ways to present findings to the community of non-experts was made by the 17th-century European museums, which displayed diverse archaeological objects (e.g. Musée du Louvre, Musei Vaticani, Uffizi di Firenze, British Museum, etc.). Their growth in the last quarter of the 19th century has undoubtedly left a mark on the relationship between those who manage the dominant discourse and those who receive it (e.g. Farro 2016; Kristiansen 2012; Podgorny 2009).

More specifically in the early 1970s the first to use the same computer resources applied to archaeology were the same manufacturers (Vanhoutte 2013) who published guidelines on how to use computers, with repercussions for diverse local archaeologies. Examples include the implementation of a computer methodology that aimed to define chronologies for the Argentinian Northwest (Lahitte 1970) or the implementation of semi-automatic lithic artifact classification named TILCO and DELCO (Bellelli et al. 1985–1987; Guraieb and García 1985–1987).

Undoubtedly, from these first contacts between computer science and archaeology, digital records were created of interest for both archaeological research and public interest. However, what happened to these primary data? Obviously, the obsolescence of the equipment, the advance in software programming and the proliferation of networks have made many of these systems inoperable today. As an example, McDavid (2004) presents in her analysis of the use of the internet in the practice of public archaeology some obsolete internet browsers. She also characterizes relationships among different actors through discussion forums, static web pages, collaborative web pages, and more. Just fifteen years ago, McDavid failed to foresee the emergence of social networks and mobility in the use of data, including the proliferation of applications (apps) in mobile devices of diverse formats, from notebooks to tablets, to smartphones. Of course, we also do not know what technologies will be in use in the next fifteen years, but if we understand that changes and innovation will occur, then being reflexive on this type of archaeology practice must be continuous. In this sense, various theoretical developments such as Digital Humanities (Huggett 2012, Terras et al. 2013), the 'Humanidades Digitales' (Rio Riande et al. 2015) or, more generally, digitization and digital sciences (Lauzikas 2007) take on a leading role.

We will try to see how these developments have affected practices within Argentinian archaeology and how it has been influenced by international experiences. We begin with a subject of particular interest to the discipline: the conservation of objects studied by archaeologists and by researchers in the humanities and social sciences in general. In the 1970s the idea appeared of transforming physical objects into digital objects through their description mediated by a set of metadata. With this began the digitization of collections, supported by museums around the world. One of the pioneers was the British Museum through its digital catalogue. It and other institutions focused on trying to define controlled vocabularies and sets of standardized metadata for the management of these large inventories, such as the Getty vocabularies or the UNESCO thesaurus, amongst others.

The 'digital' in Argentinian archaeology

In Argentina, interest in this type of initiative began in the humanities. In the early 1990s projects from the Ravignani Institute of the University of Buenos Aires Digitization Center (Feldgen et al. 2002), the Library of the Faculty of Humanities of the Universidad Nacional de La Plata (Borrel et al. 2015) and other various historical archives of universities and official bodies are good examples. Perhaps the NAyA (Noticias de Antropología y Arqueología, Anthropology and Archaeology News) initiative, created in 1996, was at the time the most innovative in its use of the internet as a means of disseminating archaeological advances.

In any case, the history of the computerization of archaeological collections began long before the governing bodies of science and technology in Argentina became interested in this subject. Towards the end of the 1990s, three institutions began to implement this type of initiative: the Ethnographic Museum (Facultad de Filosofía y Letras, Universidad de Buenos Aires), the Institute of Archaeology and Museum (Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán), and the Museum of Anthropology (Facultad de Filosofía y Humanidades, Universidad Nacional de Córdoba). All these initiatives led to products that facilitated collections management, generally focussed on complete artifacts and, in several cases, created a digital catalogue.

On this basis, the development of facilities dedicated to the digitization of collections of scientific interest began to be defined in three institutional spaces that over time provided infrastructure, methodologies and training to various digital initiatives: the National System of Digital Repositories – SeCyT (Secretary of Science and Technology), the PLIICS-CONICET (National Research Council Interactive Research Platform for Social Sciences) (Izeta and Cattáneo 2018; Leff and Pluss 2015; Pluss and Leff 2013) and the Open Access and Digitization Program of Collections in Social Sciences and Humanities-MINCyT (Borrel et al. 2015). In particular, PLIICS has promoted the development of capacity at the individual and institutional level for the creation, use, storage, stewardship, preservation and dissemination of digital data generated by archaeological projects.

Following this intermittent progress towards digitization, and in line with the implementation of PLIICS, we began the project 'Support for the computerization of documentary archives and collections of the IDACOR-Museum of Anthropology (FFyH, UNC), CONICET'.

After years of experience in the design and implementation of a technical repository, our interest is focused on recognizing the need for interaction with other repositories and institutions that develop these processes. A search for possible partners was undertaken and the ADS (Archaeology Data Service, UK) appeared as a potential collaborator, which ultimately allowed us to get involved in international projects such as ARIADNEPlus. This e-infrastructure will enable us to link our data to others around the world, opening up access to Argentine data, and allowing cross-search and support for the development of our repository.

Furthermore we were able to position the digitization and computerization of archaeological collections as the area inside the institution receiving the highest funding (proportionally even more than the research area), based on the grants received from various non-governmental organizations such as the Williams Foundation, Bunge and Born Foundation–CONICET and federal funding by PLIICS-CONICET. As a by-product, training in the creation of digital data for use by archaeologists and non-archaeologists was also achieved.

All of the above, along with the concepts of open source software, open access, and open science (Willinsky 2005), has provided the first opportunity to publicly make available not only archaeological interpretations: the dominant message, but also the primary material (the 'raw' and unprocessed data, if such exists) from which the discourse of professional archaeologists is generated. This arrives in time to begin a process of democratization of science and knowledge, although we admit that this will not reach everyone as the infrastructure and resources to access technology is not available to all individuals (see McDavid 2004; Richardson 2013). In any case, preserving these digital data and the channels of communication and the results generated in the interaction of the different actors involved will allow the construction of a corpus of information accessible to those interested in these issues.

Suquía: The archaeology digital thematic repository

As noted above, the use of digital data by archaeologists is now an everyday task (Costopoulos 2016). However, a lack of a digital approach in Argentinian archaeology has been identified. That is why, based on experience in projects focused on this problem, the Digital Archaeology Program of the Anthropology Museum (PAD) was formed in 2015. The goal of the PAD is to cover a vacancy within regional archaeology that has to do with the need to generate, preserve and disseminate archaeological data in digital format. For this, we rely on an important volume of digital objects like data and metadata from archaeological objects and complementary material associated with them (field notes, notebooks, publications, among others). At the same time, an important component of this type of practice is the use of all available resources to achieve communication between digital data generators and consumers. Therefore, the PAD can be found in various digital spaces (mainly internet) such as social networks (e.g. Facebook, Twitter, Instagram), digital institutional repositories, and web pages (static and dynamic). It should be noticed that these resources are used by various research teams including Arqueología de La Pampa Norte, Patrimonia, Arqueología de la región del Salado, Proyecto Arqueológico Miriguaca, Arqueología de Ambato-UNC and Proyecto Arqueológico Ongamira, among other Argentinian examples. In most cases these spaces are used to present news, introduce the professional group and provide some contact information. At this point it is remarkable how in the last year the interest in digital formats has advanced within the discipline, either as a means

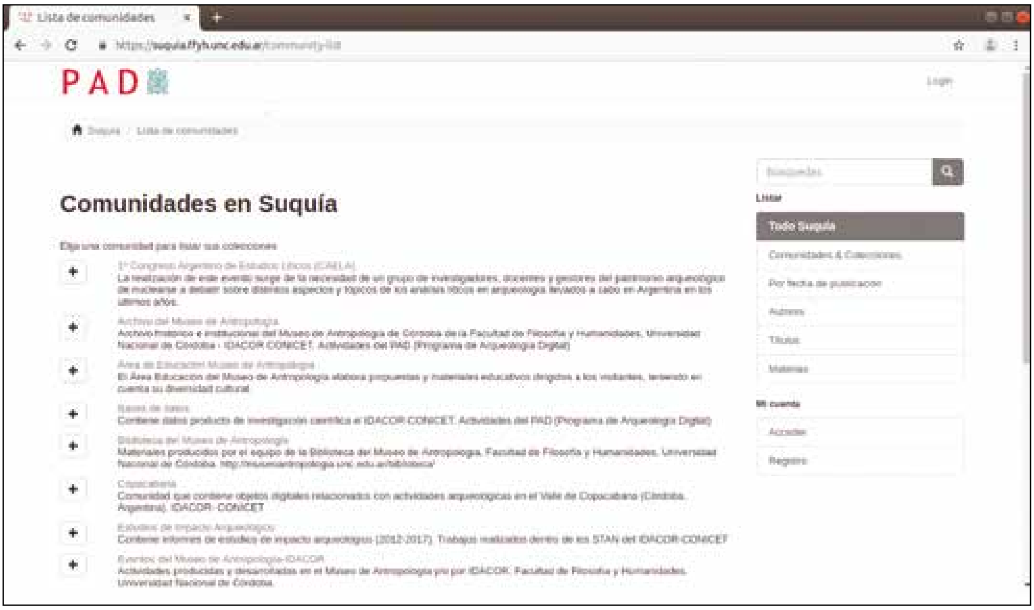


Figure 1: Snapshot of the Suquia Repository Online Public Access Catalog (OPAC). Some of the communities are shown

for communication (following the model of the 'authorized heritage' discourse or 'top-down' model as defined by Belford 2013) or as a means by which to generate new archaeological knowledge.

However, the existence of archaeology digital thematic repositories has not yet been widely developed in Argentina and general in South America. However, this kind of infrastructures exists in other countries like the Archaeology Data Service (ADS, UK; Richards 2017), the Digital Archaeological Record (tDAR, USA; Kintigh 2006) and Open Context (USA; Kansa 2010) among others (see Meghini et al. 2017).

In the framework of an Open Science project, the accessibility of primary data is of fundamental importance. In this sense, having adequate means becomes indispensable, which is why in 2016 an Institutional Digital Repository was created focusing on archaeology known as Suquía.[1] This was built according to the FAIR principles (Findable, Accessible, Interoperable, Reusable). Therefore, we chose a series of repository attributes such as the use of a universal metadata scheme (Dublin Core), free software with an active support community (DSpace) and compliance with the requirements of national laws regarding the deposit of primary data.

The Suquía Repository contains various types of information that have a common feature: difficult access by the community in their original format. Databases, reports, presentations at scientific events, images of excavations, and images of archaeological collections, are just some of the data types contained in the repository. Among

1 https://suquia.ffyh.unc.edu.ar/

these are collections such as the Heritage Reserve of the IDACOR - Museum of Anthropology which hosts collections of several Argentinian regions such as the Northeast, Patagonia, Córdoba and the Northwest, among others; the Publications of the Institute of Archaeology, Linguistics and Folklore and the Institute of Anthropology; and those of archaeological sites or areas worked by current projects (e.g. Alero Deodoro Roca Sector B; Parque Natural Ongamira 1; and Copacabana). There are also a variety of other digital objects that correspond to spreadsheets, photos, text files, GIS files, inventories, databases, manuscripts, posters, brochures, books (in their various stages). To date there are over 2000 digital objects available under Creative Commons licenses for use and reuse both for archaeological research and for those interested in the subject.

An Argentinian repository federation

Given the above scenario, our next goal was to transfer our experience to different Argentine institutions to build a network that connects these repositories. PLIICS and other programs, together with private foundations, are encouraging the creation of new repositories, which will eventually lead to a consortium or Federation of digital repositories oriented to archaeology. Fifteen institutions representing the entire Argentinian country are involved in the creation of this network, as suggested by the authors along with Julian Richards of the ADS in a meeting held in CONICET in October 2017 where a number of archaeologists, science managers, and members of the third sector discussed this topic. The Digital Archaeology Program together with the Williams Foundation and CONICET (National Research Council of Argentina) have begun working on a program aimed at generating interest in repositories, offering the possibility of generating new spaces or including new data in those already available. This task has just begun, but it seems to be promising and, without a doubt, the way of seeing, using and preserving archaeological data has changed at the national level.

In this same line, inclusion in other networks beyond South America has been achieved, including ARIADNEplus. In this case, we are also the only representatives of Latin America. However, we believe that this situation of uniqueness is a challenge to overcome and with it, we can offer a geopolitical vision for archaeological data from the global south. In other words, we can make the production of knowledge and archaeological theory visible from the periphery of the academic world.

Digital repositories and public outreach

As for digital archaeology as public archaeology, it should be noted that according to Bonnin (2015), the development of public archaeology in Argentina has been consolidated in recent years. This is undoubtedly true in some academic institutions and research centres. However, in order to measure how developed this is within archaeology, a survey was undertaken of the subjects applied for in the last three CONICET national calls. We assume that this represents the current topics of interest for active archaeological researchers. The result is that less than 6% of the applications are related in some way to Public Archaeology. However, if we reflect on the definition of public archaeology (e.g. Merriman 2004) in which we must assume that all Archaeology is public (as we proposed above that all archaeology is digital) then only a very small portion of archaeologists define or implement projects or actions aimed at this disciplinary field. However, we should notice this type of practice and approach from archaeology has been implemented relatively recently in Argentina (Salerno et al. 2016). In fact, much of the development of public archaeology is due to local issues: the relationship between archaeologists and communities (in the broad sense, but also specifically aboriginal communities); the transfer of archaeological knowledge through formal and non-formal education; the need to intervene in archaeological impact studies; and, the need to use new forms of dissemination based on the interrelation with diverse interest groups. Fabra et al. 2015 identify key issues for public archaeology in Argentina: education, heritage, treatment and restitution of human remains, rescue archaeology, multivocality (community archaeology?), Sociology of archaeology, tourism and archaeology, among other topics. However, in all this analysis it is not possible to find references to the role of digital media (recording, conservation, communication) in the construction of these themes.

However, the rise of archaeological digital repositories is becoming a game changer, especially as new ways of networking at the national and international level are available. In this respect, ARIADNEPlus is laying the foundations for Argentinian archaeologists to start practising different modes of thinking and doing archaeology from a global perspective.

Conclusion

Undoubtedly, the new field that opens up within archaeology will demand more development, both in practice and in reflexivity on the theoretical aspects that underlie it. However, the current reality reveals an active community that is generating digital information and is eager to share it. In fact, a survey carried out within the framework of the PLIICS in 2012 showed that of 730 Argentinian researchers from the social sciences and humanities 61% have digitized the primary data product of their research. Of this same sample, 87.5% answered that these digital data can be released to the public. This demonstrates that researchers (including archaeologists) are willing to share their digital data.

There is also legislation (Act 26899: Creation of Digital Access Repositories Open Access, Own or Shared) that is changing the rules in requiring free access to digital information obtained from the research process by archaeologists who receive state funding. Undoubtedly, digital archaeology will take a leap in quantitative and qualitative development in terms of digital information available and in the number of institutional or thematic repositories that are created in the not too distant future.

In summary, it can be observed that there has been progress in applying digital products in the research process and work is being done to make digital information available to researchers and the general public through web applications. Although this process is still under development it must be recognized that Digital Archaeology is here to stay and in Argentina it will become a process that improves interaction between archaeologists and non-archaeologists, advancing multi-vocality in archaeological interpretations.

Acknowledgements

Several institutions and people walked with us in the last seven years in order to achieve a digital repository dedicated to archaeology. Here we want to thank some of them. CONICET-PLIICS, Fundación Williams, Fundación Bunge & Born, Mirta Bonnín, Andrés Laguens, Eduardo Pautassi, Silvia Mateo, Thiago Costa, Julián Mignino, Isabel Prado, Paloma Laguens, Jorge Tezón, Laura Leff, Ricardo Pluss, Martín Oliver, and Julian Richards.

References

Belford, P. 2011 'Archaeology, Community, and Identity in an English New Town', *The Historic Environment: Policy & Practice,* 2(1), 49–67.

Bellelli, C., Guraieb, A.G., and García, A. 1985–1987 'Propuesta para el análisis y procesamiento por computadora de desechos de talla lítica (DELCO – Desechos líticos computarizados)', *Arqueología Contemporánea,* II(1), 36–53.

Boellstorff, T. 2012 'Rethinking Digital Anthropology' in H.A. Horst and D. Miller (eds) *Digital Anthropology,* Oxford: Berg Publishers, 39–60.

Bonnin, M. 2015 'Prólogo' in M. Fabra, M. Montenegro and M.E. Zabala (eds) *La Arqueología Pública en Argentina. Historias, tendencias y desafíos en la construcción de un campo disciplinar.* Jujuy: Ediunju, 1–3.

Borrel, M., Fuente, M. V. and González, C. 2015 'Una propuesta metodológica de relevamiento para iniciar proyectos de digitalización y preservación' in G. Rio Riande, L. Cantamutto and L. Striker (eds) *Actas de las I Jornadas de Humanidades Digitales*, Ciudad Autónoma de Buenos Aires: Editorial de la Facultad de Filosofía y Letras Universidad de Buenos Aires, 69–84. doi:10.5281/zenodo.2875811

Costopoulos, A. 2016 'Digital Archaeology is here (and has been for a while)', *Frontiers in Digital Humanities* 3:4. doi:10.3389/fdigh.2016.00004

Dallas, C. 2015 'Curating Archaeological Knowledge in the Digital Continuum: from Practice to Infrastructure' *Open Archaeology,* 1(1), 176–207. doi:10.1515/opar-2015-0011

Evans, T.L. and Daly, P.T. 2006 *Digital Archaeology: Bridging Method and Theory.* London: Routledge.

Fabra, M., Montenegro, M. and Zabala, M.E. (eds) 2015 *La Arqueología Pública en Argentina. Historias, tendencias y desafíos en la construcción de un campo disciplinar.* Jujuy: Ediunju.

Farro, M. 2016 'Natural History Museums in Argentina, 1862–1906', *Museum History Journal,* 9:1, 121–134. doi:10.1080/19369816.2015.1118257

Feldgen, M., Clúa, O., Boro, F. and Santos, J.J. 2002 'Argentinean historical heritage project' *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries.* ACM: New York, 392.

Guráieb, A.G. and García, J.A. 1985–1987 'Sistema para el procesamiento por computadoras de instrumentos y artefactos líticos. TILCO (Tipología lítica computarizada)', *Arqueología Contemporánea* 2 (1), 23–35.

Huggett, J. 2012 'Core or Periphery? Digital Humanities from an Archaeological Perspective', *Historical Social Research / Historische Sozialforschung,* 37(3), 86–105.

Izeta, A.D. and Cattáneo, G.R. 2018 ¿Es posible una Arqueología Digital en Argentina? Un acercamiento desde la práctica. *Actas Congreso Internacional Humanidades Digitales: Construcciones locales en contextos globales.* Asociación Argentina de Humanidades Digitales: Buenos Aires. doi:10.5281/zenodo.1239201

Kansa, E.C. 2010 'Open Context in Context: Cyberinfrastructure and Distributed Approaches to Publish and Preserve Archaeological Data', *SAA Archaeological Record* 10(5), 12–16.

Kintigh, K. 2006 'The Promise and Challenge of Archaeological Data Integration', *American Antiquity*, 71(3), 567–578. doi:10.1017/S0002731600039810

Kockelman, P. 2013 'The anthropology of an equation. Sieves, spam filters, agentive algorithms, and ontologies of transformation', *HAU: Journal of Ethnographic Theory* 3(3), 33–61. doi:10.14318/hau3.3.003

Kristiansen. K. 2012 'Archaeological communities and languages' in R. Skeates, C. McDavid and J. Carman (eds) *The Oxford Handbook of Public Archaeology,* Oxford: Oxford University Press, 461–477.

Lahitte, H.B. 1970 *Arte y arqueología. El análisis documental sobre piezas de la cultura santamariana.* Monografía N° 5. Buenos Aires: Museo Etnográfico Municipal e Instituto de Investigaciones Antropológicas.

Lauzikas, R. 2009 'Digitization as a Science', *Archeologia e Calcolatori* 20, 247–259.

Leff, L. and Pluss, R. 2015 'Plataforma Interactiva de Investigación en Ciencias Sociales' in G. Rio Riande, L. Cantamutto and L. Striker (eds) *Actas de las I Jornadas de Humanidades Digitales,* Ciudad Autónoma de Buenos Aires: Editorial de la Facultad de Filosofía y Letras Universidad de Buenos Aires, 137–142. doi:10.5281/zenodo.2875811

Lowenthal, D. 1985 *The Past is a Foreign Country.* Cambridge: Cambridge University Press.

McDavid, C. 2004 'Towards a More Democratic Archaeology? The Internet and Public Archaeological Practice' in N. Merriman (ed) *Public Archaeology,* London: Routledge, 159–187.

Meghini, C., Scopigno, R., Richards, J., Wright, H., Geser, G., Cuy, S., Fihn, J., Fanini, B., Hollander, H., Niccolucci, F., Felicetti, A., Ronzino, P., Nurra, F., Papatheodorou, C., Gavrilis, D., Theodoridou, M., Doerr, M., Tudhope, D., Binding, C. and Vlachidis, A. 2017 'ARIADNE: A Research Infrastructure for Archaeology'. *Journal on Computing and Cultural Heritage (JOCCH) - Special Issue on Digital Infrastructure for Cultural Heritage, Part 2.* 10.3. doi:10.1145/3064527

Merriman, N. 2004 'Introduction: Diversity and Dissonance in Public Archaeology' in N. Merriman (ed) *Public Archaeology.* London: Routledge, 1–17.

Morgan, C. 2016 'Video games and archaeology', *SAA Archaeological Record* 16(5): 9–15.

Pluss, R. and Leff, L. 2013 'Plataforma Interactiva de Investigación para las Ciencias Sociales' in *Anales de las 42JAIIO, Jornadas Argentinas de Informática, – 7º Simposio Argentino de Informática en el Estado,* 174–188. http://42jaiio.sadio.org.ar/proceedings/simposios/Trabajos/SIE/14.pdf (Last accessed 12 May 2016)

Podgorny, I. 2009 *El sendero del tiempo y de las causas accidentales: los espacios de la prehistoria en la Argentina, 1850–1910.* Rosario: Prohistoria Ediciones.

Richards, J.D. 2017 'Twenty Years Preserving Data: A View from the UK', *Advances in Archaeological Practice,* 227–237. doi:10.1017/aap.2017.11

Richardson, L. 2013 'A Digital Public Archaeology?', *Papers from the Institute of Archaeology,* 23(1), 1–12. doi:10.5334/pia.431

Rio Riande, G., Cantamutto, L., and Striker, L. 2015 *Actas de las I Jornadas de Humanidades Digitales.* Ciudad Autónoma de Buenos Aires: Editorial de la Facultad de Filosofía y Letras Universidad de Buenos Aires. doi:10.5281/zenodo.2875811

Salerno, V.M., Picoy, M.C., Tello, M., Pinochet, H.C., Lavecchia, C., and Moscovici Vernieri, G. 2016 'Lo "público" en la arqueología argentina', *Chungara, Revista de Antropología Chilena.* doi:10.4067/S0717-73562016005000015

Terras, M., Nyhan, J. and Vanhoutte, E. (eds) 2013. *Defining Digital Humanities. A Reader.* Farnham: Ashgate Publishing Limited.

Vanhoutte, E. 2013 'The Gates of Hell: History and Definition of Digital Humanities Computing' in M. Terras, J. Nyhan, and E. Vanhoutte (eds) *Defining Digital Humanities. A Reader*. Farnham: Ashgate Publishing Limited, 119–153.

Willinsky, J. 2005 'The unacknowledged convergence of Open Source, Open Access, and Open Science', *First Monday* 10(8). http://firstmonday.org/ojs/index.php/fm/article/view/1265 (Last accessed 12 May 2016)

Zubrow, E. 2006 'Digital Archaeology: The Historical Context' in T. Evans and P. Daly (eds) *Digital Archaeology: Bridging Method and Theory.* Routledge: London, New York, 10–31.

Prospects and potential for the comprehensive database of archaeological site reports in Japan

Yuichi Takata*, Akihiro Kaneda** and Dessislava Veltcheva***

NARA: Nara National Research Institute for Cultural Properties, Nara, Japan
*ytakata@nabunken.go.jp
**a-kaneda@nabunen.go.jp

University of Oxford, UK
***bgrdessislava@gmail.com

Abstract

A vast number of archaeological site reports is published annually in Japan. As archaeologists rely on accumulated data in their research, it is vital that they continue publishing these reports. However, the quantity of information is already so colossal that it is humanly impossible to grasp it in its entirety and a searchable database is needed. For that purpose, the Agency of Cultural Affairs has asked the Japanese local governments to upload all the site reports related to sites under their jurisdiction to the Comprehensive Database of Archaeological Site Reports in Japan *(CDASR)*[1] *in PDF format. As a result, the information for a searchable database is accumulating rapidly. In this paper, firstly we will introduce the way excavations are conducted in Japan and how this is related to the accumulation of data. In the second part, we will focus on the nature and functions of the CDASR.*

Keywords: Japan; site reports; NLP; site summaries; thesaurus

The Nara National Research Institute for Cultural Properties

The *Comprehensive Database of Archaeological Site Reports in Japan* (henceforth CDASR) is operated by the Nara National Research Institute for Cultural Properties (NARA). NARA is a national institute whose primary function is to conduct comprehensive research of cultural properties. It carries out various activities, including:

1. providing training programmes for cultural property experts from local governments and foreign researchers;
2. participating in various joint research projects with domestic and foreign institutions;
3. collecting and publicizing information concerning cultural properties (Tanaka 1984).

NARA has a massive library that stores around 400,000 books and 5000 magazines on cultural properties. It receives a staggering 8000 books annually.

1 https://sitereports.nabunken.go.jp/en

Excavations in Japan

The governmental policy towards excavation is based on the presumption that archaeological sites will be destroyed to give space for development projects. The details of this policy can be found in the *Law for the Protection of Cultural Properties* (1950). A summary is given below.

Article 95 states that detailed maps of all registered archaeological sites of each prefecture of Japan must be made and that these maps must be publicly accessible. There were 465,021 registered sites marked on the maps in 2012 (The Agency for Cultural Affairs the Cultural Properties Department Monuments and Sites Division 2017).

Article 93 and 94 state that if a developer wishes to conduct development work that includes digging up the ground on registered sites, they must notify the prefectural administration about their plans. In 2014 there were 50,859 application forms submitted for that purpose. The annual number submitted is on the rise.

Article 93 states that if the developer plans to damage a registered site during construction, the local government must request a pre-construction excavation from the developer. In 2015 there were 8,184 such excavations. In the last fifteen years, there was a steady number of c. 8000 pre-construction excavations annually.

Article 99 states that local governments (both prefectural and municipal) have the right to conduct excavations. The local governments employ an archaeological administrator whose work is to supervise excavations related to development projects. In 2015, there were 5,724 such administrators: 1,889 on the prefectural level and 3,835 on the municipal level. These specialists are university graduates who have majored in history or archaeology and are then hired either as archaeological administrators or curators. As the number of development projects changes according to the state of the economy, the number of excavations and archaeological administrators fluctuates as well.

The local government conducts the excavations as a general rule; however, in some cases, a commercial archaeological company may be hired instead. The developer is obliged to bear the expense of the excavation in each case. In 2017 the total cost of pre-construction excavations was 60.5 billion yen. Excavations were at their peak in 1997, when a total of 132.1 billion yen was expended. The annual funds have halved in the last 20 years.

A site report containing all the data gathered from the site should be published within three years of an excavation. It is obligatory for the developer to cover the cost of everything up to the publication of the final report, including the processing of the unearthed materials (For details see Tanaka 1984).

The current situation

Site reports act as a replacement for the actual sites that get altered or destroyed in development projects. Therefore, the secure storage of the reports is essential. To prevent the loss of site reports in case of catastrophe, 300 copies of every report are printed and distributed among various institutes all over Japan. These institutes include the National Diet Library, local public libraries, and universities and research institutes with faculties related to archaeology or historiography. By dispersing the copies, the chances are increased that some of them will remain intact.

Around 1,700 site reports are published annually. For example, there were 1,859 in 2010, 1,820 in 2011, 1,672 in 2012, 1,739 in 2013, 1,654 in 2014 and 1,451 in 2015. The total number of site reports published since the start of the twentieth century is estimated to be somewhere between 100,000 and 200,000. NARA is working on making a complete catalogue for every prefecture to define the exact number. Based on the trends in the data processed already, Takata estimates that the total number of site reports must be c.125,000 (Takata 2019). This equals to roughly 15.57m pages with 9.7 billion characters and 9.5m images (including drawings, photographs, charts and graphs). However, so far, only a small number of catalogues are finished, so this is still a very early estimate.

The accumulation of information is vital for archaeology and historiography. Our knowledge of the past can only be deepened by accumulating new pieces of knowledge. Therefore, researchers generally think that the more information, the better. However, as the data amass, it becomes harder to find the exact information one seeks.

The history and policy of the various databases related to site reports

The titles of the site reports usually do not convey much about the excavations themselves. It would not be possible to begin with, because the site reports contain information about many different finds from various time periods. As the title is not descriptive enough, researchers previously needed to read through the whole reports to be able to tell if the reports contained the information they were searching for or not. This was very inefficient as while some reports can be as short as a dozen pages, others can be over a thousand pages. (The number of pages varies depending on various factors including whether the report is a preliminary or a full site report; or the size of the excavation.) To solve this problem, the Agency for Cultural Affairs asked the local governments to start attaching summaries to their reports in 1994 (Morimoto 2017).

The summary includes the name, location (address), position (latitude and longitude), size, type, age(s) of the site; the date and reason for the excavation; and brief lists of the structural remains and excavated materials found. This made identifying the contents of the reports much easier.

However, the circulation of the summaries was very limited because they were attached to the site reports, and those were scattered around the country. In short, if one had no access to the actual report, one could not read the summary either. To amend this, the Agency for Cultural Affairs asked the local governments to start submitting the summaries to NARA on an annual basis. In turn, NARA organized the accumulated metadata and published it online in the form of a database.

The role and functions of the CDASR

The site summaries are an effective tool for grasping the contents of the site reports. However, there are still problems that a simple database of summaries cannot solve:

1. The contents of the summaries are heavily influenced by the knowledge and interest of the excavator. Because site summaries only contain brief descriptions, the terminology and the included information is up to the personal interests of the excavator.
2. The summaries are only useful for grasping the contents of the reports, they do not replace the reports themselves. Researchers cannot base their work on mere outlines. They need to know the details of the excavations.
3. The limited accessibility of the site reports creates inequality among the researchers. Researchers who belong to institutions that have extensive collections of site reports are at an advantage over those who need to borrow them from other institutions.

To resolve this situation, 21 national universities joined forces under the name of *Zenkoku iseki shiryō ripojitori purojekuto* [Nationwide Excavation Site Data Repository Project] in 2008 and started sharing the full text of the site reports on the internet. Their goal was to make the invaluable information contained in the reports easily accessible to anyone. In June 2015, NARA took over this project and combined the individual data of the 21 universities into a monolithic database. And with that, the *Comprehensive Database of Archaeological Site Reports in Japan* was born (Takata 2016). Although NARA manages the CDASR itself, NARA is not alone in this project. The CDASR is a joint effort with various universities, local governments, museums, public foundations, and scientific societies.

The policy of the Agency of Cultural Affairs on the digitization of site reports

Eventually, it became evident that having digital versions of the whole site reports publicly available on the internet was a working solution to the problem of limited access to these reports. As such, more and more research facilities decided to follow this practice. This brought up the question of the ideal format of the site reports themselves.

There are two essential requirements for site reports:

1. Be as sustainable as possible. The primary function of the site reports is to conserve invaluable data about sites that have been destroyed.
2. Be as accurate as possible. Researchers need accurate data for their work, including not only the text but the graphs, maps, drawings and photographs as well.

The Agency of Cultural Affairs started actively sharing its opinion on these matters in 2017. The summary can be found in the three volumes of *Maizo-bunkazai hogo gyōsei ni okeru dejitaru gijutsu no dōnyū ni tsuite (hōkoku)* [Introduction of Digital Technology to the Protection of Buried Cultural Properties] (The Agency for Cultural Affairs: Cultural Properties Department Monuments and Sites Division 2017). Volume one is about the introduction of digital cameras (March 2017), volume two is about the digitization of site reports (Sept 2017), and volume three is about the digitization of primary sources (in press). (NARA assisted in the making of these policies by providing data regarding the technological side of the problems.)

Volume two defines the pre-requisites of a site report as follows:

1. it must be in a format that is long-lasting;
2. images must be in high-quality;
3. it must be easily accessible to anyone.

After confirming these pre-requisites, the volume goes on to describe the possible benefits of digitization. It considers the merits and disadvantages of three possible formats for the site reports: hard copy, high-quality PDF, and low-quality PDF. Hard copies are the most secure, have the best possible image quality and provide unquestionable authenticity. However, because their circulation is limited, public access is problematic. Also, a full-text search is not possible. Both versions of the PDF make full-text search possible. High-quality PDF provides excellent image quality, but the file size is too big for practical usage. The low-quality PDF files do not provide the needed quality in the images; however, their small file size makes them ideal for practical usage. The conclusion is that as both the digital versions and the hard copy have their merits and demerits, none can replace the other, and must, instead supplement each other. Namely, low-quality PDF should be used as an index that guides the users to the hard copies. Lastly, the report closes with the statement that: 'The CDASR of NARA is a system that has already overcome the various problems related to making low-quality PDF files public [online], and therefore everyone is advised to participate [in their endeavour].' As an outcome of this, institutes from all over Japan began applying to join the project. As of May 2019, there are 912 registered institutions.

The search system and database integration

The database for the site summaries and the database of the full-text site reports at NARA were initially two distinct endeavours. However, in June 2019 the site summary database was merged into the CDASR (i.e. the full-text database). At the same time, two other site summary databases previously managed by different institutions were also integrated.

As a result, by 11 June 2019 the CDASR contained a total of 125,641 site summaries and 61,758 pieces of bibliographical information for 1,583 organizations. (The bibliographical data includes information not only about site reports but about some other type of documents, such as site maps, as well. However, their number is negligible compared to that of the site reports.) Out of the 61,758 documents listed in the bibliographical data, 23,435 have full-text PDF files available.

One might think that site summaries became redundant as we can now perform a full-text search. However, a full-text search gives too many irrelevant results. And as we have pointed out already, the title of the reports does not contain enough information about the sites. This is where the site summaries are useful, as they provide detailed metadata for combined searches. Therefore, having a database where the site summaries, the bibliographical data and the actual site reports are all linked together is indeed a meaningful and practical solution. This combination gives users a fast and accurate tool to find what they need.

Further functions and potential of the CDASR

Using keywords to search for something presumes that the user knows the right terminology. As such, it is not the best method for those who do not possess specialist knowledge in archaeology. At NARA we want to provide alternative search methods that anyone can use. This is primarily because we believe that providing easy access to archaeology is important to gain the understanding of Japanese society towards cultural properties. As such, we always try to provide intuitive search methods and functions for CDASR users.

Bubble charts

There are 1.8 billion characters worth of text data in the CDASR. We took that data and extracted the most common 70,411 terms related to archaeology and cultural properties and built an internal thesaurus. We have also classified the terminologies into three separate categories: excavated materials; structural remains; or other (Takata and Nagashima 2018). Then we developed various functions based on this thesaurus. One of them is the ability to show bubble charts.

Since April 2017, we have provided two kinds of bubble charts. One shows the frequency of the terminologies, the other shows their regional characteristics. In both types, the user can click on any of the words to send a query to the database. In the first type, the bigger the bubble, the more commonly the word is used. The bubbles are also represented in different colours according to the category to which they belong. Besides the nationwide bubble chart, there are separate charts for the individual prefectures as well.

The second type – given its nature – is only available for the individual prefectures. This type of chart is made by showing the terminology that is frequently used in a given prefecture and at the same time is uncommon in the other prefectures.

Both of the charts are based on natural language processing (NLP) technology. There are many creative ways these charts can be used. We will give two examples here.

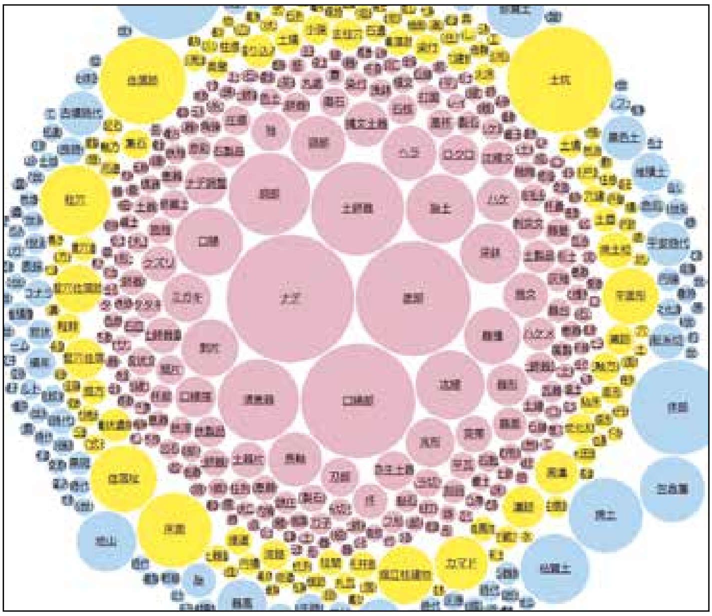


Figure 1: Bubble chart showing terms frequently used in reports

Example 1:
If we take a look at the bubble chart that represents the unique terminologies of Fukuoka prefecture, we can see that one of the peculiar words is *kamekanbo* (jar-burial)[2]. And indeed, this type of burial was something especially common in the northern parts of Kyūshū, where Fukuoka prefecture is located.

Archaeologists were certainly aware that jar-burials were peculiar to this area. However, with this chart we can arrive at the same conclusion without having any previous knowledge of such matters. Nor was the chart made by researchers; it was automatically generated based on the data.

Example 2:
In the nationwide chart of most frequent words, we can see that terms related to excavated materials are the most common. Within that, terminology related to pottery are especially common, and amongst those, words describing features that help date pottery are the most frequent[3]. For example, *kōenbu* (rim shape) frequently comes up. From this, the user can surmise that rim shapes are important. And indeed, they are often used to estimate the age of a site.

These were just two examples of what can be done with visual representations of the data. We are certain that there are many more creative uses for the bubble charts and are looking forward to see what users themselves will come up with.

2 https://sitereports.nabunken.go.jp/en/visualization/term/pref/40
3 https://sitereports.nabunken.go.jp/en/visualization/term.all

Recommending similar reports

Using the same thesaurus, we had also implemented a system that automatically recommends similar site reports to the one the user is currently viewing. In order to achieve this we compiled the statistics for how many times the 70,411 selected terms are used in each PDF file. We then defined the 40 most common terms in any given site report as the representative group of words for that report. The automatic recommendations are then based on the similarities between the representative group of words for every PDF. We believe this is a very useful function for anyone interested in archaeology, as comparing similar resources is very important.

Multilingual features

It is difficult for foreign researchers to gain access to Japanese archaeological reports. Even if a researcher has a good command of the Japanese language, the lack of unification in the terminologies causes difficulties when one is searching for a report about a given topic. For example, Japanese researchers varyingly use *ishikiriba, ishikirichōba, ishichōba, saisekiba, or saisekichōba* to refer to a "quarry". To solve this problem, in August 2016 we implemented a Japanese-English dictionary of archaeological terms and a database of Japanese synonyms within the built-in CDASR thesaurus. Therefore, if a user enters the word 'quarry' in English, all the possible Japanese translations are queried in the CDASR.

Connecting public events to site reports

In Japan it is considered very important to share research results about cultural properties with the public. Therefore, excavators give a public lecture about their findings in the final stage of the excavation, and once the data processing is completed, the excavated materials are exhibited in a museum. In addition, once the historical significance of an excavation becomes clear, a public symposium will be held where the researchers will explain the details to the public. Cultural properties are the properties of the nation; therefore, it is vital to convey the results of the excavations. In this manner, public events related to excavations are held thoroughout the year in Japan.

These events can be registered in the CDASR. As there are many users of the database, registering the events provides significant publicity. In addition, since September 2016, the CDASR automatically searches for the matching site report for the event based on the words used in the description.

There are many practical uses for the automatic linking of site reports to these events. For example, a user can easily access the site report about excavated material displayed in an exhibition. Also, the system automatically recommends public events to the user which are related or similar in topic to the site report that the user is currently viewing. By linking the public events and the site reports together, the chances that viewing one leads to viewing the other are significantly raised.

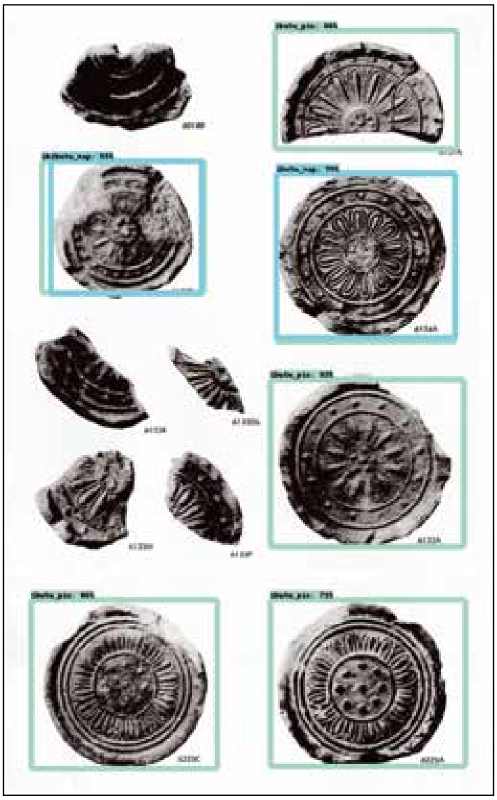


Figure 2: Automatic extraction of round eave tile images by machine learning

Image search

Finally, we would like to introduce a function that is still in the early stages of development. In archaeology images and drawings are crucial. To be able to search for images in a text box, the pictures need to be linked to representative keywords. However, description of pictures can easily alter the meaning of what is shown. Even if pictures are described by professionals, this problem cannot be circumvented. Also, if the user does not know the right terminology, they will be unable to search for the image they want. A search query using images could solve this dilemma. Therefore, currently we are experimenting with an image search based on machine learning (ML) (Takata and Nagashima 2018). However, the digital copies of site reports usually consist of scanned versions of the hard copies, containing both images and text. As such, they are made up of unstructured data that is difficult for a computer to process. Therefore, first we need to find a method with which we can automatically extract images of the structural remains and excavated materials from the files. Currently we are still at the level where we are only experimenting with the auto-recognition of images of the same type of unearthed materials, for example, round eave tiles.

CDASR usage summary

520,000 files were downloaded from the CDASR in 2015; 840,000 in 2016; 970,000 in 2017; and 1.41m in 2018. These numbers are already too large to be generated only by users employed in positions related to cultural properties. Therefore, it is safe to assume that the CDASR is being used by a wide range of users. Also, it is clear that it is rapidly gaining in popularity.

The impact of participating in ARIADNEplus

Historically, the CDASR is entirely a domestic project built to satisfy the needs of Japanese researchers and excavators. However, after the authors met with Julian Richards from the ADS in February 2017, and again in February 2018, things have changed drastically.

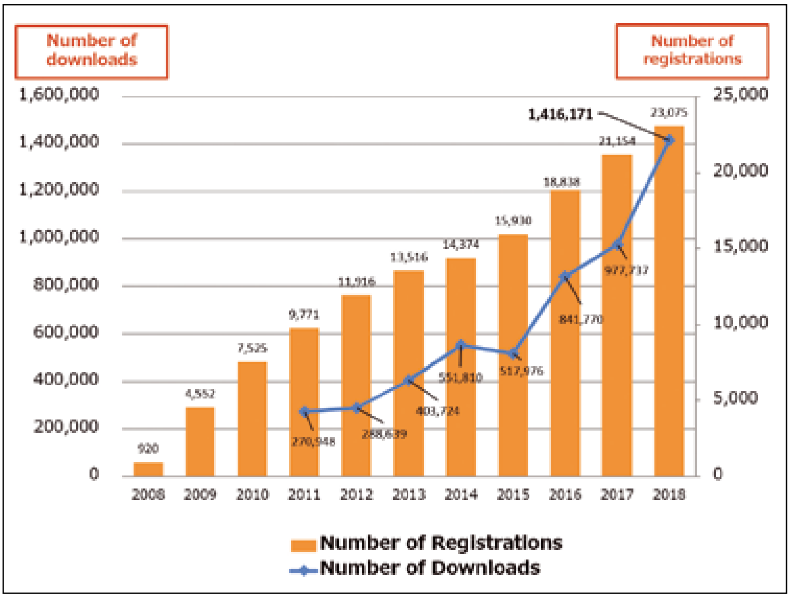


Figure 3: Trends in usage status and number of data by year

In these brief meetings the authors learnt much about practices employed abroad, especially about the guidelines described in the ADS *Guides to Good Practice*.[4]

By officially joining ARIADNEplus, NARA is hoping not only to learn much about the international developments, but also to be able to provide new insights based on our experience with the CDASR. Our first mission is to link together European and Japanese archaeological data. We give an example of what this could mean practically. In 2013 at the excavation conducted at the Katsuren Castle ruins (Uruma city, Okinawa prefecture) excavators unearthed bronze coins which originated within the Roman Empire (4th century) (City of Uruma 2016). If we manage to link the CDASR with ARIADNEplus, then this Japanese excavation will be found when someone searches for 'Roman coin' in ARIADNEplus. However, to be able to do that, first we must conduct extensive data cleansing to get rid of the unevenness of the data found in the CDASR.

Conclusion

Digital data about cultural properties is expected to grow dramatically in the next few years. To be able to make use of this, we have to organize it and find ways that allow us to use it in a practical manner. We also have to solve the problem of how to store the data in a manner that ensures it does not get affected by natural or other types of calamities. It is the responsibility of the current generation to conserve the data for the future. Given that there were 1.41m downloads from the CDASR in 2018, we can safely say that

4 http://guides.archaeologydataservice.ac.uk/g2gpwiki/

Japanese citizens are interested in the archaeological data it holds. Furthermore, culture knows no national borders. The cultures of local groups have always had a long-lasting impact on the culture of other groups. ARIADNEplus allows us to work across national boundaries and with that it can bring new insights to the history of humanity. We therefore believe it is a very important project and wish to be a part of it.

ACKNOWLEDGMENTS

Thanks are extended to Peter Yanase for his assistance in translating this text from Japanese to English.

REFERENCES

Tanaka, M. 1984 'Japan' in H. Cleere (ed) *Approaches to the Archaeological Heritage*. Cambridge: Cambridge University Press, 82–88.

The Agency for Cultural Affairs: Cultural Properties Department Monuments and Sites Division 2017 *Maizo Bunkazai Kankei Tokei Siryo [Statistical Survey on the Buried Cultural Properties],* Tokyo: The Agency for Cultural Affairs the Cultural Properties Department Monuments and Sites Division.

Available http://www.bunka.go.jp/seisaku/bunkazai/shokai/pdf/h29_03_maizotokei.pdf (Last accessed: 25 June 2019)

Morimoto, S. 2017 *Iseki Joho Kokan Hyojun no Kenkyu [Standardization for exchanging information on archaeological sites]* 4, Nara: Nara National Research Institute for Cultural Properties.

Takata, Y. 2016 'Construction of a Comprehensive Database of Archaeological Site Reports in Japan', *BULLETIN Nara National Research Institute for Cultural Properties 2016.* Nara: Nara National Research Institute for Cultural Properties, 14–15.

The Agency for Cultural Affairs the Cultural Properties Department Monuments and Sites Division 2017 *Introduction of digital technology to the protection of buried cultural properties,* Tokyo: The Agency for Cultural Affairs the Cultural Properties Department Monuments and Sites Division. Available http://www.bunka.go.jp/seisaku/bunkazai/shokai/pdf/hokoku_12.pdf (Last accessed: 25 June 2019)

Takata, Y. 2019 'Estimate data volume of excavation report' *Bunkazai no tsubo* 7, Nara: Bunkazai hohoron kenkyukai, 4–5.

Takata, Y. and Nagashima, Y. 2018 'Visualization of Archaeology Big Data and Practical Example of Improvement of Accessibility – Image Recognition by Machine Learning and Statistical Natural Language Processing Technology' *Bunkazai no tsubo* 6, Nara: Bunkazai hohoron kenkyukai, 12–15.

City of Uruma 2016 'Katsurenjo ato kara roma teikoku osuman teikoku jidai no koin wo hakken' [Katsuren Castle Ruins in Okinawa Uruma city had found Roman period bronze coin and Ottoman empire bronze coin] *Koho uruma [Public relations magazine of Uruma city]* 140, Uruma: City of Uruma, 2. Available https://www.city.uruma.lg.jp/sp/userfiles/files/16uruma11p2.pdf (Last accessed: 25 June 2019)

Innovation and impact of the ARIADNE initiative

Guntram Geser

Salzburg Research, Austria
guntram.geser@salzburgresearch.at

Abstract

This chapter first gives an overview of the innovation and impact of the data infrastructure project ARIADNE in the archaeological sector in Europe, and presents four views on the outcomes. These are the footprint of ARIADNE in the sector, the high-level recognition of its integrating effects, positive results on the national level, and the digital infrastructure made available to the whole research community. The follow-up project ARIADNEplus aims to take the next steps in enabling data sharing and use for archaeological research across institutional and national as well as disciplinary boundaries. The chapter describes these steps and challenges and opportunities ahead. Specifically addressed is the need in many countries for a state-of-the-art data repository for depositing and making available archaeological research data, providing virtual research environments and tools on top of the data integration and search platform, and demonstrating research efficiency and innovation enabled by the new ARIADNE platform.

Keywords: digital research infrastructure; innovation; impact; evaluation

Evaluation of the ARIADNE integrating activity

The initial 4-year ARIADNE project (until January 2017) was an Integrating Activity funded under the 7th Framework Programme of Research and Development (FP7) of the European Union. The second round, ARIADNEplus, funded from January 2019 to December 2022 under the EU Horizon 2020 Programme is also such a project. An Integrating Activity aims to integrate within the European Research Area (ERA) the community of a field of research. It has to be carried out based on the I3 (Integrated Infrastructures Initiatives) Model which requires that it combines three lines of activities: Networking, aimed to bring together research organizations and research infrastructure providers to work on common objectives; Transnational Access, offering researchers physical and remote digital access to research infrastructures; and Research and Technological Development focused on improving the accessible facilities, systems and knowledge resources.

The overall objectives of the first round of the ARIADNE initiative have been to build a community of archaeological institutions in Europe interested in making their data findable and accessible through a digital research infrastructure. The infrastructure should aggregate and integrate records of data items from their repositories and databases, and provide a portal for discovering and accessing items in the distributed sources. Support for the data sharing community focussed on guidance in the preparation of what nowadays is commonly called FAIR data, particularly using common data models and vocabularies to enable interoperability for search and access.

For the evaluation of the impacts of ARIADNE a set of project-specific indicators of success was defined for the different I3 activities. In addition, a number of broader impacts of Integrating Activities expected by the FP7 Work Programme for Research Infrastructures had to be considered. These concern contributions to EU high-level goals, e.g. coordinated evolution of research infrastructures or promotion of industrial innovation, and are formulated in a rather general way. The expectations also do not distinguish between different types of infrastructures, e.g. between large natural science laboratories and a digital infrastructure for sharing research data as developed by ARIADNE. Therefore these expectations had to be interpreted and specified so that they could be applied as indicators of success of ARIADNE.

The elaborated set of indicators comprises 13 impact areas and 41 quantitative and qualitative indicators, as ARIADNE specific 8 areas and 28 indicators, and as programme related 5 areas and 13 indicators. The ARIADNE Impact Report addresses all impact areas and gives a detailed account of ARIADNE's achievements (ARIADNE 2017a). In this paper it is not possible to present all achievements of ARIADNE, only to point out and describe key outcomes.

ARIADNE's achievements in brief

The ARIADNE Impact Report states that the project achieved good results in all evaluation areas, and highlights that it:

- Accomplished its goal to provide a digital infrastructure and services for searching and accessing archaeological data in repositories and databases of institutions in different European countries;
- Increased interoperability of datasets based on a common model (ARIADNE Catalogue Data Model), improved vocabularies (e.g. vocabulary mapping tools), and other methods;
- Implemented a European-level data portal providing advanced search capability for 'what' (subjects), 'where' (location) and 'when' (cultural chronology / date ranges).
- Made available additional high-value services (e.g. 3D artefact and landscape services), and demonstrated advanced capability in making data better accessible and useful (e.g. fieldwork reports through metadata extraction with natural language processing methods);
- Achieved a large 'footprint' in the sector regarding the numbers of institutions and researchers that have been informed and involved, including potential providers of additional datasets.

Among the broader impacts, for example, is ARIADNE's contribution to the coordinated development of digital Research Infrastructures (RIs) for the humanities, cultural heritage and archaeology, through knowledge exchange with all European-level RIs as well as major national projects in these fields. Regarding interoperability of digital RIs, ARIADNE especially promoted the development and alignment of dataset catalogues (with the ARIADNE Catalogue as a reference example), the use of common ontologies (in primis the CIDOC Conceptual Reference Model), and other data description and integration standards.

It is also worth noting that the assumption of the Research Infrastructures Programme that RIs of all disciplines could enable industrial innovation does not hold for archaeology and the humanities and social sciences in general. Industrial businesses also play no leading role, if any, in the building of digital RIs of Integrating Activities, because these are being developed collaboratively by domain and technological research organizations (EPIRIA 2014, 60). ARIADNE is an exemplary case of such collaboration.

The core of the ARIADNE project has been the building of a European-level platform where dispersed archaeological data resources can be registered, shared, discovered and accessed (Aloia et al. 2017). Such a platform did not exist before and its implementation arguably is the project's key innovation for the archaeological community in Europe (and beyond). The ARIADNE Impact Report concludes that the project not only had a strong impact, but that it could become a lasting impact, especially by exploiting the high potential for further advances provided by the data sharing and access infrastructure.

While the results of ARIADNE may be outstanding, there is no common scheme (and therefore also no database) which allows us to monitor and compare the results of Integrating Activities, despite the fact that across all disciplines over 90 such projects have been funded under the 7th Framework Programme. The reason arguably is that significant differences between scientific disciplines, types of research infrastructures and services, research instruments and data, etc. do not allow the application of a common evaluation framework.

However, as demonstrated by the Impact Report, ARIADNE's integrating activities moved the field of Archaeology from a Starting to an Advanced Community in terms of the Framework Programme for Research Infrastructures. This has been recognized by including Archaeological Data Infrastructures for Research in the Horizon 2020 Work Programme 2018–2020, in the call for further activities of Advanced Communities of all disciplines. The proposal for ARIADNEplus was submitted and it appears that its current status and proposed next steps were convincing.

Views on project impact

Four views on project impact can help make clear what ARIADNE achieved and ARIADNEplus will build upon, expand and enhance: the 'footprint' of ARIADNE in the sector in terms of institutions and researchers reached and involved, the high-level recognition of its integrating effects, positive results on the national level as confirmed by project-external organizations, and the community research infrastructure that ARIADNEplus will enhance with new services and tools.

The ARIADNE footprint

ARIADNE addresses the archaeological research community, particularly those with an interest in sharing and using data through digital infrastructure and services. Many stakeholders, institutions and individuals have been reached by ARIADNE and involved in project activities. ARIADNE involved over 60 European archaeological heritage institutions (including seventeen partners) from 26 countries and ten other European projects through cooperation agreements and memoranda as well as cooperation on an informal basis. Several of the institutions are now new formal partners in ARIADNEplus. From the international outreach four partners show that interest to join the ARIADNE initiative has also been raised in other world regions: the Israel Antiquities Authority, the Instituto de Antropología de Córdoba (CONICET-IDACOR) in Argentina, the Arizona State University (Center for Digital Antiquity, tDAR repository) in the United States, and the National Research Institute for Cultural Properties (NARA) in Japan.

A large footprint has also been achieved through broad dissemination activities and direct participation of institutions and individuals in different ARIADNE activities such as surveys, professional development and training, (co-)organized conference sessions and workshops, presentations at other events, etc. Considering only the direct participation over four project years, ARIADNE reached 10,500 scholars, students and early stage researchers, practitioners and others. A total of 14,014 project-external participants were counted, with an assumed 25% participation of people in more than one activity. The figure of 10,500 is 30 times larger than the membership of the Computer Applications and Quantitative Methods in Archaeology (CAA) organization/conference (350), 5 times larger than the membership of the European Association of Archaeologists (2,000+), and over 30% of the number of archaeologists working in Europe (33,000), estimated by the Discovering the Archaeologists of Europe project (DISCO 2014).

Regarding the ARIADNE data portal, the indicator of success was 1,100 users in the last project year, 800 anonymous and 300 registered/authenticated users, although registration and authentication for using the advanced data access services was not implemented. The portal was brought online in January 2016, launched officially on the 30th of March 2016 at the CAA conference in Oslo, and received over 10,800 visitors by January 2017. After the funded period of the ARIADNE initiative, a lack of funds did not allow incorporation of more and additional types of data, development of new services, and promotion activities to keep the momentum. However, the data portal has been kept online to demonstrate the data integration, search and access capability. Also of course the underlying digital infrastructure has been maintained by the ARIADNE partner Institute of Information Science and Technologies (ISTI) of the National Research Council of Italy and already ported to their Cloud-based D4Science platform.

High-level recognition

The core institutions of both the archaeological domain and the research infrastructures domain in Europe acknowledged ARIADNE's leading role in building infrastructure and services for sharing and accessing archaeological data. The European Archaeological Council (EAC) strongly encouraged organizations to participate in the ARIADNE initiative. The EAC is comprised of heads of national services responsible under law for the management of the archaeological heritage in the Council of Europe member states. In their Amersfoort Agenda, setting the agenda for the future of archaeological heritage management in Europe, the Council emphasises *'the need to share, connect and provide access to archaeological information with the help of digital technologies. The key to this aspiration is to improve collaboration – we need to share rather than exchange. It is essential to encourage the development of European data-sharing networks and projects in the field of archaeology. The ARIADNE project is an excellent European initiative in this regard and participation in this project should be strongly encouraged'* (Schut et al. 2015, 21).

The recognition of ARIADNE as the core initiative for sharing of datasets across Europe has been confirmed by the participation of Leonard de Wit, former President and now Honorary Member of the Board of the European Archaeological Council, in the advisory board of ARIADNEplus.

The European Strategy Forum on Research Infrastructures (ESFRI) in their Roadmap 2016 acknowledged ARIADNE's role as the leading integrator of archaeological research data infrastructures: *'In the archaeological sciences the ARIADNE network developed out of the vital need to develop infrastructures for the management and integration of archaeological data at a European level. As a digital infrastructure for archaeological research ARIADNE brings together and integrates existing archaeological research data infrastructures so that researchers can use the various distributed datasets and technologies'* (ESFRI 2016, 175).

This recognition stems not only from the achievements of ARIADNE regarding the mobilization and integration of archaeological infrastructures such as repositories of research data as described above. ARIADNE also contributed to the coordinated development of digital infrastructures for the wider cultural heritage and humanities research community. Coordination activities included exchange of knowledge and best practices with other digital research infrastructure initiatives at the European level such as CENDARI (history), CLARIN (languages), DARIAH (arts & humanities), IPERION-CH (heritage conservation science), as well as with major projects at the national level. ARIADNE partners supported this exchange by organizing and contributing to core conferences, executive meetings and workshops of lead developers of digital infrastructures. Among the good practices promoted by ARIADNE were the development and alignment of dataset catalogues (with the ARIADNE Catalogue as a reference example), and the use of common ontologies (e.g. CIDOC Conceptual Reference Model) and thesauri for the description and integration of datasets. Thereby ARIADNE contributed to coordination, cross-fertilization and synergies among major projects.

Impacts on the national level

Complementing the high-level recognition of ARIADNE, institutions and individuals at the national level confirmed the appreciation and impact of the project results. In an international online survey ARIADNE collected user expectations and requirements of nearly 700 researchers, directors of research institutes and repository managers for the data infrastructure and services the project was developing (ARIADNE 2014). Statements regarding existing issues and expectations from ARIADNE for example included:

'Ariadne seems like a great idea, it is such a pity that so many repositories and data are not shared by wider communities, for developing new ideas, teaching, creating new collaborations, sharing the competence of each other.' – data manager, Norway.

'This is a splendid initiative - and I hope that it will reap the fruits we all wish!' –research director, Malta.

'The ARIADNE project addresses major issues of archaeological data. Many archaeologists are waiting for the results of this project.' – researcher, France.

These were statements about 10 months after the start of ARIADNE in February 2013. In the final project conference (December 2016) partners presented statements collected from project-external cultural heritage authorities, directors of research institutions and others with a focus on the impact of ARIADNE on the national level (ARIADNE 2016), for example:

'The ARIADNE project made it possible for the Archaeological Map of Bulgaria to become known on the international level through improved data management, mapping and sharing. All this is useful for both national and international researchers.' – Bulgarian Ministry of Culture.

'The HNM Archaeology Database became the most comprehensive online database in Hungary and has a profound effect on teaching, research, data management and informing developers and the public.' – Forster Gyula National Centre for Cultural Heritage Management, Hungary.

'The on-line Inrap documentary data on the ARIADNE platform represents a considerable step forward for the French archaeological community. The next step is putting online information systems containing the primary data. We hope to contribute jointly with the Consortium Memory of Archaeologists and Archaeological Sites (MASA) from the Very Large Facility Huma-Num.' – Xavier Rodier, Ingénieur de Recherche, CNRS, France.

'From TII's perspective, our project with the Discovery Programme, ARIADNE and the Digital Repository of Ireland, is very significant, as it establishes a framework for the validation, long term curation and dissemination of our data sets especially the

archaeological excavation reports. The expertise developed through ARIADNE, has helped to ensure that this data set can easily be integrated with other archaeological data sets, be it in Ireland or abroad.' – Ronan Swan, Head of Archaeology, Transport Infrastructure Ireland.

These statements illustrate the impact of ARIADNE through enhancing the capability of archaeological heritage and research institutions to prepare and share through the ARIADNE dataset catalogue and portal data for discovery, access and use by research groups in Europe and worldwide. The preceding chapters in this book describe institutional and country-level impacts already achieved by partners through ARIADNE and further progress aimed for by these and new partners in ARIADNEplus.

The pipeline for data sharing and access

As the name of ARIADNE, Advanced Research Infrastructure for Archaeological Data Networking in Europe states, the core of the initiative is providing infrastructure for networking of archaeological data. The outstanding achievement of ARIADNE here is a fully functional pipeline to harvest, integrate and make searchable on a portal, data records from repositories and databases of institutions across Europe and beyond.

The pipeline aggregates the data records into the ARIADNE catalogue, and feeds the search portal with records that are integrated based on general standards, e.g. WGS84 coordinates for locations, and domain vocabularies such as the Getty Arts & Architecture Thesaurus, the PeriodO system for cultural periods, and others. The data portal then provides advanced search options such as multi-lingual subjects-based search (including term suggestion); map-based search, including indication of available records when zooming into the map; timespan-based search with a visual interface for selecting date ranges. Furthermore, within a data record selected by the user, pointers to records that are thematically similar or refer to the same or nearby locations are provided.

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 such as dendrochronology data, etc. that are accessible in a repository or specialized database. 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 served by the portal.

This difference between item-level access versus collection-level access is due to the technical setup of some data collections which make it difficult to provide records of individual 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 required contextual information. Therefore, in ARIADNEplus for each new and updates of some of the already present data collections in the ARIADNE catalogue the best integration approach will be defined taking account of the content and technical setup of the collection.

Realizing the full potential

ARIADNE built a solid basis for taking the next steps in data sharing and use for research across institutional and national as well as disciplinary boundaries. ARIADNE established a common platform for archaeological data sharing, discovery and access, and seeded it with representative datasets from repositories and databases of project partners. The digital infrastructure and services span the whole chain from data aggregation to search and access services for the integrated data. In addition, expertise has been developed to help guide new partners in the preparation of what nowadays is commonly called FAIR data, particularly using common data models and vocabularies to enable interoperability for data search and access. In January 2019 ARIADNEplus started with a consortium of 41 partners while the initial ARIADNE project had 23 partners. The consortium now comprises 37 partners from 23 European countries and one each from Argentina, Israel, Japan, and the United States. Thus the ARIADNE initiative is present now not only more strongly around Europe but also in other world regions. What will be the next steps in ARIADNEplus and challenges and opportunities ahead?

The next steps in brief

ARIADNEplus now aims to aggregate and integrate datasets from a wide range of archaeological domains of research, and provide research tools and services in addition to data search and access services. The Cloud-based D4Science platform maintained by CNR-ISTI will power the ARIADNEplus data pipeline as well as provide several new or enhanced services and tools. 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.

Furthermore, ARIADNEplus will offer Virtual Research Environments that, in addition to data discovery and access, will provide new services and tools for different types of data and tasks. For instance, these will include geo-spatial/GIS data services, tools to annotate texts and images, Natural Language Processing of documents to extract specific information they contain. Existing services for visual data objects (e.g. 3D models) will also be enhanced and new ones explored. Several pilots will be developed to test and demonstrate innovative uses of the new digital research platform.

Challenges and opportunities ahead

In its current phase the ARIADNE initiative will face challenges but also has the chance to enable significant advances of the archaeological enterprise in different respects.

Some of these challenges and potential advances are addressed in the sections that follow.

Promoting data repositories

The objective of the ARIADNE data infrastructure is to allow researchers and other users to discover and access archaeological data held and shared by repositories across Europe and beyond. But many archaeologists in European and other countries do not yet have available a digital repository for depositing and making 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.

Benchmarks for national-level archaeological data repositories exist, for example, the ARIADNE partners Archaeology Data Service (UK) and the E-Depot for Dutch Archaeology of Data Archiving and Networked Services - DANS (Netherlands). 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).

ARIADNE inspired project partners in other countries to promote the building of archaeological repositories or collections in their country. In smaller countries some successes have been achieved. 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 (Transport Infrastructure Ireland 2017).

However, in Europe much effort will be necessary to create more data archiving solutions so that archaeologists can safely deposit and make available their data to the research community and other users. 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.[1] 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.

[1] http://www.seadda.eu

The ARIADNE initiative will not provide its own central data repository, its role is to enable finding and accessing through its infrastructure data that is being shared through existing repositories. However, ARIADNEplus will help developers of new repositories plan participation in the research infrastructure at an early stage. For example, this will involve following best practices in data organization, and description of resources using domain-specific vocabularies so that FAIR item-level data records can be easily aggregated and integrated in the common pool of the ARIADNE initiative.

Providing virtual research environments

Building on the data sharing and access system established by ARIADNE, among the next steps of ARIADNEplus is to provide Virtual Research Environments (VREs) for e-archaeology on the Cloud-based D4Science platform. In addition to data discovery and access such environments provide more specific services and tools which research communities can use 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. Advanced VREs make tools available in a highly integrated way to support research workflows.

ARIADNE already offered some e-research services like the visual media and landscape services, which enable effective online publication and exploration of images (e.g. Reflectance Transformation Imaging - RTI) and 3D models of objects and landscapes. In ARIADNEplus these services will be enhanced and new ones explored, e.g. visualization in 3D of the layers of an excavation and the related documentation. Also planned is provision of geo-spatial/GIS data services, Natural Language Processing of documents to find and extract specific information, and tools to annotate texts and images (e.g. fieldwork reports, artefact or laboratory images).

In ARIADNE the state of e-archaeology in different fields of research, perceived difficulties, and requirements for progress towards innovative solutions have been investigated (ARIADNE 2017b). The study results suggested that there is much potential for ARIADNE to provide VREs, with the proviso that the data infrastructure and services will have to take account of the multi-disciplinarity of archaeological research, particularly different data standards and vocabularies that are being used by different research communities.

ARIADNEplus aims to incorporate datasets from a wide range of research communities, including environmental archaeology, bio-archaeology, archaeometry and dating, epigraphy, among others. This requires standardized 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 experts.

The records can then be aggregated and integrated for using different tools for data search (e.g. Linked Data based search), visualization and exploration based on the records (e.g. finds distribution maps based on location data), and item-level access and study with data-specific tools (e.g. for 3D models). The interest of archaeological researchers in using the different tools envisaged by ARIADNEplus is currently being investigated as part of an online survey. Finding out which tools to combine in Virtual Research Environments will be investigated with researchers in workshops and experimental VREs on the D4Science platform. VREs will also be used in pilots undertaken by project partners to demonstrate the innovative capabilities of ARIADNE services and tools for archaeological research as well as public/community archaeology.

Demonstrating research efficiency and innovation

Frameworks proposed for the evaluation of research infrastructures include different dimensions of impact, including scientific, technological, economic, social and environmental impacts (e.g. European Commission 2010, 43–48; Griniece et al. 2015; RIFI-FenRIAM 2011). However, in the evaluation of the impact of virtual research infrastructure mainly effects in the scientific and technological dimensions can be demonstrated, for example, impacts of improved access to research data, available new research tools and environments, support of networking and collaborative resarch.

Economic, social and environmental impact indicators can hardly be applied to virtual research infrastructures. These indicators are for contributions to economic prosperity and growth, job generation, quality of life, social cohesion and environmental sustainability. Such indicators mainly concern the building and operation of major single-sited research infrastructures (e.g. large natural or life sciences facilities), and must be demonstrated as positive effects on the regional economy and society.

Addressing the potential of digital research infrastructures to generate economic impacts, the authors of an impact framework for such infrastructures note, *'Indicators for the impact of e-Infrastructures on economy are difficult to assess as there is no direct reporting of realised competitive advantages or economic growth due to infrastructure access. Also, even if e-Infrastructure projects generate new jobs in their affiliated institutions such numbers are of negligible size … Overall, involved parties can only realise productivity gains if the infrastructure is easily accessible, works efficiently and stimulates innovative activity'* (Leimbach et al. 2012, 70).

This is an argument for, not against, digital research infrastructures. In the domain of archaeological research small teams of national-level data repositories, e.g. Archaeology Data Service (UK) or the E-Depot for Dutch Archaeology (Netherlands), together with an also small group operating the ARIADNE data sharing infrastructure could aggregate and provide access to a large part of the digital record of archaeology in Europe. From such an increase in access significant effects in research efficiency and innovation in transnational research can be expected. In the case of the Archaeology Data Service, a national-level data repository, the increase in research efficiency of the users has been calculated to be worth at least five times the costs of data curation and access (Beagrie and Houghton 2013). Similar effects may be achieved in the future at the European level and internationally by the ARIADNE data infrastructure and services.

In the ARIADNE user needs survey (November–December 2013) of 498 responding researchers 74% considered it as very or rather important having easy access to international data(sets), while 72% were less or not satisfied at all with the situation in this regard (ARIADNE 2014, 100). While most archaeological researchers arguably work in a national or regional context there is a need to find, access, and use easily research data from other countries for comparative research and broad synthesis. Improved access to international data is critical for innovation and progress in archaeological research, because many fundamental research questions transcend modern political boundaries and concern regions extending all over Europe and beyond.

Conclusion

The ARIADNEplus platform is a community asset with high potential for advances in research efficiency and innovative research across institutional and national as well as disciplinary boundaries. However, to realise its full potential sustained operation of the platform beyond the current ARIADNEplus project will be necessary, especially to integrate many more datasets from repositories and databases across Europe and beyond. Therefore, institutional stakeholders in archaeological research data archiving, sharing and access should place the ARIADNE initiative on their agenda for participation and support.

References

Aloia N. et al. 2017 'Enabling European Archaeological Research: The ARIADNE E-Infrastructure', *Internet Archaeology* 43. doi:10.11141/ia.43.11

ARIADNE 2014 First Report on Users' Needs, D2.1, April 2014. Available http://legacy.ariadne-infrastructure.eu/resources-2/deliverables/ (Last accessed: 8 July 2019)

ARIADNE 2016 Final Conference: Success stories from partners and the research community. Available http://legacy.ariadne-infrastructure.eu/resources-2/presentations/ariadne-final-conference/ (Last accessed: 8 July 2019)

ARIADNE 2017a Final Impact Evaluation Report, D2.5, January 2017. Available http://legacy.ariadne-infrastructure.eu/resources-2/deliverables/ (Last accessed: 8 July 2019)

ARIADNE 2017b Report on E-Archaeology Frameworks and Experiments, D17.1, January 2017. Available http://legacy.ariadne-infrastructure.eu/resources-2/deliverables/ (Last accessed: 8 July 2019)

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. Available http://repository.jisc.ac.uk/5509/ (Last accessed: 8 July 2019)

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 (Last accessed: 9 July 2019)

EPIRIA 2014 Evaluation of Pertinence and Impact of the EU Research Infrastructure Activity in FP7, Final Report by Technopolis Group and Empirica. Available http://dx.doi.org/10.2759/62665

ESFRI - European Strategy Forum on Research Infrastructures 2016 Strategy Report on Research Infrastructures: Roadmap 2016, StR-ESFRI and ESFRI Secretariat. Available http://www.esfri.eu/sites/default/files/20160309_ROADMAP_browsable.pdf (Last accessed: 9 July 2019)

European Commission 2010 A vision for strengthening world-class research infrastructures in the ERA, Report of the Expert Group on Research Infrastructures, Luxembourg: Publications Office of the European Union. Available http://dx.doi.org/10.2777/79690

Griniece E., Reid A., and Angelis J. 2015 Guide to Evaluating and Monitoring Socio-Economic Impact of Investment in Research Infrastructures, Technopolis Group, Tallinn, Estonia. Available https://www.technopolis-group.com/31627/ (Last accessed: 9 July 2019)

Leimbach T. et al. 2012 Development of impact measures for e-Infrastructures, Fraunhofer ISI & ZEW, Karlsruhe/Mannheim. Available https://publications.europa.eu/en/publication-detail/-/publication/a73ad408-277a-4bd2-8eb3-5cf7cfd60f9c (Last accessed: 9 July 2019)

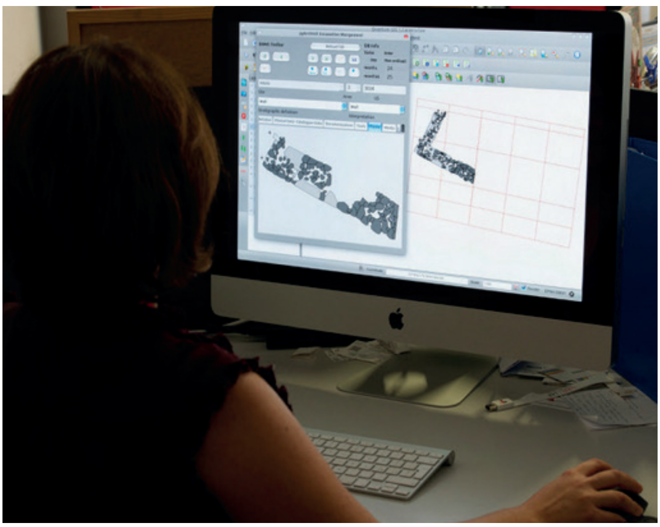
McManamon F.P., Kintigh K.W., Ellison L.A., and Brin A. 2017 'tDAR: A cultural heritage archive for twenty-first-century public outreach, research, and resource management', *Advances in Archaeological Practice* 5, Special Issue 3, 238–49. doi:10.1017/aap.2017.18

RIFI-FenRIAM - Research Infrastructures: Foresight and Impact 2011 FenRIAM Full Guide: Proposal for a Foresight-enriched Research Infrastructure Impact Assessment Methodology. Available http://www.fenriam.eu/doc/FenRIAM%20full%20guide.pdf (Last accessed: 9 July 2019)

Schut P.A.C., Scharff D., and de Wit L.C. (eds) 2015 Setting the Agenda – Giving New Meaning to the European Archaeological Heritage, European Archaeological Council, Namur, Belgium. Available https://www.europae-archaeologiae-consilium.org/eac-occasional-papers (Last accessed: 9 July 2019)

Transport Infrastructure Ireland 2017 Digital heritage collections launch, TII website, 19 June 2017. Available https://www.tii.ie/news/archaeology/tii-digital-heritage-coll/ (Last accessed: 9 July 2019)

Research e-infrastructures, digital archives, and data services have become important pillars of 21^{st}-century scientific enterprise. The archaeological research community was an early adopter of digital tools for data acquisition, organization, analysis, and presentation of research results of individual projects. However, the provision of services and tools for data sharing, discovery, access, use and re-use has lagged behind. This situation is being addressed by ARIADNE, and its follow-on project ARIADNEplus. This volume introduces ARIADNE and provides national perspectives from ARIADNE and ARIADNEplus partners on the current and anticipated impacts of this international collaboration in their own countries and beyond. The publication was funded by the European Commission under the H2020 Programme, as part of the ARIADNEplus project.



ARCHAEOLINGUA