
Article:
Meghini, Carlo, Scopigno, Roberto, Richards, Julian Daryl orcid.org/0000-0003-3938-899X et al. (17 more authors) (2017) ARIADNE : A Research Infrastructure for Archaeology. ACM Journal on Computing and Cultural Heritage. ISSN 1556-4711

https://doi.org/10.1145/3064527

ARIADNE: A Research Infrastructure for Archaeology

Journal:	*Journal on Computing and Cultural Heritage*
Manuscript ID	JOCCH-16-0084.R1
Manuscript Type:	Digital Infrastructure for Cultural Heritage
Date Submitted by the Author:	n/a
Complete List of Authors:	Binding, Ceri; University of South Wales, Faculty of Computing, Engineering and Science Tudhope, Douglas; University of South Wales, Faculty of Computing, Engineering and Science Cuy, Sebastian; Deutsches Archaologisches Institut Doerr, Martin; FORTH, Institute of Computer Science Theodoridou, Maria; FORTH, Institute of Computer Science Fanini, Bruno; Istituto per le Tecnologie Applicate ai Beni Culturali del Consiglio Nazionale delle Ricerche Felicetti, Achille ; PIN scrl Niccolucci, Franco; PIN scrl Fihn, Johan; Swedish National Data Service Meghini, Carlo; Istituto di Scienza e Tecnologie dell'Informazione del Consiglio Nazionale delle Ricerche,
Keywords:	Digital Infrastructures, Metadata, classification schema, ontologies and semantic processing for CH multimedia repositories

ARIADNE: A Research Infrastructure for Archaeology

CARLO MEGHINI and ROBERTO SCOPIGNO, Consiglio Nazionale delle Ricerche ISTI, Pisa, Italy
JULIAN RICHARDS, Department of Archaeology, University of York, York, United Kingdom
GUNTRAM GESER, Salzburg Research, Salzburg, Austria
SEBASTIAN CUY, Deutschen Archäologischen Institut, Berlin, Germany
JOHAN FIHN, Swedish National Data Service, Gothenburg, Sweden
BRUNO FANINI, Consiglio Nazionale delle Ricerche ITABC, Montelibretti, Roma, Italy
HELLA HOLLANDER, Koninklijke Nederlandse Van Wetenschappen, Data Archiving and Networked Services, The Hague, The Netherlands
FRANCO NICCOLUCCI and ACHILLE FELICETTI, PIN, Polo Universitario Città di Prato, Prato, Italy
FEDERICO NURRA, Institut National de Recherches Archéologiques Préventives, Paris, France
CHRISTOS PAPATHEODOROU and DIMITRIS GAVRILIS, Digital Curation Unit, IMIS, ATHENA Research Centre, Athens, Greece
MARIA THEODORIDOU and MARTIN DOERR, Institute of Computer Science, FORTH, Heraklion, Crete, Greece
DOUGLAS TUDHOPE and CERI BINDING, Faculty of Computing, Engineering and Science, University of South Wales, Pontypridd, South Wales

Research e-infrastructures, digital archives and data services have become important pillars of scientific enterprise that in recent decades has become ever more collaborative, distributed and data-intensive. The archaeological research community has been an early adopter of digital tools for data acquisition, organisation, 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, the Advanced Research Infrastructure for Archaeological Dataset Networking in Europe. 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 paper describes the current landscape of data repositories and services for archaeologists in Europe, and the issues that make interoperability between them difficult to realise. The results of the ARIADNE surveys on users' expectations and requirements are also presented. The main section of the paper describes the architecture of the e-infrastructure, core services (data registration, discovery and access) and various other extant or experimental services. The on-going evaluation of the data integration and services is also discussed. Finally, the paper summarises lessons learned, and outlines the prospects for the wider engagement of the archaeological research community in the sharing of data through ARIADNE.

CCS Concepts: •**Information systems** → **Data management systems;** •**Data management systems** → *Information integration;* •**Information integration** → Federated databases;

Additional Key Words and Phrases: Archaeology, Cultural Heritage, E-infrastructure, Computer Graphics, Data Standards, SKOS, CIDOC-CRM

ACM Reference Format:

C. Meghini, R. Scopigno *et al.* ARIADNE: A Research Infrastructure for Archaeology *ACM J. Comput. Cult. Herit.* 99, 99, Article 1

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

Contact address: F. Niccolucci, PIN, Polo Universitario Cittá di Prato, Piazza Giovanni Ciardi, 25, 59100 Prato, Italy

 1556-4673/2016/00-ART1 $15.00
DOI: 0000001.0000001

(Month 2016), 29 pages.
DOI: 0000001.0000001

1. INTRODUCTION

In the last 20 years, e-infrastructures have become ever more important, if not crucial for the conduct and progress of research in all branches of scientific enterprise. Increasingly collaborative, distributed and data-intensive research requires the sharing of resources (data, tools, computing facilities) via e-infrastructure as well as other support for effective co-operation among research groups. Moreover, there is the expectation that with large datasets ("big data"), e-infrastructure and advanced computing techniques, new scientific questions can be tackled.

The archaeological research community has been an early adopter of various digital methods and tools for data acquisition, organisation, analysis and presentation of research results of individual projects. Lagging behind other research fields, *e.g.* the natural and life sciences, however, is the provision of e-infrastructure and services for data sharing, discovery, access and (re-)use. The consequence is a high level of fragmentation of archaeological data and limited capability for collaborative research across institutional and national as well as disciplinary boundaries.

This situation has been addressed by ARIADNE, the Advanced Research Infrastructure for Archaeological Dataset Networking in Europe. This e-infrastructure initiative is being promoted by a consortium of archaeological institutes, data archives and technology developers, and funded under the European Union's 7th Framework Programme. ARIADNE enables archaeological data providers, large and small, to register and connect their resources (datasets, collections) to the developed e-infrastructure, and a data portal provides search, access and other services across the integrated resources.

ARIADNE integrates resource discovery metadata using various controlled vocabularies, *e.g.* the W3C Data Catalogue Vocabulary (adapted for describing archaeological datasets), subject thesauri, gazetteers, chronologies, and the CIDOC Conceptual Reference Model (CRM). Based on this integration the data portal offers several ways to search and access resources made available by data providers located in different countries. ARIADNE thus acts as a broker between data providers and users and offers additional web services for products such as high-resolution images, Reflectance Transformation Imaging (RTI), 3D objects and landscapes. Employing such services in research projects or for content deposited in digital archives will greatly enhance researchers' capability to publish, access and study archaeological content online.

ARIADNE therefore represents a substantial advance for Archaeology; in particular it provides a common platform where dispersed data resources can be uniformly described, discovered and accessed. It is also an essential step towards the even more ambitious goal of offering archaeologists integrated data, tools and computing resources for web-based research that creates new knowledge (e-archaeology).

The next section describes the current landscape of data repositories and services for archaeologists in Europe, and the issues that make interoperability between them difficult to realise. The results of the ARIADNE user surveys undertaken to match expectations and requirements for the e-infrastructure and data portal services are then presented. The main part of the paper describes ARIADNE's overall architecture, core services (data registration, discovery and access) and other extant or experimental services. A further section presents the on-going evaluation of the data integration and set of services. Finally, the paper summarises some lessons already learned in the integration of data

resources and services, and considers the prospects for the wider engagement of the archaeological research community in the sharing of data through the ARIADNE e-infrastructure and portal.

2. THE ARCHAEOLOGICAL RESEARCH COMMUNITIES AND THEIR REQUIREMENTS

2.1 Existing infrastructures, standards, best practices, services and data

Most European countries have provision for the documentation of archaeological sites and monuments through national or regional databases. Despite being created for management purposes, these can also be invaluable research tools, although public access is rarely provided. Several initiatives have begun to integrate archaeological datasets under a common portal, often national in scope. Some have responded to the need for research-focused services such as digital preservation and open access. The best known is the UK's Archaeology Data Service (ADS), established in 1996[1]. The ADS is the mandated place of deposit for archaeological research data for a number of UK research councils and heritage organisations and makes all of its holdings freely available for download or online research. The ADS currently provides access to over 36,000 unpublished fieldwork reports and over 1000 data rich digital archives. It was the first archaeological digital archive in Europe, but there have been related initiatives in several other European countries, although so far these are concentrated in Northern Europe and Scandinavia. In 2007 the ADS was joined by EDNA, the e-depot for Dutch archaeology, which was established as part of DANS (Data Archiving and Networked Services)[2]. In 2007 agreements to deposit archaeological data at DANS were formalised in the quality standard for Dutch archaeology, making archaeology one of the largest components of the digital resources hosted by DANS. By 2016 DANS provides access to over 21,000 reports and 4,000 excavation archives with collections growing daily. The Swedish National Data Service (SND), based at the University of Gothenburg, has also extended its collection policy to focus on Archaeology[3]. It archives a number of archaeological reports, including over 450 GIS files with excavation data from Östergötland. A second Swedish infrastructure, the Strategic Environmental Archaeology Database (SEAD) is based at Umea University, and focuses upon access to data pertaining to environmental archaeology[4]. After a three year preparatory phase, begun in 2012, the German Archaeological Institute (DAI) is now developing IANUS, a digital archive for German archaeology[5]. ARIADNE has provided additional impetus for other countries to develop their own infrastructures, for example, the "Archaeological Map of the Czech Republic" by the Institute of Archaeology of the Czech Academy of Sciences [Kuna et al. 2015] and the "Archaeology Database" of the Hungarian National Museum[6]. In Austria, the Institute for Oriental and European Archaeology in collaboration with the Austrian Center of Digital Humanities, both at the ARIADNE partner Austrian Academy of Sciences, are developing a repository. Outside Europe, the United States has the best developed archaeological digital repository, tDAR, hosted at Arizona State University on behalf of the Digital Antiquity consortium[7]. Open Context, hosted by the Alexandria Institute, provides an alternative option although its focus is digital data publication rather than preservation[8].

There are other infrastructures where the focus is upon networked access rather than digital preservation. Classical archaeologists are relatively well provided for in this regard. Fasti Online provides a database of archaeological excavation projects for Classical Archaeology since the year 2000. The

[1]http://archaeologydataservice.ac.uk
[2]http://www.edna.nl/
[3]http://snd.gu.se/en/
[4]http://www.sead.se/
[5]http://www.ianus-fdz.de/
[6]http://archeodatabase.hnm.hu/en
[7]http://www.tdar.org
[8]http://opencontext.org

project originated in Italy, but now includes a further nine countries[9]. At the level of artefacts rather than excavations, Arachne is a major resource. Arachne is the central object database of the German Archaeological Institute (DAI) and the Archaeological Institute of the University of Cologne[10]. It provides archaeologists and classicists with an online research tool for quickly searching hundreds of thousands of records on objects and their attributes. Both Fasti Online and Arachne also supply data to Pelagios, an initiative supported by the Mellon Foundation, to use Linked Open Data to aggregate information about the Classical World[11]. Finally, although primarily aimed at the general public rather than researchers, there have been a number of European projects, including CARARE[12], LoCloud[13] and 3D-ICONS[14], which have aggregated archaeological data for the European cultural portal, Europeana. These tend to focus on image data but provide a useful resource for research[15].

Many of the existing research infrastructures already recognise that whilst modern Europe is highly politically and institutionally fragmented, archaeological research questions often transcend modern political boundaries. It is unrealistic that such data will ever be brought together in a single database and, in any case, it is better maintained at national or regional level where there is ownership and often a legal responsibility to maintain archives. Therefore we should look to options for interoperability which allow cross-searching of distributed resources. Such an approach was advocated as long ago as 1992 [Hansen 1992] and the 2002 ARENA project provided an early exemplar of semantic interoperability at European level in the archaeological sector [Kenny and Richards 2005]. ARIADNE therefore seeks to provide a bridge between existing national services, and to foster new ones where they do not so far exist. It differs from the existing national infrastructures in that it seeks to provide an integrating layer that exists independently of any one service, and it should allow the development of new research questions that transcend national data sets. However, in order to integrate services on the European level and beyond there are a number of issues related to differences in classifications, and vocabularies, metadata, and different languages which make interoperability difficult to realise.

(1) Past cultures and modern political borders rarely correspond; hence researchers carrying out investigations that span sites located in different countries face a number of problems when trying to place their discoveries in a broader context. They would like to easily compare features and items of their site with those of sites in other countries, yet these will usually be documented in a different language and in a different way.
(2) Thematic datasets, on the contrary, may span different regions. Yet in many cases they are unrelated to the context (*e.g.,* a pottery database does not enable users to access other data concerning the fieldwork context in which the pottery was found).
(3) Harmonization of vocabularies and metadata structures among different, but similar, datasets is usually modest. When it exists, it is more often the result of good archaeological practice than a design feature of the databases involved. This affects terms, names, geographical names, and time periods.

Providing researchers with the ability to pose questions at pan-European scale does not mean that there will always be single European answers. The importance of specific historical circumstances should not be underestimated: the limes of the Imperial Roman frontier system in Germania might be

[9]http://www.fastionline.org/
[10]http://arachne.uni-koeln.de/
[11]http://pelagios-project.blogspot.co.uk/
[12]http://www.carare.eu/
[13]http://www.locloud.eu/
[14]http://www.3d-icons.eu/
[15]http://www.europeana.eu/portal/

culturally and temporally equivalent to the Hadrian's Wall milecastles and the frontier fortlets of the Roman East, but the sheer scale of regional variation means that local factors will influence the particular form that these fortifications take. However, in order to appreciate the role of local circumstances one also needs to compare data sets that cross modern boundaries. During the Neolithic many European cultures developed megalithic tombs as means of commemorating their dead. Scholars who limit their research to the monuments of only of Britain or Ireland, Denmark, France, or Spain will derive partial answers. Both archaeological and linguistic groups transcend political borders demarcated in the modern world. Nonetheless, the cultural context and different historical traditions within which archaeology has operated in the different European countries highlights the perils, as well as the benefits, of harmonization. It will no doubt be easier to achieve interoperability in some areas rather than others.

Some national systems have benefitted from decades of investment in thesauri development and controlled vocabularies; others have grown organically and suffer from a lack of standardisation. Metadata are the key factor to guarantee interoperability among different data collections via mappings to common standards. They must be rich and specific enough to provide researchers with information useful and relevant for specific research questions. They must be simple to create and maintain, through automatic recording of machine created or transformed data, or the use of standardized procedures and tools (thesauri and taxonomies, among others) when data are manually generated. The challenge here is to reconcile these apparently conflicting requirements and overcome the tension of simplicity *vs.* richness and interoperability/generality *vs.* specificity. This requires testing the effectiveness of metadata in research practice and expert evaluation of adequacy: a joint effort of archaeologists and information scientists. The STAR project's research demonstrator showed the potential of combining the CIDOC CRM (extended for archaeology) with SKOS subject vocabularies, in order to achieve semantic interoperability between diverse UK archaeological datasets [Binding et al. 2008]. At a high level substantial advances have been achieved through an increased compliance of archaeological metadata schemas with CIDOC-CRM. Within ARIADNE much effort has been invested in mapping between different national or regional-based time periods and subject classifications. At the level of individual file types and of those metadata required to enable digital preservation and data reuse, the online series of Guides to Good Practice initiated by the ADS has seen widespread adoption[16]. The Guides have been further developed in collaboration with the US-based Digital Antiquity consortium, and have been taken up by IANUS, with enhancements by ARIADNE partners.

2.2 Identification of user requirements

ARIADNE carried out several research activities to identify users' requirements for the e-infrastructure and portal services of the project. The objective was to ensure that ARIADNE addresses the existing and emerging needs of the archaeological research community in Europe and beyond. The research comprised an extensive literature review, 26 interviews with ARIADNE partners and other stakeholders, two online questionnaire surveys with participation of over 600 archaeological researchers and repository managers, a survey of 25 content/data portals, and contributions by ARIADNE Special Interest Groups. Here we present selected study results with a focus on the surveys that allowed the project to acquire a good understanding of user needs and expectations of the ARIADNE data infrastructure and services, which are being developed accordingly.

2.2.1 *Online questionnaire surveys.* Two international online surveys conducted in November/December 2013 collected needs and requirements of researchers and repository managers [ARIADNE

[16]http://guides.archaeologydataservice.ac.uk

2014] (pages 69–143). The survey questionnaires mainly presented questions with sets of predefined answer options, most with a free text field for collecting comments (which were used by many respondents). The results of the survey of archaeological researchers made clear that in most countries they lack an appropriate data repository and services for finding and accessing relevant data. The selected results presented below are based on between 470 and 590 survey responses per result.

The majority of researchers agreed that they often do not know what is available, because research data are scattered across many places and different databases. Consequently 95% considered it to be very or rather important to have a good online overview of available data sets. About the same percentage required data sets to be available online and in an uncomplicated way, not "limited to specific persons/communities" or "kept in private collections of other researchers". 75% of respondents thought that it is important to have easy access to international data sets, which signals a high interest in data that allows for comparative studies and integrative research.

Furthermore, 60% of the researchers said that their organisation (university, research institute or other) does not have an institutional repository that is managed by dedicated staff, and 66% perceived a lack of international archives. Indeed, most institutional repositories manage only documents. Consequently the survey found that data were made available through an institutional repository only in a few projects or not at all by 67% of the researchers. The figures for national and international repositories were 76% and 83%, respectively.

Most researchers wanted ARIADNE to create a data portal that allows an overview of existing archaeological data resources and search across the resources, using novel mechanisms for data discovery and access. Asked which services they would benefit from most ("very helpful"), researchers responded: a portal that makes it more convenient to search for existing archaeological data that is stored in different archives/repositories (79%); innovative and more powerful mechanisms for data discovery and access (63%); a directory of European archaeological databases and repositories (52%); services for geo-integrated data (58%).

Thus capability to search and "mine" distributed digital archives for relevant data was appreciated most. There was much less interest in typical features of Web 2.0 platforms such as content filtered based on tags or ratings provided by other users. Only 29% of the respondents considered such features as "very helpful". Researchers appreciate effective mechanisms that save time in identifying relevant data (*e.g.,* clear licensing information); what they typically do not like is resources pre-selected by others.

The results of the additional online survey of managers of data repositories are only indicative due to the small sample of 52 respondents. The main concern of the data managers is the quality of metadata, but they would also appreciate higher awareness of good practice in data management (*e.g.,* available guides and recommendations) among researchers. Moreover, the data managers expected much more than the researchers better data access through improvements in data/metadata extraction and indexing as well as Linking Data. But Web 2.0 features were also ranked last amongst this group.

2.2.2 *Survey of existing data portals.* Many further insights for the development of the ARIADNE portal services have been acquired through a survey in November/December 2014 of various websites by a panel of 23 archaeological researchers and data managers involved in the project [ARIADNE 2015b]. The panel members served as "lead users" because they make intensive use of searchable archives and other websites and have a good understanding of the state-of-the-art and potential solutions that might serve their requirements even better.

The survey evaluated 25 archaeological websites, giving access to content/data of more than one institution or project, and some existing data portals of other domains. Most of the websites/portals were "international" in that they provide access to content/data from research not only in one country. The

survey participants looked for good practices and gave recommendations for services of the ARIADNE portal. The 34 suggestions of the survey report were then evaluated by 28 experts in order to focus on the most relevant services in the short to medium term [ARIADNE 2015a] (pages 278–289).

The highest scores were received by highly functional portals, *e.g.,* with regard to overview of searchable data and portal navigation, and search and filter functionality based on geo-location (maps) and date ranges/chronologies. High relevance was also attached to deploying Linked Open Data to integrate information within the portal and to link to external resources. Furthermore, providing interfaces to allow external applications to exploit available data, metadata and terminologies was considered as important. Indeed, the ARIADNE infrastructure and portal should not be an "island" but enable added value in the wider information ecosystem of archaeology and beyond.

Some suggested portal features were not ranked highly. These features concern personalized portal services (*e.g.,* alerts on possibly relevant new data), linking of online professional information (*e.g.,* researcher profiles) or networking and discussion on the portal. Portals for the latter exist (*e.g.,* Academia.edu, ResearchGate and others) and are used by many archaeological researchers. Clearly the service portfolio of the ARIADNE portal should meet core requirements of data discovery, access, visualization and re-use. There is little scope to invest limited funds on specific services that are not appreciated, are provided by other portals, or may run ahead of the needs of broad user segments.

The latter includes support for online research work (e-research), which is not an immediate need of the archaeological research community, but may emerge when more open data becomes available through digital archives and novel services provided by e-infrastructures. However, some specific services provided by the ARIADNE portal (*e.g.,* for 3D models of artefacts, buildings and landscapes) can be seen as first components of a future virtual research environment for archaeologists.

2.2.3 *Requirements for Visual Services.* To complement the survey described above, the project organized a workshop specifically aimed at gathering a clear view over the user needs related to visual data technologies and services. The results of this activity made clear that the community was already intensively producing visual data (2D, 3D, videos, terrains) and that the status of the related enabling technologies was considered sufficiently consolidated. Conversely, we discovered that one of the major limitations perceived was the lack of knowledge and tools for easy sharing of those visual resources and to support remote visual analysis (web-based publication and visualization). In response to these needs, two services have been designed and implemented as part of the ARIADNE infrastructure: the Visual Media Resources and the Landscape Services, both described in Section 3.4.

2.2.4 *Summary of requirements.* The user requirements surveys equipped the project with a good understanding of user needs and expectations from the ARIADNE data portal. The core service portfolio of the portal should allow

—an overview of available archaeological data resources, *i.e.* a catalogue of such resources,

—search across records of research repositories located in different countries,

—filter records based on subjects, geo-location (maps) and date ranges/chronologies to access the most relevant resources, and

—support special research content, particularly visual media such as 3D models.

The survey respondents clearly saw these services as priorities because often they do not know if relevant data is available and looking for such data takes a lot of time. With regard to visual media researchers expressed a strong interest in services that support easy sharing of such research resources. Respondents were less interested in:

—Web 2.0 features such as content filtering based on tags or ratings provided by other users,

—linking of online professional information (e.g. researchers' profiles),

—support of expert networking and discussion on the portal.

Obviously respondents did not expect much from user annotations or rating of content, and saw little sense in duplicating services already provided by Academia.edu, ResearchGate and others.

3. THE ARIADNE RESEARCH INFRASTRUCTURE

The Section describes the ARIADNE Research Infrastructure, starting with its architecture, and proceeding with its main services.

3.1 Rationale and Overall Architecture

Integration of data created by archaeological research and in the Cultural Heritage domain in general, is a highly complex process. This complexity is mainly due to the fact that, although they are often very similar to each other, the various institutions that create and use such information have to maintain different types of collections that are documented in different ways, using different languages and different metadata schemas for their encoding. Very often, the way information is organized is influenced by the vision derived from related disciplines or by specific objectives related to the places and periods under study. However, managing this information in an interoperable way has become a vital necessity to ensure efficient use in order to unlock its full potential and to bring a significant contribution to the advancement of archaeological research. This can only take place in an integrated environment where different data are mutually interpretable and able to be consumed as if they were stored in a single archive. The retrieval of meaningful information on both factual and space/temporal level will thus be ensured.

Integration in ARIADNE required the identification of formats, standards and services already in use by the content providers in charge of supplying content to the project. Descriptions of these analyses were collected in various ways and encoded using a data model, the ARIADNE Catalogue Data Model (ACDM), developed by ARIADNE specifically to produce a detailed, formal and unambiguous representation of the archaeological information of the legacy archives (and described in detail in Section 3.2.1). Integration usually means a series of complex operations that takes place on multiple levels and at multiple depths. The core of any activity of this type is the identification of key elements, common traits that can identify objects and conceptual entities that could then be described through a common language.

The top level of this integration starts at the conceptual level, where these fundamental elements can be detected in each archive and captured in accordance with the famous "who, what, where, when" paradigm, in order to identify people, objects, places and time periods, elements of crucial importance especially in archaeology. Careful analysis of these elements demonstrated that integration based on these profiles was possible, if preceded by an appropriate reduction of the concepts themselves to a common shared language. ARIADNE has therefore devoted part of its activities to the identification of those key features and the proper encoding using existing and already well-accepted international standards and terminological tools.

Definition and encoding of key elements and high-level entities has constituted the basis for the creation of the ARIADNE Catalogue, a core resource intended to store metadata and other valuable information concerning the archaeological archives and services connected to them. The Catalogue and the detailed descriptions it contains, constitutes the core of the whole integration process, since it provides all the support necessary for the retrieval and analysis of integrated archaeological information and the resource discovery facilities.

The subjects to which the various datasets relate (*e.g.*, excavations and archaeological surveys, monuments, burials, pottery and the like), which constitute the "what" strand in our model, are described using terms drawn from the Art and Architecture Thesaurus (AAT) of the Getty Research Institute[17]. The AAT forms the spine for the whole framework of terms in ARIADNE, not only with regard to the general subjects, but also for every other typological, morphological and functional description of archaeological objects and activities connected to them. The use of a shared thesaurus required a mapping of each terminological resource already in use by content providers to the AAT concepts.

Integrating spatial entities (the "where") was also straightforward since many archaeological archives already contain detailed spatial data in a standard format. ARIADNE has recommended the use or the conversion of the spatial coordinates in WGS84 format to enable the browsing of archives through geographical tools[18]. Specific resources, like the GeoNames gazetteer[19], were used to obtain spatial coordinates starting from simple names of places in the case where these were the only geographic information present. As for the use of the historical names that a location may have had in the past, an invaluable collaboration with the Pelagios project was established in order to get geographic information from Pleiades (a thesaurus of past places built on bibliographic database) and deploy it in Linked Open Data format to unambiguously identify places of the past[20].

Of particular interest was the time-based integration (the "when") including information concerning dates, times, time intervals and periods abundantly present in archaeological archives. The sharing of dates expressed in numeric format poses no problem, these being unambiguous once the appropriate schema information is given. It should, however, be noted that very often time indications in databases only appear as simple names, without any reference to absolute dates; this may rise ambiguities in an integrated perspective, *e.g.*, The Iron Age in Anatolia has a very different time span from the Iron Age in the British Isles. It is evident, therefore, that the temporal definition of an "age" in the absolute sense is impossible without a precise spatial reference.

An obvious and immediate solution to the problem of periodization was to convert each period in absolute time spans by specifying start and end dates. However, this would not solve the semantic overlaps resulting from the need to keep the original time stamps as part of the documentation. Collaboration with the PeriodO project[21], whose aim is to manage collections of periods built as intersections of documented events on specific geographical areas, facilitated the solution of this issue.

A deeper stage of interoperability has been reached with the integration of individual records coming from the legacy archaeological archives; this is what ARIADNE has defined as "item-level integration". Preparatory activities towards this goal include a broad conceptualisation, mappings and conversions of archaeological information and the construction of a repository with semantic capabilities to perform complex queries on aggregated data.

The implementation of these features is based on the definition of mappings able to capture and express the semantic richness of archaeological data. Mappings are performed within the project through specific tools which allow individual partners to track complex correspondences between the entities contained in their databases and conceptual classes provided by the CIDOC CRM and its extensions (CRMarchaeo in primis, see Section 3.5). Conceptual mappings for each partner, applied to real data, enable the creation of semantic representations for individual items in RDF, in order to form a com-

[17] http://www.getty.edu/research/tools/vocabularies/aat/
[18] http://earth-info.nga.mil/GandG/wgs84/
[19] http://www.geonames.org
[20] http://pelagios-project.blogspot.co.uk; http://pleiades.stoa.org/places
[21] http://perio.do

plex graph of relationships ready to be viewed, queried, integrated with semantic technologies and published in Linked Open Data format.

The integration platform designed and implemented by ARIADNE (shown in Figure 1) appears, in its final form, as a complex modular system, providing advanced interfaces and features and an architecture able to interact with distributed archives in a transparent way. The system is able to query and extract integrated information concerning legacy archives, to present them to users in a coherent way by means of advanced services and tools to visualize, analyse and possibly use them as part of subsequent queries.

All the operations are constantly driven by the Catalogue, which, in addition to detailed descriptions of the original files, contains data related to digital provenance and the complete record of all the "addresses" through which legacy data can be browsed and harvested. Catalogue information is used to address queries to those archives, which contain the information the user is interested in. A set of additional services, deployed on top of the integrated framework, will provide users with advanced features for using data in different ways, such as advanced visualisation and landscape analysis for the definition of use cases and scenarios potentially different from the ones in which the same data were created.

The access point to the whole infrastructure is the ARIADNE Portal[22], which represents the highest level of the architecture. Through it, users are able to browse, query, analyse all the available information, discover and activate the services, and trigger all the features provided by the system. Advanced interfaces for querying the item-level semantic network are also provided, so as to obtain relevant information about objects, places, events, people and types according to semantic criteria and to retrieve and display them in a user-friendly and meaningful way.

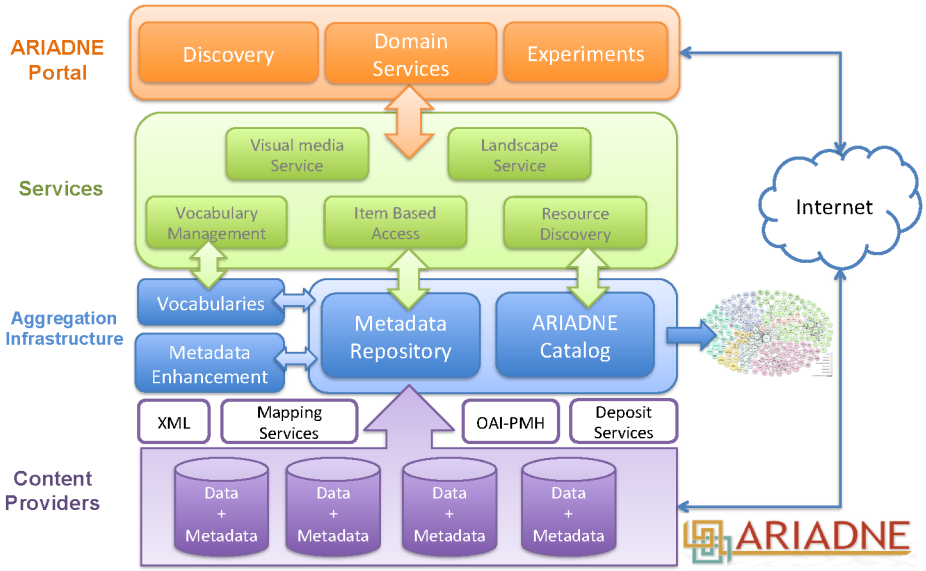

Fig. 1. Architecture of the ARIADNE Infrastructure

[22]http://portal.ariadne-infrastructure.eu

3.2 Resource discovery

Resource discovery is the basic service of the ARIADNE Research Infrastructure, allowing researchers to (a) discover the data and services that populate the ARIADNE information space, (b) obtain basic information about them, and (c) access them. This service hinges on the ARIADNE Catalogue, a collection of descriptions of the resources, structured according to the ACDM. The descriptions in the Catalogue are computed by the ARIADNE Aggregation Infrastructure, which takes the original descriptions from the holding institutions and transforms them into valid ACDM records. The rest of this Section gives the basics of the ACDM (Section 3.2.1), then presents the ARIADNE Aggregation Infrastructure (Section 3.2.2), and concludes with the search functionality enabling discovery, divided in querying (Section 3.2.3) and browsing (Section 3.2.4). The current contents of the Catalogue are illustrated in Section 4.1.

3.2.1 *The ARIADNE Catalogue Data Model.* The main goal of ARIADNE project is "to bring together and integrate the existing archaeological research data infrastructures so that researchers can use the various distributed datasets and new and powerful technologies as an integral component of the archaeological research methodology". In order to achieve this goal, it is necessary to (i) gather information about the existing data resources and services in the archaeological domain, and (ii) to implement advanced search functionalities across this information in order to support the discovery of resources that make good candidates for integration. As a necessary step towards the realization of the former objective, a data model is needed for representing archaeological resources that come in three different types: Data Resources, including the resources that are containers of data such as databases and collections; Language Resources, including the resources related to the formal languages used in Data Resources, such as vocabularies and metadata schemas; and Services, including the resources offering some kind of functionality in the archaeological domain.

The ACDM was built around the DCAT vocabulary [Maali and Erickson 2014], which was expanded by adding classes and properties that were needed for best describing the ARIADNE assets. Its adoption therefore places ARIADNE in an ideal position publish archaeological data resources as Open Data. As illustrated in Figure 2 the central notion of the model is the class *ArchaeologicalResource,* which uses terms of the DCAT vocabulary, to which it adds properties for specifying the access policy and the original identifier of the resource. The class, as noted above, is specialized in:

(1) *DataResource,* whose instances represent the various types of data containers owned by the ARIADNE partners and lent to the project for integration.
(2) *LanguageResource,* having as instances vocabularies, metadata schemas, gazetteers and mappings (between language resources).
(3) *Services,* whose instances represent the services owned by the Ariadne partners and lent to the project for integration.

Each of these classes is described in some detail below.

3.2.1.1 *DataResource.* This class has as instances the archaeological resources that are data containers such as databases, GIS, collections or datasets. Hence its subclasses are *Collection, Dataset, Database,* and *GIS.* A Dataset is defined as a set of homogeneously structured records that are not managed through a Database Management System, whereas we define a collection as an aggregation of resources, called the items in the collection. Being aggregations, collections are akin to datasets, but with the following important difference: the items in a dataset are data records of the same structure. In contrast, the items in a collection are individual objects different from records (*e.g.*, images, texts, videos, *etc.*) or are themselves data resources such as collections, datasets, databases or GIS; for in-

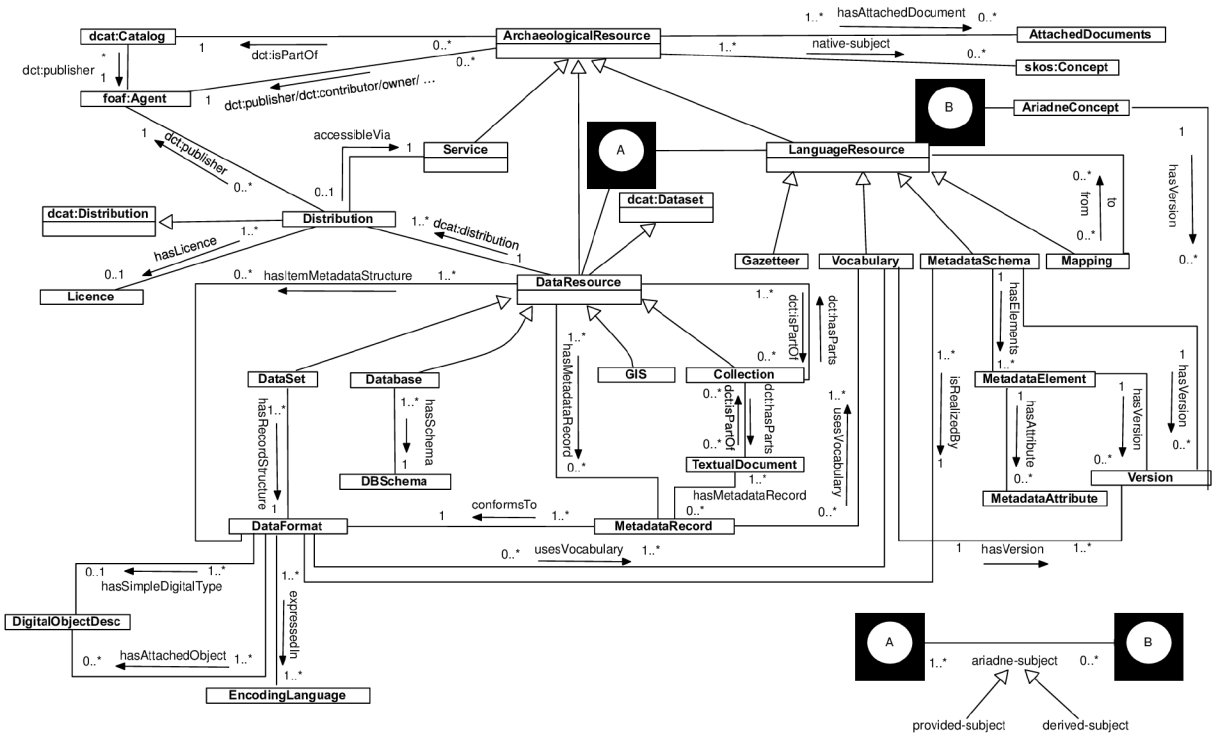

Fig. 2. The ACDM as UML diagram.

stance, a collection may include a textual document, a set of images, one or more datasets and other collections.

Two important attributes of this class are *dct:temporal* and *dct:spatial,* giving the spatial and temporal coverage of each instance data resource. The attributes will be used for establishing the degree to which two data resources are worth integrating. Moreover, the attribute *archaeologicalResourceType* associates any archaeological resource with one or more common resource type, drawn from the following: *Fieldwork archives, Event/intervention resources,* such as unpublished fieldwork reports (the "grey literature"), *Sites and monuments databases or inventories, Scientific datasets,* such as databases of radiocarbon dates, *Artefact databases or image collections,* and *Burial databases.*

The association *ariadne-subject* associates any data resource with a subject from the AAT. A resource has at least one ariadne-subject. We further specialize this property into:

—provided-subject. Associates any data resource with zero, one, or more manually specified subjects drawn from the AAT;

—derived-subject. Associates any data resource with zero, one, or more subjects, automatically derived from mapping local vocabularies to the AAT.

At least one, possibly both, of the above elements must be provided.

Moreover, the property *native-subject* associates any data resource with a subject from a vocabulary in use by the original owner of the resource.

3.2.1.2 *LanguageResource.* This is the class of all language resources described in the Catalogue for the purposes of re-use or integration within the ARIADNE community. A language resource is a resource of a linguistic nature, whether in natural language (such as a gazetteer) or in a formal language (such as a vocabulary or a metadata schema). It also includes mappings, understood as associations between expressions of two language resources that may be of a formal (*e.g.*, sub-class or sub-property links) or an informal (*e.g.*, natural language rules) nature. *LanguageResource* has as instances vocabularies, metadata schemas, gazetteers and mappings (between language resources). Its subclasses are:

—*MetadataSchema.* This subclass has as instances metadata schemas used in the archaeological domain. A metadata schema may be realized by one or many data formats, while a data format can be the realization of exactly one metadata schema.

—*Mapping.* An instance of this class represents a mapping between two language resources.

—*Gazetteer.* This is the class of gazetteers, *i.e.*, geographical indexes or dictionaries.

—*Vocabulary.* An instance of this class represents a vocabulary or authority file, used in the associated structure. The instances of this class define the ARIADNE vocabulary Catalogue.

3.2.1.3 *Service.* The modelling of the services to be integrated by ARIADNE is at a preliminary stage of development. The goal is to provide the primitives for describing the services for which integration or re-use can be envisaged. A preliminary survey has identified the following categories:

—*Stand-alone services.* Tools to be downloaded and installed on one's machine.

—*Web services.* Web accessible services with an API; the services developed by the ARIADNE project fall in this category.

—*Services for humans.* Web accessible services with a GUI only.

—*Institutional services.* Services offered by some institution and that must be negotiated via a personal interaction with representatives of that institution in order to be accessed.

We did not find a shared ontology to express the characteristics of services as discussed above. The best approximation is the ontology adopted in DBpedia for describing software, so we defined the *ARIADNEService* class as a specialization of the *DBpedia-Software* class.

3.2.2 *The aggregation infrastructure.* The ARIADNE Catalogue aggregates metadata, such as descriptions for datasets, metadata schemas, vocabularies, *etc.* utilising the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[23]. Content aggregation is inherently a content-driven task. This raises the importance of the data model which needs to be robust and flexible so as to be able to aggregate information for different domains and schemas. Therefore the metadata and object repository aggregator (MORe)[24] has been utilised and customized [Isaac et al. 2013] in ARIADNE. The MORe aggregator has been used effectively in numerous projects and provides an easy and flexible way of aggregating metadata from multiple sources and in multiple formats.

MORe aggregates dataset items that consist of seven datastreams:

(1) The *administrative* metadata stream, which contains information about the provider, package, and general the history of the item.
(2) The *technical* metadata, which contains technical metadata regarding the contents of the item.
(3) The *native* metadata, which contains the source representation (*e.g.*, the native metadata as they were initially harvested).

[23]http://www.openarchives.org/pmh/
[24]http://more.dcu.gr/

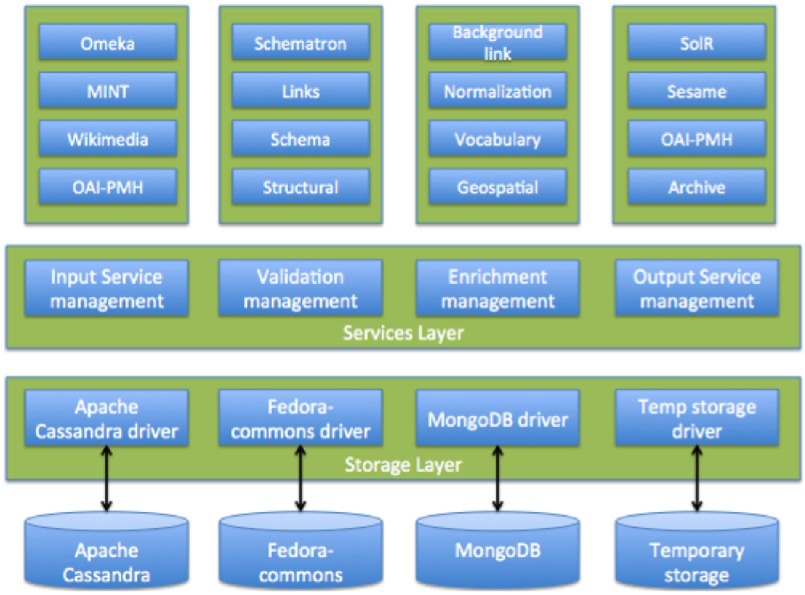

Fig. 3. The architecture of MORe aggregator.

(4) The *enriched native* metadata, which contains a representation of the enriched version of the native metadata.
(5) The *target* metadata, which contains the representation according to the target schema.
(6) The *enriched target* metadata, which contains a representation of the enriched version of the target metadata.
(7) The *preservation* metadata, which is a log of PREMIS events.

The overall architecture of MORe (see Figure 3) includes the following major elements:

A storage layer.. This layer provides an API that allows attaching virtually any CRUD based storage technology. For each storage technology a driver implementation is required and currently the Apache Cassandra, Fedora-commons and Temporary drivers have been implemented.

A services layer. This layer consists of a number of core services, including:
—Harvest, for harvesting content from multiple sources.
—Ingest, for ingesting content into the appropriate storage.
—Validation, for validating content.
—Indexing, for indexing specific elements.
—Quality, for measuring metadata quality.
—Transform, for transforming content from one schema to another.
—Enrichment, for enriching content using specific enrichment micro-services.
—Publish, for publishing aggregated content to a specific target.

A set of micro-services.. Some of the above services follow the micro-services architecture in order to increase the flexibility of certain tasks. One of the most important aspects of MORe is that it employs a number of curation/enrichment micro-services that can enrich metadata in various ways. Indicative micro-services that are have been integrated/developed in MORe are:

—Geo-coding: A geo-coding as well as a reverse geo-coding micro-service that is based on Geonames.
—Rule based thematic enrichment: A subject collections micro-service that allows the user to create thematic collections of concepts encoded in SKOS.
—Automatic thematic enrichment: A vocabulary matching micro-service that identifies SKOS concepts based on title-, description-, and subject-related information found in each metadata record.
—Wikipedia & DBPedia automatic enrichment: A background links service that automatically identifies Wikipedia and DBPedia entries, based on title-, description- and subject-related information found in each metadata record.
—Language identification: identifies languages based on a title or description using Apache Tika.
—Thesauri mappings: allows loading and managing SKOS concepts mappings from SKOSified subject terms to a target SKOS thesaurus.

3.2.3 *Querying the ARIADNE Catalogue.* Users can discover resources via the ARIADNE Portal (see Figure 4), which also provides access to the ARIADNE services.

Fig. 4. The initial page of the Ariadne Portal

The Portal was built using version 5 of Laravel[25], an open source, PHP-based, web application framework. At the heart of the portal lies the ARIADNE Catalogue comprising descriptions of all the resources in the ARIADNE information space, according to the ACDM data model described in Section 3.2.1. The ARIADNE Catalogue includes descriptions of millions of resources, as detailed in Section 4.1.

The general discovery functionality is a free text search accessible from both the portal entry page as well as from a bar located in the menu of the portal. The free text search enables a full text search on all metadata fields of the ACDM. The entry page search also gives the option of specifying a number of facets to specialize the search. The available facets are:

—*Resource type.* Every resource in the portal is categorized with a resource type. The type can be any of the options given in Section 3.2.1: Fieldwork archives, Event/intervention resources, Sites and monument databases or inventories, Scientific datasets, Artefact databases or image collections, or Burial databases;

—*Native Subject.* Subjects from a vocabulary used by the original owner of the resource. Associated with the skos:Concept class;

—*Derived Subject.* Subjects derived from mapping native subjects to Getty AAT vocabulary terms;

—*Keyword.* Keywords or tags describing the resource;

—*Contributor.* The agent responsible for describing the resource in the Catalogue;

—*Publisher.* The agent responsible for making the resource accessible;

—*Place.* Place names the resource is connected with;

—*Period.* Time periods the resource is associated with;

—*Rights.* Access rights connected to the resource;

—*Language.* Language of the resource.

These facets are also available on the search result page to refine the search result by only displaying more specific items.

The underlying storage and search engine is elasticsearch[26], a Lucene[27]-based open source search engine ideal for a service like ARIADNE Portal, providing near real-time search on resources within provided indices. elasticsearch has the capability to be run as a distributed system by dividing the included indices into “shards” which in turn can have one or more replicas. This approach facilitates an automatic load balancing that has been built into the system.

elasticsearch provides a JSON-style query language used to execute queries on the stored documents. This query language provides facilities such as, among others, full-text queries, term-level queries as ranges, exist, wildcards, fuzzy search, and geo queries.

The searchable content is stored as denormalized documents in a Javascript Object Notation (JSON), ingested into the data store via the aggregation infrastructure described above. The JSON structure has been derived from the ACDM model and structured as to serve the search and discovery interface optimally.

Two separate indices have been created in elasticsearch to accommodate the portal:

—First and foremost is the resource index where metadata for all resources have been included.

—The second index is the AAT index which includes AAT subjects, as well as mappings of these terms to native subjects from data provider thesauri.

[25] https://laravel.com/
[26] https://www.elastic.co/products/elasticsearch
[27] https://lucene.apache.org

3.2.4 *Browsing the ARIADNE Catalogue.* In addition to searching for specific topics through a full-text search interface, users can also visualize and filter the contents of the Catalogue along geospatial, temporal and thematic dimensions, thereby being able to explore and dig into the available information resources.

3.2.4.1 *Where - Map based browsing.* The "where" section of the browsing interface is realized as a full-screen map layout based upon OpenStreetMap[28] and implemented with the help of Leaflet[29]. The main challenge that had to be resolved in the implementation process was to develop a view that would provide both a dynamic visualization of vast amounts of geographical data and the ability to narrow down the visible resources in order to be able to pick out the specific datasets the user is interested in.

Therefore the resources are first visualized as a heatmap that represents resource density. This view dynamically changes to markers representing single locations when the user has reduced the result set by filtering or zooming.

The implementation of the dynamic heatmap is realized with the help of elasticsearch's aggregation feature. This allows the creation of buckets that cluster similar resources based on indexed field values. The particular index used in this aggregation is based on the geohash representation of geographical coordinates and accelerates access to geographically similar objects. By being able to do this accumulation on the server, based on indexes already present for the search functions, we were able to greatly reduce the cost of data transmission and processing in the browser. This enabled us to visualize millions of datasets without major lag or performance issues for the user.

3.2.4.2 *When - Timeline browsing.* A similar approach was taken in the realization of the "when" section of the browsing interface. The particular implementation involves the creation of dynamically defined buckets that cover date ranges distributed over a logarithmic scale. These buckets are then visualized as an area graph that represents the distribution of the dates connected to the archaeological resources over time. The visualisation, which is based on D3.js[30], also makes use of the zoom metaphor users are acquainted with from map interfaces and allows drilling down into smaller date ranges for increased details. User can then select date ranges to be used as a starting point for a search in the Catalogue.

3.2.4.3 *What - Subject browsing.* The "what" section aims to present yet another starting point for discovering the contents of the ARIADNE Catalogue. Its purpose is to provide a summary of the different thematic aspects of the registered resources and to offer an exploratory entrance into the available subjects built upon the unifying mapping provided by the AAT. In addition, the thesaurus data collected in the subject index is used to provide auto-completed suggestions for the search field. These can then be used by the user to discover resources connected to a particular theme present in the common thesaurus.

The combination of Where, When and What browsing covers the most common research questions posed by archaeologists. For example, a prehistorian can identify all Bronze Age sites with barrow burials along the North Atlantic seaboard, or a medievalist can search for all known market sites in the Mediterranean, dated from AD 1200-1600. In several regions the power of ARIADNE is the ability to link resources from several data providers. For example, in the British Isles the data provided by the ADS includes records for all known medieval buildings in England, and any text reports for buildings surveys, whereas the north-western European dendrochronology database provided by DANS provides information about tree-ring dates for specific buildings in the same area. This is the first time that

[28] https://www.openstreetmap.org/
[29] http://leafletjs.com/
[30] https://d3js.org/

these data sets have been combined. In most cases the user is able to drill down to the primary resource held by the data provider to recover further information, and to download reports and data sets.

3.3 Vocabulary resources and services

For subject access, the ACDM *ArchaeologicalResource* class has two kinds of subject property. The property, native-subject, associates the resource with one or more items from a controlled vocabulary used by the data provider to index the data. However, there are a large number of partner vocabularies in several different languages. Cross search and semantic interoperability is rendered difficult, as there are no semantic links or mappings between the various local vocabularies. Standard ontologies for metadata schemas, such as the CIDOC CRM, do not have vocabulary coverage so there is a need to complement the ontology with the terminology contained in subject vocabularies. Trivial differences in spelling variations or different synonyms for the same concept can result in failure to find relevant results. This problem is exacerbated when subject metadata may be in different languages, which is clearly the case when providing an infrastructure for European archaeology. Not only may useful resources be missed when searching in a different language from the subject metadata but there is also the problem of false results arising from homographs where the same term has different meanings in different languages. For example, "vessel" has different archaeological meanings in the English language, while "coin" is French for "corner", "boot" is German for "boat" and "monster" is Dutch for "sample".

The established solution to this problem is to employ mapping between the concepts in the different vocabularies. However the creation of links directly between the items from different vocabularies can quickly become unmanageable as the number of vocabularies increases. A scalable solution to this mapping problem is to employ the hub architecture, an intermediate structure where concepts from the ARIADNE data provider source vocabularies can be mapped [ISO 2013]. In the portal, retrieval based on a concept from one vocabulary (in a search or browsing operation) can use the hub to connect to subject metadata from other vocabularies, possibly expressed in other languages. In the ACDM, *ariadne-subject* is used for shared concepts from the hub vocabulary (the AAT), which have been derived via the various mappings from source vocabularies. This underpins the MORe enrichment services augmenting the data imported to the Registry with mapped hub concepts (see section 3.2.2). These *derived-subjects* in turn make possible concept based search and browsing in the ARIADNE Portal (see section 3.2.3). The mappings from the various source vocabularies to the AAT underpin a multilingual capability in the Portal via concept-based retrieval.

The AAT was chosen as an appropriate hub vocabulary, following a prototype mapping and retrieval exercise involving five ARIADNE vocabularies (in three different languages). The AAT had recently been made available as Linked Open Data by the Getty Institute[31], which fit well with ARIADNE's strategy for semantic interoperability. The AAT linked data is expressed in the standard SKOS RDF representation and the appropriate representation for the mappings is via SKOS mapping relationships[32]. The next step was to produce the mappings from the subject vocabularies employed to index the various datasets selected for the ARIADNE Catalogue. This is not a trivial exercise. It requires domain experts to make quality mappings, who may not have expertise in computing semantic technologies. The vocabularies themselves vary from a small number of keywords from a picklist for a particular dataset to standard national vocabularies with a large number of concepts.

Two different tools were developed to support the domain experts doing the mapping between vocabulary concepts, oriented to different contexts for the vocabularies. An interactive mapping tool was

[31] Getty Vocabularies as Linked Open Data. http://www.getty.edu/research/tools/vocabularies/lod/
[32] http://www.w3.org/TR/skos-reference/#L4138

developed for ARIADNE oriented to major vocabularies already expressed as Linked Data via local or national initiatives. The mapping tool generates SKOS mapping relationships in JSON and other formats between the source vocabulary concepts and the corresponding AAT concepts. To assist the production of quality mappings, the mapping tool displays the source concepts and the AAT concepts side by side, together with contextual evidence and allows the person making the mappings to browse related concepts in either vocabulary to fine tune the mapping. The mapping tool is a browser based application working directly with linked data, querying external SPARQL endpoints directly [Binding and Tudhope 2016]. A pilot mapping exercise was performed by ADS on UK HeritageData vocabularies. Analysis of results informed an iteration of the mapping guidelines and the mapping tool user interface. The mapping guideline revisions included recommendations on the appropriate SKOS mapping relationship to employ in different contexts and when appropriate to specify more than one mapping for a given concept.

The second mapping tool was oriented to cases where the source vocabularies were not expressed as linked data and included simpler "flat list" vocabularies. Since many of the simpler vocabularies were already available or easily expressed in spreadsheet format, the most flexible solution was to design a standard spreadsheet with example mappings that domain experts could use to specify the mappings. A CSV transformation produced the RDF/JSON format required by the Catalogue. The spreadsheet was accompanied by a set of guidelines informed by the pilot mapping exercise (together with support from the vocabulary team on problematic mappings or precedents from other partner mappings). In some cases, data cleansing was required before the mapping exercise could proceed. The mapping template contained a tab to record metadata for the mapping. In future work, making the mappings available as outcomes in their own right, with appropriate metadata for the mappings would be desirable, as more than one mapping may be produced for large vocabularies.

The information from the mapping tool is passed to MORe which associates it with the provider of the vocabulary. It updates the property *derived-subject* and enriches an ACDM record (see Figure 5), adding a broader term, or a skos:altLabel to correlate a term using the "use for" relationship, or adds multilingual labels (skos:prefLabel and skos:altLabel) in order to facilitate multilingual search.

Prototype experiments have shown the potential of working with the URI identifiers of AAT concepts [ARIADNE 2016]. Search results can include the specified concept and all hierarchically descendant concepts in accordance with the AAT hierarchical structure. Using the URI identifier for the concept avoids the problem (discussed above) common with multilingual data of terms that are homographs in different languages. Working at the concept level also makes possible hierarchical semantic expansion, making use of the broader generic ("IS-A") relationships between concepts in a hierarchically structured knowledge organization system, such as the AAT. Thus a search expressed at a general level can (if desired) return results indexed at a more specific level, for example a search on *settlements* might also return *monastic centers*.

3.4 Visual services

As pointed out at the end of Section 2, ARIADNE included two services in its infrastructure: the Visual Media Service[33] and the Landscape Service[34]

3.4.1 *The ARIADNE Visual Media Service.* The ARIADNE Visual Media Service [Ponchio et al. 2015] is a resource providing easy publication and presentation of complex visual media assets via a web browser. It is an automatic service that allows the user to upload visual media files to an ARIADNE server and to transform them into an efficient web format, making them ready for web-based

[33]http://visual.ariadne-infrastructure.eu/
[34]http://landscape.ariadne-infrastructure.eu/

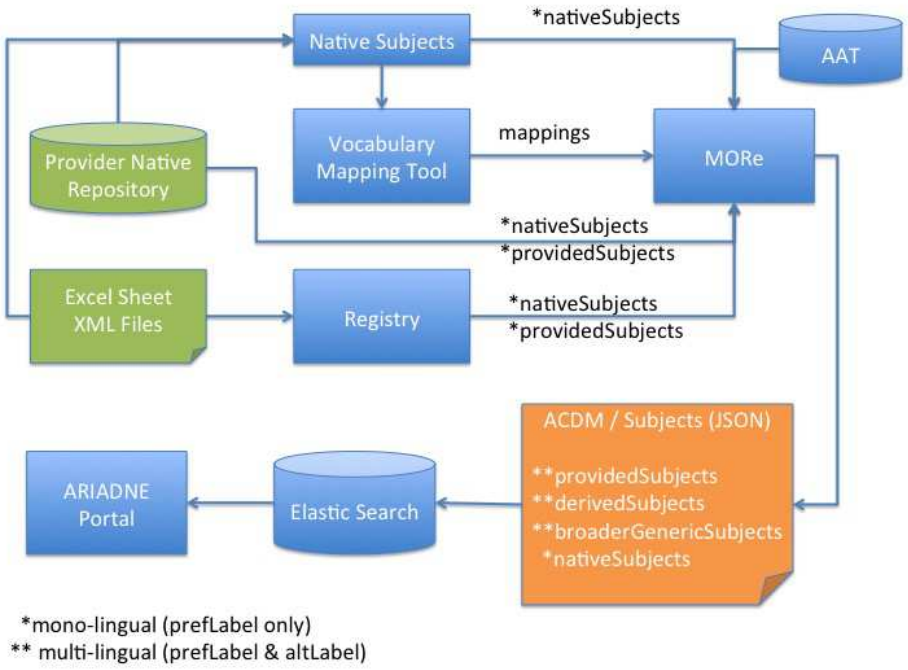

Fig. 5. MORe enrichment

visualization. The user is asked only to fill up a small form and to upload the raw file; all processing required to transform the data in a web-compliant and efficient format is done in an automatic manner by an ARIADNE server.

The service supports the publication on the web and browsing of three types of visual media:

—High-resolution 2D images (input images are converted in a multi-resolution format and can be browsed in real time, zooming in and out);

—Reflection Transformation Images (RTI), also known as Polynomial Texture Maps (PTM) images, *i.e.*, dynamically re-lightable images [Mudge et al. 2008];

—3D models (triangulated meshes, point clouds and textured models).

For each media type, automatic conversion to an efficient multi-resolution representation is supported, offering data compression, progressive transmission and view-dependent rendering; each data type has a specific web-browser, implemented using Web-GL and appearing in a standard web page (see Figure 6).

The page can be personalised by changing the navigation paradigm and the style of the page. Moreover, tools for creating sections and for taking point-to-point measurements are available, and they can be added to the visualization page.

3.4.2 *The ARIADNE Landscape Service.* The Landscape Service is a set of online services for the processing, management and publication of large, multi-resolution 3D interactive terrain datasets within a collaborative workflow (see Figure 6). The goals within this service are: (1) aid and support 3D landscape reconstruction tasks and projects in Virtual Archaeology, and (2) provide online services

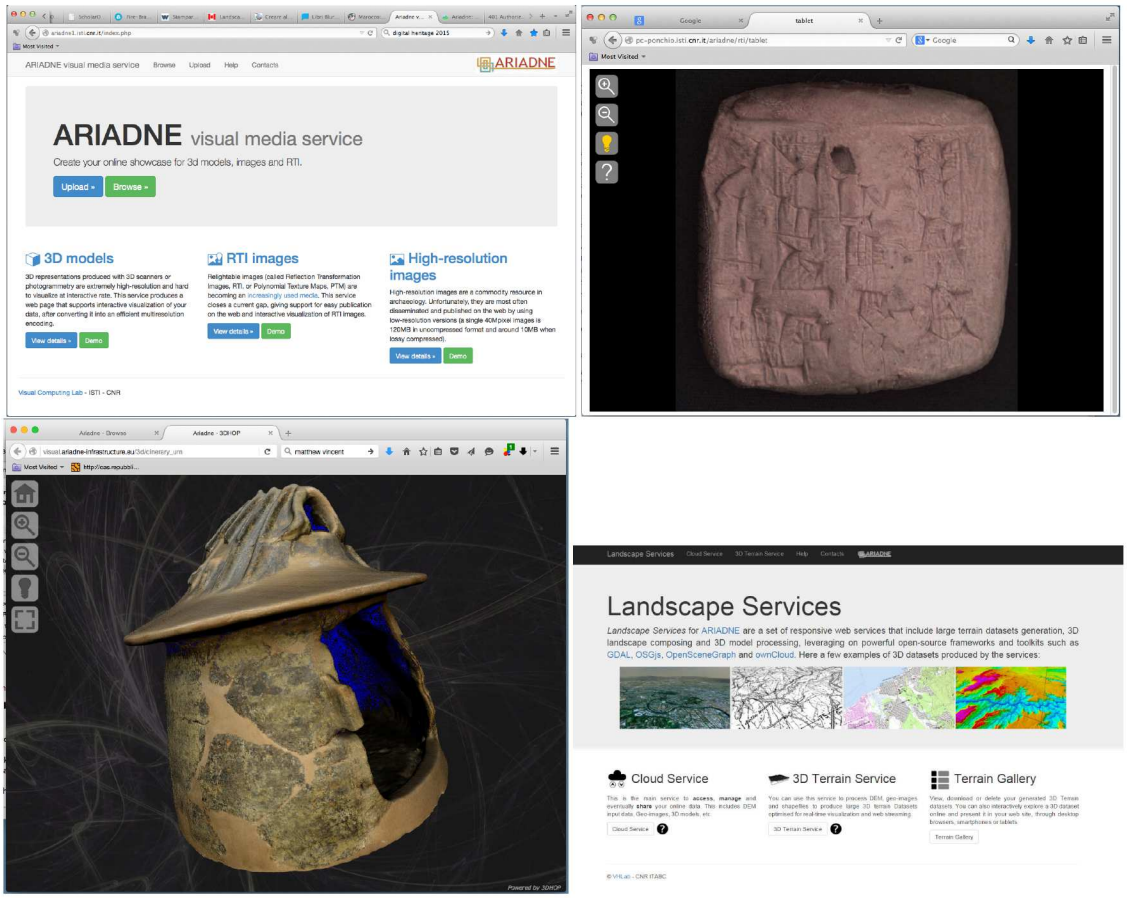

Fig. 6. Some snapshots from the services (from top to bottom, left to right): the home page of the Visual Media Service; an example of an RTI image visualized with the RTI browser; an example of 3D model visualized with the provided 3D browser; the home page of the Landscape Services.

for dissemination of interactive landscapes, through several devices. The Landscape Services are designed for responsiveness, thus adapting to both desktop and mobile devices such as smartphones and tablets. Data management is performed through a cloud service, allowing fine-grained access control on input/output data and collaborative approaches among research institutions and professionals, with specific focus on input DTMs/DEMs, imagery and shape files. The Terrain service allows generation and publication of 3D datasets, by presenting different options to control format, resolution and dissemination segment; the service will then take care of multi-resolution, geometry/texture compression and much more. The Gallery service allows to control, update or delete their projects.

The WebGL Front-End provides a high level of customization and several features including:

—paged multi-resolution on desktop and mobile browsers for efficient streaming;
—camera and Point-of-View management;
—embed options;
—metadata presentation;

—support for touch and multi-touch devices;

—multi-texturing and spherical panoramas.

3.5 Item-level integration

Among the emerging needs of the archaeological research community is the capability to answer a research question by using relevant information from several available heterogeneous sources. This can be achieved only with the integration of the rich, structured information from all the heterogeneous sources through a common, consistent representation of data that have a potential bearing on questions beyond their local context of creation and use, so that directly and deep-indirectly related facts can be filtered out effectively from the mass in order to support further interpretation by the researcher.

In order to address the complexity of archaeological data integration, the main challenge for ARIADNE was to develop a global, extensible schema in the form of a formal ontology that allows for integration without loss of meaning. The CIDOC CRM [Doerr 2003][35] was chosen as the backbone of the ARIADNE Reference Model and a suite of extensions was developed to address the complexity of archaeological data integration (Figure 7 left)[36].

The CIDOC CRM (ISO21127) is a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information. It was developed by interdisciplinary teams of experts, coming from fields such as computer science, archaeology, museum documentation, history of arts, natural history, library science, physics and philosophy, under the aegis of the International Committee for Documentation (CIDOC) of the International Council of Museums. It started bottom up, by re-engineering and integrating the semantic contents of more and more database schemata and documentation structures from all kinds of museum disciplines, archives and recently libraries as empirical base. The CIDOC CRM contains the most basic relationships to describe what happened in the past at a human scale, *i.e.*, people and things meeting in space-time, parts and wholes, use, influence and reference. More detailed kinds of discourse require extensions.

At its core the CIDOC CRM provides generalizations for both the human activities and products that are the subject of archaeological and historical research, and for the research activities themselves, such as excavations. Therefore in the framework of the ARIADNE project specific extensions were developed:

—CRMarchaeo to describe the archaeological process by abstracting five European national standards for creating archaeological records;

—CRMba to describe archaeological analysis of built structures by abstracting the most comprehensive Italian standards as well as other European national standards;

—CRMsci to describe processes of analytical investigations by abstracting ten different processes from different techniques, such as remote sensing, thermoluminescence, dendrochronology, LIBS chemical analysis;

—CRMgeo to bridge the concepts of matter, evidence and cause in the CRM with the OPENGIS standards by developing a model of space-time volumes that applies symmetrically to objects, material places, periods and events compatible with the understanding of modern physics and geographic practice.

[35]Version 6.2 available from http://83.212.168.219/CIDOC-CRM/Version/version-6.2

[36]The reference documents and RDFS encodings of the CRM extensions are available from http://www.ariadne-infrastructure.eu/Resources/Ariadne-Reference-Model

All those extensions and the independently developed CRMinf for scholarly arguments and CRMdig for digitization processes were mutually harmonized and integrated, and the CIDOC CRM was adapted to be compatible with the newly introduced enhanced spatio-temporal notions in close collaboration with CIDOC. All extensions have been approved or are under review for approval by CIDOC. Together, they form the ARIADNE Reference model and provide almost complete coverage of the epistemological processes and archaeological research at the level of evidence and inferred material facts. They do not cover typology and theory building, as well as statistical abstractions, which may be the subject of future work.

Having defined the ARIADNE Reference Model (RM), integration is accomplished by creating an advanced knowledge base (target, aggregation database) based on the common reference model. The integrated knowledge base is the aggregation of several existing archaeological databases that were transformed by mapping their individual schemata (source schemata) into the ARIADNE RM (target schema). The mapping process was supported by the X3ML Mapping Framework ensuring the integrity of the initial data and preserves their initial "meaning". The X3ML framework [Minadakis et al. 2015] consists of a set of components that assist the aggregation process for information integration following five major steps:

(1) *Syntax Normalization.* A first step used to normalize the source records. It exploits local syntax rules and produces a new source schema definition that will be mapped to the ARIADNE RM.
(2) *Schema Matching.* This step produces mappings from the schema of each source dataset to the common ARIADNE RM. The mappings obtained during schema matching preserve as much as possible the meaning of the source schemas fields. To this end, a close collaboration is required amongst domain experts, who know the semantics of the source schemas, experts of the ARIADNE RM, and the IT experts who guide the others on using the mapping specification tools; these tools include both the language for encoding the mappings (X3ML) and the software for creating and managing the mappings [37].
(3) *URI Generation Specification.* This step aims at defining the functions assigning an appropriate URI to each resource found in the source datasets. Domain experts contribute their expertise on namespaces as well as any policy on naming that is in place in the institution where the integrated dataset will be deployed; the task of IT experts is to properly configure the tools so that the chosen URIs are generated.
(4) *Terminology Mapping.* This step produces mappings from the thesaurus used by each source dataset to a common thesaurus that is used by the aggregator database. It is similar to the schema matching step, and requires the same tight collaboration amongst different experts.
(5) *Transformation.* This is the final step that transforms every record of the source datasets to a set of appropriate RDF triples, subsequently stored in a knowledge base hosted in the ARIADNE infrastructure Linked Data cloud.

In order to demonstrate the item-level integration process of archaeological datasets, ARIADNE has chosen as a use case the numismatics field, a highly standardized field with widely available data. Five datasets have been selected. Four of them have been mapped to the ARIADNE RM and transformed to RDF using the X3ML framework, while the fifth had been already in CIDOC CRM RDF form compatible with the ARIADNE RM. As a common thesaurus for the aggregated knowledge base, the nomisma.org[38], as the most appropriate resource in the numismatics, and the AAT thesauri have been adopted. The datasets are:

[37] http://www.ics.forth.gr/isl/3M
[38] http://nomisma.org/

—The dFMRÖ archive: Digitale Fundmünzen der Römischen Zeit in Österreich is an online MySQL database of the Numismatic Research Group of the Austrian Academy of Sciences [39]. The dFMRÖ archive was chosen as the first hands-on exercise to map a relational database schema to CIDOC CRM, since it represents a large class of well-defined traditional databases.

—Two numismatic archives from the COIN project[40]:

—The FWM archive, a subset from the Department of Coins and Medals of the Fitzwilliam Museum Database, recording information on medals and coins of different types and age, discovered during excavations or coming from various acquisitions or donations, currently kept by the Fitzwilliam museum.

—The SAR archive, originally a Microsoft Access DB, created for the cataloguing of archaeological finds of monetary type managed by the Archaeological Superintendence of Rome, coming from public and private collections and from archaeological excavations made in the city of Rome and its immediate surroundings.

—The iDAI.field archive of coins taken from the Pergamon project of the German Archaelogical Institute (DAI).

—MuseiD-Italia collections, the digital library integrated in CulturaItalia[41]. MuseiD-Italia includes several collections of coins from Italian museums already in CIDOC CRM form.

The mapping and transformation workflow is presented in Figure 7. The ultimate goal of the integration of the diverse coin datasets is to create an environment where users will be able to specify queries that will be evaluated on the common aggregated repository and will be able to combine results coming from the different datasets. The ARIADNE portal provides a main access point to the integrated repository and an intuitive interface guides the user to formulate queries, browse the results and refine the search with facet view taking advantage of the principles of the Fundamental Categories and Fundamental Relationships as defined in [Tzompanaki and Doerr 2012; Tzompanaki et al. 2013]. Research questions that are supported include:

—Origin - Where does this coin come from?

—Tracking - How did it arrive here?

—Chronology - First/last appearance

—Practical/symbolic value, incidents - Why is it deposited here?

—Political message - Why was it produced (*i.e.*, “minted”)?

—Economic stability, power - Why was it widely used/not used?

—Statistics - Material versus nominal value.

Some queries might appear trivial if answered by each dataset separately; however they become important if they can be answered by the aggregated repository. Results from our first experimental aggregated repository are quite promising [Felicetti et al. 2015].

3.6 Natural Language Processing services

The archaeological domain generates vast quantities of text, including journal articles and not formally published reports of fieldwork or specialist analysis (grey literature). This text information is frequently difficult to access and opaque to computer based tools for cross searching or meta-analysis.

[39] http://www.oeaw.ac.at/antike/index.php?id=358
[40] http://www.coinproject.com/
[41] http://www.culturaitalia.it/

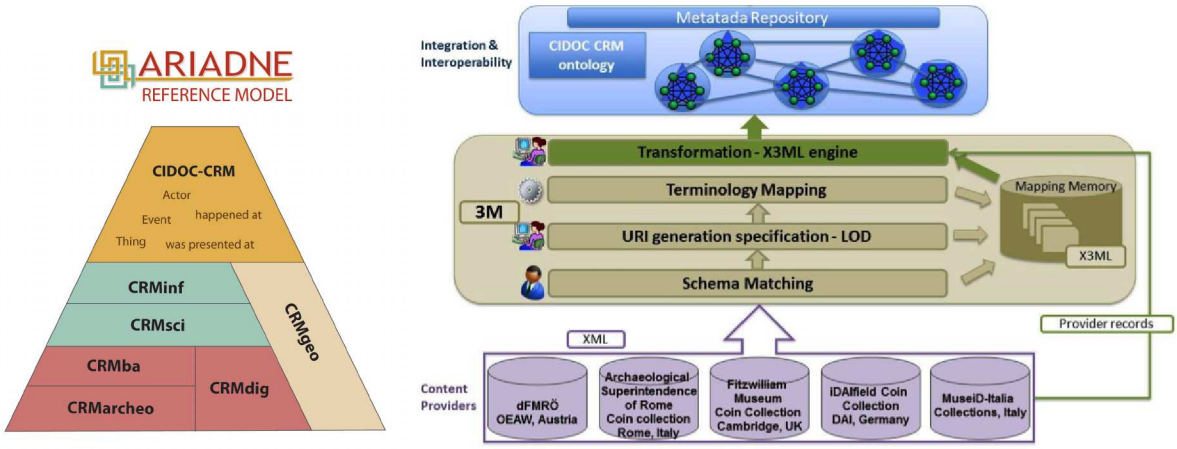

Fig. 7. The ARIADNE Reference Model and the mapping and transformation workflow

This has become recognised as a significant problem for archaeological research. ARIADNE is addressing this issue, particularly as regards grey literature by experimenting with text analysis methods based on Natural Language Processing (NLP) techniques for information extraction. The ultimate aim within ARIADNE is to extract additional relevant subject metadata from these reports and express it using the same ontological (CIDOC CRM) and vocabulary standards as being used for describing archaeological datasets within the Catalogue. This is a long-term goal beyond the reach of the immediate project. Nonetheless, some initial investigations point the way for further research.

Information Extraction is a specific NLP text analysis technique which extracts targeted information from context. This technique analyses textual input to form a new textual output capable of further manipulation. There are two types of NLP information extraction techniques, rule-based and machine learning [Richards et al. 2015]. The aim within ARIADNE is to investigate both approaches; each has its respective strengths and weaknesses and we have explored both with regard to their usefulness within the archaeological domain.

Rule-based techniques have been employed with available archaeological vocabularies from Historic England (HE) and Rijksdienst Cultureel Erfgoed (RCE). This builds upon previous work with the grey literature digital library from the ADS, which proved capable of semantic enrichment of English language grey literature reports conforming both to archaeological thesauri and corresponding CIDOC CRM ontology classes representing archaeological entities, such as Artefacts, Features, Monuments Types and Periods. The current pilot system has achieved some promising semantic enrichment of Dutch grey literature reports, for example artefacts such as "pottery/*aardewerk*" (via the RCE *Archeologische artefacttypen* vocabulary) and other concepts including time periods. Work extending the techniques to develop a Swedish language pipeline is underway. The resulting NLP tools will be available via the ARIADNE Portal.

English language NLP research has investigated the issue of negation detection in archaeological grey literature reports, with a view to distinguishing a finding of evidence (for example) of Roman activity from statements reporting a lack of evidence, or no sign of Roman remains. A technique previously used in the biomedical domain was adapted to archaeological vocabulary and writing style. Evaluation on rules targeted at identifying negated cases of four CIDOC-CRM entities gives promising results, Recall 80% and Precision 89% [Vlachidis and Tudhope 2015].

Table I. Contents of the ARIADNE Catalogue (as of March 2016)

Data Resources		Data Resource Properties		Data Resource Types	
				Sites and monuments databases or inventories	1,529,498
		Spatial	98%	Event/intervention resources	51,820
Datasets	1,534,375	Temporal	100%	Artefact databases or image collections	40,726
Collections	43,182	Native Subject	97%	Scientific datasets	4,904
Textual Documents	50,807	Derived Subject	48%	Fieldwork archives	1,340
Total	1,628,364	Publisher	100%	Burial databases	76

4. EVALUATION

The effectiveness of the ARIADNE Infrastructure in providing services to its research community will be evaluated in time, by measuring the quantity and quality of usage of the services by archaeological researchers. However, during the project lifetime an initial evaluation is underway. This Section describes this evaluation, focussing on two central respects: the adequacy of the Catalogue, which provides an overview of the ARIADNE information space and supports the discovery functionality of the infrastructure; and the evaluation the remaining services.

4.1 Contents of the ARIADNE Catalogue

The ARIADNE consortium consists of 24 partners in 16 countries including Sweden, United Kingdom, Ireland, Germany, Austria, Hungary, Czech Republic, Slovenia, France, the Netherlands, Italy, Spain, Greece, Cyprus, Romania and Bulgaria. The ARIADNE discovery service has been developed to create a single global access point which provides open access to integrated archaeological information and supports researchers and professionals, educators and students as well as the wider interested public.

After ingesting the metadata of the ARIADNE consortium into the Catalogue, this service has been evaluated several times. Adaptations of the ARIADNE ACDM took place in order to maximise the effect that the services created by ARIADNE will have on the different stakeholders. The different metadata schemas have commonalities which allowed mapping to each other. Crosswalks to establish and provide an integrated approach can be made. Improved thesauri are helping to overcome linguistic barriers by linking related terms expressed in different languages.

Table I gives an overview of the current contents of the Catalogue. All descriptions provided by the ARIADNE partners could be mapped to the ACDM and therefore inserted into the Catalogue. This proves that (1) the ACDM is rich enough to accommodate all such descriptions while at the same time supporting the rich discovery and browsing services outlined in Section 3.2 and 3.2.4, respectively; (2) the MORe aggregation infrastructure is powerful enough to implement and correctly apply the several mappings involved in the aggregation process, and to cope with the variety of data provisions mechanisms exhibited by the participating institutions. So, both the ACDM and the aggregation infrastructure have accomplished their mission.

The numbers shown in Table I are already significant, covering a large percentage of the data made available by the ARIADNE partners. Further additions are expected before the end of the project. Above all, it is expected that opening the Catalogue to the whole archaeological community will bring other descriptions, further enlarging the ARIADNE information space.

4.2 Service evaluation

The French National Institute for Preventive Archaeological Research (INPRAP) is in charge of testing the services produced within ARIADNE Project. The evaluation will be implemented in two complementary methods: predefined testing scenarios and open evaluation questionnaires. The aim of the questionnaire, related to a specific service, is to determine whether the service meets the expectations

of the users. The questionnaire will provide precise questions about usability of the service; open comments about the service (usability, request for improvements and so on); a note given to the service (*e.g.*, from 1 to 5); quantitative data about usage (*e.g.*, number of downloads, number of running processes, number of files uploaded, *etc*). A quarterly analysis of the data results is planned. The evaluation process for any service requires the full availability of the service, in a stable version. It also requires the availability of a comprehensive set of data appropriate to using the service.

The project just completed a first testing phase, in which a panel of 30 testers evaluated three main services developed within the project, namely the Portal, the Visual Media Services, and the Landscape Factory. The first results have shown a great interest in the services. The tests showed a high approval rating: 3.86/5.0; 4.14/5.0; 4.0/5.0, respectively. The main problems highlighted by testers are related to ergonomics of the system (GUI and need for more explanation) and the request of some specific tools to add in the Visual Media and Landscape Services, *e.g.*, measure tools, metadata management, export tools. Services strengths are undoubtedly speed and great potential, reported by many testers.

A second testing phase upon the whole system will be conducted from July to December, 2016. This testing phase will focus on the usability of the ARIADNE service infrastructure taken as a whole, the ARIADNE portal being the entry point.

5. CONCLUSIONS AND OUTLOOK

This paper has presented the ARIADNE Infrastructure, an ongoing European initiative that aims to create a single information space, where the data and the services owned by European archaeological institutions can be discovered and accessed through a single search facility. After reviewing the current landscape of Research Infrastructures in archaeology, a set of requirements have been distilled and addressed by technological developments. At the heart of the ARIADNE Infrastructure is the ARIADNE Catalogue, which describes the elements of the ARIADNE information space and supports their discovery and access. The Catalogue is the result of a major effort; first, an adequate data model has been created and validated by the members of the consortium. An aggregation infrastructure has then been set-up for populating the Catalogue, by transforming the descriptions collected from the contributing partners. Finally, the discovery functionality has been implemented by relying on the Catalogue contents and on the state-the-art search engine provided by elasticsearch. Thanks to this effort, the archaeological community has a unique access point where its resources can be found. Having tested the infrastructure with a broad range of archaeological content, the Catalogue and its supporting technology will be opened up for contributions from the broader archaeological community, thereby becoming a global knowledge source for archaeology.

ARIADNE has also started to address the item-level integration of archaeological data, by conducting Linked Data experiments on data related to coins. For this, it has relied on the pivotal role of a well-known ontology for cultural data integration, the CIDOC CRM, and on the associated technology for mapping and transformation. The paper has described the experiment in some detail, since it is key to an important future development, namely the creation of knowledge aggregations where researchers can find answers to research questions spanning several datasets. The experiment has had encouraging results, and will form an important item for the further agenda. ARIADNE has also begun to tackle the sharing of services, making services on visual data and on reference resources accessible through the infrastructure.

Nonetheless, it is clear that much remains to be done before Archaeology has a mature research infrastructure.

—Data integration needs to be undertaken more systematically, by making the available tools and resources available to the community through a state of the art Virtual Research Environment,

where domain experts can convene to collaboratively develop the necessary transformation rules and to apply those rules to create integrated data.

—A permanent conduit needs to be created through which archaeologists can channel their requirements to the relevant technological research and development communities, who can then respond with matching technology.

—The work of researchers in archaeology needs to be supported in a more substantial way, by endowing the infrastructure with the ability to understand and manage the knowledge generation process. This support requires tracking the provenance of the data found in the infrastructure, and the possibility of defining, sharing and executing complex workflows for processing the data.

All of this is within the reach of current ICT technology. ARIADNE has been made possible through European funding which has allowed archaeologists and information scientists throughout Europe to collaborate on solving some major issues. We look forward to working together to fulfill the next stage of our vision.

REFERENCES

Project ARIADNE. 2014. *First Report on Users Needs*. Deliverable D2.1. Retrieved October 26, 2016 from http://www.ariadne-infrastructure.eu/Resources/D2.1-First-report-on-users-needs

Project ARIADNE. 2015a. *Preliminary Innovation Agenda and Action Plan*. Deliverable D2.3. Institution. Retrieved October 26, 2016 from http://www.ariadne-infrastructure.eu/Resources/D2.3-Preliminary-Innovation-Agenda-and-Action-Plan

Project ARIADNE. 2015b. *Second Report on Users Needs*. Deliverable D2.2. Retrieved October 26, 2016 from http://www.ariadne-infrastructure.eu/content/view/full/1188

Project ARIADNE. 2016. *Report on Thesauri and Taxonomies*. Deliverable D15.1. Retrieved October 26, 2016 from from http://www.ariadne-infrastructure.eu/Resources/D15.1-Report-on-Thesauri-and-Taxonomies

C. Binding, K. May, and D. Tudhope. 2008. Semantic interoperability in archaeological datasets: Data mapping and extraction via the CIDOC CRM. In *Proceedings 12th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2008) (Lecture Notes in Computer Science)*, Vol. 5173. Springer Verlag, Aarhus, 280–290.

C. Binding and D Tudhope. 2016. Improving Interoperability using Vocabulary Linked Data. *International Journal on Digital Libraries* 17, 1 (2016), 5–21.

M. Doerr. 2003. The CIDOC Conceptual Reference Model: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* 24, 3 (2003), 75–92.

A. Felicetti, P. Gerth, C. Meghini, and M. Theodoridou. 2015. Integrating heterogeneous coin datasets in the context of archaeological research. In *Proc. of the Workshop: Extending, Mapping and Focusing the CRM, TPDL 2015*. Poznan, POL.

H.J. Hansen. 1992. European archaeological databases: problems and prospects. In *Computing the Past. Computer Applications and Quantitative Methods in Archaeology*, Andresen et al. (Ed.). Aarhus, 229–37.

A. Isaac, V. Charles, K. Fernie, C. Dallas, D. Gavrilis, and S. Angelis. 2013. Achieving Interoperability between the CARARE schema for Monuments and Sites and the Europeana Data Model. In *Proceedings of the International Conference on Dublin Core and Metadata Applications, DC-2013*. Lisbon, Portugal, 115–125.

ISO. 2013. *Information and documentation - Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies*. Technical Report 25964-2:2013. ISO - International Organization for Standardization. Retrieved October 26, 2016 from http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53658

J. Kenny and J. Richards. 2005. Pathways to a Shared European Information Infrastructure for Cultural Heritage. *Internet Archaeology* 18 (2005). Retrieved October 26, 2016 from http://dx.doi.org/10.11141/ia.18.6

M. Kuna, J. Hasil, D. Novák, I. Boháová, L. Čulková, P. Demján, D. Dreslerová, M. Gojda, I. Herichová, D. Křivánková, O. Lečbychová, J. Mařík, J. Mařková-Kubková, M. Panáek, J. Podliska, A. Pokorná, J. Řihošek, E. Stuchlíková, M. Suchý, J. Válek, N. Venclová, and L. Haišmanová. 2015. Structuring archaeological evidence. The Archaeological Map of the Czech Republic and related information systems. Institute of Archaeology of the Czech Academy of Sciences, Prague. Retrieved October 26, 2016 from https://www.academia.edu/25951973/Archaeological_Map_of_the_Czech_Republic_Structuring_archaeological_evidence_

F. Maali and J. Erickson. 2014. *Data Catalog Vocabulary (DCAT)*. Recommendation. World Wide Web Committee (W3C). Retrieved October 26, 2016 from https://www.w3.org/TR/vocab-dcat/

N. Minadakis, Y. Marketakis, H. Kondylakis, G. Flouris, M. Theodoridou, M. Doerr, and G. de Jong. 2015. X3ML Framework: an effective suite for supporting data mappings. In *Proc. of the Workshop: Extending, Mapping and Focusing the CRM, TPDL 2015*. Poznan, POL.

M. Mudge, T. Malzbender, A. Chalmers, R. Scopigno, J. Davis, O. Wang, P. Gunawardane, M. Ashley, M. Doerr, A. Proenca, and J. Barbosa. 2008. Image-Based Empirical Information Acquisition, Scientific Reliability, and Long-Term Digital Preservation for the Natural Sciences and Cultural Heritage. In *Eurographics 2008 - Tutorials*, M. Roussou and J. Leigh (Eds.). The Eurographics Association. DOI:http://dx.doi.org/10.2312/egt.20081050

F. Ponchio, M. Potenziani, M. Dellepiane, M. Callieri, and R. Scopigno. 2015. ARIADNE Visual Media Service: easy web publishing of advanced visual media. In *Proc. of Computer Applications and Quantitative Methods in Archaeology (CAA) 2015*. Archaeopress, Oxford, UK.

J. Richards, D. Tudhope, and A. Vlachidis. 2015. Text Mining in Archaeology: Extracting Information from Archaeological Reports. In *Mathematics in Archaeology*, J. Barcelo and I. Bogdanovic (Eds.). CRC Press, Chapter 12.

K. Tzompanaki and M. Doerr. 2012. *Fundamental categories and relationships for intuitive querying CIDOC-CRM based repositories*. Technical Report TR-429. ICS-FORTH.

K. Tzompanaki, M. Doerr, M. Theodoridou, and I. Fundulaki. 2013. Reasoning based on property propagation on CIDOC-CRM and CRMdig based repositories. Online Proceedings for Scientic Workshops. (2013).

A. Vlachidis and D. Tudhope. 2015. Negation detection and word sense disambiguation in digital archaeology reports for the purposes of semantic annotation. *Program: electronic library and information systems* 49, 2 (2015), 118–134.

Received February 2016; revised July 2016; accepted October 2016