Humboldt-Universität zu Berlin
Institut für Bibliotheks- und Informationswissenschaft

Berliner Handreichungen
zur Bibliotheks- und Informationswissenschaft

Heft 359

A Digital Registry for Archaeological Find Spots and Excavation Documentation in IANUS

von
Wibke Kolbmann

A Digital Registry for Archaeological Find Spots and Excavation Documentation in IANUS

von
Wibke Kolbmann

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 359

Kolbmann, Wibke

A Digital Registry for Archaeological Find Spots and Excavation Documentation in IANUS / von Wibke Kolbmann. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2014. - 57 S. (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 359)

ISSN 14 38-76 62

Abstract:

Grey literature (site notebooks, reports etc.) and research data in archaeology are invaluable sources of information currently lacking a central reference registry in Germany. This paper discusses requirements and the underlying data model of a registry to be developed for find spots and archaeological excavation data within the IANUS project at the German Archaeological Institute. This registry is to collect information on archaeological investigations data for a finding aid service. The focus for this registry will be based on the collection of metadata about primary data and grey literature, not on secondary data or on publications. Starting with the acquisition of basic metadata needs drawn from the IANUS mission and project charter. A review of already existing projects and initiatives in this field (EDNA, tDAR, ADS, Open Context) provides more details about which information should be captured during a registration of research data for a long term digital preservation archive. Finally recommendations for the data model of this registry are drawn from the evaluation of existing generic and archaeology-specific metadata standards (Dublin Core, EDM, LIDO, ADeX, CARARE).

Diese Veröffentlichung geht zurück auf eine Masterarbeit im postgradualen Fernstudiengang M. A. Bibliotheks- und Informationswissenschaft (Library and Information Science) an der Humboldt-Universität zu Berlin.

Online-Version: http://edoc.hu-berlin.de/series/berliner-handreichungen/2014-359

Table of Content

1 Introduction

Already in 1971 at the beginning of the use of personal computers Robert G. Chenhall (Chenhall 1971) expressed the idea of a data infrastructure in archaeology to share information.

> For a number of years, a small group of archaeologists scattered around the world have been struggling toward what seems an impossible dream: the creation of a world-wide, computer-oriented data bank. The motives for engaging in this struggle are undoubtedly as varied as the number of persons involved.

He also mentioned the problems that would be serious obstacles for the realisation of such an infrastructure.

> (1) in spite of the claims of computer manufacturers, computer hardware and software have only very recently been adequate for this kind of project,
>
> (2) organizational facilities have not been available to support any sort of a consistent and unified project, and, perhaps most important,
>
> (3) archaeological theory has been so diffuse and ill-define that no one has been able to say with assurance what should be recorded in such a data bank.

Furthermore he raised the question as to how such a system could be implemented.

> First, are we talking of one massive computer bank, or a network of subsystems located in different geographic regional centers? The network idea is, for many reasons, probably the better alternative. […]
>
> A second series of questions, for which I am not prepared to suggest answers at this time, involves the whole matter of organization and financing.

Forty years later Julian D. Richards (Richards 2012) summarizes the current state of affairs. The European Strategy Forum on Research Infrastructures (ESFRI) and the Seventh Framework Programme of the European Commission (FP7) have provided funding for scientific research infrastructures across a range of disciplines. For the Arts and Humanities the projects DARIAH (Digital Research Infrastructure for the Arts and Humanities) (DARIAH-EU 2013) and ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe) (ARIADNE 2013) have been established. Their goal is to facilitate multilateral initiatives leading to the better use and development of research infrastructures, at EU and international levels. ARIADNE thereby aims to integrate the existing archaeological research data infrastructures so that researchers can use the various distributed datasets. This is because many archaeological research questions cannot be bound to modern political boundaries. Richards is convinced that this will only be possible if such an infrastructure builds on trusted and secure national data repositories for long term preservation as it is unrealistic

that all archaeological data in Europe will ever be brought together in a single database. Such data is therefore better maintained at the national or regional level where there is ownership and a legal responsibility to maintain archives.

Whereas in the Netherlands, the United States and the United Kingdom such national repositories are already in place Germany is unfortunately still lacking an organisation and national repository for long term digital preservation of archaeological data. The German Research Foundation has now provided funding for the first phase of the IANUS project (IANUS 2013) which aims to define a concept for a Research Data Centre for Archaeology and Ancient Studies. IANUS started its work in April 2012 coordinated by a team of two people at the German Archaeological Institute in Berlin. The German Archaeological Institute (DAI 2013) is also partner in DARIAH-DE and ARIADNE and will align the activities in IANUS with these projects to ensure best possible interoperability with their services.

This paper will discuss the requirements and the underlying data model of a registry to be developed for find spots and archaeological excavation documentation data. The registry will be one component of the services offered by IANUS. It is to collect information on archaeological investigations data for a finding aid service. The focus for this registry will be based on the collection of metadata about primary data and grey literature, not on secondary data or on publications. A thorough discussion of GIS, CAD and virtual reality as means of data analysis and visualisation of findings in archaeology is therefore out of scope of this paper.

The discussion follows the methodology for a metadata life cycle for digital libraries suggested by Chen, Chen and Lin (Chen, Chen, and Lin 2003). Starting with the acquisition of basic metadata needs drawn from the IANUS mission and project charter and from the results of a workshop of the Interoperability Working Group in IANUS. A review of already existing projects and initiatives in this field provides more details about which information should be captured during a registration of research data for a long term digital preservation archive and what kind of workflows proved to be best practice. The main tasks of an archive can be simply stated in three words — selection; preservation; access. Therefore the following chapters will evaluate metadata standards for collection, project/study and item level descriptions to investigate deeper metadata needs for providing access to archaeological research data. In this way considering the special requirements in the context of a national repository for both academic institutions, whose research is based on funding and therefore has to respect conditions for data access made mandatory by the German funding organisations, and Federal State Departments of Archaeology more concerned with the administration of historic preservation are considered. Finally a strategy for the data model of the registry should be identified. Firstly to meet the needs for transferring data into a long-term digital preservation archive and building up a finding aid for archaeological research data in Germany. Secondly to achieve interoperability with the well-known metadata standards discussed as a basic step towards a network for data exchange amongst archaeological research data centres in Europe. The results of this thesis constitute the groundwork for the development of a first prototype of the registry.

2 The IANUS Project and the Scope of a Registry for Find Spots and Excavation Documentations

The German Research Foundation's funding programme “Information Infrastructures for Research Data” which provides the funding for the IANUS project defined some goals for the applicants that have an impact on the design of the registry to be discussed here. The programme aims to support science and the humanities in drawing up and implementing specific and needs-orientated digital infrastructures for the improved handling of research data and research data repositories.

Research data, for the purposes of this funding, is to be understood as digital and electronically-storable data collected in the course of a research project, for example, as a result of source research, experiments, measurements, surveys or interviews.

Projects that received funding aim to develop discipline-specific forms of organizations to secure, store and enable re-use of research data and build a repository if necessary. In addition, these projects, taking into account the overall objectives, have to provide discipline-specific guidelines for the management of research data and measures for their implementation. They should develop guidelines for cataloguing and mechanisms for processing research data incorporating the requirements of the wide variety of research data related to the form, amount and the usage scenarios. They also need to take into account the scientific and legal interests of scientists and researchers for their need for accessibility and citability of data. Therefore they have to determine metadata standards as well as quality control procedures to sustain interpretability of the data.

Is there a need for digital archiving in archaeology in Germany? What is required for the re-use of data in archaeology? A survey by Harrison Eiteljorg (Eiteljorg 2002) revealed that much of the information produced by archaeological research is locked in unstructured written reports scattered in libraries, museums, repositories, and offices around the world. The data that underlies these reports is stored on computer cards, magnetic tapes, floppy disks, CDs, DVDs and external drives and is in danger of degradation. At the same time the technology to retrieve the data and the human knowledge to make them meaningful also disappears.[1]

It is also important to distinguish the phases in which digital data in archaeology is captured during fieldwork, post-excavation, publication and when it is archived (Backhouse 2006). Each phase requires a different approach for storing and enabling accessibility as the data captured differs. During fieldwork the data produced contains digitally transformed context sheets, photography, CAD maps from initial surveys and site layouts. However due to the sometimes difficult to access sites manual capturing of data on paper and later on transferring it into a database is not an unusual procedure. Database analysis, virtual reality models and GIS are usually done in the post-excavation phase. As transferring data from paper and data analysis both take place at the same time, namely post-excavation, the distinction between primary and secondary data in archaeology is

1 Unpublished surveys conducted by Dr. Thomas Götzelt and Rainer Komp at the German Archaeological Institute also confirm this conclusion for the situation in Germany.

slightly blurry. Publication of results in archaeological research has been quite costly due to pages of illustrations and photographs attached to them. Archaeological research also only reaches a very small readership. New forms of linking publications with raw data, maps etc. are seen to be a chance to make them more affordable and therefore increasing the visibility of archaeological research (Richards 2006; Babeu 2011, p. 64). Whereas archiving of paper-based archaeological reports follows clearly defined workflows, these are missing for the archiving of digital data and what is most precarious is the missing contextual information for the digital data. It is almost impossible for an outsider to understand the meaning of data collections or different file versions without any explanation. This of course is counterproductive for reuse of data and for new evolving methods of quantitative data analysis in archaeology.

A blind test for reusing archaeozoological datasets organised by the Open Context team, used to evaluate documentation needs for legacy data, gave insight into how researchers approach the task of analysing unknown data collections (Atici et al. 2012). The researchers began their analysis by taking an inventory of the database, and checking for misspellings, mismatches and errors. They were missing contextual and methodological information to help them with judging the integrity of the dataset. The minimum of contextual information the researchers defined was time and place of recording. Other information they thought to be useful were, the name of the original analyst, decoding of data and elements as well as how identification of the species were derived. All the blind analyses by the researchers produced diverse analytical results confirming that interpretations depend on many analytic choices. In this way, access to primary data is needed to replicate scientific results.

2.1 The IANUS Project — Its Objectives and the current Situation of digital Data Preservation in German Archaeology

The German Archaeological Institute applied for a grant in the above mentioned funding programme. The concept for the project is shortly outline in the next section.

As the resources of the project are limited it will focus on the long term digital preservation of primary data as recommended by the German Research Foundation (DFG - Deutsche Forschungsgemeinschaft 2009) and on dissemination of, and easy access, to the data archived. In the initial phase of the project a comprehensive portfolio analysis of existing long term digital preservation archives in the field of archaeology, the social sciences and related natural sciences is planned. In addition, the compilation of the discipline-specific needs in archaeology and ancient studies to build a research data centre will be undertaken. IANUS intends to set up two testbeds, one will be the registry for find spots and excavation documentation.

The starting position for IANUS is strongly influenced by the heterogeneous fields of activity in archaeology which might have tempted Chenhall to assume that archaeological theory has been diffuse and ill-defined. There is a broad spectrum of specialized sub-disciplines whose needs have to be considered: archaeological research about diverse cultures and regions, the classics, ancient history to archaeometry, archaeozoology and computational archaeology just to name a few (Renfrew and Bahn 2008; Bernbeck 1997). This also includes supporting data preservation for a broad variety of methods to reconstruct cultural and natural formation processes: excavations (stratigraphy), historical building research, prospections and surveys (inspections, test excavations), patterning, analysis of finds/findings (e.g. pottery analysis), photogrammetry, different science and art based methods of dating objects, climate and landscape history (geology, geomorphology, hydrology) and anthropological investigations (skeletons; food pattern; diseases; genetics) resulting in primary data in heterogeneous formats. photographs, aerial photographs, satellite images, texts, databases, survey data, point clouds, 3D reconstructions and models, videos, vector drawings to large-and small-scale scans.

An additional challenge to meet will be merging information of born digital data with digital surrogates as there has been a change in the means used for documenting within archaeology. In the early days of archaeological research plaster casts, drawings, squeezes and written reports were used. Now we observe a rapid growth in digital collections of object-related and spatial information in databases, using GIS, statistical analysis etc. New methods for analysing primary data have evolved, such as quantitative analysis. However as there is only one chance of cataloguing archaeological investigations, because excavations and surveys destroy the unique contexts and archaeological evidence, these new methods also depend on old data that is currently widely distributed, partly non digital and therefore inaccessible.

A first step towards improved accessibility to primary data is by enabling resource discovery via a finding aid service to improve knowledge of data collections. The archaeological community in Germany will have to agree on a minimal standard to

describe collections of primary data and single items. Here IANUS can build on the preceding work of the Association of State Archaeologists. In 2005, the association started to work on the data exchange standard called ADeX (Archäologischer Datenexport-Standard). Due to its focus on historic preservation issues such as localization, mapping and protection criteria of sites, this standard does not support more complex queries about, for example, chronological correlations. In this way, ADeX will need to be extended or integrated with another standard to allow for the capturing of more detailed and broader descriptive information on archaeological investigations. A suggestion of how this could be done is part of this essay. Another issue to be discussed is access rights to archaeological data. Although the German Research Foundation is requiring that all research data created by projects with public funding is to be published in open access this is not possible in all circumstances for archaeological data as there are restrictions regarding intellectual property rights of for instance satellite images or LIDAR scans.

2.2 The Registry — a Technical Component in the IANUS Infrastructure

The definition of the registry should follow the example of the Information Environment Service Registry (IESR) project (Powell, Black, and Collins 2011). Here a registry is understood to be a repository of specialized data that describes datasets and the functionality of various network-addressable services. In addition, the registry contains service metadata, interface descriptions, and other information that users need in order to access and use the data and services. The digital registry for find spots and excavation documentation in the IANUS infrastructure will be the tool to collect metadata on archaeological sites and collections of primary data, for uploading of data and datasets to be transferred to the long term digital preservation archive and for information on interface descriptions of other repositories with archaeological data. The ingestion of metadata and data starts here and the tool should guide data providers through the process of registering and depositing their content.

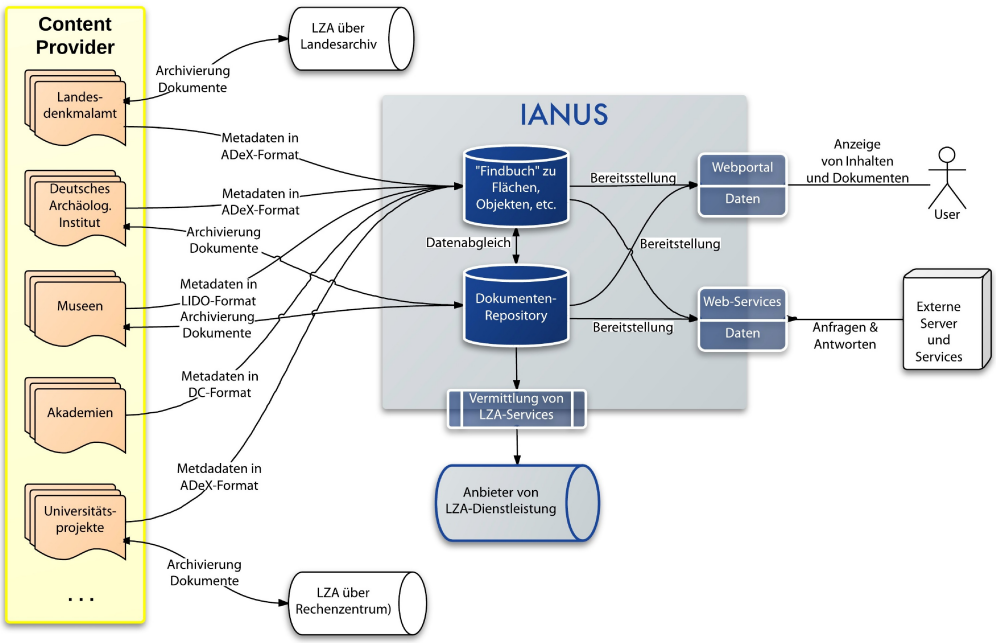

Fig. 1: Basic concept of the data flow in the IANUS infrastructure, Felix F. Schäfer, 2012

Furthermore the IANUS infrastructure will provide centralized file-sharing and backup of dynamic data for ongoing projects and for long-term preservation of selected, static data. Standardised metadata and file formats will facilitate retrieval and free dissemination of open (published) data and raw-data via an online portal. Additionally data provision via web services for data exchange and interoperability

with aggregators such as DARIAH-DE and the German Digital Library is also planned. The registration process has to obtain all the information needed for offering these services. In the next section the evaluation of existing registry services for scholarly content and research data should deliver a more detailed insight into the workflows and the description levels required for registration.

3 Existing Registry Services for Research Data

The concept and data model of the IANUS registry for find spots and excavation documentation can build on the experience of already existing registries for research data. The choice of the registries introduced in the following section is based on whether IANUS may contribute content to these registries and therefore has to meet their requirements to be interoperable with them. This is the case for the collection registry of DARIAH-DE and re3data.org,. In addition there is a focus on registries for long term preservation archives in the field of archaeology. Here the e-Depot Nederlandse Archeologie (EDNA), The Archaeological Record (tDAR) and the Archaeological Data Service (ADS) were chosen because these preservation archives work closely together and are well advanced towards the described network of shared archaeological data. Open Context should also be discussed although it is not quite in line with the other registries as it initially focused on data discovery only but has just recently entered in providing long term preservation for data.

The evaluation of these registries focuses on their overall mission and how it compares with the mission of IANUS, as well as on the definition of the workflows to deposit data and how the division of work between content/data provider and data curation of the service is organised. Collecting and merging information and data from many different partners is always a challenge regarding the willingness and the previous knowledge of the people involved. How do these archives therefore support content providers? How did they adjust their registry concept and procedures based on user feedback?

Furthermore, it is of interest as to how these services enable resource and data discovery and on which level in terms of data granularity they give access to the stored content. Do they provide access to the data as received? How do they achieve interoperability with other data sources on repository, schema and record level?

Funding organisation require research projects that data produced with public funding money is published open and with free access. The provenance of archaeological data is very heterogeneous. Some data providers in archaeology are not only bound to the requirements of funding organisation but also to the regulations of the countries where the excavation or investigation is carried out. Others have to consider historic preservation policies or are based on research carried out with third party data and are therefore subject to the intellectual rights applied by these parties. Which paradigm for rights management will prove to be sustainable?

3.1 Interdisciplinary Registry Services

Currently there are two interdisciplinary registry services for research data and research repositories under development in Germany. The scope and sustainability of these registry services is still in flux. Nevertheless IANUS will need to prepare to deliver content to these registries as this is one possible way to open up archaeological data for interdisciplinary use. Interdisciplinarity in research

activities is seen to foster innovation. That this is especially true for archaeological research is evident by the tremendous input new technologies and adoption of methods from natural science have, for example, in building chronology in archaeology.

3.1.1 Collection Registry of the 'Digital Research Infrastructure for the Arts and Humanities' project (DARIAH-DE)

DARIAH (DARIAH-EU 2013) is being established as a European Research Infrastructure Consortium (ERIC). It is to facilitate long-term access to, and use of, all European Arts and Humanities (A+H) digital research data. The DARIAH infrastructure will be a connected network of people, information, tools, and methodologies for investigating, exploring and supporting work across the broad spectrum of the digital humanities. The core strategy of DARIAH is to bring together national, regional, and local endeavours to form a cooperative infrastructure, which operates through a European wide network of Virtual Competency Centres (VCC). Currently there are four VCCs: VCC1 e-Infrastructure, VCC2 Research and Education Liaison, VCC3 Scholarly Content Management and VCC4 Advocacy, Impact and Outreach.

DARIAH-DE (DARIAH-DE 2013) is the German national contribution to DARIAH-EU. It is funded by the Federal Ministry of Education and Research (BMBF). Currently, 17 partner institutions from the fields of information technology as well as the arts and humanities are involved in DARIAH-DE, including universities, data centres, research institutions, one academy, one commercial partner, and one library.

The technical core components of the DARIAH-DE e-infrastructure are distinguished as generic services and demonstrators and academic services.

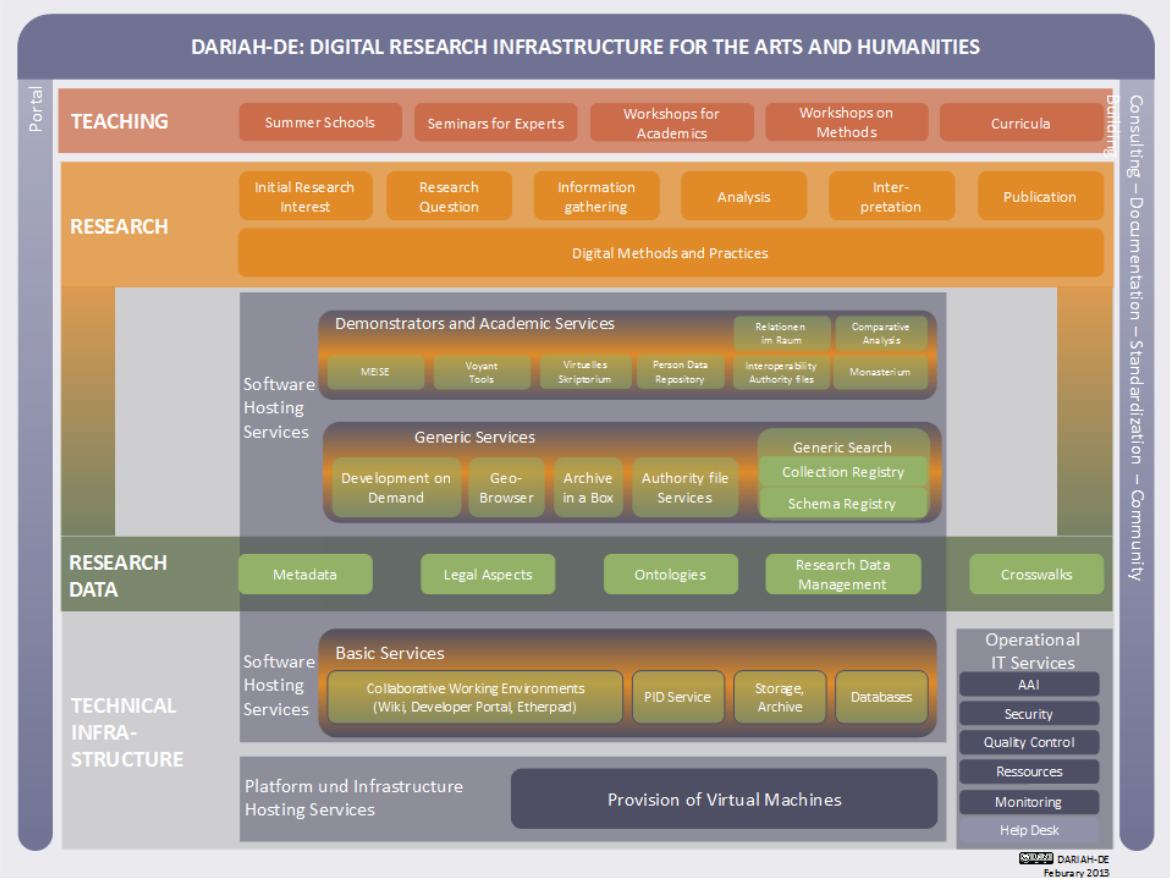

Fig. 2: Chart of DARIAH-DE activities and services, DARIAH-DE 2013

The collection registry, the schema registry and the generic search are generic services and together form a unit. DARIAH-DE does not intend to undertake data integration on the level of data of a uniform standard. The idea is to collect information on data collections at collection level together with information on interfaces, APIs to access the data. The schema registry is to collect information on the metadata schema used for the registered collections and includes a crosswalker which should allow users to pick data collections and the respective schema and individually create specific mappings to extract and convert data for re-use.

The need for collection level description information particularly for DARIAH-DE additionally focuses on:

- Enabling the interfacing of data repositories with services and applications,
- Improving accessibility of sensitive data for research purposes within a trusted registry due to differentiated access rights,
- Monitoring the growth of the partner network.

The data model of the collection registry of DARIAH-DE is based on the Dublin Core Collection Application Profile (DCCAP) (Dublin Core Metadata Initiative 2013).

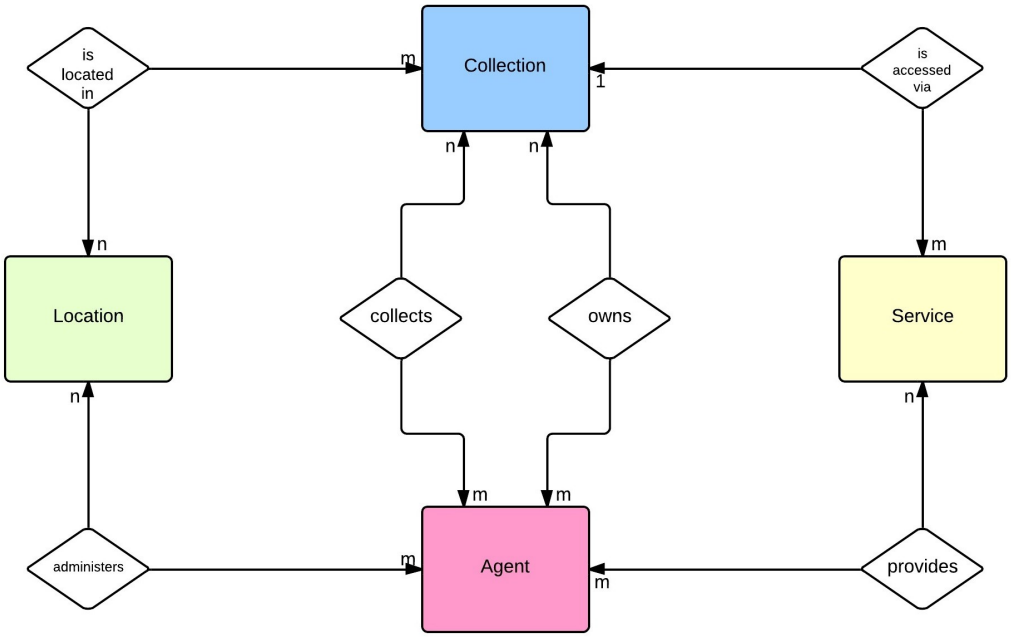

Fig. 3: Entity relationship model of entities adapted from the Dublin Core Collection Application Profile for the DARIAH-DE Collection Registry

A survey of the European DARIAH partners proved that most of them could support the Dublin Core standard for collection description as a minimum standard. As they already have implemented this standard it will not require more or redundant effort from them for data collection description. A first prototype of the collection registry was presented at a workshop on metadata standardization at the Leibniz-Institut für Europäische Geschichte in Mainz. Currently it is still necessary to enter collection descriptions manually but it is planned to harvest them from registered repositories if possible. The DARIAH-DE partners have already started to register about 70 collections so far.

3.1.2 r3data.org

re3data.org is a project funded for two years from the German Research Foundation (2012-2014). Its mission statement reads as follows:

> The goal of re3data.org is to create a global registry of research data repositories. The registry will cover research data repositories from different academic disciplines. re3data.org will present repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. In the course of this mission re3data.org aims to promote a culture of sharing, increased access and better visibility of research data. (re3data.org 2013)

The partners of re3data.org are the Berlin School of Library and Information

Science, the German Research Centre for Geosciences (GFZ) and the Karlsruhe Institute of Technology (KIT) Library. As the registry for research data repositories to be developed is not focused on a specific domain of sciences like DARIAH-DE for the Arts and Humanities and as the partners from re3data.org have a background in information science and natural science the data model is focused more on gathering information on interfacing data and access rights and not so much on the description of the data collection content (Vierkant et al. 2012).

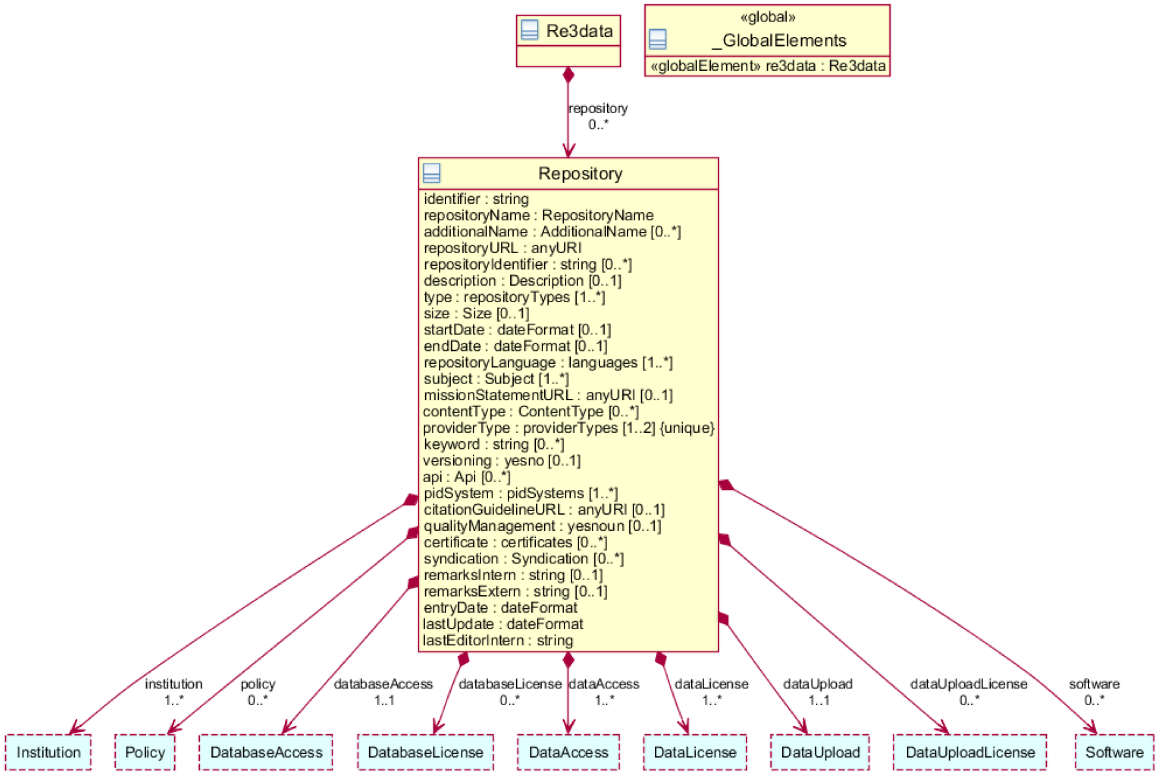

Fig. 4: re3data.org, Vocabulary for the Registration and Description of Research Data Repositories

re3data.org offers a search based on the metadata collected about repositories. Users can also suggest the registration of new repositories that are missing. The description of the repositories is done by the editorial team of re3data.org itself. Currently 180 repositories with a full description are registered and 160 repositories are still in the process of being described (Scholze, Goebelbecker, and Pampel 2013).

re3data.org also plans to improve the import and export of repository descriptions based on XML and Dublin Core. Furthermore they want to provide statistics based on the gathered descriptions which might be considered as a service to improve research evaluation based on impact factors on data output and not only publications.

3.2 Archaeology-specific Registry Services

The granularity of data collection and research data repository descriptions of interdisciplinary registry services is rather coarse. Can we assume that discipline-specific research data registries and long term archives, due to a better understanding of the use cases for data re-use offer more comprehensive and useful information on data collections?

Unlike the interdisciplinary research data registry projects described, the organisations that will be introduced in the next section, have an additional task of long term digital preservation and of providing sustainable access to research data in an interoperable way. Long term digital preservation of data builds on different strategies: refreshing, replication, migration, emulation, persistent identification and metadata management.

Refreshing and replication are two methods of bit preservation. Refreshing is the regular copying of data to the same media. Contrary to migration the binary data formats are not changed. Replication is the procedure of multiple storage of the same data in several different locations and the synchronization of these data sources (Ulrich 2010).

Migration is converting data between storage types, file formats, or computer systems. At the beginning of the depositing process of data into a digital preservation archive it is mostly necessary to migrate data to a preservation format. Often the conversion processes of digital objects are then connected with modifications, so that they remain usable. Most of the time this leads to loss of data (Funk 2010a).

The emulation model for preservation envisages archiving the digital object together with metadata and all data and information that is necessary to run the digital objects in a new environment using emulation software that imitates or emulates the old hardware and software environment (Funk 2010b).

Persistent Identifiers in digital preservation are needed for referencing to a digital object, a set of files or record (Schroeder 2010). This is especially important for citability of resources, data or datasets.

Which of these strategies do archaeological digital preservation archives follow? What is the effect of this on the design of their registry?

3.2.1 e-Depot Nederlandse Archeologie (EDNA)

After a pilot project from 2004 to 2006, in which requirements for the digital preservation of archaeological data were gathered, the e-depot for Dutch archaeology (EDNA 2013) was established as part of DANS (Data Archiving and Networked Services), and funded by KNAW, one of the main Dutch Research Councils. The e-depot stores the digital files with research data of Dutch archaeologists. These are files with the primary archaeological data of excavations, regional explorations and material studies. Since 2007 excavation methodology and documentation in the Netherlands are legally bound to a strict procedure laid down in the quality standard for Dutch archaeology (Willems and Brandt 2004). It defines an archaeological heritage management cycle also including the depositing of data and documentation.

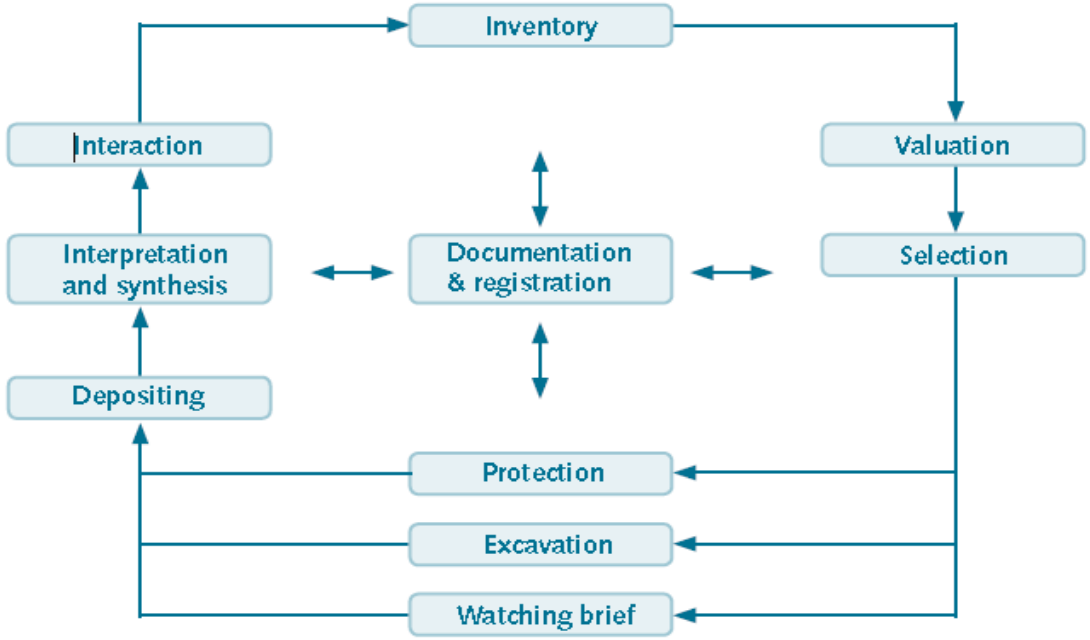

Fig. 5: The archaeological heritage management cycle, (Willems and Brandt 2004)

According to the quality standard, digital data has to be supplied to the State Archaeological Service within 4 weeks after the final report of a campaign or project. Products of the digital archive of a project should contain standardized reports, reference data, additional data and observations. EDNA has defined a list of preferred and acceptable file formats that can be deposited (see 10.1).

For depositing data EDNA offers the tool, EASY which guides content providers through the process of uploading their digital archive (DANS Data Archiving and Network Services 2013). The process of describing the project and uploading the digital archive with its resources is relatively straight forward due to most of the

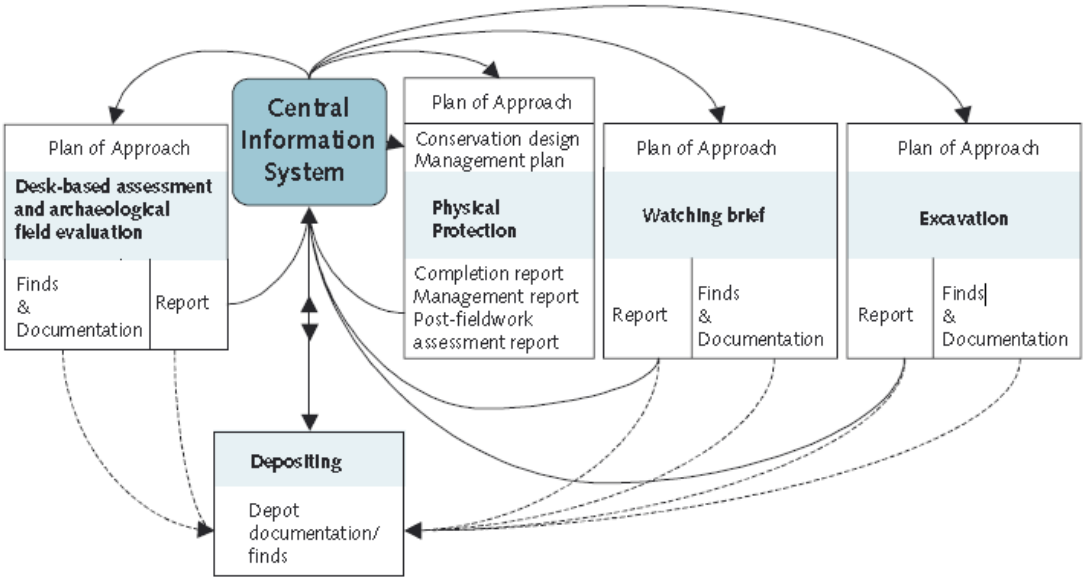

Fig. 6: Standardized process reports are part of the digital project archive.

contextual and methodological information being stored in the standardized reports and a mandatory codebook. The metadata element set for project and resource collection description is based on Dublin Core (DANS Data Archiving and Network Services 2011). After the login a new deposit has to be created and classified with the discipline associated with it. The content provider has to fill in six different forms:

1. Primary information and intellectual ownership – information on data collector and rights holder (Title*, Alternative title, Creator*, Contributor, Date created*, Rights holder, Publisher)
2. Research project — Descriptive information about the entire project (Description*, Subject (ABR Complex), Temporal coverage (ABR), Spatial coverage (point, box), Identifier (PID Scheme), Relation (Title, Url)
3. Digital Files — Descriptive information on digital files (Type (DCMI resource type), Format (Internet Media Type), Language (ISO 639), Source, Date (ISO 8601))
4. Upload files
5. Rights — access rights
6. Overview and submitting

The codebook contains in total, a list of 47 elements for the description of a deposit. EDNA offers four levels of access rights to choose from: open access -free for all registered users, restricted: archaeology group — only accessible by the user group of archaeologists, restricted: request permission — access is granted by the rights holder on personal request and other access. The access rights applied on project level are inherited by all resources. Only projects where the access right 'open access' has been applied, can be associated with a standard licence such as Creative Commons. In this way, only these projects and resources are available for data exchange in an automated and machine-

processable way.

All submitted files are checked and converted to a sustainable preservation format by the data curators of EDNA. The supplied data files (project and file metadata, codebooks and digital resources) are made available online through the EDNA search portal and are downloadable for registered users. A full integration of all data is not feasible for the archive therefore only the provided metadata on project description and data files is used for retrieval. By the end of 2011, EDNA provided access to over 17,000 reports.

3.2.2 The Archaeological Record (tDAR)

tDAR is hosted at the Arizona State University and has been supported since 2009 by a grant from the Digital Antiquity consortium from the Andrew W Mellon Foundation (McManamon, Kintigh, and Brin 2010). The mission of tDAR is (tDAR 2013a):

> One goal is to expand access to digital files related to a wide range of archaeological investigations and topics, e.g., archives and collections; field studies of various scales and intensities; and historical, methodological, synthetic, or theoretical studies. In order to accomplish this goal, Digital Antiquity maintains a repository for digital archaeological data.

In the Trans-Atlantic Gateway Project (TAG 2013) tDAR and the Archaeological Data Service (ADS) worked on the interoperability between their repositories developing an infrastructure to enable basic cross-search of Dublin Core compatible metadata records for digital resources covering the archaeology of the USA and UK. A similar undertaking had been the mission of the project ARENA — Archaeological Records of Europe —Networked Access for data from European countries (ARENA 2013; Kenny and Richards 2005; Austin 2003; Ross, Janevski, and Stoyanovich 2007).

In 2011 tDAR integrated the National Archaeological Database (NADB) with a significant number of reports generated by archaeological investigations for public projects across the United States. Grey literature therefore constitutes a large part of the content of the repository.

tDAR offers a web-based tool for creating metadata descriptions of resource collections, projects of archaeological investigations and digital resources, and for uploading these resources. All documentation of the tDAR development is publicly accessible as is the data dictionary describing the data model. It is very close to Dublin Core for the collection and project level descriptions (tDAR 2013b). On the level of the resource description there is a significant difference to the model of EDNA. Here tDAR requests users to enter the information online and not to provide a codebook together with the digital project archive. tDAR is also working a lot with controlled vocabularies for metadata descriptions offering value lists to classify sites, cultural, archaeological investigation and material types. It also allows users to describe and upload own ontologies as owl or rdf files. Although not required collections, projects and resources can be nested as needed.

Any one may register to use tDAR but only approved users can add resources. In

order to register, it is necessary to agree to accompany any use of tDAR information resources with proper citations of the data contributors, which is evidence of the citability of all resources. tDAR used DOI (Digital Object Identifier) for persistent identifiers. Digital resources may be registered as public with open access or private only available to users explicitly permitted access (tDAR 2013c). Rights can be assigned via a project or a collection to enable inheritance of these rights to the resources associated with it. However, it also is possible to change rights individually for each resource afterwards. In addition, individual fields (e.g., site location) may be flagged as confidential, with additional permission needed for access.

tDAR requires more help from content providers regarding the preparation of data than EDNA. Therefore it accepts only a limited choice of file formats (see 10.1). Uploaded databases are converted to a standard relational database format that is updated and maintained in the long term. Copies of the files, as submitted, are kept but are not updated as the software changes.

Currently users can search tDAR for digital documents, images, spreadsheets, data sets, ontologies and sensory data based on the metadata provided. Still under investigation is how file types such as GIS, GPS, CAD, 3D images, could be made available via visualization online. Registered users may download files if rights are granted accordingly.

3.2.3 Archaeological Data Service (ADS)

The Archaeological Data Service in York has the longest experience in digital preservation of archaeological data of all the digital archives described here. It was established in 1996 as one of the five discipline-based service providers making up the UK Arts and Humanities Data Service (AHDS). Since that time it provides support for digital preservation and re-use, for research, learning and teaching for the Archaeology and the Historic Environment sector. Furthermore it takes the lead in promoting standards and guidelines for best practice in the creation, description, preservation and use of archaeological information. It has a very detailed collections policy with a content development strategy and clearly defined administrative workflows. Its content development focuses on data from UK archaeologists and the archive gives priority to digital data for the archaeology of the British Isles. There is no thematic or chronological limit to the content. The ADS follows a layered collections policy providing access to digital resources archived and within the responsibility for preservation by the ADS, to resources that are catalogued and disseminated by the ADS but another institution maintains the long-term preservation of the digital resource and to digital finding aids and metadata of non-digital resources.

The ADS supports digital archiving of a vast variety of data types: electronic texts, fieldwork ‘grey literature’, electronic journals, PhDs, bibliographic finding aids, databases, geophysics, aerial photographs, topographic survey, including LiDAR, and 3-D laser scanning, buildings survey, visualisation such as 3-D reconstructions, VRML, GIS, moving images and still image collections.

To strengthen its position as repository for academic and high quality

archaeological data the ADS made agreements with funding agencies to encourage recipients of grant to offer their datasets for deposit to the ADS.

In their start-up phases, both EDNA and tDAR consulted the ADS for advice, adopting many of their documentation standards. This partly explains the dominance of Dublin Core within all three digital archives. The structure of ADS has grown historically, which is why it just recently changed to use a web-based tool for depositing data (SWORD 2013; SWORD-ARM 2013). Additionally to the depositing procedure, similar to the one of EDNA and tDAR, this tool contains an e-licensing system to assist depositors in applying licences to the resources and thereby helps streamlining the process of publishing data. A costing tool allows content providers to evaluate the financial outlay for archiving. The ADS-easy system (ADS-easy 2013) is integrated within the existing ADS infrastructure, so that project and file-level metadata can be uploaded directly to the collections management system.

All the content of the ADS is free of use to research and educational purposes. The ADS maintains an online catalogue 'ArchSearch' for search and retrieval of its resources and for cross-disciplinary search contributes to the AHDS gateway.

In the last couple of years, the ADS was the driving force for improving interoperability and exchange of archaeological data and for opening up deep web content through data integration to uniform standards. They participated in the development of the FISH Interoperability Toolkit (Forum on Information Standards in Heritage 2013) to convert data to the MIDAS XML schema, a common format for the storage, processing and exchange of historic environment information (English Heritage 2012).

In the STAR (Semantic Technologies for Archaeological Resources) and the STELLAR (Semantic Technologies Enhancing Links and Linked data for Archaeological Resources) projects, the ADS together with the University of Glamorgan and English Heritage are doing research on introducing semantic and knowledge-based technologies to archaeological data based on the English Heritage Centre for Archaeology ontological model CRM-EH, an extension of the CIDOC Conceptual Reference Model (Cripps et al. 2004; May, Binding, and Tudhope 2008; Binding, May, and Tudhope 2008; Tudhope et al. 2008).

3.2.4 Open Context

Open Context (Open Context 2013) is a web-based data sharing platform for archaeological data. It is based at the Alexandria Archive Institute and has started to cooperate with the California Digital Library to provide long-term citation and preservation for the data collected (Kansa 2010a; Kansa 2010b). The concept of Open Context is based on data integration to the ArchaeoML standard developed by David Schloen for the University of Chicago OCHRE project (Schloen and Schloen 2012, 339). It is an item-based information model, where individual atomic units of observation are related to each other and to their descriptive attributes. It contains 20 general categories: Bibliographies, Dictionaries, Locations and Objects, Maps and Plans, Persons and Organizations, Predefinitions, Presentations, Projects, Queries, Relationships, Resources, Sets

and Tables, Styles, Taxonomy, Texts, Thesaurus, Timelines, User Accounts, Values Variables and Writing Systems. Eric Kansa, leading the technology development of Open Context, is convinced that data integration can make search and retrieval of archaeological information easier, as it can increase the speed and efficiency of routine research such as finding comparative collections and materials. He argues for the choice of ArchaeoML:

> Because ArchaeoML is so simple and generalized, it is less specific in expressing semantic relationships than some other formal ontologies, including the CIDOC-CRM. This makes Open Context's semantic relations much more ambiguous, requiring some background knowledge to interpret and properly use its datasets. Nevertheless, we chose to implement ArchaeoML because of its simplicity, long history of development, and implementation in current excavation databases. Many other ontologies, though sophisticated and useful, require much more specialized expertise to apply and cost more to implement and use.

Open Context request content providers to help with the mapping of their data and provides a mapping tool, called “Penelope” that assists content providers through the process of uploading content into the system (Kansa and Kansa 2007).

Data for import should be in Microsoft Excel tables. To facilitate data import, Penelope guides content providers through a step-by-step process to classify each field in their legacy data table according to ArchaeoML. At the end of the process a URI is assigned to each unique entity (typically, sites, excavation contexts or individual artefacts and ecofacts) thereby providing a foundation for citing and linking to other data sets.

Metadata to facilitate searches, information retrieval, and comprehension of contributed data must also be created. Each project entered into Penelope requires an abstract with introductory information describing the project goals, key findings, as well as methods and recording systems, key words, approximate calendar date ranges, location information, additional notes and descriptions of fields. Additional background descriptions of specialist analyses should be submitted as text files. Each field of every dataset must be described to aid interpretation as the editorial team of Open Context is also involved in improving the data quality by data cleansing and the mappings, and therefore need to have a deeper understanding of the data provided.

Archaeologists often manage their data in relational databases, therefore Open Context requests the primary and secondary keys in each table. For depositing images content providers are asked to prepare a separate table listing each image file name, an image description and the number/identifier provided in the associated dataset of the object or place the image describes.

Open Context requires at least one geographic reference for each project. Sensitive location information can be provided with reduced precision. Users are informed of this manipulation, so that they can contact the content provider to obtain precise location data.

In general all content in Open Context is released in public domain or with a Creative Commons (creativecommons.org) copyright licence. This is to avoid a more complex rights and access management.

3.3 Lessons Learned

Interdisciplinary registries are interested in harvesting metadata from repositories such as IANUS in the future therefore the registry for find spots and excavation documentation has to provide for an interface giving access to the collection descriptions metadata. All registries described here make use of collection and project descriptions based on the Dublin Core standard. Even re3data.org plans to enable metadata contributions on repository description based on this standard. This is the minimum standard that the IANUS registry for archaeological data will have to comply with. Contextual and methodological information on the archaeological investigation from which the data derives, should also be captured at both a collection and project level.

Important additional information is the access rights to a collection of digital resources and data. Unfortunately there is no shared concept for the registries described. While Open Context and EDNA are reducing complexity to access rights by making depositors publish their data in the public domain or apply the rights to a whole collection. Additionally they offer them the option to either give free access, access only for the user group of archaeologists or access by request. tDAR on the contrary offers depositors to embargo resources for a certain time or even decide for access rights on field level of data. It is just a common practice for all archives to give free access to the metadata of collection and resource description.

For all preservation archives, it is essential to encourage depositors to document their data in as much detail as possible and to support them by offering a web-based registration tool to guide them on how to provide this information. All digital archives for archaeological data migrate provided files to a preservation format and therefore need to have a deeper understanding of the data contained in these files.

Databases are a rather modern method of documenting archaeological investigations in the long history of this discipline. By far the most information on archaeological activities is locked in unstructured written reports which is why grey literature constitutes the largest part of the content of all archives.

Data integration to a uniform standard is only performed by ADS and Open Context. As it requires deep knowledge of the standard which the data is transformed to the mapping is usually not possible for depositors so there is no need for mapping files as part of the information stored in the registry. The archaeological community is of different opinion whether data integration to a uniform standard is desirable. While Eric Kansa (Kansa and Bissell 2010) and Julian Richards (Richards 2009) argue for it, they also acknowledge that there is no one objective archaeological record and that standardization should not lead to reducing diversity in methodological approaches in archaeology and in recording systems Harrision Eiteljorg (Eiteljorg 2010) critically mentions that data integration to a uniform standard might provide excellent access to single objects or data facts but lacks utility to finding common excavation contexts from multiple projects.

4 Requirements for a Registry of Archaeological Data in Germany

Preservation of research data is not an end in itself. IANUS will be a service provider for a community of stakeholders particularly users and producers of research data. As it cannot preserve all research data, it has to define a collections policy of what data types and formats it wants to curate and preserve and what the acquisition concept is about. It clearly has a discipline-specific focus but its responsiblity of dissemination of the data goes beyond the archaeological community.

Criteria influencing the acquisition concept could be relevance, uniqueness and value of the data, quality of data and documentation or technical aspects whether IANUS is able to handle a specific file format for example. All this requires some prerequisites on the design of the registry.

A first set of requirements for the registry could be drawn from the evaluation of the digital archives for archaeological data in other countries. What needs be taken in consideration with regard to the specific situation of archaeology and research in Germany currently? Who are the stakeholders and how do the given structures impact on workflows for depositing and dissemination of research data and for documentation needs?

4.1 Stakeholders

In the following section the stakeholders that have an impact on setting up the registry of find spots and excavation documentation will be introduced. Who are they? How important is their influence? Similarly, how will IANUS prioritize their interests concerning the initial concept for the infrastructure to be developed for digital preservation of archaeological data?

4.1.1 Content Providers

The group of content providers for the registry of find spots and excavation documentation includes the 16 Federal State Departments of Archaeology, archaeological excavation firms and academic researchers of universities and non-university research institutions (German Archaeological Institute, Max Planck Institute for the History of Science etc.).

The task of the Federal State Departments of Archaeology is to ensure the proper collection, preservation and study of historical monuments within Germany. The State Offices procure, evaluate, process and convey all the information necessary for this and decide on public funds for preservation. In detail this comprises professional consulting of all stakeholders (researchers but also building contractors for example), the preparation of expert reports, the acquisition and exploration of archaeological monuments of culture, the implementation of rescue and research excavations and projects, the creation of a basis for conservation of archaeological sites and monuments, the maintenance of expert information

systems and archives and the publication of the results of archaeological work (public relations, academic publications). As the boundaries of ancient cultures do not match today's boundaries of federal states the interest in IANUS activities is focused on establishing a registry of sites and find spots associated with their cultural and temporal coverage. This would allow better cooperation between the different federal state departments. The expert information systems in place at the federal state departments are usually closed systems that are not meant to be given access to (Nds. Landesamt für Denkmalpflege 2013). The federal state departments main requirement for the registry is therefore to provide references of the resources in their archives (digital or not) to the outside by collection descriptions. Access to the resources themselves can only be granted by single request.

Archaeological excavation firms are usually contractors of the federal state departments and need to comply with their policies of excavation documentation (Bayerisches Amt für Denkmalpflege 2010; Brandenburgisches Landesamt für Denkmalpflege und Archäologische Landesmuseum, Abt. Bodendenkmalpflege 2006; Generaldirektion Kulturelles Erbe (GDKE) Rheinland-Pfalz und Direktion Archäologie (DA) - Speyer 2007; Landesamt für Denkmalpflege Hessen, Abt Archäologie und Paläontologie 2005; LVR-Amt für Bodendenkmalpflege im Rheinland 2011; Verband der Landesarchäologen in der Bundesrepublik Deutschland 2006; Wiegmann 2005). Every federal state has its own policy and it is their responsibility to collect the documentation from the excavation firms therefore IANUS will not have to deal directly with this group of stakeholders.

The situation for academic researchers of universities and non-university research institutions is even more diverse. The same is true at both the thematic and regional scope of cultures they do research on, as well as on the methodology and documentation used. Similarly divers is how they archive their data. if for example by self-archiving, depositing data at their university's data centre or in a repository with a more discipline-specific orientation (ArboDat 2013). The project-based, meaning time limited organisation of their research activity, adds to the diversity in technical solutions for collecting and storing data as well as redundant efforts in implementing systems. The main interest of these stakeholders in IANUS is the support in depositing data in the way to satisfy the conditions the funding institutions have made mandatory and that will enhance their reputation through visibility, citation and recognition of their research outputs. Additionally this will improve the possibility of success in application for research funding in the future for them. However the procedure of depositing data should not lead to an overhead of administration occupying to much time and keeping them from their research activity.

4.1.2 Data Curators of IANUS

The data curators or editorial team of IANUS should not be forgotten as a group of stakeholders. Their main interest in the registry is to access information on depositors allowing them to contact them if necessary. Secondly it is important to them to receive detailed description on the contents of deposits so that they can continue working with it either for migrating files to preservation formats,

validation, virus check or data integration etc. Furthermore as IANUS has only limited resources in personal and equipment the registration process as part of the registry should be self-explanatory that there is only a limited need for further training or schooling of depositors by IANUS.

4.1.3 Content Users

Among the group of content users of the registry are students, archaeologists, researchers in the field of Digital Humanities and of other researchers.

Students and researchers in archaeology may want to find and access data and resources related to the topic they work on for patterning, chronological analysis etc. or are looking for an unpublished collection of data and excavation documentation from federal state departments they could work on in their thesis. In the field of Digital Humanities researchers may be more interested in accessing single data facts they can run a quantitative analysis on or with which they could semantically enrich text sources with object information. Therefore the re-use of the registry content depends heavily on the granularity and quality of metadata and data as well as on the level of access rights granted.

4.1.4 Funding Organisations

Funding organisations are interested in a research data registry to control the fulfilment of funding conditions. The German Research Foundation has just recently published an update on the recommendations for digitization which also effects research data (DFG - Deutsche Forschungsgemeinschaft 2013). Here they defined technical specifications for digital resources or surrogates and specifications for metadata required if a project receives funding. The recommendations are mainly focussed on text and image items. In comparison with the needs in an archaeological context the recommendations lack of a broad range of other digital formats and resources. The DFG recommends the standards METS/MODS for text, METS/TEI for manuscripts, EAD or SAFTXML for archival material and LIDO for objects for the cataloguing of digital surrogates. Collections or inventories should be described with a minimum standard based on the Dublin Core Collection Application Profile or according to the above mentioned standards. The metadata should be made available freely on the internet. It is even recommended to write an article in German and English with the description of a collection of digital resources in Wikipedia. The registry in IANUS could then be used to check if projects have deposited their data conforming to the recommendations.

By the support of establishing research data centres as IANUS funding organisations expect to have an increase in cost efficiencies from shared data services. These centres are to ensure more coordinated and coherent services for data curation and preservation through skilled staff as a single project can provide for. Additionally funding organisations are interested in promoting knowledge transfer to other sectors requiring from IANUS to invest in interoperability for metadata exchange. All this will increase the visibility of outputs from public

funding of research.

At last research evaluation is also a topic for research data centres to be aware of. The German council for Science and Humanities (Wissenschaftsrat) published a paper with recommendations on a core set of elements to be captured for the assessment and management of research performance (“Empfehlungen zu einem Kerndatensatz Forschung” 2013). Although its definition of research output is still focussed on publications it can be expected that soon these recommendations will be updated with respect to data output of research. The minimum standard to comply with here is the project description containing elements for name of applicant of the project, speaker and project manager, title of the project and sub-projects associated with it in German and English, name of funding organisation, funding code, start and end date of the project.

4.1.5 Memory Institutions—Libraries and Museums

Dissemination of the knowledge of data collections will be a key task of IANUS if it wants to meet the expectations from researchers and funding organisations in enhancing the visibility and the impact of their work. Although IANUS has plans to develop a portal with a search interface for discovery of references to sites, collections, projects and resources, it also should think of the infrastructure libraries and museums offer. Especially academic libraries have started to include references to databases to their online catalogues. The dissemination of collection level descriptions on data should be one service offered by IANUS interfacing directly with the catalogues of the library networks and libraries or via an interdisciplinary registry service for research data and repositories such as the collection registry of DARIAH-DE or re3data.org.

Museums hold themselves large collections of digital and non-digital archaeological research data. They are content providers and users for IANUS and play a significant role in the dissemination of archaeological data. Large parts of the documentation of excavations abroad from German researchers can be found in their archives. With their activities publishing and visualizing archaeological knowledge in exhibitions and virtual exhibitions online they contribute to the visibility of archaeological research to a broader audience not only researchers. IANUS can build on the standardization efforts from the museum domain and can contribute to the refinement of vocabularies. It will benefit from the translation work done on internationally used thesauri if it prepares to publish linked data.

4.2 Comprehensive Documentation Needs & Workflows

It has to be emphasized that the focus on comprehensive documentation needs for the IANUS registry of find spots and excavation documentation is on the description of digital resources and their context of recording not on the description of archaeological objects. Data integration to a uniform standard as performed by Open Context and the ADS is to be understood as a downstream step to the registration. The process of registering resources is planned to be

accelerated by the possibility to inherit information from the collection description. At the beginning IANUS will limit the selection of resources and formats accepted and sets the scope of the resources on primary data. Later it can gradually broaden the variety of resources and formats. Therefore resources not accepted yet are software, any sort of publication (PhD, e Journal article etc.) and secondary data (visualization etc.).

4.2.1 Find Spot & Site Description

Any archaeological investigation starts with the identification of a place of archaeological interest. Therefore it is important to allow accessing research data via the localization of find spots and sites. In 2005, the Association of State Archaeologists established a working group to define a standard for data exchange with the goal to integrate archaeological sites into the GDI-DE (Geo Data Infrastructure Germany) and to improve the interoperability of data from the expert information systems from the federal state departments for archaeology (AG Modellierung der Kommission „Archäologie und Informationssysteme“ im Verband der Landesarchäologen der Bundesrepublik Deutschland 2007). They defined the standard ADeX (AG Modellierung der Kommission “Archäologie und Informationssysteme” im Verband der Landesarchäologen der Bundesrepublik Deutschland u. a. 2011). Its focus is on cultural heritage preservation, describing areas of archaeological interest and landscape protection areas. IANUS builds on this standard for find spot and site description.

Documentation Needs	
Elements	**Explanation**
Identifier from Source System	Identification number
Type of ID	For example ADeX-ID, ID of Federal State Department etc.
Title	Name of site
Type of Site Investigation	ADeX defines three types: Archäologiefläche, Untersuchungsfläche, Schutzfläche
Contact Person	Agent, person with the authority to provide more information on the site.
Site Registrar	Agent who registered the find spot or site.
Access Rights	Should follow the example of EDNA: Public, User group of archaeologists, By request. If not public direct geographical references are not shown.
Comment on Site	Free text

Spatial Coverage	
Direct Geographical Reference	
Coordinate Reference System	In Germany uniformly ETRS89 and UTM will be used in the future. ADeX list, Anlage 4
Point	
Latitude	north-south position
Longitude	east-west position
Bounding Box	
minX	Top Left
minY	Top Right
maxX	Bottom Left
maxY	Bottom Right
Precision	
Error Accuracy	Indication of the accuracy of the geo-reference supplied, in meters
Precision Comment	Free text
Indirect Geographical Reference	
Municipality Identifier	Identification number of municipality
Municipality Name	Name of municipality
Type of Ancient Site Use	
Classification of Site Use	ADeX list, 3.1.4
Site Use Description	Free text
Site Use Comment	Free text
Temporal Coverage	
Classification of Period	Instead of the ADeX list it is recommended to use the Datierungssystematik (Bibliotheksservice Zentrum Baden-Württemberg 2005)
Description of Site Dating	Free text
Temporal Coverage Comment	Free text
Relations	
Associated Find Spot & Site	A site can include several smaller units
Associated Project	Zero or more projects can be associated with a site
Associated Collection	Zero or more resource collections can be associated with a site
Associated Resource	Zero or more resources can be associated with a site

4.2.2 Project & Study Description

Project or study descriptions are necessary to explain the context of recording of the digital resources. When digital resources are removed from their initial context, implicit and referential information gets lost. The documentation needs for projects and studies are based on the core elements defined by the German council for Science and Humanities and include information that is part of the application forms of the German Research Foundation and the Ministry of Education and Research. Unfortunately these organisations do not provide access via an API to their data (Ebert et al. 2012). This would simplify the registration process and reduce redundant capturing of information if it would already be possible to query and harvest relevant information for the IANUS registry.

Resource collections and resources associated with a project that is still ongoing should be embargoed by default. Until three years after the project concluded access rights are granted by request to ensure that the creators of these resources and data have enough time to prepare own publications first.

Documentation Needs	
Elements	**Explanation**
Project or Study Title	Name of project/study in German and English
Acronym	Abbreviation of project title
DFG Classification of Discipline	List of disciplines, see (DFG-Deutsche Forschungsgemeinschaft 2013)
Name of Applicant	Agent, only for projects
Name of Project Speaker	Agent/Person, only for projects
Name of Project Manager or Head of archaeological investigation/study	Agent/Person, responsible for the project/ study and contact person for further information
Start Date	Date format, ISO 8601
End Date	Date format, ISO 8601
Name of the Funding Organisation	Agent/Organisation
Funding Code	Identifier provided by the funding organisation
Description of Research Interest	Free text, for DFG-funded projects insert summary
Description of Method	Free text
Archaeological Investigation Type	List of tDAR
Project References	BibTeX-file with references to project-related publications

Relations	
Associated Project	A project can have several sub-projects
Associated Find Spot & Site	Zero or more sites can be associated with a project
Associated Collection	Zero or more collections can be associated with a project
Associated Resource	Zero or more resources can be associated with a project

4.2.3 Resource Collection & Digital Repository Description

It will not be possible to always associate a project or study with resources, especially for old digital resources of unknown origin. Alternatively it will be possible to associate resources with a collection in the registry. As IANUS will also include research data collections of other repositories in its catalogue the data model for the registry needs to include a level for collection and repository description. The outlined documentations needs for resource collection and repository descriptions take account of aspects of the Dublin Core Collection Application Profile (DCCAP) and the specifications of re3data.org. Furthermore they consider information that is necessary to meet the DFG requirement of creating a Wikipedia article on a collection.

Documentation Needs	
Elements	**Explanation**
Type	Collection, Repository
Collection/ Repository Identifier of Source System	Identification Number of Source System
Collection or Repository Title	Title in German and English
Description	Description in German and English
Resource Access Rights	Information about who can access the collection or repository resources. Should follow the example of EDNA: Public, User group of archaeologists, By request.
Resource Licence	Licence that is inherited to all resources associated with the collection or repository.
Accrual Policy	A policy governing the addition of items to a collection or repository.
Custodial History	A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation.
Collector	Agent, An entity who gathers (or gathered) the items in a collection together.
Rights Holder	Agent, rights holder of collection or repository
Contact Person	Agent, person responsible for maintaining the collection or the repository
Is Accessed Via	Collection or Repository URL

Collection References	BibTeX-file with references to collection-related publications
Relations	
Associated Collection & Repository	A collection or repository can be associated with several other collections and repositories
Associated Find Spot & Site	Zero or more find spots and sites can be associated with a collection and repository
Associated Project	Zero or more project can be associated with a collection or repository
Associated Resource	Zero or more resources can be associated with a collection or repository

4.2.4 Digital Resource Description

In the initial phase of IANUS the focus for the digital resources will be on text, images and datasets. These digital resources are longest in use in archaeology and more likely to be in danger of degradation than newly recorded digital resources. IANUS will then gradually broaden the accepted types of resources and formats according to its capacities.

Documentation Needs	
Elements	**Explanation**
General	
File Name	File name captured by upload.
Resource Title	Name of resource
Resource Description	Description of file content
File Format	File format captured by upload; IANUS will publish a list of accepted file formats[2]
Software application	Software application used to create file
Software version	For example Word 3.0
Resource Type	Text, Image, Dataset; captured by upload according to file format
Resource Access Rights	Information about who can access the resource. Should follow the example of EDNA: Public, User group of archaeologists, By request.
Creator	Agent, person that created the resource
Rights Holder	Agent
Resource Licence	Licence under which the resource is published

2 The list of accepted file formats is still under discussion at the time of writing.

Datasets	
Table/ Worksheet Name	Name of Table
Table/Worksheet Purpose	Purpose of Table
Row Number	Number of Rows
Primary Key	Unique identifier to each record in the table
Foreign Key	Field that matches a candidate key of another cross-referenced table
Field Name	Name of Field
Description of Field	Codes or terminology used
Data Type	Data type and field length
Relations	
Associated Resource	Zero or more resources can be associated with a resource

4.2.5 Agent Description

To store contact information for the IANUS data curators and for content users that want to request access to a restricted information on a site or resource it is necessary to create an address book.

Documentation Needs	
Elements	**Explanation**
Name	Organisation or Person Name (order first name, last name)
Position	Job title
Email	Email address
Agent Type	Person, Organisation
Role	Contact Person, Collector, Creator, Project Applicant, Project Manager/ Head of archaeological investigation, Project Speaker, Rights holder, Site Registrar
Street	Street name and house number
Postcode	Text
City	Text
Country	ISO 3166
Telephone	E.123, international
Logo	Image for use in the portal
Website	Link to the homepage of the agent

Contact Access Rights	Information about who can access the contact information. Should follow the example of EDNA: Public, User group of archaeologists, By request.
Relations	
Associated Organisation	An organisation can have several sub- organisations
Associated Person	Zero and more persons can be associated with an organisation

5 A Uniform Standard for Data Integration in IANUS — Evaluation of Interdisciplinary and Archaeology-specific Metadata Standards

Although the community of archaeologist is divided on the question of data integration there are some parties requiring it. As described the German Research Foundation defined in its recommendations for digitization some metadata standards for data exchange that should be supported. Especially European projects focus on standardization work for data exchange. Therefore IANUS has to provide data integration as a service for its data providers to improve dissemination of data. In the long term it will not be enough to just provide collection level descriptions. The mapping of legacy data to more sophisticated metadata standards and ontologies requires deeper insight to both the archaeological domain and to information science. It cannot be expected from archaeologists to have the knowledge for this mapping work.

Some of the metadata standards for data exchange are especially interesting for the cultural heritage domain and are introduced in the next section. In which context are these metadata standards used? How does this context meet the support of archaeological research? How can IANUS more efficiently provide data integration to these metadata standards for data exchange? On the other hand it is important that archaeologists can understand the standard IANUS will do data integration to. The terminology used for the standard should reflect the terminology archaeologists are familiar with. Otherwise this will turn out to be a barrier for re-use of data.

5.1 Interdisciplinary Metadata Standards

In the evaluation on interdisciplinary metadata standards three standard are discussed: Dublin Core (DC), the Europeana Data Model (EDM) and Lightweight Information Describing Objects (LIDO). Dublin Core was chosen because it is prominently in use by other digital archives of archaeological data. EDM is the data model created for Europeana the largest network for exchange of digital resources of cultural heritage objects in Europe. LIDO is an exchange format from the museum domain. As museums are beyond the group of stakeholders of IANUS it is important to understand the infrastructure for data exchange in the museum domain.

5.1.1 Dublin Core

The Dublin Core standard resulted from a workshop in Dublin, Ohio in 1995 (Baker 2012). Host of the workshop was OCLC Online Computer Library Center, Inc.. It does therefore not surprise that participants were mostly from the library domain. The initial intention for the definition of the Dublin Core standard was to specify a set of metadata elements broad and generic enough to describe a wide range of electronic objects and improve search for these objects online. The new standard was composed of fifteen elements now known as Simple Dublin Core. The element set is very similar to a library cataloguing record. Due to its flat and rigid format it is very popular and widely in use. The example of the ADS that uses Dublin Core application profiles to capture collection and resource descriptions for digital preservation, shows that the influence of Dublin Core goes beyond the library domain. However the simple structure of Dublin Core leads to ambiguity in the interpretation of the elements. Its is therefore less semantically interoperable. This might be the reason why the ADS performs crosswalks from DC to MIDAS XML and CRM for English Heritage (CRM-EH). Nevertheless as the examples of the projects Trans-Atlantic Gateway Project and ARENA — Archaeological Records of Europe show IANUS will have to be able of providing Dublin Core for collection level description. It is also important for the metadata exchange with the national registries for research data such as the collection registry of DARIAH-DE and re3data.org.

5.1.2 Europeana Data Model (EDM)

The Europeana Data Model (Europeana v1.0 2012) was created to make cultural heritage objects discoverable via web resources and is the standard for structuring the data that the project Europeana is ingesting, managing and publishing in its portal (Europeana 2013).

Each of the different heritage sectors represented in Europeana uses different data standards. EDM is designed in a generic approach to transcend the information perspectives of the different domains in Europeana—the museums, archives, audiovisual collections and libraries. One intention of the creators of EDM is to provide a model that:

> … can be seen [...] as an anchor to which various finer-grained models can be attached, making them at least partly interoperable at the semantic level, while the data retain their original expressivity and richness. (Isaac 2011)

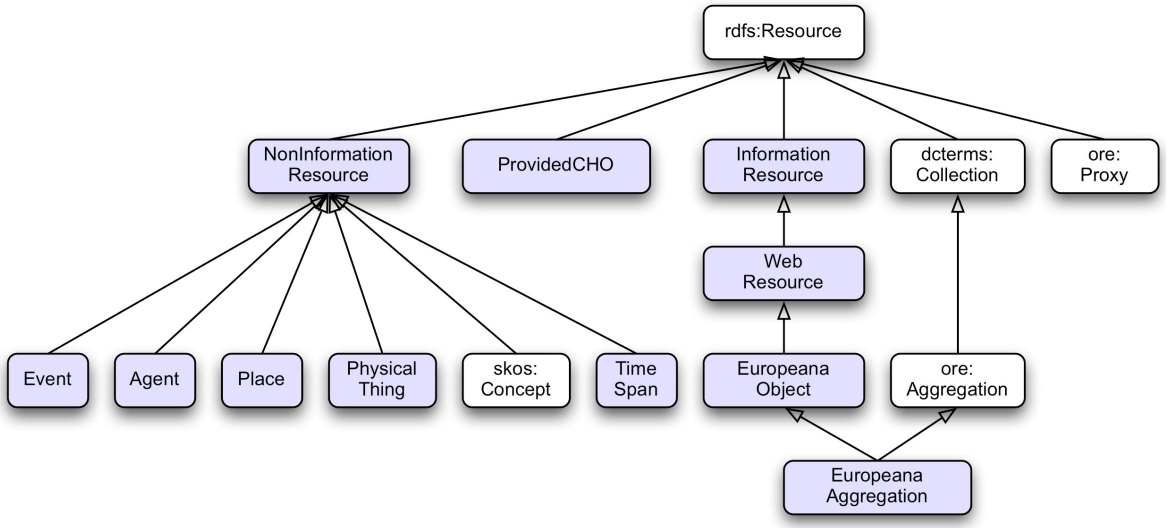

Fig. 7: The EDM Class hierarchy. The classes introduced by EDM are shown is light blue rectangles. The classes in the white rectangles are re-used from other schemas; the schema is indicated before the colon. (Europeana v1.0 2012)

EDM re-uses from the following namespaces:

- The Resource Description Framework (RDF) and the RDF Schema (RDFS))
- The OAI Object Reuse and Exchange (ORE)
- The Simple Knowledge Organization System (SKOS)
- The Dublin Core namespace for elements

The creators of EDM themselves point out that EDM is a very complex data model. The provision of content to Europeana therefore passes through several steps. First the content provider maps the legacy data to a domain-specific exchange format such as Lightweight Information Describing Objects (LIDO),

Encoded Archival Description (EAD), Metadata Encoding & Transmission Standard (METS) or Machine-Readable Cataloging (MARC). These switching schemas allow then to channel crosswalking among the multiple individual schemas to EDM. This happens by schema mapping of the exchange format to EDM.

By the time of writing Europeana provides access to almost 27 million digital objects of cultural heritage contributed from more than 2000 European institutions. It is therefore the largest network for exchange of digital resources of cultural heritage objects in Europe. With associated projects such as Europeana Cloud, for example, Europeana increases its influence into new fields of activity. Europeana Cloud will provide content, metadata, a linked storage system, tools and services for researchers and a new platform—Europeana Research.

For dissemination purposes it is important for IANUS to contribute to the Europeana network. In the following LIDO and CARARE as exchange formats to prepare data for EDM are introduced.

5.1.3 Lightweight Information Describing Objects (LIDO)

The LIDO standard evolved from museumdat, Categories for the Description of Works of Art (CDWA), Cataloging Cultural Objects (CCO) and Spectrum (ICOM International Committee for Documentation 2013). It is designed as a standard for the museum domain more precisely for the event-orientated description of museum objects. It therefore has a focus on moveable objects.

Descriptive and administrative elements of a LIDO record	
-Object Classification-	*-Events-*
Object/Work Type (mandatory)	Event Set
Classification	*-Relations-*
-Object Identification-	Subject Set
Title/Name (mandatory)	Related Works
Inscriptions	
Repository/Location	*-Administrative Metadata-*
State/Edition	Rights
Object Description	Record (mandatory)
Measurements	Resource

Museums are part of the stakeholder group of IANUS. LIDO will be the exchange format for contributions to IANUS of museum data.

On the other hand contributions of IANUS to Europeana will go via the German Digital Library (DDB) the national aggregator of digital resources of cultural heritage in Germany that delivers data to Europeana. The DDB has defined LIDO as one of the excepted exchange formats in which data can be provided to it. Therefore IANUS itself will have to map data to LIDO.

5.2 Archaeology-specific Metadata Standards

LIDO as a museum standard already meets most of the documentation needs in archaeology but with its focus on moveable objects it lacks some options to capture information on sites and architecture. Another associated project of Europeana the Connecting ARchaeology and ARchitecture in Europeana (CARARE) project created a standard for the specific needs in archaeology.

5.2.1 Connecting ARchaeology and ARchitecture in Europeana (CARARE)

The CARARE schema (Papatheodorou et al. 2011) defines a set of standard elements which are drawn from existing standards from the archaeology and architecture domain. It is a harvesting schema which is based on MIDAS Heritage, LIDO and the CIDOC CRM. Initially the intention for the design of the CARARE schema was to mediate between the native metadata collected by the partners of the CARARE project (CARARE 2013) and Europeana. Therefore it is the best choice for IANUS as exchange format to which it can map all legacy data. A mapping to DC (ESE) (Tsalapati, Trivela, and Tzouvaras 2011), EDM (Papatheodorou,, Kakali, and Tsakonas 2011) and LIDO (Tzouvaras, et al. 2011) is already available so that it is possible to support also the required other formats for collection and item description needed for data dissemination in IANUS.

The root element of a CARARE record is the CARARE wrap. This contains the four global wrappers which hold descriptive and administrative metadata for:

- Heritage Asset Identification—the monument, historic building, archaeological landscape area, shipwreck, artefact or ecofact being described.
- Digital resource—digital representations and information sources (images, texts, videos, audio, 3D models) of the heritage asset described.
- Activity—events in which the heritage asset has taken part (its creation, adaptation, reconstruction, field investigation, research and any historical events).
- Collection—the collection in which the data being provided forms part, e.g. a national monuments inventory, museum collection, etc.

A mapping from ADeX to CARARE is possible without significant information loss (see 10.2). This will allow for an automated transformation of ADeX records to CARARE during upload to the registry.

6 Conclusion

The discussion proved that there is a need for professionalising digital archiving of archaeological data especially for long term digital preservation in Germany. The IANUS project will take on the task of developing a concept for a German research data centre of archaeological data. In this way the registry for archaeological find spots and excavation documentation forms an important component of the technical infrastructure of IANUS. It is to collect information on archaeological sites and excavation data for a finding aid service.

The evaluation of existing research data registries provided a first set of requirements for the design of the registry. Consequently, the registry needs to ask content providers for the description of site, collection, project, agent, data type and format information to contextualize a resource. This will prevent that the digital resources uploaded to the digital preservation archive loose referential information that is necessary to understand them. Collection level descriptions can also be used to help content providers to speed up the process of registering digital resources. For this purpose the registry needs to support the functionality to pass on parts of collection information to digital resources. Access rights next to licensing of data is still a difficult issue for all registries. In archaeology some information on sites such as coordinates are sensitive. Therefore it must be possible to restrict access to this information. To avoid a complex system of access rights and user management it is recommended to leave three options to the content providers for access rights: public, for the group of archaeologists only and by request. This should be sufficient for the protection of information. A minimum standard for collection level description shared of all registries is based on Dublin Core. First efforts of data sharing in archaeology made use of collection information in the Dublin Core format. IANUS should therefore prioritize the preparation of collection level description of archaeological data before it starts with data integration.

There are special needs for documentation defined by the stakeholders of IANUS to be captured within the registry. A short analysis of their interests helped to specify more requirements. Information relevant to identify data deposits of projects is needed by funding organisations for research evaluation and the control of fulfilment of funding conditions. Researchers that deposit data are especially interested in the dissemination of the knowledge of their data collection to increase visibility, citation and recognition of their research outputs. Libraries, museums and networks such as the German Digital Library and Europeana can help in the dissemination of this knowledge. However if IANUS wants to exchange data with these institutions and projects it has to support certain metadata standards. This draws attention to the question of data integration in IANUS.

Dublin Core (DC), the Europeana Data Model (EDM) and Lightweight Information Describing Objects (LIDO) are the standards required from funding organisations and the above mentioned aggregators to comply with. For the researcher in archaeology it is high barrier to gain deeper knowledge of all these standards. Consequently IANUS should offer the service and do the mapping of deposited data to the mentioned standards for the researchers. The most efficient way to perform crosswalks of legacy data to a number of standards is to first map it to an

exchange format and then map this exchange format on the schema level to other standards. The CARARE schema created by another project in the network of Europeana can serve this purpose. Mappings of the CARARE schema to DC, EDM and LIDO are already available. It builds on standards from the archaeology and architecture domain. This has the additional benefit that archaeologist will understand it better as they are familiar with the terminology used in it. Hopefully this will encourage re-use of data.

7 Acknowledgements

Thanks are due to Dr. Felix F. Schäfer and Maurice Heinrich the project coordinators of IANUS for offering background information about the strategies and concepts of the project and to the members of the interoperability workgroup in IANUS Irmela Herzog (researcher, LVR-Amt für Bodendenkmalpflege im Rheinland), Regine Stein (head of IT, Bildarchiv Foto Marburg), Arne Weiser (lecturer, Hochschule für Technik und Wirtschaft Berlin) and Tom Brughmann (PhD student at the University of Southampton, member of the Archaeological Computing Research Group) for helpful input to the definition of requirements and scope of the registry. A special thanks goes to Sally Chambers (Secretary-General, DARIAH-EU Coordination Office) for her professional support and for improving my writing.

8 List of Figures

9 References

ADS-easy. 2013. “ADS E-archiving System.” Accessed May 16. http://archaeologydataservice.ac.uk/easy/.

AG Modellierung der Kommission „Archäologie und Informationssysteme“ im Verband der Landesarchäologen der Bundesrepublik Deutschland. 2007. “Ergebnisprotokoll.” http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Archaeologie-Informationssysteme/Dokumente_AIS_Modellierung/Abgrenzung.pdf.

AG Modellierung der Kommission „Archäologie und Informationssysteme“ im Verband der Landesarchäologen der Bundesrepublik Deutschland, R. Göldner, Irmela Herzog, and R. Wanninger, ed. 2011. “ADeX Version 2.0 - Archäologischer Daten-Export, Standard für den Austausch Archäologischer Fachdaten.” http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Archaeologie-Informationssysteme/Dokumente_AIS_ADeX/ADeX_2-0_Doku.pdf.

ArboDat. 2013. “ArboDat - Das Hessische Datenarchiv Für Botanische Großreste von Archäologischen Ausgrabungen.” Accessed May 19. http://www.hessen-archaeologie.de/Archaobotanik/Datenbankprogramm/datenbankprogramm.html.

ARENA. 2013. “ARENA -Archaeological Records of Europe - Networked Access.” Accessed May 14. http://ads.ahds.ac.uk/arena/.

ARIADNE. 2013. “Ariadne - Advanced Research Infrastructure for Archaeological Dataset Networking in Europe.” Accessed May 13. http://www.ariadne-infrastructure.eu/.

Atici, Levent, Sarah Whitcher Kansa, Justin Lev-Tov, and Eric C. Kansa. 2012. “Other People’s Data: A Demonstration of the Imperative of Publishing Primary Data.” *Journal of Archaeological Method and Theory* (April 11). doi:10.1007/s10816-012-9132-9. http://www.springerlink.com/index/10.1007/s10816-012-9132-9.

Austin, Tony. 2003. “Dublin Core Metadata Schema: ADS Interpretation and Mapping for the ARENA Project Partners.” In . http://elib.sbras.ru:8080/jspui/handle/SBRAS/105.

Babeu, A. 2011. “Rome Wasn’t Digitized in a Day”: Building a Cyberinfrastructure for Digital Classics. http://www.clir.org/pubs/abstract/pub150abst.html.

Backhouse, Paul. 2006. “Drowning in Data? Digital Data in a British Contracting Unit.” In *Digital Archaeology - Bridging Method and Theory*, 50–58. Abingdon, Oxon: Routledge.

Baker, Thomas. 2012. “Libraries, Languages of Description, and Linked Data: a Dublin Core Perspective.” *Library Hi Tech* 30 (1) (March 2): 116–133. doi:10.1108/07378831211213256.

Bayerisches Amt für Denkmalpflege, ed. 2010. “Vorgaben Zur Dokumentation Archäologischer Ausgrabungen in Bayern.” http://www.blfd.bayern.de/medien/vorg_doku_arch_ausg.pdf.

Bernbeck, Reinhard. 1997. *Theorien in der Archäologie*.

Bibliotheksservice Zentrum Baden-Württemberg. 2005. “Datierungssystematik.” http://museum.zib.de/museumsvokabular/documents/datierung-bsz.pdf.

Binding, Ceri, Keith May, and Douglas Tudhope. 2008. “Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction Via the CIDOC CRM.” In *Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries*, 280–290. ECDL ’08. Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-87599-4_30.

Brandenburgisches Landesamt für Denkmalpflege und Archäologische Landesmuseum, Abt. Bodendenkmalpflege, ed. 2006. “Richtlinien zur Grabungsdokumentation.” http://www.bldam-brandenburg.de/images/stories/PDF/richtlinien%20zur%20grabungsdokumentation.pdf.

CARARE. 2013. “CARARE - Connecting ARchaeology and ARchitecture in Europeana.” Accessed May 21. http://www.carare.eu/.

Chen, Y.N., S.J. Chen, and S.C. Lin. 2003. “A Metadata Lifecycle Model for Digital Libraries: Methodology and Application for an Evidence-based Approach to Library Research.” Edited by IFLA. http://metadata.teldap.tw/bibliography/proceeding/cp030805.pdf.

Chenhall, Robert G. 1971. “The Archaeological Data Bank: A Progress Report.” *Computers and the Humanities* 5 (3) (January): 159–169. doi:10.1007/BF02402182.

Cripps, P., A. Greenhalgh, D. Fellows, K. May, and D. Robinson. 2004. “Ontological Modelling of the Work of the Centre for Archaeology.” *CIDOC CRM Technical Paper*.

DAI. 2013. “Deutsches Archäologisches Institut.” Accessed May 13. http://www.dainst.org/de/.

DANS Data Archiving and Network Services, ed. 2011. “Archeologische Metadata -Uitleg Voor Het Invullen van de Metagegevens van Archeologische Datasets.” http://www.edna.nl/sites/default/files/Archeologische%20Metadata.pdf.

———, ed. 2013. “Data Deponeren in Zeven Stappen Instructies Voor Archeologische Data.” Accessed May 1. http://www.edna.nl/sites/default/files/Deponeerinstructies%20Archeologie%20NL%20DEF.pdf.

DARIAH-DE. 2013. “DARIAH-DE.” Accessed May 13. http://de.dariah.eu/.

DARIAH-EU. 2013. “DARIAH - Digital Research Infrastructure for the Arts and Humanities.” Accessed May 13. http://www.dariah.eu/.

DFG - Deutsche Forschungsgemeinschaft, ed. 2009. “Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten.” http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf.

———. 2013. “DFG-Praxisregeln ‘Digitalisierung’.” http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung_2013.pdf.

———. 2013. “DFG Classification of Disciplines.” Accessed May 20. http://www.dfg.de/en/dfg_profile/statutory_bodies/review_boards/subject_areas/index.jsp.

Dublin Core Metadata Initiative. 2013. “Dublin Core Collections Application Profile.” Accessed May 13. http://dublincore.org/groups/collections/collection-application-profile/.

Ebert, Barabara, Alexander Kujath, Joachim Holtorf, Karsten Holmberg, and Thomas Rupp. 2012. “Erfahrungen aus der Einführung des Forschungsinformationssystems Pure an Der Leuphana Universität Lüneburg.” In *Forschungsinformation in Deutschland: Anforderungen, Stand und Nutzen existierender Forschungsinformationssysteme*, 65–78. iFQ-Working Paper 10.

EDNA. 2013. “E-depot Nederlandse Archeologie, a Collaboration Between DANS and RCE.” Accessed May 2. http://www.edna.nl/.

Eiteljorg, Harrison. 2002. “The Archaeological Data Archive Project Ceases Operation.” *CSA Newsletter* XV (2). http://csanet.org/newsletter/fall02/nlf0201.html.

———. 2010. “Aggregation for Access Vs. Archiving for Preservation.” *CSA Newsletter* 23 (2) (September). http://csanet.org/newsletter/fall10/nlf1005.html.

“Empfehlungen zu einem Kerndatensatz Forschung.” 2013. Drs. 2855-13. Berlin. http://www.wissenschaftsrat.de/download/archiv/2855-13.pdf.

English Heritage, ed. 2012. “MIDAS Heritage – The UK Historic Environment Data Standard, V1.1.” http://www.english-heritage.org.uk/publications/midas-heritage/midas-heritage-2012-v1_1.pdf.

Europeana. 2013. “Europeana - Think Culture.” Accessed May 22. http://europeana.eu/.

Europeana v1.0, ed. 2012. “Definition of the Europeana Data Model Elements - Version 5.2.3.” http://pro.europeana.eu/documents/900548/bb6b51df-ad11-4a78-8d8a-44cc41810f22.

Forum on Information Standards in Heritage. 2013. “FISH Interoperability Toolkit.” Accessed May 15. http://www.heritage-standards.org.uk/.

Funk, Stefan E. 2010a. “Migration.” In *Nestor Handbuch - Eine Kleine Enzyklopädie Der Digitalen Langzeitarchivierung - Version 2.3*. urn:nbn:de:0008-20100617189.

———. 2010b. “Emulation.” In *Nestor Handbuch - Eine Kleine Enzyklopädie Der Digitalen Langzeitarchivierung - Version 2.3*. urn:nbn:de:0008-20100305134.

Generaldirektion Kulturelles Erbe (GDKE) Rheinland-Pfalz, and Direktion Archäologie (DA) - Speyer, ed. 2007. “Ausgrabungs- und Dokumentationsrichtlinien Version 1.01.” http://download.gdke-rlp.de/archaeologie/richtlinien_ausgrabung.pdf.

IANUS. 2013. “IANUS - Forschungsdatenzentrum Archäologie & Altertumswissenschaften.” Accessed May 13. http://www.ianus-fdz.de/.

ICOM International Committee for Documentation. 2013. “What Is LIDO.” Accessed May 22. http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/.

Isaac, Antoine. 2011. “Europeana Data Model Primer V1.0.” Edited by Europeana. http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5.

Kansa, Eric C. 2010a. “Publishing Data in Open Context: Methods and Perspectives.” *CSA Newsletter* XXIII (2) (September). http://csanet.org/newsletter/fall10/nlf1001.html.

———. 2010b. “Open Context in Context - Cyberinfrastructure and Distributed Approaches to Publish and Preserve Archaeological Data.” *The SAA Archaeological Record* 10 (5) (November): 12–16.

Kansa, Eric C, and Ahrash Bissell. 2010. “Web Syndication Approaches for Sharing Data in ‘Small Sciences’ Domains.” *Data Science Journal* 9 (July 8). https://www.jstage.jst.go.jp/article/dsj/9/0/9_009-012/_pdf.

Kansa, Eric C., and Sarah Whitcher Kansa. 2007. “Open Content in Open Context.” *Educational Technology Magazine* 47 (6) (November 1): 26–31.

Kenny, Jonathan, and Julian D. Richards. 2005. “Pathways to a Shared European Information Infrastructure for Cultural Heritage.” *Internet Archaeology*. http://intarch.ac.uk/journal/issue18/6/toc.html.

Landesamt für Denkmalpflege Hessen, Abt Archäologie und Paläontologie, ed. 2005. “hessenArchäologie- Richtlinien zur Grabungsdokumentation.” http://hessen-archaeologie.de/Download/LfDHessenGrabungsrichtlinien2005.pdf.

LVR-Amt für Bodendenkmalpflege im Rheinland, ed. 2011. “Prospektions- und Grabungsrichtlinien für drittfinanzierte archäologische Maßnahmen.” http://www.bodendenkmalpflege.lvr.de/denkmalschutz_prakt_bodendenkmalpflege/richtlinien_2011.pdf.

May, K., C. Binding, and D. Tudhope. 2008. “A STAR Is Born: Some Emerging Semantic Technologies for Archaeological Resources.” *Proceedings Computer Applications and Quantitative Methods in Archaeology CAA2008*.

McManamon, Francis P., Keith Kintigh, and Adam Brin. 2010. “Digital Antiquity and the Digital Archaeological Record (tDAR): Broadening Access and Ensuring Long-Term Preservation for Digital Archaeological Data.” *CSA Newsletter* XXIII (2) (September). http://csanet.org/newsletter/fall10/nlf1002.html.

Nds. Landesamt für Denkmalpflege. 2013. “ADABweb.” Accessed May 19. http://www.denkmalpflege.niedersachsen.de/portal/live.php?navigation_id=12669&article_id=55567&_psmand=45.

Open Context. 2013. “Open Context: Data Publication for Cultural Heritage and Field Research.” Accessed May 15. http://opencontext.org/.

Papatheodorou, Christos, Phil Carlisle, Christian Ertmann-Christiansen, and Kate Fernie. 2011. “The CARARE Metadata Schema, v 1.1.” http://www.carare.eu/eng/Resources/CARARE-Documentation/CARARE-metadata-schema.

Papatheodorou,, Christos, Constantina Kakali,, and Giannis Tsakonas. 2011. “D2.2.3_3.4 EDM-CARARE Metadata Mapping.” http://www.carare.eu/eng/Media/Files/D2.2.3_3.3-CARARE-DC-ESE-metadata-mapping.

Powell, James E., Krista Black, and Linn Marks Collins. 2011. “A Semantic Registry for Digital Library Collections and Services.” *D-Lib Magazine* 17 (11/12) (November). doi:10.1045/november2011-powell.

Re3data.org. 2013. “About | Re3data.org.” Accessed May 13. http://www.re3data.org/about/.

Renfrew, Colin, and Paul G. Bahn. 2008. *Archaeology: Theories, Methods and Practice*. Thames & Hudson.

Richards, Julian D. 2009. “From Anarchy to Good Practice: The Evolution of Standards in Archaeological Computing.” *Archeologia e Calcolatori* 20: 27–35.

Richards, Julian D. 2006. “Electronic Publication in Archaeology.” In *Digital Archaeology - Bridging Method and Theory*, 213–225. Abingdon, Oxon: Routledge.

———. 2012. “Digital Infrastructures for Archaeological Research: A European Perspective.” *CSA Newsletter* XXV (2) (September). http://csanet.org/newsletter/fall12/nlf1202.html.

Ross, Kenneth A., Angel Janevski, and Julia Stoyanovich. 2007. “A Faceted Query Engine Applied to Archaeology.” *Internet Archaeology* 21 (April). http://intarch.ac.uk/journal/issue21/3/toc.html.

Schloen, J. David, and Sandra R. Schloen. 2012. *OCHRE - An Online Cultural Historical Research Environment*. http://ochre.uchicago.edu/sites/ochre.uchicago.edu/files/uploads/OCHRE%20Manual.pdf.

Scholze, Frank, Hans-Jürgen Goebelbecker, and Heinz Pampel. 2013. “Re3data.org - Registry of Research Data Repositories” presented at the 5. Kongress Bibliothek und Information Deutschland,, March 11, Leipzig. http://www.opus-bayern.de/bib-info/volltexte//2013/1382/pdf/scholze_re3data_11_3_13_1.pdf.

Schroeder, Kathrin. 2010. “Persistent Identifier (PI) - Ein Überblick.” In *Nestor Handbuch - Eine Kleine Enzyklopädie Der Digitalen Langzeitarchivierung - Version 2.3*. urn:nbn:de:0008-20100305164.

SWORD. 2013. “SWORD - Simple Web-service Offering Repository Deposit.” Accessed May 16. http://swordapp.org/about/.

SWORD-ARM. 2013. “SWORD_ARM: SWORD & Archaeological Research Data Management.” Accessed February 15. http://archaeologydataservice.ac.uk/research/swordarm.

TAG. 2013. “Transatlantic Archaeology Gateway.” Accessed May 14. http://archaeologydataservice.ac.uk/TAG/intro.jsf.

tDAR. 2013a. “The Digital Archaeological Record.” Accessed May 14. http://www.tdar.org/.

———. 2013b. “Data Dictionary.” Accessed February 22. https://dev.tdar.org/confluence/display/TDAR/Data+Dictionary.

———. 2013c. “Access Rights Matrix.” Accessed May 14. https://dev.tdar.org/confluence/display/TDAR/Access+Rights+Matrix.

Tsalapati, Eleni, Despoina Trivela, and Vassilis Tzouvaras. 2011. “D2.2.3_3.3 CARARE-DC (ESE) Metadata Mapping.” http://www.carare.eu/eng/Media/Files/D2.2.3_3.3-CARARE-DC-ESE-metadata-mapping.

Tudhope, D., C. Binding, K. May, and E. Heritage. 2008. “Semantic Interoperability Issues from a Case Study in Archaeology.” In *Semantic Interoperability in the European Digital Library, Proceedings of the First International Workshop SIEDL*, 88–99. http://hypermedia.research.glam.ac.uk/media/files/documents/2008-07-05/SIEDL08-Tudhope-v3.pdf.

Tzouvaras,, Vassilis, Eleni Tsalapati, Nasos Drosopoulos, and Despoina Trivela. 2011. “D2.2.3_3.2 LIDO-CARARE Metadata Mapping.” http://www.carare.eu/eng/Media/Files/D2.2.3_3.3-CARARE-DC-ESE-metadata-mapping.

Ulrich, Dagmar. 2010. “Bitstream Preservation.” In *Nestor Handbuch - Eine Kleine Enzyklopädie Der Digitalen Langzeitarchivierung - Version 2.3*. urn:nbn:de:0008-20100305123.

Verband der Landesarchäologen in der Bundesrepublik Deutschland, ed. 2006. “Ausgrabung und Prospektion - Durchführung und Dokumentation.” http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Grabungstechniker/grabungsstandards_april_06.pdf.

Vierkant, Paul, Shaked Spier, Jessika Rücknagel, Jens Gundlach, David Fichtmüller, Heinz Pampel, Maxi Kindling, et al. 2012. “Vocabulary for the Registration and Description of Research Data Repositories. Version 2.0” (December). doi:doi:10.2312/re3.002. http://ebooks.gfz-potsdam.de/pubman/item/escidoc:76875/component/escidoc:76874/re3data_vocabulary_v2-0.pdf.

Wiegmann, Mario. 2005. “Zeichenrichtlinien des Landesamtes für Denkmalpflege und Archäologie Sachsen-Anhalt.” *Archäologie in Sachsen-Anhalt* (3): 287–306.

Willems, W.J.H., and R.W. Brandt. 2004. *Dutch Archaeology Quality Standard*. Den Haag. http://www.dans.knaw.nl/sites/default/files/file/archief/KNA_uk.pdf.ADS-easy.

10 Appendices

10.1 Accepted File Formats from Archaeological Preservation Archives

Type of data	EDNA	*tDAR*	ADS	Open Context
Text documents	***Preferred format(s):*** PDF/A (.pdf) ***Acceptable format(s):*** OpenDocument Text (.odt) MS Word (.doc, .docx) Rich Text File (.rtf) PDF (.pdf)	Rich Text Documents (.rtf) Microsoft Word (.doc, .docx) PDF Documents (.pdf)	***Preferred format(s):*** OpenDocument Text (.odt) MS Word (.doc, .docx) ***Acceptable format(s):*** Rich Text Documents (.rtf) MS Word (.docm) OpenOffice.org 1.0 (.swx) .html, .xhtml, .xml, .sgml	MS Word PDF
Plain text	***Preferred format(s):*** Unicode TXT (.txt, … ***Acceptable format(s):*** Non-Unicode TXT (.txt, …)	Text Documents (.txt)	Text Documents (.txt)	
Spreadsheets	***Preferred format(s):*** PDF/A (.pdf) Comma Separated Values (.csv) ***Acceptable format(s):*** OpenDocument Spreadsheet (.ods) MS Excel (.xls, .xlsx)	Comma Separated Values (.csv) MS Excel (.xls, .xlsx)	***Preferred format(s):*** Comma Separated Values (.csv) MS Excel (.xls, .xlsx) OpenDocument Spreadsheet (.ods) ***Acceptable format(s):*** Lotus 1-2-3 - 123,WK4, WK3, WK1, WKS	
Databases	***Preferred format(s):*** ANSI SQL (.sql, …) Comma Separated Values (.csv) ***Acceptable format(s):*** MS Access (.mdb, .accdb) Dbase III or IV (.dbf)	Comma Separated Values (.csv) MS Access (.mdb, .accdb, .mdbx) Tab Separated Values (.tab)	***Preferred format(s):*** MS Access (.mdb, .accdb) OpenDocument Database (.odb) Delimited text ***Acceptable format(s):*** Dbase (.dbf)	Filemaker, Access, and Open Office, as well as comma separated value files
Statistical data	***Preferred format(s):*** SPSS Portable (.por) SAS transport (.sas) STATA (.dta) ***Acceptable format(s):*** R		***Preferred format(s):*** Delimited text ***Acceptable format(s):*** SPSS - SAV, POR, SPO SAS - SAS7DBAT, SAS MS Excel - XLS OpenDocument Spreadsheet - ODS SYLK - SLK MS Access - MDB xBase - DBF	
Pictures (raster)	***Preferred format(s):*** JPEG (.jpg, .jpeg) TIFF (.tif, .tiff)	Tagged Image File Format (.tiff, .tif) Graphics Interchange Format (.gif) JPEG Image (.jpg, .jpeg) Bitmap Image (.bmp) PICT Image (.pict) Portable Network Graphics (.png)	***Preferred format(s):*** Uncompressed Baseline TIFF v.6 - TIF ***Acceptable format(s):*** RAW format Portable Network Graphics (.png) Joint Photographic Expert Group (.jpg) Graphics Interchange Format (.gif) Bit-Mapped Graphics Format (.bmp) PhotoCD (.pcd) Photoshop (Adobe) (.psd)	
Pictures (vector)	***Preferred format(s):*** PDF/A (.pdf) Scalable Vector Graphics (.svg) ***Acceptable format(s):*** Adobe Illustrator (.ai) PostScript (.eps) PDF (.pdf)		***Preferred format(s):*** AutoCAD - DWG Drawing Interchange Format - DXF Scalable Vector Graphics - SVG ***Acceptable format(s):*** Adobe Illustrator - AI	
Video	***Preferred format(s):*** MPEG-2 (.mpg, .mpeg, …) MPEG-4 H264 (.mp4) Lossless AVI (.avi) QuickTime (.mov)		***Preferred format(s):*** MPEG1 & 2 - MPG, MPEG MPEG4 - MPG4 ***Acceptable format(s):*** DivX - DIVX, AVI	
Audio	***Preferred format(s):*** WAVE (.wav) ***Acceptable format(s):*** MP3 AAC (.mp3)		***Preferred format(s):*** Broadcast Wave Format .bwf Waveform Audio WAV Audio Interchange AIF ***Acceptable format(s):*** SUN au AU	
Computer Aided Design	***Preferred format(s):*** AutoCAD DXF version R12 (.dxf) ***Acceptable format(s):*** AutoCAD other versions (.dwg, .dxf)		***Preferred format(s):*** X3D VRML Java3D QTVR	
Geographical Information	***Preferred format(s):*** MapInfo Interchange Format (.mif/.mid) ***Acceptable format(s):*** ESRI Shapefiles (.shp and accompanying files) MapInfo (.tab and accompanying files) Geographic Markup Language (.gml)		***Preferred format(s):*** ESRI Shapefile - SHP + SHX + DBF Geo-referenced TIF Image - TIF + TFW Geography Markup Language - GML ***Acceptable format(s):*** ESRI Grid MapInfo Interchange Format - MIF + MID Spatial Data transfer standard - DDF MOSS - EXP Vector product Format - VPF	
Geophysics			***Preferred format(s):*** Raw xyz data: TXT, CSV Rendered images: TIF ***Acceptable format(s):*** Raw data: DAT REP (Contors)	

10.2 Mapping ADeX — CARARE

ADeX	CARARE
ADEX_ID	activity/recordInformation/id
BEZEICHNG	activity/appellation/name activity/spatial/locationSet/namedLocation
FLAECH_ART	*activity/eventType*[3]
ERFASS_DAT	activity/temporal/timeSpan/startDate
AENDER_DAT	activity/temporal/timeSpan/startDate
ANSPRECHP	activity/actors/name activity/actors/roles/archaeologist
DAT_QUELLE	activity/actors/name activity/actors/roles/creator
BERECHTIG	activity/recordInformation/rights/accessRights/conditions
COPYRIGHT	activity/recordInformation/rights/reproductionRights/contacts/name activity/recordInformation/rights/reproductionRights/contacts/roles/rights holder
ZUSATZ	
KOO_REFSYS	activity/spatial/spatialReferenceSystem
X_KOORD	activity/spatial/geometry/quickPoint/x
Y_KOORD	activity/spatial/geometry/quickPoint/y
X_VON, X_BIS	activity/spatial/geometry/boundingBox/minX activity/spatial/geometry/boundingBox/maxX
Y_VON, Y_BIS	activity/spatial/geometry/boundingBox/minY activity/spatial/geometry/boundingBox/maxY
GENAUIGK	activity/spatial/geometry/storedPrecision
GENAUIGK_T	
GDE_KENN	
GDE_NAME	activity/spatial/locationSet/geopoliticalArea activity/spatial/locationSet/geopoliticalAreaType/municipality
TYP_GROB	heritageAssetIdentification/characters/heritageAssetType
TYP_FEIN	heritageAssetIdentification/characters/heritageAssetType
TYP_ERLAEU	
DAT_GROB	activity/temporal/periodName heritageAssetIdentification/characters/temporal/periodName

3 The element Flächenart in ADeX describes die event type but is a compound word of the activity and the area. This mapping is not an exact match.

ADeX	CARARE
DAT_FEIN	activity/temporal/periodName heritageAssetIdentification/characters/temporal/periodName
DAT_ERLAEU	