Data Structures in Archaeology - Information for the Future

Content

Key Words: **archaeology, data structure, digital archive, Saxony.**

File: Paper_en_DataStructures_V0.5.doc
saved: 25.07.2014 15:27
printed: 25.07.2014 15:27

state of content: July 2014
draft: R. Göldner

1 Motivation

The Archaeological Heritage Office Saxony (Landesamt für Archäologie Sachsen - LfA) is responsible for protection, documentation and presentation of archaeological heritage in Saxony. While it is our priority to preserve archaeological monuments for future generations whenever possible by means of management, we have also developed an effective rescue program in order to salvage as much information as possible if destruction is inevitable. A further important aspect of heritage management is the integration of research programs and the publication of the results.

So we have quite a lot of information to be stored and to be provided for heritage protection and for scientific research. And all these information has to be preserved for very long terms, for future generations, ranging from office documents to CAD/GIS and 3D data.

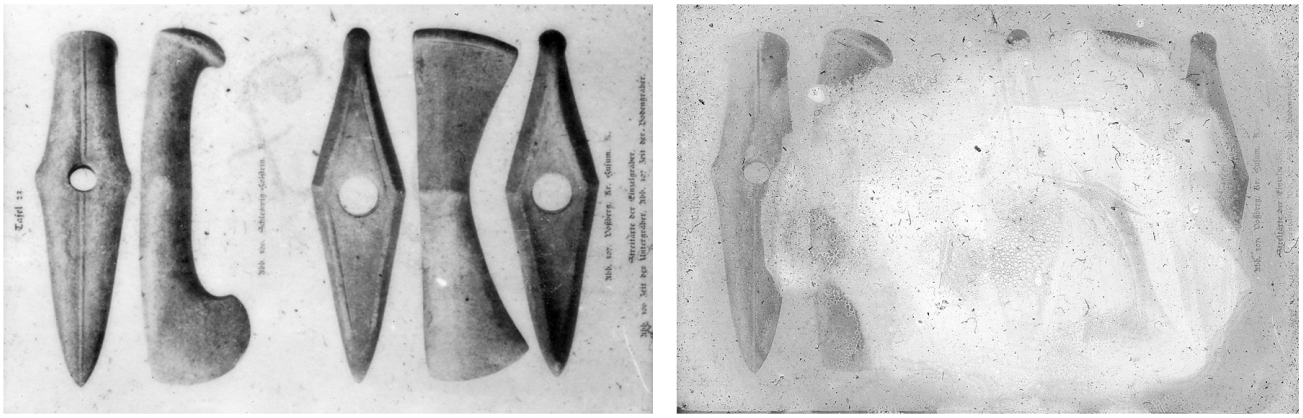

Figure 1 – Process of aging at analogue media: left – xero-copy (1990th) at paper, right – original slide (2011).

2 Archaeological Data Structures in Saxony

2.1 Overview

As an overview model on our information we use two kinds of view, a rather archaeology-specific view and a more type driven view. Both views help us to differentiate important properties of the data and to concentrate attention on main features.

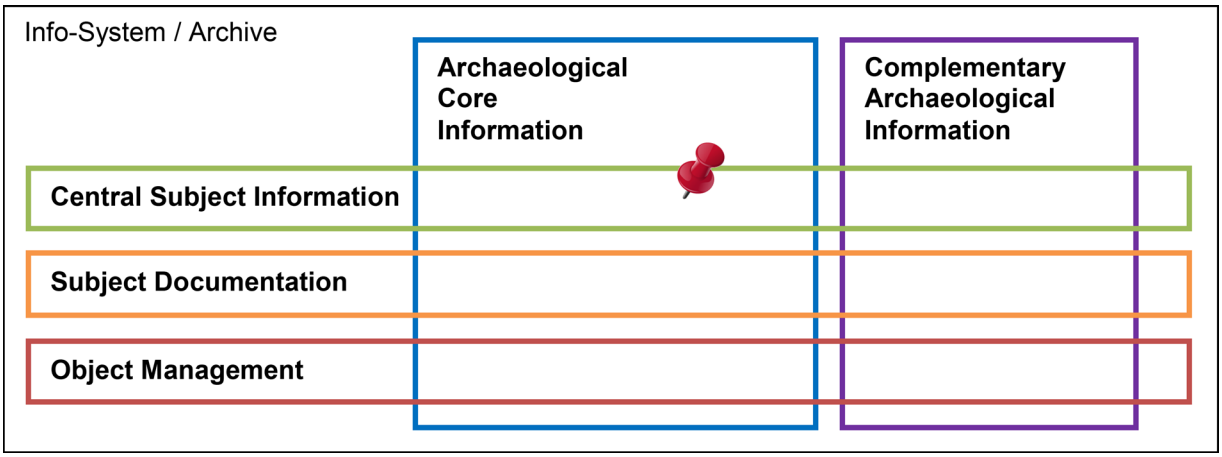

Figure 2 – Structure Overview

2.2 Archaeology-specific Structure

The first view is an archaeology-specific structure:

- **Archaeological Core Information** contains **geo-referenced activities, sites and monuments**. Almost all other information can be referenced to one or more items of Archaeological Core Information.
- **Complementary Archaeological Information** contains further information, e.g. description of finds and contexts, excavation documentation, 3D-scans etc.

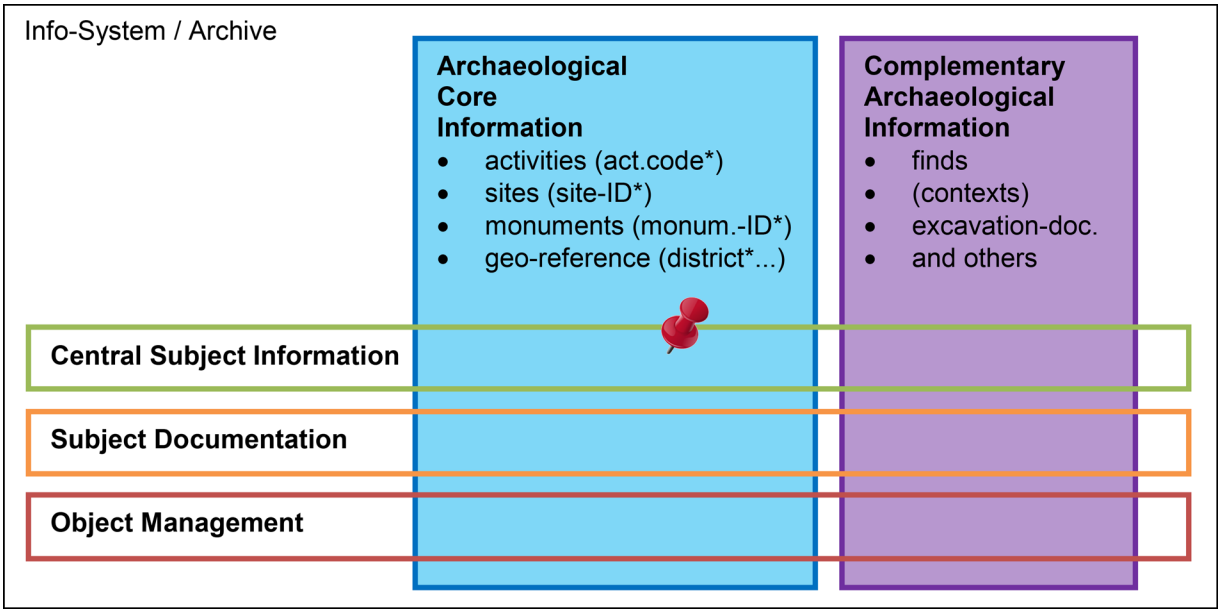

Figure 3 – Archaeology-specific Structure

2.2.1 Archaeological Core Information

The elements of Archaeological Core Information (geo-referenced activities, sites, monuments) are central points of orientation and reference for almost all further information of the LfA. The element-classes can be described as follows:

- **Archaeological Activity**:
  measure, performed at a specific archaeological site by archaeological means and methods, that brings out archaeological results (finds, information, …)
  typical attributes: activity-code (key), -type, -date, …
- **Archaeological Site**:
  site/location/area, where archaeological activities take place (essential categories are: find spot, excavation site, negative site, survey site, …)
  typical attributes: site_ID (key), -category, -name, …
- Archaeological Monument:
  known site/area, that is under heritage protection by law (SächsDSchG), normally based on archaeological sites, contain information needed for heritage protection
  typical attributes: monument_ID (key), -type, foundation document, reasoning, description, …
- **Geographic Reference**:
  describes spatial location and geometry of archaeological items directly by GIS-features (polygon, line, point, …) or coordinates or indirectly by naming a location (district, address, …)
  typical attributes: GIS_feature_ID, coordinates, district_code, …

The key attributes of the core information are socalled "**Core-IDs**". They require a high level of uniqueness, consistency and accessibility. So they build the main axis of our information system and are able to link together almost all kinds of archaeological information.

Geographic data is an essential component of Core Information, which is closely linked with all other information. Using geographic information systems (GIS) makes it quite easy to record, manage, process and

evaluate this multiple information. The results may be published by web services (e.g. web map service, web feature service) that allow a broad access via internet.

2.2.2 Complementary Archaeological Information

The Complementary Archaeological Information especially contains documentations on finds and contexts, accompanied by further archaeological documentations.

2.3 Type-specific Structure

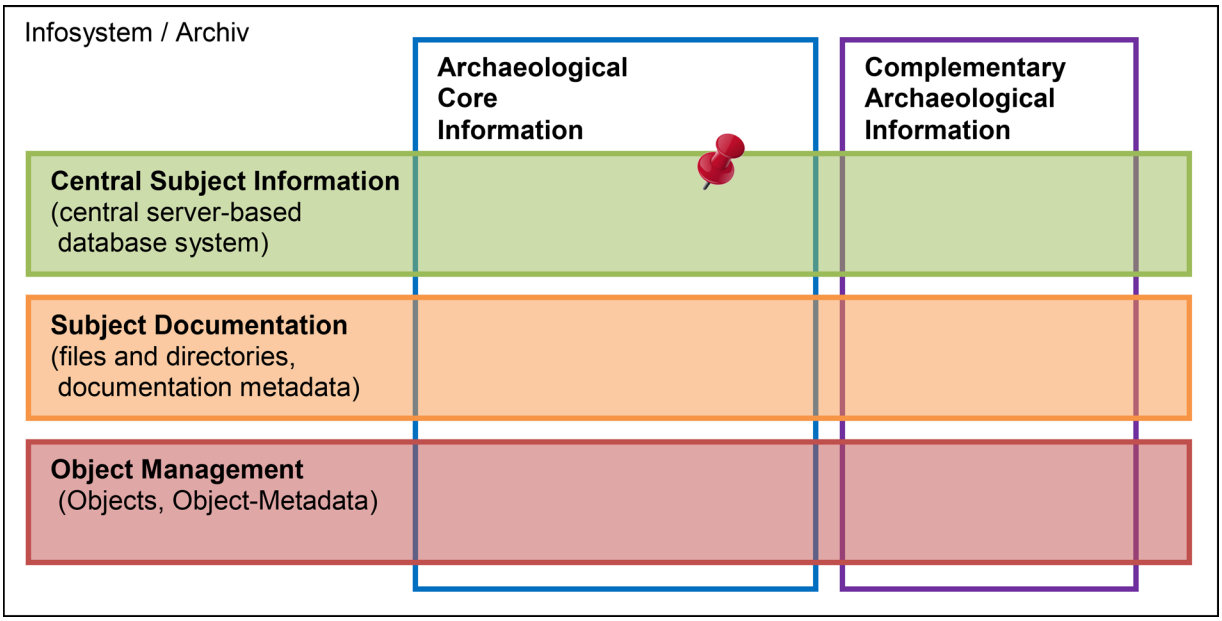

Figure 4 – Type-specific Structure

2.3.1 Central Subject Information

Central Subject Information is a group of important, well-structured archaeological information with high relevance to many business processes of LfA. They are stored and managed in a central server-based database system. They provide an overview of all archaeological issues in Saxony and are starting point for inquiries to further information, especially to subject documentation and object management.

2.3.2 Subject Documentation

Subject Documentation supplements central subject information with more detailed but less structured information stored in files, directories and meta-datasets. This could be analogue documents (paper, film, …) or digital documents (office, images, geo-data, …). To support finding and accessing those documents they have to be linked with the central subject information (especially core information).

2.3.3 Object Management

Object management has a kind of exceptional position, because it deals with physical objects (finds). These finds have to be registered in an inventory that stores all information needed to manage import, export and curation.

3 Information for the Future

3.1 Archiving of digital Research Data

The Archaeological Heritage Office Saxony (Landesamt für Archäologie Sachsen - LfA) acts as research institution. So, more and more scientific results arise in digital form. As they are information on cultural heritage they have to be preserved and archived. This applies to excavation documentation too, which takes the place of the original heritage situation. For that reason LfA plans to permanently preserve and archive digital documents and data from the research area.

But information technology mostly stands for short innovation cycles that act contrary to archiving interests. This requires special means to archive digital data in addition to bit-stream preservation as usual.

3.2 Requirements on Archiving

The aim of archiving digital data is to permanently guarantee the availability of content and functionality, or in other words, to permanently preserve usability.

Due to the scientific focus we need great flexibility. Archiving has to be integrated into the central data structures of LfA and it has to support functional interoperability between central subject information (database) and subject documentation (archive). And, contrary to the management rules of administrative documents (that doesn't help us here) we concentrate on the "Open Archival Information System" (OAIS, ISO standard 14721:2012).

On the basis of the OAIS-method it is profitable to reduce the amount of different data formats to avoid trouble in migrating data to newer format types. Besides some very simple formats PDF/A is often suggested to archive digital documents.

So we have two important aspects to discuss, preservation planning (OAIS) and archive data formats.

3.2.1 Preservation Planning

To support permanent preservation of archaeological research data we need a system that operates due to the OAIS standard. This includes the following basic functionality:

- Ingest
- Archival Storage
- Data Management
- Administration
- **Preservation Planning**
- Access

Preservation planning means to permanently monitor the corresponding information technology (especially data formats and application and presentation software) and to permanently plan, perform and observe preservation tasks (e.g. migration steps).

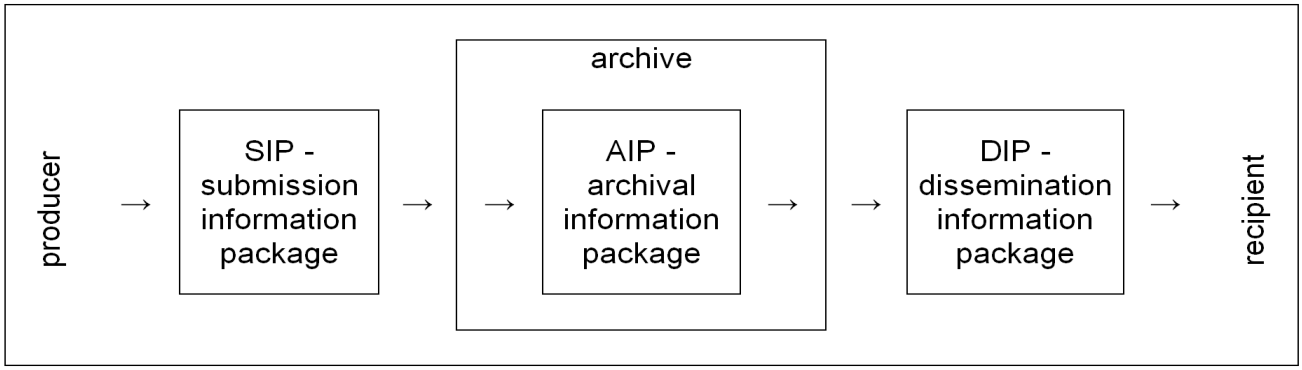

Figure 5 – Use of information packages in OAIS

Preservation planning at one hand refers to technological progress in digital archiving itself. At the other hand it supports the realisation of specific preservation methods. E.g., the renewal of old AIP formats has to be organised, including questions of format, integrity, authenticity and so on. So Preservation planning is a conceptual issue of central importance. It mainly focussed on the migration method of archiving but is open to other approaches.

To support permanent preservation of archaeological research data we need a system that operates due to the OAIS standard.

3.2.2 Data Formats, PDF/A

OAIS doesn't limit data formats, but migration effort increases with a greater diversity of formats and with shorter life cycles, so it would be interesting to reduce the amount of formats to some special types that are well-suited for archiving. Regarding this the following formats are suggested:

- for images: TIFF or JPEG
- for unformatted text: TXT
- for formatted text: PDF/A
- for databases, CAD/GIS and other structured data (in addition): PDF/A

As formats are independent from the OAIS standard they can be considered and determined independently from a special archiving software in most cases. So you can start thinking about how to create TIFF, JPEG and above all PDF/A without having an archiving software.

To create PDF/A take into account the following tasks:

- generative creation of PDF/A level A (Accessible Conformance, unique visual reproduction and content based structure)
- conversion of standard office documents (e.g. doc, xls, ppt, …) to PDF/A level A
- conversion of any documents to PDF/A level B (Basic Conformance, unique visual reproduction)

3.3 Aspects of a first Implementation

As server databases inherently provide functionality to permanently preserve its content data, they don't need much attention at the moment. So we can focus the subject documentation, where we state a lack of archiving activities.

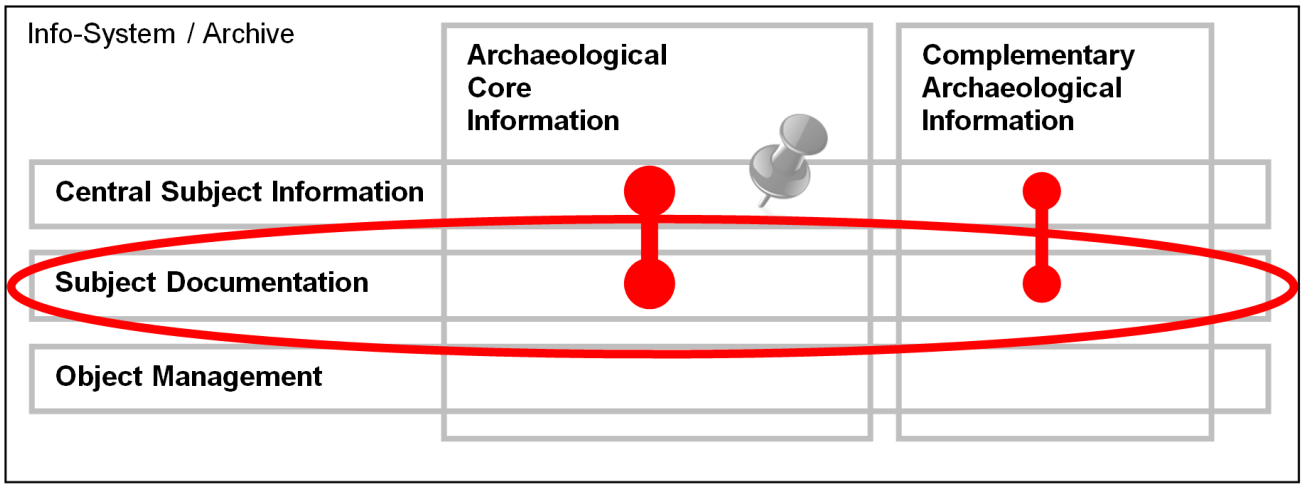

Figure 6 – Data structure and focus of archiving.

As defined in the data structure, it is necessary to link the archive packages from subject documentation to data sets of central subject information.

For a first implementation we address the following archive branches:

- **General Archive** (unspecified document types, standard set of metadata including core-IDs)
- **Sites and Monuments Archive** (digital "Ortsakte": text + images + scans, specific set of metadata)
- **Excavation/Project Archive** (excavation documentation: text + geo-data, specific set of metadata)
- **Image Archive** (test bed to check principles of functional interoperability with "Diathek" database)

4 Résumé

Data structures are models of reality. They are used to build information from data and to manage and preserve this information. Especially the availability of GIS and geo-data is an important incentive to develop realistic models and data structures.

Digital archiving becomes respectable. Methods and standards gain acceptance and first systems are commercially available. So the time has come to try first steps in this new terrain.

To which extend our current information will be available to future generations of scientists? The answer will depend on data structures and archiving strategies.

5 References

R. Göldner, Dokumentations- und Informationssystem Archäologie – Studie zur Datenstruktur (http://www.archaeologie.sachsen.de/download/LfA_DIA_Studie_Datenstruktur_V1_1a.pdf, 28.05.2014)

R. Göldner et al., Kleiner Ratgeber zur Archivierung digitaler Daten. In: Archäologie und Informationssysteme. Vom Umgang mit archäologischen Fachdaten in Denkmalpflege und Forschung. Arbeitshefte zur Denkmalpflege in Niedersachsen 42. Niedersächsisches Landesamt für Denkmalpflege (Hameln 2013), 86-104 (http://www.landesarchaeologen.de/verband/kommissionen/archaeologie-und-informationssysteme/projektearbeitsgruppen/veroeffentlichung-ais/, 28.05.2014)

R. Göldner et al., Ratgeber zur Archivierung digitaler Daten (http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Archaeologie-Informationssysteme/Dokumente_AIS_Archivierung/Ratgeber-Archivierung_Vorab-V0.07.pdf, 28.05.2014)

nestor – Kompetenznetzwerk Langzeitarchivierung. (http://www.langzeitarchivierung.de, 28.05.2014).

nestor Handbuch: Eine kleine Enzyklopädie der Langzeitarchivierung. (http://nestor.sub.uni-goettingen.de/handbuch/index.php, 28.05.2014).

OAIS – Open Archival Information System. Referenzmodell für digitales Langzeitarchiv. (http://public.ccsds.org/publications/archive/650x0m2.pdf, 28.05.2014).

PDF/A Competence Center. (http://www.pdfa.org, 28.05.2014).