archaeoDox

Information management from the dig into the archive

Jörg RÄTHER[1] | Eicke SIEGLOFF[2]

[1] Archaeological Museum Hamburg, Germany | [2] State Archaeological Department of Schleswig-Holstein, Germany

Abstract: The design of an archaeological database or the decision for a particular application often depends on special scientific interests or the requirements of a single project. The requirements regarding such an application are complex. It needs to reflect a wide range of contexts. The preservation and care of field monuments with the need to document everything from a find notice to a longtime excavation it will need to document everything - from prehistoric sites to development led urban archaeology
It needs to gathering homogenously structured archaeological information through both staff members and external contractors. Under varying operational conditions, the management of the increasing amount of digital files is a main task. Meeting those requirements archaeoDox was designed as a common relational database for scientific information like the documentation of finds, contexts, stratigraphic relations etc. and on top of that as a document management system. It builds up a file system based on the documentary steps of each project which can be searched directly within the application linked to the scientific information.

ArchaeoDox was developed by derbrill IT service in cooperation with the Archaeological Museum Hamburg and the State Archaeological Department of Schleswig-Holstein. In both organisations archaeoDox serves as a standard application at archaeological excavations and for the revision of archived documentary and the storage of digitalized analog files. The technology of archaeoDox is open source and a community edition of it was published as free software under GPLv3 in April 2014. A general aim of the development of archaeoDox was a simplification of working processes within a data-centric architecture. archaeoDox can be applied as a client-server-application as well as a stand-alone-application on site and a geo-data extension is currently planned. archaeoDox is also considered to be a necessary step towards the possibility of long term digital preservation[1] later on.

Keywords: excavation database, document management system, open source/ free software

_____

[1] For further general information on long term digital preservation see
http://www.landesarchaeologen.de/fileadmin/Dokumente/Dokumente_Kommissionen/Dokumente_Archaeologie-Informationssysteme/Dokumente_AIS_Archivierung/Ratgeber-Archivierung_Vorab-V0.07.pdf;
http://guides.archaeologydataservice.ac.uk/g2gp/Main.

Initial situation

The bandwidth of archaeological sites from palaeolithic to modern times can be overwhelming. Trying to put this diversity into a piece of excavation-software is a huge task. To remain flexible in data entry is a challenge in itself. Yet it is crucial to incorporate the different situations and the depths of accuracy of data being collected into the concept and design of an excavation database. Simple observations on building sites under high pressure of time, preliminary investigations and longer lasting projects need to be reflected in the data model and user interface. The necessity to display rural and urban scenarios adds to the complexity. Meanwhile database software like Microsoft Access has enabled almost everybody to set up own databases. Those are in use for either single or multiple excavation projects or specific scientific questions. But often the main focus does not lie on building up really homogenous data structures, at least not across projects or institutions.

While excavating, a soil archive will be destroyed und thus needs to be transformed into a documentation and finds archive. Only complying with conventions data can be conserved in a uniform and reusable manner. All documents and finds must be converted into an achievable state; documentation must be completely transparent without gaps and be clear and comprehensive. Resulting data should be uniform regardless the person filing it. Without meeting those standards, it will not be possible at all to maintain the archive of our cultural heritage. Once destroyed is destroyed forever – not only using shovels, but also by accidentally deleting data or careless backing up, misplacing files etc.

Besides find objects, the documentation is the only remaining source of information left for posterity. The only source archaeological information can be extracted and re-evaluated from.

Although this is obvious to most parties involved, way too often the structure of filing systems is not exactly prescribed and lies in the hands of a processor. They try their best to file the data where and how they can see fit. If database-software is used, in most cases it could certainly manage a single project, maybe including document-management on a hard disk. A cross-project overview on technical and scientific data and related documents should be rather difficult.

Due to the lack of a central repository with homogenous data structures it will become more and more difficult to gain a complete overview on the ever growing amount of excavation projects and their documentation data. This is regardless of them being in the own institution, own district or beyond that. A central repository would also drastically increase the quality of the results of scientific data queries and reduce the time spent to do the research.

archaeoDox - Concept and technical background

Facing the same initial situation, the development of archaeoDox followed a different mindset. First of all, major focus is the homogenous data-structure. A combination of a relational database and a document managing system is used. Technical and scientific data is collected and the linked documentation files are transmitted into a structured filing system that is expanding automatically during the course of the excavation.

.

ArchaeoDox is a client-server system. All data being collected on the client is synced into a central repository and a global filing system. When working in network mode the connected clients can work collaboratively. Using the web component, currently under construction, it is possible to browse through data and documentation of all processed archaeological activities using a web browser. You can also choose to use the client directly on the central repository.

After processing more than 400 archaeological activities with archaeoDox in Schleswig-Holstein since 2011, there is a homogenous filing structure for millions of data files likes photos etc. in a respective activity folder. There are always two subfolders placed in each of them: the backup-folder which holds a variety of automatically created output lists and an XML file. It can be used to exchange data, but also to fully restore the data for the excavation campaign. The second subfolder „documentation steps" contains a flat hierarchy of folders that holds the filing system for related documents. With every step in documentation carried out in the field, the filing system grows by one new folder that only needs to be filled with the documents. They will automatically be renamed by the system. The official in charge does not need to care about the naming conventions anymore.

Data being collected using the client, which of course also can be carried out offline, will be synced into a main repository as shown in a schematically overview on Fig. 1. Using the connect module any machine can work as the data repository on a large excavation campaign in a local network. In this way an entire excavation team can collaboratively work on the same campaign. Thus the complete architecture is scalable. Viewed from a different angle, this can also mean that a branch office autonomously collects data in its own network and afterwards delivers mentioned data to a common main data repository. This is specifically interesting looking at cross-county linear projects and of course scientific research. Using interfaces, data can be exchanged with external systems. The whole system is based on efficient open source software and is available as a free community edition under GPL license since spring 2014.
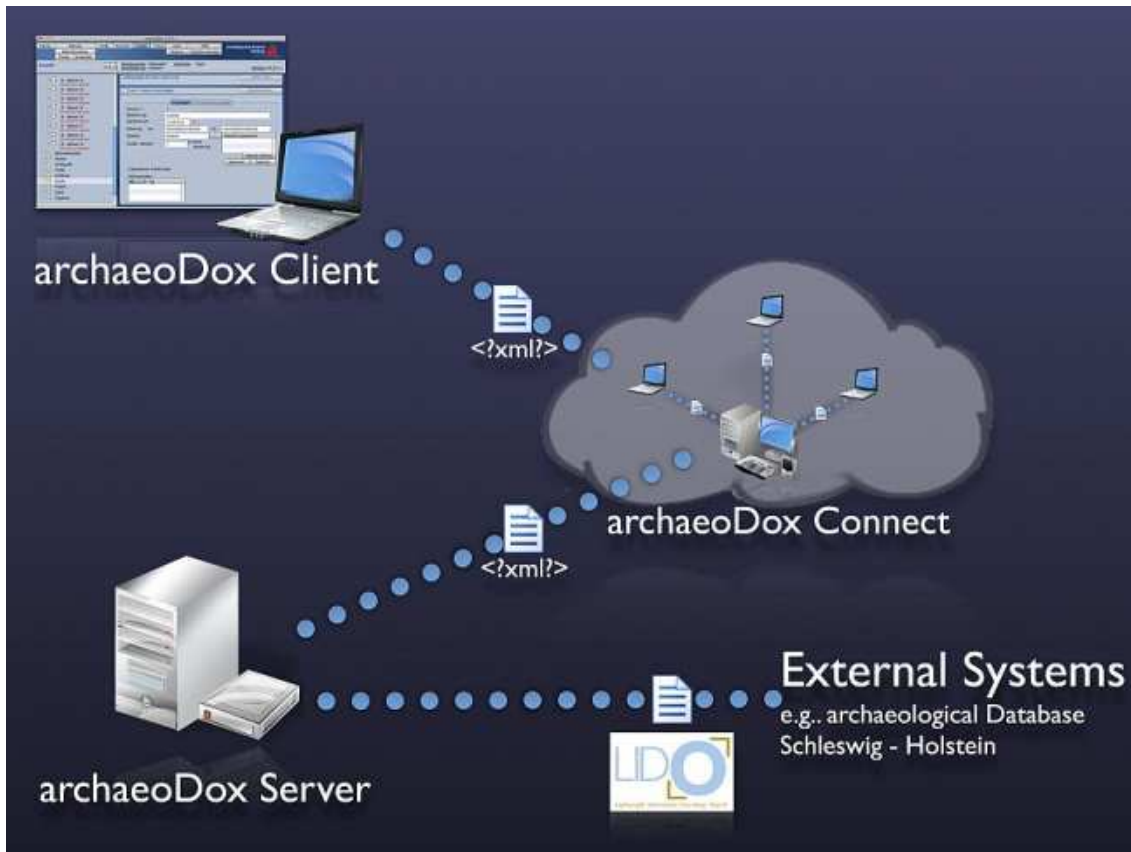
Fig. 1 – scalability of the archaeoDox-system-components

In order to distinguish the different archaeological activities a distinct number range is of crucial importance. Even though on the surface Schleswig-Holstein and Hamburg are using the same numbering convention, data integrity is always ensured through the use of so-called UUIDs, universal unique identifier. The concept of using UUIDs is carried out through all database entities. Thus every data-entry is always uniquely identifiable.

Additional to mandatory setting the geographical context, the official in charge needs to decide which module to use for collecting data. Both modules are based on single-context-recording. The module „urban excavation" is strictly based on stratigraphic units and thus fitting best for typical layer based excavations like in historic city centers or complex palaeolithic sites. The other module „field excavation" is based on features, fitting for example for excavations on prehistoric rural settlements or burial grounds. The underlying data-structure is exactly identical and gets very close to the general CRM- and the CRMarchaeo-data model[2].

──────

[2] http://www.ics.forth.gr/isl/index_main.php?l=e&c=711.

.

archaeoDox – In brief

In Fig. 2 the main data input screen is shown. Three major components are displayed. There is a tree-view on the left hand side, a list view and navigation in the center and on top a quick navigation using buttons. The buttons and nodes in the tree view correspond to existing technical units in an excavation, like for example so called areas. An area means nothing else than any section in an excavation. In this case eleven areas in the terrain have been entered and recorded in archaeoDox.
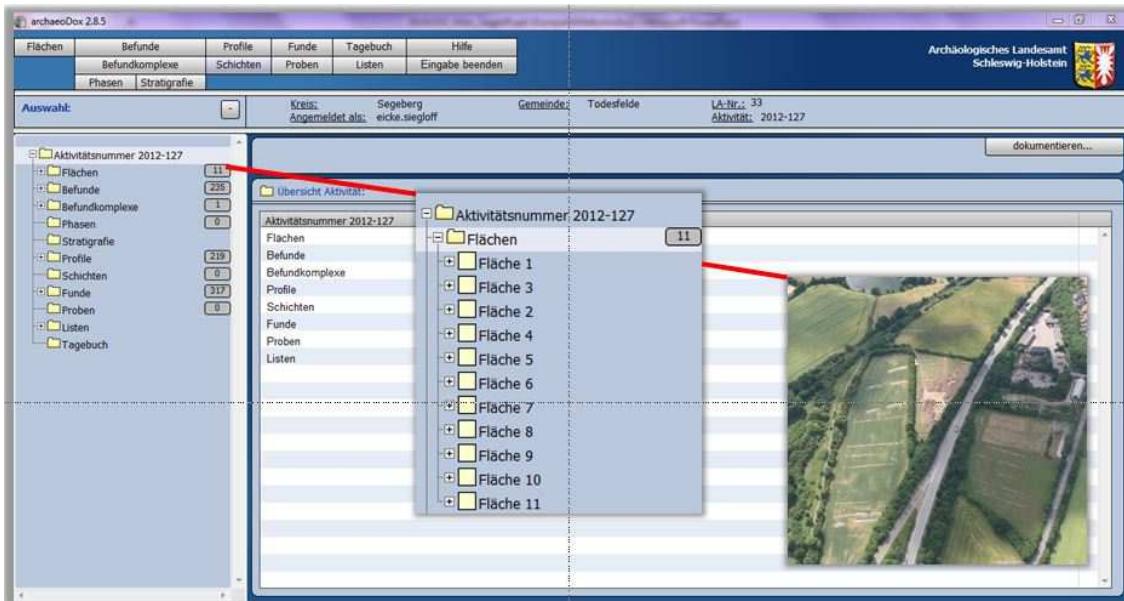

Fig. 2 – screenshot of the main data-input-screen (areas)

The same holds true for all other entities, such as features. They can be recorded in batch processes to minimize input time. Displayed on Fig. 3 is the main view of a feature with its various information, such as the associated excavation area as the feature's context. Thesaurus driven fields are used for dating and feature naming. Features can generally be grouped into so-called feature-complexes, for example the grouping of several post-holes into a footprint of a building. Stratigraphic references such as „older than" or „younger than" can be entered using the stratigraphy-submodule. Looking at the tree-view one can see that the displayed feature has been excavated in two plana. In the bottom right hand corner all related find objects are listed. The list fields always work as a quick navigation, too, in this case in order to jump to the data of one of the related finds. The same holds true for all the related documentation steps. In the shown example it becomes clear that the feature has been documented in five steps. archaeoDox increases a counter for the documentation-steps of each activity. Verbal descriptions are shown directly in the main screen of the feature. Using the Button "*dokumentieren*…" the documentation can be completed anytime.
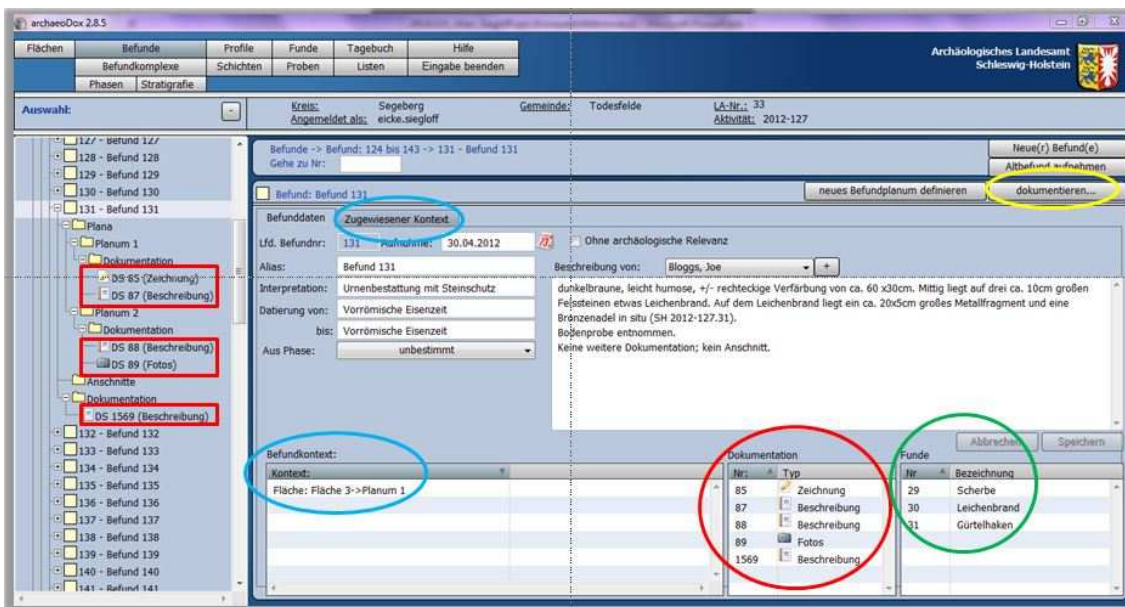
Fig. 3 – screenshot of the main data-input-screen (features)

When setting up a new documentation step a choice is given which type of documentation was carried out and should be entered, like photos, measurements, drawings, etc. The core information, which context is being documented, is already given just by pressing the button. Now additional metadata can be entered. Due to archaeoDox also being a document management system governing hybrid filing systems, there are for example input fields for the page numbers of analogue drawings. Wherever possible and where it makes sense thesaurus driven fields are used.

This extremely simple relationship hierarchy enables a simple context based folder and file naming as mentioned before and particularly also multiple allocations. If for example a photo contains an overview of multiple features it can be assigned to all of them. No redundant documents are being created. Each allocated feature will show the photo documentation step in its overview and redundancies are avoided. Simplifying the overview in the folders, optional tags can be defined. Entering a documentation step archaeoDox automatically creates the corresponding folder in the filing system. With the priority of keeping it simple the person in charge only needs to hit the link „open folder" and just drag and drop the desired files into it – done. The files are also displayed in the tree view and in this example have already been renamed by the system. For all documents of the type picture a preview is implemented. Every other document can be viewed in the related viewing program just by clicking a link.

For the entering of finds only object designation, material, dating, quantity and the find context are mandatory and for simplification can be entered in batch mode, too. Representing all types of entities, be they profiles, plana, samples, grids etc. also find objects can be documented via the function of documentation steps.

Generally thesaurus driven vocabularies are used for the fields object designation, material and dating. archaeoDox is capable of feeding a web-interface for data exchange to an external finds-database (Fig.4), in

.

this case the Archaeological Database Schleswig-Holstein, using the standardized LIDO XML format[3]. The user triggers this dataset-transmission by pressing a button on the startup screen.

The used vocabularies are maintained using a web based tool called xTree in which the entries are grouped by topic in multiple logical and hierarchal levels. It lets you use different kinds of synonyms for the same entry, also in other languages and with ISO-conform definitions and mappings to other vocabularies like the AAT. So for data entry or search queries no unique term is needed. Hamburg and Schleswig Holstein at the moment participate in building up a vocabulary for feature designations in cooperation with the leading authority Landschaftsverband Rheinland.
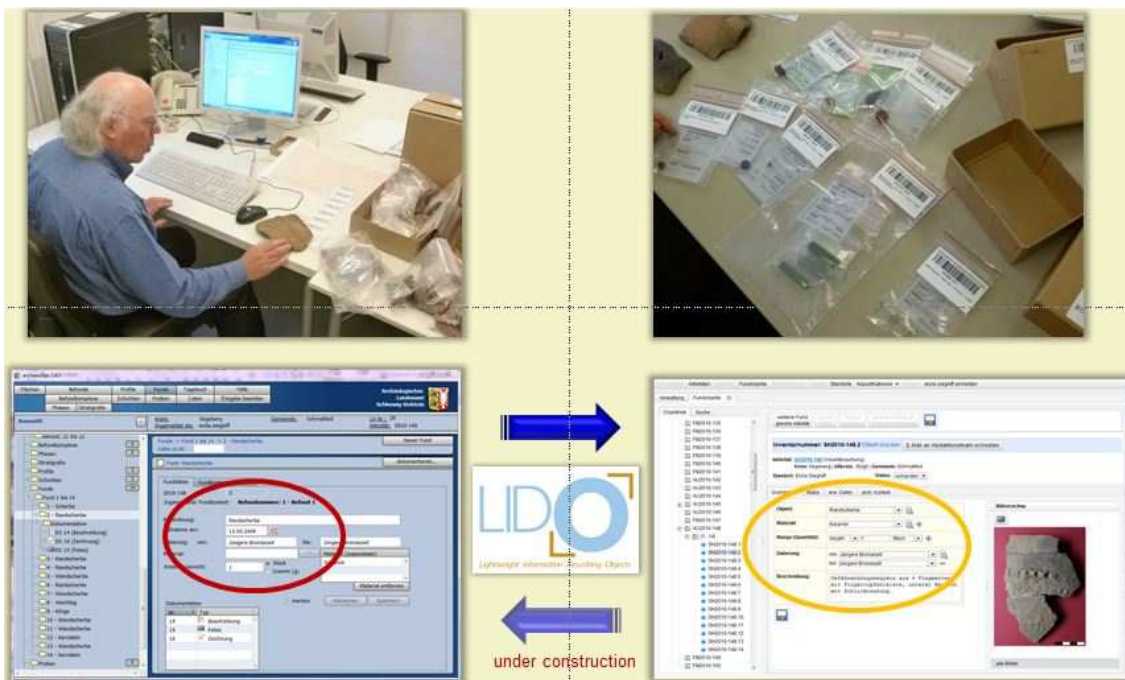

Fig. 4 – Inventarisation of find objects via archaeoDox and data-interchange to external systems (LIDO-logo by Rob Lancefield; http://www.lido-schema.org)

As a substantial benefit to the person in charge archaeoDox creates different kinds of output lists, like a finds list and a feature catalogue for the excavation report with the complete documentation cited. These lists are always accessible in archaeoDox and can also partially be used as a navigation element. On top of that a basic quality assurement list has been implemented to ensure accuracy and quality of data.

Another feature in archaeoDox is an implemented site notebook automatically set up by the documentation step list of each day. Besides that it is possible, in a sense of a classic excavation diary, to enter special occasions of the day, like weather or visitors, and additionally to enter the working hours of the team-members.

———

[3] http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/.

archaeoDox - Work in progress

The further development of the free and open-source software archaeoDox is a continuous process of refining functionality, both looking at usability and new features like a role-based user rights management, the expanding of the web-application, etc. The depths of imaginable features and possibilities seem sheer endless. Currently search and reporting functions are being implemented or expanded. These analysis-tools seem to be developed relatively late, however, laying the foundation with homogenous datasets of high quality, it is crucial to go the first steps first. Particular mention should be made of the concrete planning of an extension to deal with geodata. The development target is a bi-directional interface to an open-source desktop-GIS. There are also considerations to expand the existing archaeoDox-XML-file to a standardized exchange format[4].

The objective of the developers of archaeoDox is to deal with the everyday life in archaeological heritage management, to assure sustainability and to lay a more solid foundation for future generations of researchers – all to avoid future excavations of digital data.

———

[4]See http://www.ianus-fdz.de/it-empfehlungen/?q=node/20&qt-datenbanken=0#qt-datenbanken for further general information on databases and XML.