The e-Depot for Dutch Archaeology

Archiving and publication of archaeological data

Hella HOLLANDER

Data Archiving and Networked Services (KNAW-DANS)

Abstract: In the Netherlands, the archiving and publication of archaeological research data has led to the establishment of the e-Depot for Dutch Archaeology (EDNA) accommodated at DANS. EDNA is a collaboration between DANS and the Cultural Heritage Agency (RCE). DANS is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO).

The slogan *"Digital archaeology requires a digital memory"* was used in 2007 to bring care for digital data to the attention of Dutch archaeologists. The e-depot ensures durable archiving and unlocking of all digital documentation of the archaeological research. By 2014, DANS provides online access via EASY to more than 21,500 archaeological datasets: 18,500 reports and 3,000 large datasets consisting of data of excavations and explorations (photos, GIS, data-tables, drawings). Both the research descriptions and all data can be downloaded.

Agreements to this end have been laid down in the quality standard for Dutch archaeology. DANS ensures that access to digital research data keeps improving, through its services and by taking part in national and international projects and networks. By participating in projects such as Odyssee, CARARE, ARIADNE and DARIAH, the options for finding, accessing and re-using archaeological and other data are continuously improving. DANS stimulates cooperation between data producers and users and does research into long-term accessibility.

The existing infrastructure of the e-depot for Dutch Archaeology allows for sharing of good practices such as long-term preservation, data organisation and data dissemination for accessibility.

Keywords: Archaeological data, archiving, publication, access

Digital infrastructures and data sharing services

E-depot for Dutch Archaeology

In the Netherlands professional archaeologists follow the rules of the Quality Standard for Dutch Archaeology (KNA). This manual is describing in detail the standards and guidelines of the archaeological research, management of finds and data documentation. It is a good example of self-regulation to assure the quality of the work. As for the standards on digital documentation, the Quality Standard refers to the requirements and guidelines of DANS.

DANS stores the digital files with research data of Dutch archaeologists. These are the files with the primary archaeological data of excavations, regional explorations and material studies. The e-depot for Dutch Archaeology ensures durable archiving and unlocking of all digital documentation of the archaeological

research. The trend to make data available to users is clearly continuing, as by now more than 21,500 archaeological datasets have been archived of which 70% is freely available.

Fig. 1 – Excavation at Zaltbommel De Wildeman
Roman glass excavated in situ
Copyright Photo: ADC ArcheoProjecten

Fig. 2 – Roman glass cleaned and repaired
Dated 50-150 AD
Copyright Photo: ADC ArcheoProjecten

European Digital Infrastructures

In the past century many rescue excavations were conducted due to the booming building activities. All archaeological finds and data were collected, but time and money were missing for a proper analysis and publication. This data forms a huge potential for modern archaeologists in helping to resolve the research backlog.

Within the national Odyssee project, DANS made an inventory that led to a list of 1800 Dutch excavations that were neither analysed nor published. These projects are now offered for research. On request of archaeological researchers field-and object drawings were scanned. The NWO research program selected 32 projects to be re-analysed by Dutch Archaeologists and DANS provided the source data in digital form. Additionally, international projects like CARARE (Connecting Archaeology and Architecture in Europeana) and ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe) help to develop and expand the archaeological digital infrastructure. Within the CARARE project millions of objects in the field of archaeology and architecture were added to Europeana, the European website of cultural heritage. DANS provided the Dutch open-access publications.

ARIADNE is a research infrastructure funded by the European Commission's Seventh Framework Programme (FP7) and started in February 2013. The project responds to the growing wish to connect European national databases and archives of archaeological material so that researchers can use the

various distributed datasets with new and powerful technologies. The 4-year project addresses European access to archaeological data. Integration tools will range from metadata crosswalks and mappings to GIS and will use state-of-the-art technology such as data mining and Linked Data. ARIADNE is coordinated by PIN, a consortium formed by the University of Florence and local institutions. DANS is one of the 24 ARIADNE partners from 16 countries across Europe and will contribute the e-Depot for Dutch Archaeology and the Digital Collaboratory for Cultural Dendrochronology (DCCD) to the infrastructure.

The DCCD is hosted as a separate archive for dendrochonological data at DANS and became operational in 2011. The DCCD is the only archaeological/historical tree-ring (meta)data network existing in Europe (Jansma 2010). Belgian, Danish, Dutch, English, French, German, Latvian, Lithuanian, Polish, and Spanish laboratories have joined data in a manner that suits their shared and individual research agenda's. In its present state the DCCD contains over 50,000 measurement series of different wood species derived from over 5,000 objects and sites dating between 6000 BC and present. Ca. 50% of the DCCD collection is derived from archaeological sites and structures, including maritime archaeological sites. The remaining data are mostly historical as well, being derived from historical architecture and mobile heritage (paintings, furniture).

Most data sets are described by very detailed metadata according to the newly developed international Tree Ring Data Standard (TRiDaS), using multi-lingual controlled vocabularies (English, French, German, Dutch). TRiDaS is an XML-based international standard for Tree-Ring Data developed to address the problems and differing implementations of existing and legacy data formats and to provide a common data standard to aid the exchange of dendrochronological data and metadata between users and software applications.

The metadata is free for use and open access and is visible to any user who is not logged in. To get access to deeper information levels, user registration is necessary and different levels of access are applicable. For registered users research results such as wood species and absolute calendar dates are 100% open access. 15% of the dendrochronological time-series and associated files are fully open access (i.e. can be downloaded without contacting the authors); this percentage is expected to increase when the DCCD becomes linked to other data networks.

By participating in projects like ARIADNE, the options for finding, accessing and re-using archaeological data are continuously improving. Partners from Sweden, UK, Ireland, Germany, Austria, Hungary, Czech Republic, Slovenia, France, the Netherlands, Italy, Spain, Greece, Cyprus, Romania and Bulgaria are sharing their archaeological research knowledge and good practices for digital archiving as long term-preservation, data organisation and data dissemination.

The Digital Research Infrastructure for the Arts and Humanities (DARIAH) is providing technical frameworks, interoperability standards and sustainable storage and collaborates on these topics with ARIADNE.

Fig. 3 – St Mary's Abbey in York Photo by Hella Hollander

Sharing of good practices such as long-term preservation, data organisation and data dissemination for accessibility will improve the European archaeological digital infrastructure. By making research material digitally available, academic research will be stimulated, new scientific insights and information will be gained.

Long-term Preservation

DANS follows the OAIS Model, which stands for Open Archival Information System. Terms and conditions are followed to make long-term preservation and accessibility possible. In the recently published

Preservation Policy of DANS, the principles of sustained archiving of digital research data are outlined (DANS Preservation Policy v1.0[1]):

> *"The Archive follows a policy of active preservation with the aim of ensuring the authenticity, reliability and logical integrity of all resources entrusted to its care while providing formats suitable for research for the long term. The specific aims of the preservation policy are to:*
> *• provide authentic and reliable instances of datasets to researchers;*
> *• maintain the integrity and quality of the datasets;*
> *• ensure that digital resources are managed throughout their lifecycle (e.g. when migrations or changes in metadata are carried out) in the medium that is most appropriate for the task they perform;*
> *• ensure that the relevant level of information security is applied to each dataset;*
> *• and so to be a "trusted digital repository".*

The formulation and biannual revision of a preservation policy for the Archive are essential steps in fulfilling its strategic aims and responsibilities: it gives strategic direction both to continue initiatives which are necessary for the protection of its collections, and to meet or extend nationally and internationally agreed standards (to be) for digital preservation."

Archaeologists are aware of the importance of archiving their data in a sustainable manner, as the data coming from an excavation is unique and can only be documented once. At DANS the data is converted into a sustainable archiving format. Licence agreements and conditions of use are taking care of the rights and obligations of the data depositor and those of DANS, describing the rules for data use and copyrights.

Data Organisation

DANS applies the concept of self-archiving. This means that data producers themselves deposit their research files into EASY, the Electronic Archiving System of DANS, according to guidelines which are online accessible. Training sessions have been organized to inform and advise suppliers of data on documenting and archiving the datasets. To assure that other archaeologists can understand and re-use the deposited data, a lot of care and effort is put into ensuring that the data deposits are provided with good metadata. Searching and browsing through the metadata in EASY can quickly trace the excavation documentation for secondary analysis.

_____

[1] http://dans.knaw.nl/content/strategie-en-beleid

Deposit dataset - 5: Upload files

Deposit:	1: Import (meta-)data from other sources (optional step)
	2: Primary information and intellectual ownership
	3: Research project
	4: Digital files
	5: Upload files
	6: Rights
	7: Overview and submitting

Instructions in English or Nederlands (PDF)

Upload dataset [?]   Bladeren... Geen bestand geselecteerd. Upload   Add the dataset files

Fig. 4 – Upload screen of EASY, the online archiving system of DANS

Metadata can be discerned on three levels: project-specific metadata, file-specific metadata and metadata on the level of variables and codes. On the project metadata level an absolute minimum of seven Dublin Core fields are mandatory (Creator, Title, Description, Date created, Access rights, Date available, License). Up to twenty-four more fields can be recommended depending on the specifics of the dataset. File-specific metadata can be sent as a file list together with the dataset. For larger archaeological datasets, it is mandatory to list all of the files with a description of their content, the software used to create the file and references to codebooks necessary to understand the file. The codebook is a description of the content of the digital documents, explaining all codes and conventions.

After submitting a dataset, an archivist at DANS will process the dataset, checking the completeness and clarity and ensuring that the data files are stored in accordance to the DANS list of preferred formats; the file formats which DANS trusts to offer the best long-term guarantees for usability, accessibility and robustness. It is recommended that researchers deliver their data in the preferred format corresponding to the type of data. The preferred format for text documents, for instance, is pdf/a (.pdf)

Publication of archaeological data

The mission of DANS is to enhance permanent access to digital research data. After submitting data in EASY a persistent identifier will be assigned to the dataset, which is a unique hyperlink reference that will always lead to the intended source, independent of web address changes.

DANS operates as an open digital archive and gives support and guidance to other data archives. As much data as possible should be made freely available for (re)use in scientific research. DANS plays a key role in this process and focuses mainly on academic researchers. DANS supports the Open Access principle, with the motto 'Open if possible, protected if necessary'.

Metadata is free for use and open access. However the data itself has conditions for use and license agreements. Depositors own the data, even if a dataset is open access, the data can only be used for personal use. To get access to the data, user registration in EASY is necessary.

When depositing data the archaeological researchers decide what access level the data shall have. Researchers can choose between the equivalent of a CC BY (By Attribution) license for open access content, or to restrict access to a certain group and/or certain time (with the possibility of a temporarily embargo for up to two years).

Archaeology of the future

Progress over the years has shown that the major archaeological organisations, which had deposited their data on restricted access to the archaeology group, switch all of their access rights settings to Open Access. Of the current total of 21,500 published datasets and publications, 14,700 are set on Open Access, 6,600 on Archaeology Group Access and only 200 on Permission Request. This means that 70% of the archaeological data in EASY has now been made accessible through Open Access. The number of published archaeological datasets increases daily.

Currently, Dutch archaeologists are working on the implementation of a national protocol for uniform data exchange to improve the accessibility of archaeological data. The metadata about the finds and data is offered as an xml document to the different archives that store the digital research data (DANS), project information and publications (RCE and Royal Library) or analogue documentation and excavated objects (Provincial depots). The archives extract the relevant information out of the xml document. DANS included this new standard in its archiving system and supports future developments on digital linking of research information.

References

JANSMA, E. (2010) Preserving tree-ring data: A repository for the Low Countries. In: Driven by data. Amsterdam: Pallas publications.

DANS   http://www.dans.knaw.nl/

EDNA   http://www.edna.nl/

EASY   https://easy.dans.knaw.nl/ui/home

CARARE   http://www.carare.eu/

ARIADNE   http://www.ariadne-infrastructure.eu/

DCCD   http://dendro.dans.knaw.nl/

TRiDaS   http://dendro.dans.knaw.nl/