3.2 Past the Opening: building towards the present, on-going dissemination of Dutch archaeological data as part of the DANS archive.

Valentijn Gilissen MA
Data Manager, Data Archiving and Networked Services (DANS)

It is the year 2013 and the on-line archiving system of the Dutch data archive for sciences DANS (Data Archiving and Networked Services) holds over 17,000 archaeological publications and 3000 larger datasets (photographs, GIS, data tables, ...). Half of this total has been made accessible through Open Access. The number of published archaeological datasets increases daily. Through international projects such as CARARE and ARIADNE and by expanding its services, DANS is continuously working on improving the options for finding, accessing and re-using the data.

Ten years ago, in the year 2003, the above paragraph was far from reality. Dutch archaeologists had no guarantees for the long-term preservation and accessibility of their research data, nor a centralised system for storing and sharing actual datasets. At the time the Cultural Heritage Agency of the Netherlands (RCE) maintained already the Archis system, which offers a map-based system for registering archaeological projects, monuments, single finds and observations, but Archis was never intended for the storage or publication of data files. In fact, it was in 2003 that the foundations for the successful present (on-going) situation were laid by a small group of (ICT) archaeologists concerned with the fate of research data files. A workshop on digital archiving led Milco Wansleeben of Leiden University together with Marjan Balkstein of DANS precursor NIWI to start a pilot project of archiving a selection of research projects from Leiden University within the DARE (Digital Academic Repositories) programme. Following this, NIWI and the Cultural Heritage Agency of the Netherlands started the project EDNA (e-Depot for the Dutch Archaeology). The EDNA project ran from September 2004 to February 2006 on a subsidy from the SURF foundation; the collaborative organisation for ICT in Dutch higher education and research. The project called upon all Dutch archaeological institutions and organisations to make an inventory of their research projects. The project aimed at raising awareness and concern for archaeological data preservation. A user enquiry was held to investigate opinions and desires to be taken into account for the development of a digital depot. The enquiry showed that there was a major need for simple digital access to the Dutch research publications.

Simultaneously with the initiation of the EDNA project, a reorganisation of NIWI by the Royal Netherlands Academy for Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO) led to the foundation of DANS as the national institute for scientific data archiving. DANS was tasked to have as its main mission to enhance “permanent access to digital research data”. Additionally, DANS means to stimulate cooperation between data producers and users and does research into long term accessibility.

EASY (Electronic Archiving System) was developed in-house by DANS, as an open-source on-line archiving system through which self-archiving is made possible. Data creators can deposit their own data sets via a user-friendly interface. The project is described in metadata fields adhering to the international (Qualified) Dublin Core standard. Only a few metadata fields are mandatory, but depositors are advised to use as many fields as they can to promote the re-usability of their dataset. 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.

A dataset submission in EASY will assign a persistent identifier to the dataset, making for a unique hyperlink reference that will always lead to the intended source, independent of web address changes. EASY is qualified as a trusted repository through the implementation of the international Data Seal of Approval.

The development of EASY and the execution of the EDNA project were mutually beneficial. The user demands made apparent through EDNA, with the British Archaeology Data Service (ADS) taken as an exemplary model for policies and standards, could directly be taken into account for the programming architecture. The development of a national data archive for all sciences by DANS programmers meant that the creation of a data deposit and data access application did not need to be included in the archaeological project.

With further subsidies (the Gratema foundation, Leiden University, NWO), EDNA proceeded into its next phase as EDNAII. The project accumulated a large content by having ‘grey’ archaeological publications from Dutch municipalities, universities and commercial project bureaus scanned and ingested into EASY. Larger datasets of the archaeological projects carried out as part of big national infrastructural projects were processed at DANS and added to EASY. The content served as proof of concept to stimulate the archaeological institutions to start depositing their own datasets.

The EDNAII project ensured that the archaeological
data depot fully accommodated itself within DANS. During the EDNAII project, the Dutch official regulation for conducting archaeology (KNA, ‘Quality standard Dutch Archaeology’) was subject to revision. A new deposit obligation was incorporated in the KNA during the years 2006-2007 which states that all digital products of archaeological research need to be deposited in a trusted digital repository. DANS was appointed as the repository for (long-term preservation of) source data (datasets) while the Cultural Heritage Agency (RCE) was appointed to take care of the disclosure of new archaeological publications. The lack of an accessible digital library at the RCE led many archaeological institutions to turn to DANS EASY for depositing their research publications as well as their datasets.

Data depositors can upload their own data files into EASY during the deposit procedure. DANS supports the Open Access principle, while being aware of the fact that not all data can be freely available and without limitations at all times. Under the principle ‘Open if possible, protected if necessary’, depositors can choose the access rights to their data files themselves. Data files can be made accessible under Open Access, restricted to a user group, or restricted to users granted permission by the depositor. As the archaeological content of EASY increased, more and more depositors made the choice for Open Access as the standard access rights setting. Initially, it was feared that many data creators would opt for sharing their data only after granting explicit permission for use, being unwilling to publish their data for everyone. Because of this, an alternative access rights setting was built into EASY: the ‘Archaeology Group’ setting allows depositors to restrict their data to the user group of professional archaeologists. Users can request Archaeology group membership if they are employed by an official (commercial) archaeological organisation, the archaeological institute of a municipality or university. Students are also granted group membership.

Of the current total of 20,000 published datasets and publications, 10,000 are set on Open Access, 9,600 on Group Access and only 400 on Permission Request. Progress over the years has shown three 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.

Besides sustaining the data archive with continuous data acquisition and processing, DANS is participating in (inter)national projects in order to improve on the options for finding, accessing and re-using data. For archaeology, this includes the projects CARARE and ARIADNE:

- The recently successfully finished CARARE project (Connecting Archaeology and Architecture in Europeana) made millions of objects in the fields of archaeology and architecture accessible through the website of Europeana, the European trusted source of cultural heritage. DANS added the open-access archaeological publications to the project.
- The growing wish to connect European national databases and archives of archaeological material led to the start of the ARIADNE project. ARIADNE aims to integrate the existing archaeological research data infrastructures so that researchers can use the various distributed datasets with new and powerful technologies, as an integral component of archaeological research methodology. DANS will make the data from EDNA as well as the data from the Digital Collaboratory for Cultural Dendrochronology (DCCD) accessible through ARIADNE.

DANS is a leading partner in the project DARIAH (Digital Research Infrastructure for the Arts and Humanities). DARIAH is providing technical frameworks, interoperability standards and sustainable storage and is developing Virtual Competency Centres which will potentially be useful for ARIADNE.

In addition to participating in projects, DANS provides training and advice, and performs research into sustained access to digital information. Sharing knowledge according to the Open Access principle and advising partners who are setting up data repositories (such as the German IANUS, the Swedish SND) is part of the mission of DANS.

Within DANS, the personnel specifically working on archaeology has never numbered more than three staff members assisted by three students at the same time. Because of the embedding of EDNA within the larger organisation of DANS, good use of subsidies and essential cooperation, the vision of ten years ago was made into present day reality.