Edited by
Erzsébet Jerem
Ferenc Redő
Vajk Szeverényi

CAA 2008

Computer Applications
and Quantitative Methods in Archaeology

On the Road to Reconstructing the Past

Proceedings of the 36th Annual Conference on Computer Applications and Quantitative Methods in Archaeology

CAA 2008 Budapest
April 2–6

Budapest, Hungary

Setting a Standard for the Exchange of Archaeological Data in the Netherlands

Philip VERHAGEN[1] – Chris SUEUR[2] – Milco WANSLEEBEN[3]

[1]ACVU-HBS
[2]Vestigia BV
[3]Faculty of Archaeology, Leiden University
[1]jwhp.verhagen@let.vu.nl
[2]c.sueur@burodebrug.nl
[3]m.wansleeben@arch.leidenuniv.nl

Abstract

The introduction and growth of a commercial market for archaeology has enormously increased the amount of archaeological fieldwork done in the Netherlands. This is combined with an increasing use of digital techniques to record, store and analyse excavation and survey data. The result has been a proliferation of data formats: the various companies doing archaeological fieldwork all have developed their own databases and GIS/CAD-systems for daily use. Because of this, a national metadata standard for describing archaeological data storage was introduced in 2007. However, this standard does not yet solve the problems of data exchange between archaeological companies, heritage managers and non-archaeological parties. In this paper, we will sketch the potential of exchange standards for three main categories of data: borehole data, the national sites and monuments records, and finds that are submitted for storage in repositories.

Keywords

archaeological heritage management, archaeological fieldwork, data exchange, data standards, XML

1. Introduction

In September 2007, the Dutch parliament approved the revision of the Law on Ancient Monuments of 1988. With this step, the Valletta Convention is finally legally embedded in archaeological heritage management in the Netherlands. In practice however, 'Malta' has already been in place since 1998, and over the past ten years we have seen a transition from a government-regulated to a market-based system. Municipal authorities are now responsible for dealing with archaeology in spatial planning, and developers have to pay for archaeological research. With the establishment and ensuing growth of commercial archaeology, a need was recognized for standardizing fieldwork and reporting, as a means to guarantee a minimum quality of archaeological research. This was approached through self-regulation, and therefore the archaeological community itself established a system of quality norms (Kwaliteitsnorm Nederlandse Archeologie) in 2001. In 2005, it was decided that archaeological quality norms should be incorporated within the larger framework of soil management. This is why they are now maintained by SIKB[1], a non-profit organization that aims at harmonizing and improving procedures and technology in soil management. This is done by providing well-defined quality definitions and issuing certifications for good practice. Until recently however, standards for dealing with digital data in archaeology were absent. A major revision of the existing quality norms in June 2007, however, led to the inclusion of a generic metadata model for digital excavation data (SIKB 2007), that allows data collectors' freedom in designing their own database structures, while maintaining minimum quality norms for the way in which archaeological data are registered and stored. This model describes the various 'building blocks' to be used for both digital and paper documentation - such as lists of features and finds, and associated drawings - as well as the minimum quality standards that these data sources should possess. Given that the best way to describe and register archaeological features and finds depends on the type of archaeological site and research questions involved, it was thought impossible to impose a rigid standard (with prescribed data structures and code lists) for dealing with excavation data.

[1] Stichting Infrastructuur Kwaliteitsborging Bodembeheer; www.sikb.nl

This generic metadata model is generally seen as a major step forward in dealing with digital excavation data, especially since it was combined with the establishment of an electronic repository that will allow archaeologists to access all digitally collected data[2]. However, it does not solve the problem of exchange of digital data. In practice, getting to use someone else's data implies mapping the fields and codes used to one's own system. This is a time-consuming exercise, that has to be repeated each time a different data set is imported. And even then, data that cannot be adequately translated may be lost in the process. This is why SIKB decided to investigate this issue, especially since similar problems had been observed and tackled in soil management.

2. The need for archaeological data exchange

If we want to identify the digital data sources that need to be exchanged, we have to take a look at the process of archaeological heritage management itself. In the Netherlands, archaeological research usually starts with desktop study, and then moves up to reconnaissance survey (core sampling and/or field walking), trial trenching and finally excavation. The research in these phases is not necessarily carried out by the same parties, so within this workflow, data may have to be exchanged between various companies. The authorities and initiators are usually not very much interested in receiving and storing digital data during this whole process. However, the archaeological community itself has established three 'outlets' for digital data at the end of each phase:

- the national sites and monuments record ARCHIS2, that maintains an overview of all archaeological research and reported finds;
- the finds repositories that have the obligation to curate the physical objects found during fieldwork;
- the newly established electronic repository, that will store all digital documentation created during the fieldwork.

Obviously, the goal of archiving this documentation is to guarantee that future researchers will have a good overview of the available data, and getting the data out of the digital repositories should be as easy as possible. Given this objective, it is actually rather surprising to see that direct data exchange during the research phases is still a cumbersome exercise, and is not seen as a top priority among the creators and curators of digital data themselves.

SIKB therefore commissioned an inventory of the areas where direct digital data exchange would be beneficial and profitable. After consulting archaeological and other stakeholders, three application areas with an urgent need for digital exchange were identified: borehole data, the national sites and monuments records, and digital documentation for finds repositories. In these areas it also seemed relatively easy to achieve results, and as such they may also serve as showcases that will allow archaeologists to experience first-hand the benefits of digital exchange protocols. A fourth application area was identified in the exchange of European dendrochronological data. This project however has a much wider scope and is not discussed in this article.

3. Case 1: Borehole data

When confronted with SIKB's request for developing a structure and method for archaeological data exchange, we first had to establish the existing demand, and assess the potential of existing techniques and structures for data exchange. SIKB already had a model at hand for data exchange in soil management: the 0101-protocol. This protocol consists of an XML schema definition (XSD) that specifies how digital soil data can be exchanged (SIKB n.d.). It is, intentionally, a 'maximum standard': a very wide range of fields and value codes are defined. Parties wanting to use this protocol only have to map their own data structures to this XML-structure. Within soil management, it has proved highly successful, so much that it has even attracted international interest. Proprietary software makers can apply for a certification from SIKB when they have adapted their software to facilitate export to and import from the 0101-standard. Initially, SIKB sought to extend this protocol to also include archaeological core sampling data, as a quick and easy fix for the exchange of what is probably the largest body of archaeological digital data in the Netherlands. As archaeological core sampling is primarily done in the reconnaissance phase, data re-use is potentially rewarding in the later

[2] E-depot Nederlandse Archeologie; www.edna.nl, comparable to the Archaeology Data Service in the UK.

phases of trial trenching and excavation, but it can also be useful before starting a new reconnaissance in a neighbouring area. Exchange through XML is new within Dutch archaeology, but can be implemented relatively easily. It only involves a one-time investment for the data providers in setting up a conversion tool for their borehole management software.

However, when looking at the specifications of the 0101-protocol, it soon turned out that the overlap between the data collected for soil management purposes and for archaeological reconnaissance was minimal. Furthermore, the concept of certification for software makers did not strike a chord within the archaeological community. The Dutch market for archaeological software is too insignificant for such an initiative. So, a full integration of archaeological specifications in the 0101-protocol was not considered desirable by us and most of our colleagues. Instead, a somewhat unexpected demand emerged for the establishment of a digital repository for archaeological borehole data. This can easily be achieved by involving the curators of geological borehole data (TNO Bouw en Ondergrond), who already manage a very large database called DINO[3]. It implies integrating the archaeological borehole specifications with the geological ones. The overlap between these two categories is much larger, and in fact has already been partially achieved through the establishment of an archaeological borehole description standard (ASB) in 2005 (Bosch 2005). We demonstrated that XML exchange based on ASB is possible. A centralized storage of ASB-based XML-documents within DINO was preferred by most parties. Steps towards this goal are currently being undertaken.

4. Case 2: ARCHIS2

The Dutch national sites and monuments record ARCHIS[4] has been operational since 1991, and has become an indispensable source of information to archaeological companies, local, regional and national government and academic researchers. Since its second release in 2002 it not only contains information on registered finds and protected monuments, but also on registered research projects. In this way, it is also possible to see where archaeological research has, and equally importantly, has not resulted in any finds. ARCHIS2 is a web-GIS application with restricted access: only registered archaeologists can use it. Data input, which is obligatory for every archaeological research project, is done on-line, which becomes time-consuming with large amounts of data. The available data can be queried and viewed in the system, and reports can be printed in .pdf-format. Downloading of information through the web-interface is very limited, since only a selection of fields from specific tables are made available as a .csv-file. Even though the database architecture is based on Oracle Spatial, direct data exchange is only possible through an obsolete .dbf export format which may also be used for uploading large amounts of data.[5] Although ARCHIS2 allows a full download of all available data for the whole of the Netherlands, this is highly impracticable.

The main reason for the lack of attention to facilitating data exchange from and to ARCHIS2 has been insufficient funding. While the system itself was developed with substantial government support, its maintenance budget is very limited. The ARCHIS2 curators admit that they currently do not have the time and resources to develop an exchange protocol and implement it in the web-interface of the system. This is all the more frustrating, as there are no technical obstacles involved. We are dealing with a well established database standard, that is broadly accepted by the archaeological community, and can easily be translated into an XML-based protocol. The development of an XML exchange format also offers the potential of developing input validation tools, that are beneficial to the ARCHIS2 curators as well. At the moment they can only check a small portion of the data entered, and the amount of errors in the database is known to be substantial. A simple verification of the XML-files against a XSD scheme prior to uploading would already be an improvement.

5. Case 3: Finds repositories

In the Netherlands, archaeological finds are stored in a number of provincial and municipal repositories. Each excavator is obliged to deposit finds and documentation within two years after finishing fieldwork. These find repositories are independent

[3] www.dinoloket.nl

[4] maintained by the RACM (Rijksdienst voor Archeologie, Cultuurlandschap en Monumenten); www.racm.nl

[5] This was done for reasons of backward compatibility to the previous version of ARCHIS.

organisations that use different software platforms and maintain different standards for documentation. Data exchange with repositories is predominantly a one-way process: it is compulsory to deliver finds documentation to the repositories, but there is very little demand for digital data coming from the repositories. Over the past few years, archaeological companies have become increasingly worried about the lack of standardization in the process of deposition of finds and accompanying documentation. Companies working nationally are basically obliged to maintain different documentation protocols for different regions. The repositories themselves on the other hand are facing an enormous and increasing amount of finds coming in, and in general do not have the staff and resources to deal with it efficiently. A nationally accepted exchange protocol would be enormously helpful for this, but it will take somewhat more than converting an existing data standard into XML. First the repositories have to agree on a minimum common set of attributes, that is suited for their management applications. Secondly it implies that they will have to invest in the development of software for importing the XML data. This is probably not possible without the aid of the SIKB.

6. Conclusions

While it was our first intention to assess the (technical) potential of XML-based protocols for archaeological data exchange, it quickly became clear that this question was of minor importance. The data types we have looked into (borehole data, national sites and monuments records, and find registration in repositories) are relatively simple to translate into an XML-based protocol. Fine-tuning will be necessary at the level of the fields and value lists, but we are convinced that agreement on these issues can be achieved quickly by bringing the parties involved together. The main obstacles we have found are organisational. All curators and stakeholders involved agreed on the importance of developing exchange standards, and were willing to give it a try, but all were anxious about the financial aspects. Who will take responsibility for implementing and maintaining an exchange protocol? Obviously, SIKB can and will have to play an important role in this. This organisation can support the development of exchange protocols by providing funding, directing the development of the protocols and taking care that they are used and maintained over the long run. The management system developed for digital soil data can be used for archaeological data standards as well.

Furthermore, the relatively simple technical application of exchange protocols opens up the perspective of doing the same thing for more complex data types like excavation documentation and data analyzed by specialists, like ceramics, metalwork or botanical remains. We therefore hope that working, successful applications of exchange protocols will initiate a further harmonization of archaeological data in the Netherlands.

References

Bosch, Aleid J. (2005). *Archeologische Standaard Boorbeschrijvingsmethode, op basis van de Standaard Boor Beschrijvingsmethode versie 5.2*. Gouda: SIKB.

SIKB (n.d.). *Protocol for the exchange of soil data. SIKB 0101 ENG. Version 5.0.0 (English translation)* (http://www.sikb.nl/upload/documents/PRJ%2036/protocol%200101-version-5-0-0-ENG.pdf). Gouda: SIKB.

SIKB (2007). *Kwaliteitsnorm Nederlandse Archeologie versie 3.1. Bijlage I: Bouwstenen*. Gouda: SIKB.