

EUROPE'S ELECTRONIC INHERITANCE: THE ARENA PROJECT AND DIGITAL PRESERVATION IN EUROPEAN ARCHAEOLOGY

ABSTRACT

JONATHAN KENNY

ARCHAEOLOGY DATA SERVICE,
DEPARTMENT OF ARCHAEOLOGY, UNIVERSITY OF YORK

WILLIAM G. KILBRIDE

ARCHAEOLOGY DATA SERVICE,
DEPARTMENT OF ARCHAEOLOGY, UNIVERSITY OF YORK

As long ago as 1994 Torsten Madsen warned that as we step 'into the steep road of electronic information, we should worry about one thing only, how do we ensure that what is electronic information today remains a stable and permanent source of information tomorrow as well'. This paper will consider the progress, or otherwise, that we have made in ensuring the stability and permanence of digital archaeological archives. To consider this topic the paper concentrates on the work of the Archaeology Data Service in the United Kingdom and of the ARENA project on a European scale.

The ARENA project is hosting a day workshop on the practical process of digital preservation and considering some of the issues surrounding this work. The author of this paper will be taking part in the ARENA workshop and will then present a 'school report' on the current state of preservation work for digital archaeological archives.

INTRODUCTION

The vulnerability of data stored in digital formats has become apparent to an increasing number of archaeologists and archivists over the last ten years (Madsen 1994, Richards, Miller and Wise 1999, Richards 2002). Richards et al. (1999) illustrated awareness of the fragility of archaeological data: Although computers have been widely used in archaeology for the last 15 years or more, as recently as 1999, an extensive survey of professionals in the UK demonstrated that "little attention" was being paid to the long-term preservation of digital data.

The Archaeology Data Service (ADS) was founded in 1996 specifically with the task of providing a digital preservation service for archaeologists in UK universities (Richards 2002). That mission has included a range of research and development activities, as well as guidance and awareness raising activities (Gillings and Wise 1999, Bewley et al. 1999, Richards and Robinson 2000, Schmidt 2002, Eiteljorg et al. 2003, Fernie and Richards 2003). The work of the ADS is intended to support researchers based in the UK, wherever they chose to work. Consequently, extending expertise on preservation to partners outside the UK is important for the ADS's own mission, and it is consistent with the work of heritage agencies the world over. In 2002 an opportunity arose through the European Union through the Culture 2000 programme, to bring together a small network to research and extend expertise in digital preservation. The resulting ARENA (Archaeological Records of Europe: Networked Access) project is a path-finding initiative working on issues of data preservation and access (Kenny, Kilbride and Richards 2003). This paper summarises the work of the ARENA partners to bring forward unexpected themes and issues in digital preservation. The experience of the partners has raised issues that will become key areas of debate in the preservation of digital archives in archaeology. These issues include data quality, migration, data from "live" projects, blurring interface and preservation, possibilities for e-publication, digitisation and indexing, and advocacy. It also raises issues about the boundaries between Europe's diverse heritage agencies.

VALUE: WHY DO WE PRESERVE ARCHIVES?

The need to create an archive containing the recorded activity of an archaeological excavation is treated as a 'given' by most archaeologists. After all we work in a field that, unlike many sciences, destroys the evidence and doesn't allow for exact experiment replication. But in themselves there is no value to an archive, the value comes from its use. The first use to which an archive is put is to substantiate the interpretive claims made by the archaeologists who carried out an excavation. This has led to a close relationship between publication and archive (Richards 2002:354-357). As Richards points out this has contributed to a publication 'back log' problem, as authors attempt to integrate huge archives into publications (Richards 2002).

The problems of archaeology are unusual, but preservation is a characteristic of all viable scholarship. Research values in all disciplines dictate that scholars hand on their results to future generations that may use and refine their own work. If scholarship moves to digital forms - and the growth of e-journals, e-science and e-publishing shows that some disciplines are already very advanced in this respect (RSLG 2003) - then so there is a pressing need for that information to be preserved and transmitted in digital form too. It is the archives that keep archaeology a dynamic research pursuit as well as a development driven activity.

Richards (2002) has recently reemphasised the opportunity that is presented by the computer age to expand the value of archaeological archives, by making them accessible and searchable to a wide audience. Partners in the ARENA project have each recently released a set of archives (available online at <http://ads.ahds.ac.uk/arena/>), which will be useful for researchers in each of the partner countries - but which are also research in their own right. The most striking feature is that there is not a single model for archaeological archives. These archives presented through ARENA fall into a number of categories, representing in part the nature of the organisations that presented them. Six broad categories can be identified:

- Archaeological excavation archives. The ADS, the Danish National Agency for Cultural Heritage and the Institute for Archaeology (the FSI) in Iceland all made archives from excavations available. The archives were Danebury and Cottam in the UK (ADS), Tarraconensis landscape survey in Spain (ADS), Vorbasse, Dankirke and Helme in Denmark (The Danish National Agency for Cultural Heritage) and Hofstaðir in Iceland (FSI).
- Antiquarian archives. The Institute for Cultural Memory (cIMeC) in Romania is making available an extensive digitised archive the Archaeological Repertory of Romania (RAR).
- Catalogues. cIMeC has also made available an online catalogue, the Chronicle of Archaeological Research in Romania
- On Line Publication. The partners from Poznan Archaeological Museum in Poland have made available a summary version of the paper publication of excavations at Kowalewko.
- Historical Archives. The Poznan Archaeological Museum is also hoping to make available a set of images recording the excavations at Biskupin in the 1930s
- Landscape archives. The Norwegian partners at the Museums Project have generated a resource holding all of the documentary records of discovery and archaeological intervention for Norway broken down by individual farmsteads. This resource is being presented in relation to the archaeological landscape, allowing researchers access to documents for each farmstead in the landscape arranged in date order. For the purposes of ARENA the resource is being pioneered using the Egge and Hegge landscape.

It is clear from the archives above that the opportunities highlighted by Richards (2002) are broad and exciting. The simple, but vital, value of archaeological archives is vested in them by their use, not by their simple existence. Despite this however, archives must be preserved before they can be used. The experiences of the ARENA partners, gained in making the above archives available illustrate some of the issues involved in preservation.

DATA MIGRATION

The Danish National Agency for Cultural Heritage has worked on three archaeological sites that have run between 1970 and 2000. At all three sites (Vorbasse, Helme and Dankirke), there was no digital recording on site. The digitising of the records has been office based and has taken place over a long period of time. It soon became clear that there were data structure and quality issues as well as a variety of data formats. The process of making data available in useful formats was very time consuming, ensuring that open formats were used and that these were in the a format that was likely to be readable by users. The approach to making data available was two fold; one that the data did not need to be supplemented to make it useable and two that the format issues were best dealt with by migration. The latter is a key debate in preservation activity, migration or emulation. For all of the ARENA partners faced with this issue, migration has been the preferred option, it is easier to achieve and far more convenient for the user.

ONGOING PROJECTS AND "LIVE" DATA

For FSI in Iceland, working with a project that was still "live" posed a problem but also an opportunity. The development of the Hofstaðir archive has been interesting because excavation is continuing. New approaches to digital archives, including policy for preservation, have to feed into working practices. The archive covers many seasons of excavation raising a number of problems; there is a mix of materials in varied formats and documentation is fragmented and inconsistent. Because it is "live" the archive is constantly changing and lastly the archive needs to be connected to the post excavation process. The answers to these problems lie in the management of the archive, once more a time and resource consuming activity, ensuring the application of data standards and quality assurance. The importance of the management role on a "live" project is emphasised by the Hofstaðir experience. The FSI experience highlighted the need to remember the user when preserving data; but it is difficult to know what the user will want.

BLURRING THE LINE BETWEEN INTERFACE AND ARCHIVE: E-PUBLICATION POSSIBILITIES

The experiences of FSI and the Danish National Agency for Cultural Heritage connect preservation issues with user needs. As the Danish partners found, there was a need to decide on how to present archives. The decision to make available archives that are preserved and migrated onto useable formats, backed up by brief descriptions of the site, leave the use of the archive to the user. But attention must still be given to the user; they need some presentation of the site to decide what use they may make of the data. This can be provided online, either as a brief summary or as a more extensive online publication.

In the case of the ADS archive preparation the interface became a key issue for the preservation of the archive. The Cottam archive gives a brief overview of the project and makes preserved data archives downloadable for reuse, the basic service, properly preserved. The digital archive from the Ager Tarraconensis field survey project was created in the late 80s. It was used to create the paper publication and was then left on disc for storage. As it transpired the map data could not be migrated into a useable format. Without the maps the rest of the data made little sense, its reuse value was lost. Simply making available the data files that were migratable would not make sense to the user, even though the data might be properly preserved. A user interface was required to demonstrate the potential of the archive. This was achieved by first redigitising the map data from the paper publication. Second the scanned maps were enhanced using simple html to show the relevant distributions and spatial relationships. It is then possible to consult the data online through the maps, giving a browsable archive that demonstrates its content. From here the user can then decide to download the preserved data, including digitised maps, for their own use. This approach, inspired by necessity, blurs the line between simple preservation and developing the user interface. By serendipity the ADS experience suggested an answer to the issues

Cultural Heritage Management

raised by Aldred at the FSI, making the potential of the archive clearer to the user. Enhancement is particularly pertinent for archives that relate to a paper or electronic publication, allowing for researchers to download and work with data whilst also allowing publications to be made shorter or interactive and more accessible. The third archive presented by the ADS, the Danebury project, archive highlights this association between preservation and publication enhancement still further. The archive is linked, through simple html, to online copies of four of the five Danebury publications.

DIGITISATION, PRESERVATION AND ARCHIVE ENHANCEMENT.

The same blurring between preservation and user needs was illustrated by the experience of the Romanian partner. CIMEC found that the preservation and presentation of the catalogue of the Chronicle of Archaeological Research was relatively straight forward as there was relatively little digital data involved. The digitisation of the Archaeological Repertory of Romania (RAR) raised more issues however. To preserve the scanned documents there was also a need for documentation and indexing. It is not enough to ensure that the images are in a preservation format; they must have relevant documentation that makes sense of the content and indexing. Because this is a historic archive many of the objects or sites recorded must be identified in contemporary terms to give them a value to the user. The documentation and indexing of the digital archive in a database brings very real value to the archive, ensuring its public use and thus its preservation. It also raises an issue of investment, how deep do we go in enhancing the archive? Costs and time are the delimiters, as well as the need to define at which point to leave the use of the archive up to the user.

The archives used by the Norwegian partners at the Museums Project are documents that have been scanned, enhanced through OCR and extensively indexed using XML. Preservation work has been carried out at the same time as enhancement is added through detailed indexing. The Museum's project is taking the enhancement of the user interface still further by relating particular documents to specific landscapes. The intention here is to provide a kind of "glue" for more users to make sense of the archives in a spatial setting.

ADVOCACY

One of the fundamental points that applies to those working with preserving digital archives is the degree of awareness amongst archaeologists themselves. In Poland, partners have found that the adoption of information technology has been a slow process. The use of IT and thus the value of IT has been the preserve of a few experts. They are aware of the preservation needs of data, but their advocacy role remains important. This is illustrated by the presentation of the Kowalewko data. This important archive has very few digital components, leading to the publication of a digital version of the paper publication. This raises a question for all who attend CAA, how far has the need for digital preservation facilities been recognised? Is the value of digital data and its concomitant need for preservation still only recognised by a small expert group?

BREAKING DOWN BOUNDARIES

When looking at the experiences of the ARENA partners in the round, one may observe the breaking down of boundaries. An understanding of the place of standards in preserving digital data breaks down national boundaries, if standards are going to work for data preservation in a global sense they cannot afford to be localised. There are other boundaries to be broken down here to. The perception of the importance of preservation issues has to be spread throughout the archaeology community. This must be done hand in hand with the emphasis on the value of such archives, both as a record of an otherwise destroyed data set (the archaeology itself) and as a research tool.

CONCLUSION

Europe's digital inheritance is fragile and its longevity is not assured. All organisations, governments and individuals that decide to engage with digital technologies will have to develop strategies to maximise the lifespan of key digital assets. By providing an overview of the challenges faced and processes adopted by key cultural heritage agencies in Europe, the work of the ARENA partnership represents an important contribution to the preservation of digital data, and also provides a further opportunity to raise awareness of the problems which all will face.

REFERENCES

- BEWLEY, R., DONOGHUE, D., GAFFNEY, V. VAN LEUSEN, M. and WISE, A., 1999. Archiving Aerial Photography and Remote Sensing Data: A Guide to Good Practice. AHDS Guides to Good Practice Series, Oxbow (<http://ads.ahds.ac.uk/project/goodguides/apandrs/>).
- EITELJORG, H., FERNIE, K., HUGGETT, J. and ROBINSON, D., 2003. CAD: A Guide to Good Practice. AHDS Guide to Good Practice Series, Oxbow (<http://ads.ahds.ac.uk/project/goodguides/cad/>).
- FERNIE, K. and RICHARDS, J., 2003. Creating and Using Virtual Reality: a Guide for the Arts and Humanities. AHDS Guides to Good Practice Series, Oxbow (http://vads.ahds.ac.uk/guides/vr_guide/index.html).
- GILLINGS, M. and WISE, A. (eds.), 1999. GIS: A Guide to Good Practice. AHDS Guides to Good Practice Series, Oxbow (<http://ads.ahds.ac.uk/project/goodguides/gis/>).
- RSLG, 2003. Research Support Libraries Group Report to Higher Education Funding Council for England (<http://www.rslg.ac.uk/>).
- KENNY, J., KILBRIDE, W.G. and RICHARDS, J.D., 2003. Enter the ARENA: preservation and access for Europe's archaeological archives. In Doerr, M. and Sarris, A. (eds.), Computer Applications and Quantitative Methods in Archaeology 2002, Archive of Monuments and Publications, Hellenic Ministry of Culture:349-353.
- MADSEN, T., 1994. IT-based information processing in archaeology - a key to progress? Paper delivered at unpublished seminar on The Problems and Potentials of Electronic Information for Archaeology, British Academy, 1994, cited by Richards, J.D. et al., 1999 op cit.
- RICHARDS, J. D., MILLER, P. and WISE, A., 1999. Digital Archives in Archaeology. In Hansen, H. and Quine, G. (eds.), Our Fragile Heritage:123-130, Copenhagen: National Museum of Denmark.
- RICHARDS, J. and ROBINSON, D. (eds.), 2000. Digital Archives from Excavation and Fieldwork: Guide to Good Practice Second Edition. AHDS Guides to Good Practice Series, Oxbow (<http://ads.ahds.ac.uk/project/goodguides/excavation/>).
- RICHARDS, J.D., 2002. Digital preservation and access. In European Journal of Archaeology Vol. 5(3):343-366.
- SCHMIDT, A., 2002. Geophysical Data in Archaeology: A Guide to Good Practice. AHDS Guides to Good Practice Series, Oxbow (<http://ads.ahds.ac.uk/project/goodguides/geophys/>).