

---

# DIGITAL PRESERVATION AND ACCESS

Julian D. Richards

University of York, UK

*Abstract:* There is a crisis in the publication and archiving of archaeological field data in Europe. Computerized data are more fragile than paper archives but also more accessible via the Internet. This article explores the role of the Archaeology Data Service and examines some of the issues raised by the collection of digital data and their dissemination online and the implications for the future of archaeological publication. It discusses approaches to digital data preservation, the development of archival standards and ways of encouraging reuse. It explores the development of distributed online catalogues and archives and the need for metadata standards for cataloguing resources. Finally, it considers the role of XML as an emergent technology and introduces the European ARENA project which is developing a digital preservation and access infrastructure.

*Keywords:* archaeological information systems, digital preservation, heritage management, Internet, interoperability, metadata, online archives, resource discovery, XML, Z39.50

## INTRODUCTION

Archaeology has reached a publication crisis point in many European countries. Full publication is becoming increasingly expensive and difficult to achieve, and excavation monographs are read by few people, and bought by even fewer (Jones et al. 2001). Meanwhile museum archives are also reaching breaking point; most are running out of storage space; few can provide facilities for access and almost all report low levels of usage (Merriman and Swain 1999:259–60). Digital technology now offers the means by which the crisis may at last be overcome.

However, many archival records are also becoming increasingly fragile. Once upon a time the paper archive of site notebooks, context records, and plan and section drawings would be boxed up and put on a museum shelf. Should some future researcher need to consult them, then short of fire or flood, those records would still be available in 50 or 100 years: a little dustier, but still legible. From the late 1970s onwards a growing proportion of these records have been converted into a digital format: text files, databases, CAD files, and so on. Some types of data, such as geophysical surveys, were always digital from the outset; many field-workers are now collecting other forms of data straight into a digital format, including context descriptions, site photographs, and layer and feature

coordinates. Moreover, while some data can be printed onto paper or microfilm as a security copy, much now depend upon their digital form for their very meaning. The functionality of a GIS, or a 3D virtual model, cannot be replicated on hard copy.

Such digital data cannot just be left on the shelf if there is to be any hope of ever being able to use them again. Digital data require active curation. The computer diskette is particularly vulnerable to changes in temperature, dust or magnetism. It is also a specific storage medium that requires a particular hardware device to read it. The CD-ROM may provide a more durable means of preserving a particular sequence of binary digits, but contrary to popular belief, once the drive has been rendered redundant by the next upgrade in storage technology it will be no more secure than a 5<sup>1</sup>/<sub>4</sub> inch or an 8 inch floppy disc or a punched card or even paper tape. Furthermore the data on it may require a specific version of a software program in order to extract it, and the data held within the application may require knowledge of specific codes in order to comprehend it. Unless care is taken of each of these elements then it is likely that existing digital data will be useless within five years or less.

The computer age therefore provides archaeology with both a crisis and an opportunity. A crisis because, unless we are vigilant, our research data will be more vulnerable than they ever were had we left them in the ground. An opportunity because, by use of the computer and the Internet, it is now possible to make archives accessible and at last to enable reuse and reinterpretation of excavation archives.

In the UK the Archaeology Data Service (ADS) has taken a lead role in the preservation and dissemination of digital data. This article outlines the development of the ADS over the last five years and considers the potential impact of its activities on the research and management of archaeology. It discusses the issues associated with the archiving, access and retrieval of archaeological data and suggests some ways in which the provision of online archives may develop in the future.

## THE ARCHAEOLOGY DATA SERVICE

The ADS (<http://ads.ahds.ac.uk>) was established on 1 October 1996 with the mission to preserve, catalogue, and describe digital data generated in the course of archaeological research and to facilitate its reuse (Richards et al. 1997). These activities are inseparable as unless digital data are actively curated they will not be available to future scholars, and unless researchers are going to reuse data there is little point in expending effort attempting to preserve them. Its geographical remit was wide, covering all archaeological research conducted by those based within the UK.

### *Structure and organization*

Core funding for the ADS initially came from the Joint Information Systems

Committee (JISC), the body charged with developing information systems for UK Higher Education. From 1999 additional funding was provided by the Arts and Humanities Research Board, the principle funding body for arts and humanities research within the UK higher education sector. This reflected an interest, inherited from the British Academy, in ensuring that the archival future of the results of research funding was secure, and a realization that it is inefficient to fund fresh data collection activities when vital digital research data might already have been collected elsewhere.

The ADS was set up as part of the Arts and Humanities Data Service (AHDS). It was one of five disciplinary-defined distributed service providers, managed by a central Executive (Burnard and Short 1994). Two of the other service providers were already established in another guise, comprising the Oxford Text Archive, based in Oxford University Computing Service, and the History Data Service, based within the well-established social sciences Data Archive at Essex University. The remaining two services were also newly formed, comprising the Performing Arts Data Service in the University of Glasgow, and the Visual Arts Data Service at the Surrey Institute of Art and Design.

The assumption underlying the provision of disciplinary specific services is that subject expertise is required at every stage, from the creation of a data set, through to its preservation, and assisting with reuse. Validation of data and documentation requires knowledge of the research value of the data. This approach has so far proved correct and has also enabled the ADS to establish strong links within Archaeology which have encouraged potential depositors.

The bid to establish the ADS was developed by a consortium of university departments of Archaeology, acting together with the Council for British Archaeology (Richards et al. 1999:127–129; Wise and Richards 1999:138–139). Consortium members have retained an active involvement in the management of the ADS, and in activities such as the production of policies and guidelines.

The ADS is also distinctive in other ways. Archaeology stands out amongst the humanities for the amount of research which takes place outside of an educational establishment. Indeed, within the UK most data collection takes place within a commercial environment, with monitoring from central or local government. On the other hand there is tremendous potential demand for access to primary data from university students and researchers, and several observers had also expressed concerns that the results of current fieldwork are not being fed back into the research cycle for use by those writing the text books for the next generation. It was recognized from the outset that the ADS could not operate solely within the academic sector and that to be effective it would need to establish a niche within the complex network of existing public and private organizations operating within the historic environment sector. This was further complicated by the different systems operating within the different constituent parts of the United Kingdom. In Scotland, for instance, the National Monuments Record is maintained by the Royal Commission for Ancient and Historical Monuments of Scotland (RCAHMS) and has a statutory archival role for all archaeological interventions in Scotland (Historic Scotland 1996) although its coverage may be incomplete for private

ventures such as university-based projects. In England the former Royal Commission for Historical Monuments of England (RCHME) was merged with English Heritage in 1999 and became the English Heritage National Monuments Record. The former RCHME policy was to provide a selective archive for the paper records of projects deemed to be of national importance. At a local level the archives, including both the finds and the paper, and any digital records, were expected to go to local museums but outside major institutions such as the Museum of London there was no provision for curating digital data. In both Scotland and England the regional or county Sites and Monuments Records (SMRs), each with their own archival policies, added an additional tier of regulation, but very few explicitly referred to provision for digital data in the specifications for archaeological work which they drew up. In Wales, the Royal Commission for Ancient and Historical Monuments of Wales (RCAHMW) maintains an index level record to sites and monuments, but most archives have been retained by the four Welsh trusts who run the regional SMRs and also carry out the majority of the fieldwork. In Northern Ireland the Department of the Environment, through the Monuments and Buildings Record, licences all excavation and requires archival deposition. However, it should also be noted that irrespective of policy, when the ADS was established no National Monuments Record had archived a single digital data set, and although some were aware of the problem in some cases this had simply led to such data being turned away. From the beginning the structure of the ADS needed to reflect this complex landscape and an Advisory Committee was established, representing all the major stakeholders.

Inevitably there were some concerns from existing bodies about what role this new player would take and some initial energies were directed towards allaying suspicions and emphasizing that where there were existing bodies the role of the ADS would be to facilitate access to their collections, rather than duplicating activities. Fortunately technical developments, such as interoperability (see below), have set the trend for moves away from monolithic centralized systems and towards the distributed model, as promoted by the AHDS, and by the ADS particularly in the HEIRNET report (Baker et al. 1999). Under this model the emphasis is upon targeting specific user groups and while public bodies have been under increasing pressure to make their resources widely accessible to an audience drawn primarily from schools and the general public, the ADS has focused on the academic community, while recognizing that the professional archaeological community will also benefit from resources provided for those with a high degree of archaeological knowledge. Thus the ADS is able to provide a 'shop window' into the higher education sector for those sectors that were becoming divorced from them. During the late 1990s the JISC was developing the model of the Distributed National Electronic Resource (DNER), a loose confederation of distributed services and archives linked by adherence to common standards and protocols. In reflecting the structure of archaeological research the ADS has found itself in the special position of providing a bridge between the DNER and external networks, such as those for schools, libraries and museums.

The ADS was originally established with two members of staff but at the time of writing project funding has enabled that to be increased to eight. The staffing structure reflects the three main areas of activity: Collections Development, negotiating access to data and assisting data creators; a Curatorial section, maintaining and developing the ADS delivery mechanisms, and looking after the data once it has been deposited, including validation of data and documentation and where necessary preparing it for long term preservation; and lastly User Services, encouraging and assisting the reuse of digital data.

### *Collections policy*

For those working within a traditional academic environment the majority of funding bodies either recommend or in most cases require those creating digital data which results from their awards to be offered to the AHDS for long-term preservation. For those undertaking archaeological research that includes the British Academy, the Leverhulme Trust, the Carnegie Trust, the Wellcome Institute, the Council for British Archaeology and the Arts and Humanities Research Board. For science-based archaeology the Natural Environment Research Council has also designated ADS as the approved data archive. In addition, the Society of Antiquaries of London requires its grant-holders to offer digital research data to the ADS, unless this is precluded by specific archival requirements abroad. In practice we are aware of no other countries that have national provision for digital archiving, although the Archaeological Data Archive Project in the United States has promoted the importance of digital archiving for many years (Eiteljorg II 1995). Agreements with research funding bodies are necessary to raise the profile of digital archiving within the academic community. Surveys of Leverhulme and British Academy grant-holders (Austin 1998) revealed significant numbers of digital research products without a secure archival future. Nonetheless the academic community has proved slow to deposit data. In part this reflects the long lead-in time for research projects, and the need to plan digital archiving from the outset. However, it also reflects the low academic priority given to the creation of an ordered and well-documented archive, and the length of time required for this. Academic priorities in the UK are determined by the Research Assessment Exercise (RAE) and peer-reviewed publication is seen as the ultimate goal of every project. There is a need to accord a similar status to archives, which certainly should become more straightforward as the distinction between archive and publication becomes blurred (see below), but may require the academic refereeing of archives.

The ADS is not required to accept all the data offered to it, and some form of selectivity is appropriate. Preservation carries a cost and it is important to establish the reuse potential of any data set before expending resources on its preservation. Indeed, for some files with little reuse value hardcopy archiving with the option of rescanning, should the demand arise, may provide the most cost-effective means of archiving. The ADS Collections Policy focuses on the quality of the data, the completeness of the documentation and the reuse potential. There has also been an

effort to develop strengths in particular thematic areas, such as databases of archaeological dates.

For the commercial and public historic environment sectors there is the possibility of clear requirements to deposit the digital archive, enforced within the planning process. The curatorial sector hold the key to policing this as through their specifications for work county archaeologists can require digital data deposit. Whereas the costs of preservation for the academic sector are covered by core funding, the ADS has taken the view that commercial depositors should be required to pay a one-off deposit charge at the time of deposition. This is analogous to the box storage charge levied by museums. The ADS has therefore developed a charging policy which declares explicit per file charges, and enables contractors to budget for the creation of a digital archive. In practice it is estimated that digital archiving costs are between 1–3% of total project budget, decreasing for larger projects where there are economies of scale for large numbers of identical file formats. This additional project cost may also be reduced by savings in the publication budget as it may be more appropriate to make detailed data available via the digital archive rather than in print (see below). Significantly, for some large developer-funded projects the prospect of making the academic results of the project available by digital dissemination is very attractive. Major projects such as the Channel Tunnel Rail Link and excavations at Heathrow and Stansted conducted for the British Airports Authority by the Framework Archaeology Consortium have looked to digital publication as one of the main means of dissemination. For many commercial companies a project website may offer a more widely accessible publication and better potential for good public relations than a traditional academic monograph.

For most archaeological curators and contractors working within England the national guidelines and policies of English Heritage remain influential in determining local policies. The procedure known as MAP2 recommended in the second edition of the *Management of Archaeological Projects* (English Heritage 1991) is applied to many projects, whether or not they have direct English Heritage funding, and it is therefore significant that English Heritage has decided to re-issue its guidelines to require digital deposit. Members of the Association of Local Government Archaeological Officers may develop similar guidelines for projects they oversee, recommending deposit with a recognized digital archive. As noted above, similar policies already exist at national level in Scotland, Wales and Northern Ireland, although they may not make the digital component of the archive explicit and the policy may not always be rigorously enforced. As such approaches gain national acceptance it will make life easier for contractors who may have to operate within the jurisdiction of several authorities. It will also be necessary to develop a system of designation of approved digital archives, comparable to that which exists for registered museums (Condrón et al. 1999:Recommendation 9, 81–82). It is certainly clear that this is a rapidly changing field and that the ADS has had a major catalytic role in pushing forward policy development.

Collections Development also requires the provision of help for data creators

and potential depositors. As well as one-to-one guidance by telephone, personal visits, or email, the ADS has published Guidelines for Depositors including details of recommended file formats. It has also contributed to the AHDS Series of Guides to Good Practice with the publication of Guides for specific data types relevant to archaeological research. To date these comprise:

- GIS
- Aerial photography and remote sensing
- Excavation and fieldwork
- Geophysical data
- CAD

Further guides covering virtual reality, database design and scientific methods are at earlier stages of production. The Guides do not seek to be proscriptive about recording methods or how data should be created. Instead they make recommendations about how that data should be documented if it is to be preserved and reuse is to be facilitated. There has been interest in translating the Guides and creating local versions in other European languages and these will be linked from the ADS website as they become available.

Finally, it was necessary to develop a framework for rights management across the AHDS, to protect the legal rights of the ADS and of depositors. Depositors are expected to sign a deposit licence and to declare their copyright in the data collection, providing certain assurances that it does not contravene laws of obscenity and so forth. Indeed, they retain the copyright in their data and simply grant the ADS a non-exclusive right to distribute the data to third parties for research purposes. In return the ADS undertakes to use its best endeavours to ensure the long-term preservation of the data and to require users to observe certain conditions of use. These are set down in the common access agreement which requires users to observe the copyright in the data, not to pass on the data to others, and to acknowledge the data creator. Anyone is permitted to use data held by the ADS so long as it is for research or educational purposes, and these are defined quite broadly as purposes intended to develop knowledge and where the research output is itself destined for the public domain. Therefore reuse of data held by ADS by commercial contractors is not prevented so long as publication of their work is not limited by issues of client confidentiality.

As well as encouraging the permanent deposition of complete primary fieldwork archives the ADS also acts as an information broker for other information providers, endeavouring to provide a one-stop-shop for information about the archaeology of the UK. The ADS online catalogue, ArchSearch, contains some 400,000 index records to sites and monuments of the British Isles. This includes fields drawn from the National Monuments Record for Scotland, the Excavation Index for England, and a number of Sites and Monuments Records, including those for West of Scotland, Northumberland, Greater London, Clywd Powys, South Gloucestershire and the National Trust, as well as the York and London Archive Gazetteers. These provide basic metadata records, mapping from the local databases to a number of fields derived from the Dublin Core standard,

an internationally agreed set of 15 elements that provide the basic index fields necessary to allow users to locate and assess information resources (Miller 1996; 1999; Wise and Miller 1997). They include site name and brief descriptions, geographical coordinates, period and subject keywords, bibliographic references, and rights management details. The system was developed through the Accessing Scotland's Past project in 1998, which sought to map the Scottish National Monuments Record and a number of regional Scottish SMRs to the Dublin Core. In 1999–2000 records for England were enhanced through the OASIS project, which provided a single concorded database between the former RCHME Excavation Index and the English Heritage and Bournemouth University Archaeological Investigations Project, thereby providing a high level index to the mass of grey literature generated by developer-funded contract archaeology (Hardman and Richards forthcoming).

The records are designed to act as tools for resource discovery, and provide users with details of how to get more information if appropriate. In some cases users are able to follow a live hyperlink, for example from an index record from the Scottish National Monuments Record to the full live record in the RCAHMS online database, CANMORE. Where the ADS holds a digital archive for a site then users can also drill down within ArchSearch from the index record to richer online resources. Thus the brokered index records provide a high level backbone to the ADS catalogue such that users should find some information, no matter which site they are interested in, and as more digital archives are made available they may be able to access much more detailed information, down to complete site records.

#### *Curatorial activities*

Once data has been deposited with the ADS then it must be processed to ensure its long term preservation. As a first stage it is copied to a Unix file server attached to the University of York campus network. The file server is backed-up according to a standard 'grandfather–father–son' routine (cf. Fernie and Gilman 2000:B.33). If the data are not already held in recommended open file formats they may require migration, and it is also necessary to check that appropriate metadata have been recorded. As well as metadata for resource discovery this will also include more detailed information to enable reuse (such as how the data were collected, and any codes used). All files are logged in the Collections Management database which includes information about formats and version numbers. In some cases it is necessary to take a copy of the file in a compressed format for data delivery over the Internet, and a second copy in an expanded preservation format. In the case of images, for example, an uncompressed TIFF file is held offline, whilst a lower resolution JPEG file is made available for download online.

There are a number of strategies for digital preservation (Beagrie and Jones 2002; Hendley 1998; Ross 2000; Russell 2000). The three main ones are hardware preservation, hardware emulation and migration. Hardware preservation requires the maintenance of original hardware in order to keep software applications running on it. Clearly this can be expensive and require a high level of

technological expertise. As time goes on costs increase as more and more antiquated machines have to be preserved and it is really only a solution of last resort. Hardware emulation also tries to keep old versions of software applications in running order, but it does this by emulating old operating systems on new computers. Of course, as operating systems constantly develop it becomes necessary to have emulations running within emulations and the whole business can become extremely complex. The ADS therefore favours the third strategy of migration. This approach relies upon the assumption that it is the information content rather than the look and feel of a particular application that is important. In the case of archaeological data we feel this is justifiable, although it may not be appropriate for all disciplines. Where possible, data are converted to open file formats, such as comma delimited ASCII files. For other datatypes, such as CAD for example, they are converted to standard exchange formats, such as DXF files, and will require migration to new versions as formats develop. This strategy requires most investment of labour at the point of deposition and it is the expectation that the bulk copying of files to new versions will be relatively easy to automate, although a sampling strategy for validation of files will be essential. As part of migration it is also necessary to ensure regular backup and refreshment of the physical storage media. The AHDS is also developing a central facility for the deep storage offline of large quantities of digital data in preservation formats.

A user survey conducted in 1998 revealed a very low level of awareness of good digital archiving practice within archaeology (Condrón et al. 1999: 33–9). Many organizations were holding digital data, but 47% had not adopted any means of protecting the physical media (Condrón et al. 1999: Fig. 6.11). The ADS gained first-hand experience in data archaeology through work on the Newham Museum Archive (Kilbride 2000; Austin et al. 2001). When Newham Museum Archaeological Service was closed the digital data collected over the last 10 years were hurriedly dispatched to the ADS, where they were catalogued, and accessioned. The archive arrived on 220 floppy disks, containing some 6432 individual files. About 5% of the total was already corrupted by the time it arrived in York. Of the remainder, 1500 files contained site reports, or elements of site reports. There were well over 700 database files and 1200 geophysics files. Each one of these had to be recorded in turn and converted from the original proprietary format into formats recommended for long-term preservation. Some 900 files were held in formats that are unidentifiable and thus remain unreadable. However, the main problems were caused not by degradation of media or obsolete file formats, but by inadequate documentation. Thus, there are various catalogues of small finds, which though consistent and apparently correct, contain no indication of the excavation to which they relate. Given that over 150 separate excavations are represented, this is obviously problematic, making them more or less useless. In another case, a large cemetery had been recorded in great detail. Each bone had been recorded with a descriptive code, but there was no means of expanding the codes, so the thousands of records generated are worthless. From a cemetery with several hundred burials, only one patella can be identified with any certainty, surviving because it was referred to in a free-text field.

Newham is certainly not exceptional, and had the foresighted Newham archaeologists not recognized the value of their data when the Museum closed down, then all this would certainly have been lost for good. How many similar boxes of floppy discs live on the shelves of museums and contracting units? *Strategies for Digital Data* suggests there are thousands (Condron et al. 1999: Fig. 6.3).

### *Collections reuse*

There is little point in preservation without reuse. Therefore the third strand of ADS activity has focused on facilitating access and increasing dissemination.

From its inception the AHDS has taken a lead role in the promotion of metadata standards for the cataloguing of digital resources. A series of discipline-specific workshops led to the publication of a report (Miller and Greenstein 1997) which recommended the application of the 15-element Dublin Core standard (<http://uk.dublincore.org/>) as a cross-disciplinary means of resource discovery. The Dublin Core elements have been implemented as the key fields in the ADS online catalogue ArchSearch. In principle the implementation of these elements should allow users to search by resource title, creator, subject keywords, period or location (both expressed as sub-elements of the Dublin Core element 'coverage'). In practice cross-searching of distinct resources is hindered by the lack of adherence to common standards for vocabulary control. Thus, for example, archaeological period is described according to different classifications in each of the major resources indexed in ArchSearch. While the ability to accommodate different schemes for resource description was one of the attractions for the adoption of the Dublin Core, it will be necessary to develop the use of online thesauri to enable effective cross-searching, allowing the user to equate 'Norse' in Scotland with 'Viking' in England with 'Anglo-Scandinavian' in Yorkshire, for example.

In the meantime it has been crucial for the ADS to develop a search interface that allows the user to refine their query and identify the precise resources required. This work is ongoing but from April 2001 it became possible to support clickable map-based searching at a scale of 1:625,000 to allow the user to identify resources within a 1km, 5km or 10km square centred upon a given point. The provision of geospatial browsing is a critical area for development within the Distributed National Electronic Resource (DNER) as it allows users to integrate resources of various types, combined according to their geospatial references. Currently the ADS is developing tools to allow the display of hits overlain on a map-base.

Interoperability, or the ability to combine resources from distributed databases in a single search, has been a central tenet of the AHDS since its foundation. The online catalogues of the five service providers were linked in a single web 'gateway' which used the Z39.50 communications protocol to allow a single query to be addressed simultaneously to five 'target' databases, each structured in a different way using a different database management system. In response to a

query for holdings relating to Shakespeare, for example, the interdisciplinary humanities scholar should recover an electronic text of the Complete Works from the Oxford Text Archive, a video of the Royal Shakespeare Company performance of King Lear from the Performing Arts Data Service, an unattributed portrait from the Visual Arts Data Service, an historical database of sixteenth-century London from the History Data Service and the excavation archive for the Rose Theatre from the ADS.

The DNER architecture envisages a number of Z39.50 enabled gateways and portals providing access to distributed resources. The ADS has worked with other UK partners to develop the first Z39.50 portal for the Historic Environment, with initial targets including ArchSearch, the RCAHMS's CANMORE, SCRAN and the Portable Antiquities Database, under the title of HEIRPORT (Austin et al. 2002). This recognises the fact that ArchSearch will never include an index record for every resource and therefore the ADS needs to develop a means of providing integrated access to distributed resources. Indeed, many of the online indexes to sites and monuments are dynamic and records are constantly being added to national and regional monuments records. Therefore it is appropriate for each resource to be live on the server of the resource maintainer and for each to be made available online as a target to any number of gateways. It is likely that gateways will develop intended for specific user groups. These groups may be defined by prior knowledge and user needs, such as the academic sector, or the schools sector; by disciplinary area: archaeology, history, arts and humanities in general; or by user interface: map-based searching, text searching etc. In this vision for the future information landscape each information resource needs to be presented only once in order to be available from multiple 'shop windows' (Baker et al. 1999:Section 5).

Since its launch on 15 September 1998, the ADS catalogue has gained very high levels of usage. From April 1999–March 2000, there were over 218,000 catalogue requests; from April 2001–March 2002 this had grown to 2,317,000 requests. Analysis of the access statistics can tell us a lot about the profile of ADS users, although it cannot tell us the purpose of their research (Kilbride and Winters 2001). They do indicate that substantial numbers of researchers within the UK and overseas are downloading large numbers of files of data from the online archives. Nonetheless, if the next generation of UK archaeologists are to be trained to use online archives it is essential that their use is integrated within the Higher Education curriculum. The ADS established the PATOIS project (Publications and Archives in Teaching: Online Information Sources) with funding from the JISC to develop four web-based tutorials based on datasets held by the ADS. Each tutorial aims to provide an introduction to using different aspects of archaeological or historical resources as part of the core syllabus taught by archaeology departments in the British Isles. The tutorials cover:

- Use of monument inventories
- Use of excavation and fieldwork archives
- Use of electronic journals

- Use of interdisciplinary resources, using the excavations of the crypt at Christ Church, Spitalfields, as a case study.

All the tutorials will be available for use, free of charge, from early 2003 (Kilbride et al. 2002).

## THE INTEGRATED ONLINE ARCHIVE

Given the developments described so far it is clear that we are currently witnessing major changes in the ways in which we are able to access archaeological information. To consider the possible impact of online archives it is necessary to go back to first principles and the development of the discipline.

### *Publication and archives*

The idea of the excavation report as the factual and complete record of a site goes back at least as far as Pitt-Rivers and what has been described as the Cranbourne Chase tradition: 'A discovery dates only from the time of the record of it, and not from the time of its being found in the soil' (Pitt-Rivers, cited in Wheeler 1954:182).

For Pitt Rivers, publication provided an objective record of what had been discovered and it was the archaeologist's duty to publish in tremendous detail, as demonstrated by his own four massive volumes on the excavations he conducted on his estate (Pitt-Rivers 1887–1898). On the other hand a different emphasis is visible in Sir Flinders-Petrie, who argued that the contents of notebooks and listings of data were not publication, and that observations must lead to conclusions and generalizations (Petrie 1904). Notwithstanding this early plea for synthetic publication it was the Cranbourne Chase tradition that was to have the greatest influence on publication trends, although the tension between brief synthetic publication and full data presentation has periodically re-emerged. Throughout the first half of the twentieth century, publication was seen as an integral part of the excavation process. In much of the literature there was little mention of archiving: the only record considered effective was full publication – the published report and the archive were regarded as one and the same thing (Jones et al. 2001:Section 2).

During the 1960s and 1970s, however, British archaeology had to face up to a growing publication crisis, shared in many European countries. Against a background of massive growth in public spending on archaeology there was increasing archaeological activity and rescue excavation, a growing post-excavation and publication backlog, and soaring publication costs. In England the Frere Report (1975) attempted to address the crisis. It endorsed the traditional view that archaeologists are under an obligation to produce a full record of their excavations but accepted that, given the crisis, 'publication in printed form of all the details of a large modern excavation is no longer practicable' (Frere 1975:2). The report advocated a rationalization of recording and publication. Four levels of recording were held to characterize the successful completion of an excavation (Frere 1975:3):

- Level I – the site itself and the excavated finds
- Level II – the site notebooks, on-site recording forms, drawings, finds records, photographs, etc
- Level III – the processed complete archive: full illustration and description of structural, stratigraphic and artefactual/environmental data
- Level IV – a synthetic description with supporting illustrations

Hitherto, full Level III publication had been the norm, at least in theory, but refined publication at Level IV was now recommended, on condition that a Level III report was produced for archiving and was readily available on request. It was recognized that selectivity at Level IV would require a higher standard of archiving than was often practised, with all excavation records being properly organized, curated and accessibly housed. Consideration was also given to other, cheaper forms of dissemination at Level III, on request. In essence, the Frere Report responded to the publication crisis by advising a reduction in the amount of material that would go into print in monographs and journals, coupled with an improvement in the organization and curation of archives.

The Frere Report was the first attempt by a state heritage body to address systematically the principles and methods of publication. With hindsight it can be argued that Frere did not constitute a radical departure from traditional practices. All that the Report advocated was an uncoupling of an accepted standard of record (known as the Level III report) from the process of formal publication (Level IV). It was a pragmatic response to the costs of formal publication and the pressures on publication outlets (Jones et al. 2001).

Although Frere's recommendations were very influential on archaeological practice it is arguable whether they had much impact upon the backlog brought about by increasing numbers of large projects. Indeed, the high standard of preparation required by Level III meant that in many cases *more* time was required for post-excavation work than had been allocated before. The continued publication crisis led Hassall to suggest that the balance between publication and archive might shift totally in favour of the archive: '... professional advancement and success in the future ... may depend on non-publication, but deliberate non-publication backed up by a total and readily accessible archive ...' (Hassall 1984:151).

The 'backlog problem' refused to go away, and a joint working party of the Council for British Archaeology and the Department the Environment was convened under the chairmanship of Barry Cunliffe. With an emphasis on the importance of an accessible archive, and on targeted research and publication, the Report (Cunliffe 1983) marked a departure both from the traditional model, with its ideal of full excavation and full publication, and the Frere Report, which had confined the latter to Level III. The detailed description of the evidence was to be reduced to a summary, with detail confined to microfiche. The report had considerable impact but its implementation was problematic and was rejected by the CBA's own Council. With the benefit of hindsight it seems that one of the main problems was practical and stemmed from difficulties with the technology of the

1980s. No archive could truly be accessible, and the use of microfiche was universally loathed. Another difficulty was increasing theoretical debate about whether the full report actually represented a complete factual account of the site. Barrett (1987) argued that the publication crisis extended beyond report production to the ways in which archives and reports could be used and reused. Although it may be impossible to judge an excavator's general competence from a published report, it is possible:

for the reader to undertake a critical analysis of the internal logic of the report, examining the linkages between the assumptions employed, the stated record of observations, and the interpretative account. (Barrett 1987:410)

Hodder (1989) regretted that reports had become impersonal objective accounts of data. He argued that since the excavation process is interpretative from start to finish, personal factors which lead to the interpretation should, as far as possible, be written into the report rather than kept out of it. In other words, there should be *greater* integration between description and interpretation. Another perspective, criticizing the use of synthetic reports as the main format of dissemination of archaeological knowledge was provided by Shanks and Tilley (1987). They argued that such reports represented exercises in 'domination and control' by individuals seeking to impose their view of the past on their readers. It was therefore crucial to find ways to make data available to give a wider audience the opportunity to create their own interpretations. This acknowledges that data are dependent upon observation and do not exist in isolation; nonetheless they are still capable of reinterpretation within another context, so long as sufficient information is recorded about the original context of observation (Richards forthcoming).

One further Committee tried to address the publication/archives problem. *Archaeological Publication, Archives and Collections: towards a National Policy* (Carver et al. 1992) was written within the context of the introduction of developer-funding. It also took account of those developments in theoretical thinking which reflected a move away from the Cranbourne Chase tradition and away from 'preservation by record':

since the record is selective and therefore incomplete and post-excavation analysis must also, of necessity, be selective, the excavation report can only be a contemporary statement reflecting on aspects of the site: it cannot be an immutable and complete truth. (Carver et al. 1992:2.2.1)

The Committee took the Cunliffe Report one stage further and recommended that dissemination should normally be in the form of a published summary report and an accessible site archive. Once more, however, technology lagged behind and lacked the means of providing access to an archive with links between it and the summary publication. The report was effectively shelved.

Meanwhile, the publication crisis also became an archiving crisis as museums were expected to receive the physical archives from the backlog projects. There was a growing feeling that archives are important, but that their content and accessibility require reassessment (McAdam 1999).

In 1998 the CBA was commissioned to carry out a wide ranging survey of publication. This ran in parallel to the Digital Data Survey conducted by ADS, and also focused on user needs. Its recommendations reflect the fact that technology has moved on, and while they again focus on reducing the scale of conventional publication, the Publication User Needs Survey (PUNS) Report recommends alternative means of electronic publication and the dissemination of archival and specialist material in electronic format as a means round the practical problems. The introduction of digital technology provides an opportunity to shift away from pure synthesis towards making archaeological data accessible digitally (Gaffney and Exon 1999). Three recommendations (Jones et al. 2001:Section 6) are of particular relevance in the context of this article.

- Recommendation 3 recognizes that there are different user groups for different aspects of a report and suggest multiple forms and media of dissemination should be used, as appropriate for a given project. These might include a summary account produced during the project or immediately after: a synthetic journal article or monograph; Internet publication either alongside or instead of the above; and electronic availability of detailed and well-indexed structural and specialist reports. Recommendation 3 also concludes that all project archives to be placed on the Internet.
- Recommendation 7 also notes that 'new means of making detailed structural and specialists reports available are required – a need which is well answered by electronic media', notably the Internet. The CBA argue that it is important that this should not take place ad hoc and propose the establishment of a specific forum, where work can be indexed and accessed with ease, and where peer review ensures that such publications provide improved means of attracting academic recognition.
- Recommendation 8 also reflects the fact that the survey found strong support for the mounting of all archives on the Internet, supported by well-indexed and queryable databases. It recommends that all archives be made available on the Internet and suggests that funding agencies and local authorities should consider making this mandatory for projects within their remits, integrated with electronic publication of reports.

### *Digital archives*

It is therefore clear that there is a growing recognition that online technologies have developed to the point where they offer considerable potential for the dissemination of archaeological results. It becomes essential to define exactly what we mean by a digital archive and to consider what additional work it will involve.

One of the basic principles already established in the first edition of the *Excavation and Fieldwork Guide to Good Practice* (Brown et al. 1999) was that for information not already held in digital format it was unnecessary to embark upon a massive digitization programme. As more and more data are collected in digital format, and as computers begin to pervade all the activities of archaeological contractors from project management to post-excavation analysis, the proportion

of primary data available digitally will inevitably grow. Digitization of paper records may also be desirable to increase access, or as part of the preservation of rare and fragile documentary sources, but it should not be done for 'its own sake'. There has also been a recognition that different archaeological projects may deserve different levels of digital archive. Going back to the idea of preservation for a purpose it is not clear that all data should be automatically preserved as a reuse value has to be demonstrated. For instance, although an animal bones database will have clear value for comparative purposes, it is less obvious that the automatically logged total station coordinates from an earthwork survey of a minor site should be preserved indefinitely. The end product of the computer terrain model or contour plot may be important, but not necessarily the raw data points used to produce it.

Not all archaeological investigations make the same contribution to knowledge and not all merit the same level of digitization, although it is important to recognize that it is not always possible to identify this at the time and clearly necessary to have a proper assessment procedure to determine their significance. MAP2 (English Heritage 1991) drew upon the Frere and Cunliffe Reports, especially the latter, and tried to clarify the mechanisms required for the iterative reviews which Cunliffe envisaged. This was assisted by a refined terminology, which distinguished between the site archive, the research archive, summary publication, and full publication. The second edition of the *ADS Excavation and Fieldwork Guide to Good Practice* also adopts the idea of phases through which projects should pass, and the outputs required at each stage, in an attempt to categorize the equivalent levels of digital archive (Richards and Robinson 2000). Four levels of archive are identified; in ascending order of complexity these are:

- Index level;
- Assessment level;
- Research level; and
- Integrated archive and electronic publication.

Many desk top assessments or evaluations carried out as part of the planning process produce little archaeology. In these cases it may simply be appropriate to create the minimal digital archive: a single metadata index record that identifies the site and its location, those responsible for the work, the components of the paper and physical record, and where these can be found. The OASIS project (see above) aims to provide such an Index level record for all archaeological investigations in England. Where a report has been submitted then this will normally have been prepared in electronic format and it makes sense to archive this as well, ideally linked from the index record, so that it can be accessed more easily than if it were simply lodged in a filing cabinet in the local SMR.

For more significant projects that may warrant further analysis then the MAP document proposes the production of an assessment report which identifies the key findings and data sets available for further study. For projects that reach this stage, this too should be archived, along with such details of the stratigraphic and

structural sequence as have been digitized, as well as specialist reports and databases used in the production of the assessment report.

The majority of important projects will proceed beyond the assessment level to full post-excavation analysis and the production of a monograph or journal article. Large quantities of digital data will be created as part of this process, including detailed specialist reports, databases, spreadsheets, CAD phase plans and so forth. It is recommended that these should all be archived, along with the final text of the full report, to create a Research Level archive. The English Heritage sponsored DAPPER project (Digital Archive Pilot Project for Excavation Records), conducted jointly between ADS and the Museum of London Archaeology Service and Oxford Archaeological Unit created two research level archives, for the Royal Opera House site and Eynsham Abbey. Neither had been planned for digital archiving and dissemination at the outset and both are to be published by means of traditional hard copy monographs. Nonetheless the digital archives are extremely valuable additions to the dissemination of the site and each has had high levels of reuse. Within the first 24 months there have been over 10,000 visits to the Eynsham Abbey archive, and individual files, which include CAD and ARCView plans, have been downloaded on up to 30 occasions. ADS has received enquiries from as far away as the United States, where a class from New York State University were engaged in post-excavation study of the development of Anglo-Saxon London through the DAPPER archive. Nonetheless it must be recognised that in each case we are really only making available the digital residues left over from the post-excavation analysis, reflecting the post-excavation practices of each archaeological contractor. Where digital dissemination can be planned into a project from the outset then it is possible to produce a much more integrated and exciting product which may fully exploit the potential of the Internet for archaeological publication.

Therefore the ADS has developed the concept of the integrated archive, linking archive and publication and allowing users to pursue ideas found in the synthetic publication into the evidence that has been used to support these interpretations. In many cases the digital archive may simply be providing a much more effective and accessible microfiche, allowing dissemination of detailed tables, plans, photographs and supporting text, normally judged too expensive to publish. The Online Archive provides a much more usable alternative to fiche, and may also allow authors to include far more supplementary material than was ever possible with fiche. The Fyfield and Overton Down Project was published at a number of levels. A popular book, *The Land of Lettice Sweetapple* publishes the main results for the general reader (Fowler and Blackwell 1998). The Society of Antiquaries monograph (Fowler 2000) includes the archaeological evidence from the key sites for the academic reader, while in the project digital archive Peter Fowler makes available not only the text of four further monograph-length reports, but also some 100 Fyfod Working Papers, including specialist reports, draft texts, and background documentation. The traditional monograph includes a URL for the archive and basic instructions on how to use it.

Where the archive can be developed alongside the publication then it is possible to develop a more complete set of links, publishing URLs for specific files so that

readers can more easily locate the precise information they require. It is planned that the forthcoming publication of a major site in London, No.1 Poultry, will explore methods of detailed linkage, as well as providing a database search mechanism for the archive and a clickable map-base to allow the user to explore the archive spatially.

Where a site is also published electronically then the full potential of live hypertext links can be explored, allowing the user to move between publication and archive at will. A tentative and preliminary model of some of the possibilities is explored in the layered electronic publication of an Anglian and Anglo-Scandinavian farmstead at Cottam, East Yorkshire, in *Internet Archaeology* (Richards 2001) and the simultaneous release of the archive on the ADS site. At the top level the electronic publication follows the familiar model of a traditional printed report through introduction, methodology, results and discussion. Within the electronic version, however, the reader can follow hypertext links to pull up illustrative material such as plans and photographs, and can also read more detail of the archaeological findings. They can also search an online database of the finds from the excavation, field-walking, and metal-detecting. Furthermore, by clicking on further links they can seamlessly move into the archive, reading the specialist reports, the detailed stratigraphic evidence, and are able to download context and finds databases, and raw geophysics surveys and CAD files.

Although Cottam is a relatively small site the project demonstrates some of the potential of integrated online archives. This is not just a case of dumping raw undigested material on the web. It is essential that the electronic publication must still 'tell a story' and answer the research agenda set by the investigator. In the archive, however, resides the material necessary for the reader to explore alternative interpretations and multiple narratives.

Within the ADS catalogue the user is able to search for archives by title, period and subject keywords, as well as by clicking on a map of the British Isles for archives within a specific area. The Dublin Core metadata also allows them to search for archives which include particular applications types, as well as for archaeological subject. If they want a specific archive they can also go to it directly via the Project Archives section.

However, although they can download archives files onto their own desktop computers users are unable to search within an archive over the Internet. Thus a user interested in occurrences of a particular pottery type, for example, would have to search sequentially within all the pottery reports held on the ADS server.

Recent developments in means of structuring information provided on the Internet may help resolve this problem and lead to the creation of structured and searchable archives. The majority of current web content is designed for humans to read, not for computer programs to manipulate meaningfully. HTML markup tags indicate titles and headings but they do not encode the content of a web page in any structured fashion. In *Scientific American* for April 2001 Tim Berners-Lee describes what he calls 'the Semantic Web', whereby web robots will be able to harvest structured collections of information.

An important technology for developing Berners-Lee's Semantic Web is already

in place. XML provides an eXtensible Markup Language which permits the systematic tagging of the components of a text file. For excavation reports, for instance, where there is a fairly standard template for the organisation of information, XML tags could be used to identify each section. XML can also be used to create wrappers around non-text elements, and so images, databases files, CAD drawings and so on, could each be described by metadata held in XML tags. When XML tagged information is displayed over the Internet a browser application may be configured to display each component in a specific way. The tags may also be used to allow search engines to harvest particular categories of information and to build sophisticated indices. This is the principle underlying the Open Archives Initiative (<http://www.openarchives.org/>). This proposal follows the work of Steve Harnad and others in promoting self publication (Harnad 2001). Rather than researchers being dependent upon commercial publishers and thereby having to pay to read the results of the research which has already been paid for through the academic sector, they argue that researchers should mount their papers on local web servers, as Data Providers. John Hoopes (2000) has proposed that the Society for American Archaeology should develop a peer-reviewed web gateway for the dissemination of archaeological reports. A problem with self-publication is that it can be difficult to locate relevant papers if they may appear anywhere on the Internet, rather than in a specific journal. Under the OAI, XML tags are used to markup the papers with appropriate metadata keywords, so that they can then be harvested by automated search tools and indexed in the online catalogues of service providers. This provides a degree of interoperability but searches will only be as up to date as the last harvest as several online databases are not being simultaneously queried, as occurs under Z39.50 technology.

To date only limited use has been made of XML in archaeological publication. Holmen and Uleberg (1996) describe the use of SGML for encoding archaeological archive documents in the Norwegian Museum Project and have subsequently developed XML applications. Gray and Walford (1999) recommended it as a means of structuring archaeological reports and enabling comparison. David Schloen of the University of Chicago has been a major proponent and has even compiled an archaeological DTD called ArchML (Schloen 2001), although his model is heavily based on object-orientated data structures in the context of Near Eastern research. It is unlikely to gain widespread acceptance. The ADS is actively researching the use of XML markup for the dissemination of online archives, with partners in Europe through the ARENA project (Archaeological Records of Europe – Networked Access). ARENA is concerned with the conservation and presentation of the European archaeological heritage through new information technologies (Kenny et al. forthcoming). The project is being carried out with the support of the European Community through the Culture 2000 programme and has six partner organizations in Poland, Romania, Denmark, Iceland, Norway and the UK. The other partners are Poznan Archaeological Museum (Poland); CIMEC, Institute for Cultural Memory (Romania); the National Agency for Cultural Heritage (Denmark); Fornleifastofnun Islands (Iceland); and the Museum Project, Oslo (Norway). It is the purpose of ARENA to bring together some of these key

European heritage stakeholders to share and extend their skills and experience. The project has four key aims:

- Organization of initiatives for exchange of experience and the further training of professionals
- Promoting elements of the heritage
- The adapted and innovative use of new technologies, to the benefit of participants, users and the general public
- Organizing research projects, projects to raise awareness and to teach and disseminate knowledge

Each of the partners has begun work on preparing archives from some classic European excavations for online access, including Danebury (England), Vorbasse (Denmark) and Hofstaðir (Iceland). The archival files will be indexed according to theme and the metadata made available for cross-searching across the partners. In addition, the partners have agreed to make available index records for many thousands of sites and monuments available from a portal. Users will again be able to select a theme that will be translated into local languages at the portal and used to retrieve relevant records from distributed resources held locally in each of the partner organizations. Multilingual issues create a further challenge to interoperability and we have chosen to adopt a pragmatic solution of translating a limited number of top level terms. Nonetheless we hope that ARENA will demonstrate the potential of linking databases and online archives from across several countries. The curation of these resources is inevitably constrained by modern political boundaries but the archaeological research questions frequently refer to a shared European cultural heritage.

In summary, archaeological resources are increasingly held in electronic media and may even be born digital, with no paper copy. These data are at risk. Digital archiving is not just a UK problem and other European heritage organizations are becoming increasingly aware of the fragility of digital data. However, electronic dissemination also provides a joint opportunity for sharing solutions and improving access. Archaeological publication across Europe has reached an impasse and needs to look to electronic solutions which will redefine our approach to publication and archiving. As the numbers of online digital archives grows the prospect of cross-searching of their contents will not just secure the future for these valuable archaeological resources but will also truly open up archives for European online access.

## REFERENCES

- AUSTIN, T., 1998. Tip of the iceberg in a digital sea: Scale of the Academic Digital Resource. *Archaeology Data Service News* 4. Available online at <http://ads.ahds.ac.uk/newsletter/issue4.html>
- AUSTIN, T., D.J. ROBINSON and K.A. WESTCOTT, 2001. A digital future for our excavated past. In Z. Stančič and T. Veljanovski (eds), *Computing Archaeology*

- for Understanding the Past: CAA 2000:289–296. Oxford: British Archaeological Reports (International Series 931).*
- AUSTIN, T., F. PINTO, J. RICHARDS, and N. RYAN, 2002. Joined up writing: an Internet portal for research into the Historic Environment. In G. Burenhult (ed.), *Archaeological Informatics: Pushing the Envelope. CAA 2001:243–251. Oxford: British Archaeological Reports (International Series 1016).*
- BAKER, D., G. CHITTY, J. RICHARDS and D. ROBINSON, 1999. *Mapping Information Resources: A Report for HEIRNET.* Available online at <http://www.britarch.ac.uk/HEIRNET/rep1.html>
- BARRETT, J., 1987. The Glastonbury Lake Village: models and source criticism, *Archaeological Journal* 144:409–423.
- BEAGRIE, N. and M. JONES, 2002. *Digital Preservation Handbook.* Available online at <http://www.dponline.org/graphics/handbook/index.html>
- BERNERS-LEE, T., 2001. The Semantic Web, *Scientific American* 501. Available at <http://www.sciam.com/2001/0501issue/0501berners-lee.html>
- BROWN, A., D. BROWN, G. COLES, D. DAWSON, A. DODD, J. GARDINER, G. LOCK, C. LONGWORTH, N. MERRIMAN, P. MILLER, D. MURRAY, J. REEVE and A. WISE, 1999. *Digital Archives from Excavation and Fieldwork: Guide to Good Practice.* Available online at [http://ads.ahds.ac.uk/project/goodguides/excavation\\_1st\\_ed/](http://ads.ahds.ac.uk/project/goodguides/excavation_1st_ed/)
- BURNARD, L. and H. SHORT, 1994. *An Arts and Humanities Data Service.* Report of a Feasibility Study commissioned by the Information Systems Sub-Committee of the Joint Information Systems Committee of the Higher Education Funding Councils.
- CARVER, M., H. CHAPMAN, B. CUNLIFFE, T. HASSALL, M. HEBDITCH, A. LAWSON, I. LONGWORTH, R. MORRIS, D. PHILLIPSON, J. SCHOFIELD and G. WAINWRIGHT, 1992. *Archaeological publication, archives and collections. Towards a national policy.* London: Society of Antiquaries and the Museums Association (Printed as a supplement to *British Archaeological News* 7(2), York: Council for British Archaeology).
- CONDON, F., J. RICHARDS, D. ROBINSON and A. WISE, 1999. *Strategies for Digital Data. Findings and recommendations from Digital Data in Archaeology: A Survey of User Needs.* Available online at <http://ads.ahds.ac.uk/project/strategies/>
- CUNLIFFE, B.W., 1983. *The publication of archaeological excavations: report of a joint working party of the Council for British Archaeology and the Department of the Environment.* London: Department of Environment.
- EITELJORG II, H., 1995. The Archaeological Data Archive Project. In J. Huggett and N. Ryan (eds), *Computer Applications and Quantitative Methods in Archaeology 1994:245–247.* Oxford: British Archaeological Reports (International Series 600).
- ENGLISH HERITAGE, 1991. *Management of archaeological projects*, 2nd edn. London: English Heritage.
- FERNIE, K. and P. GILMAN, eds, 2000. *Informing the Future of the Past: Guidelines for SMRs.* Swindon: English Heritage.
- FOWLER, P. and I. BLACKWELL, 1998. *The Land of Lettice Sweetapple. An English Countryside Explored.* Stroud: Tempus.

- FOWLER, P., 2000. *Landscape Plotted and Pieced: Landscape History and Local Archaeology in Fyfield and Overton, Wiltshire*. London: The Society of Antiquaries of London.
- FRERE, S.S., 1975. *Principles of publication in rescue archaeology: report by a working party of the Ancient Monuments Board for England*. London: Committee for Rescue Archaeology.
- GAFFNEY, V., and S. EXON, 1999. From order to chaos: publication, synthesis, and the dissemination of data in a digital age, *Internet Archaeology* 6. Available online at [http://intarch.ac.uk/journal/issue6/gaffney\\_index.html](http://intarch.ac.uk/journal/issue6/gaffney_index.html)
- GRAY, J. and K. WALFORD, 1999. One Good Site Deserves Another: Electronic publishing in field archaeology, *Internet Archaeology* 7. Available online at [http://intarch.ac.uk/journal/issue7/gray\\_index.html](http://intarch.ac.uk/journal/issue7/gray_index.html)
- HARDMAN, C. and J.D. RICHARDS, forthcoming. OASIS: dealing with the digital revolution. In *Computer Applications and Quantitative Methods in Archaeology 2002*, Oxford: British Archaeological Reports (International Series).
- HARNAD, S., 2001. The self-archiving initiative. Freeing the refereed research literature online. *Nature* 410:1024–1025
- HASSALL, T., 1984. Publication: crisis and inaction. In R. Martlew (ed.), *Information Systems in Archaeology*: 142–152. Gloucester: Alan Sutton Publishing.
- HENDLEY, T., 1998. *Comparison of Methods and Costs of Digital Preservation*. London: The British Library (British Library Research and Innovation Report 109).
- HISTORIC SCOTLAND, 1996. *Publication and Archiving of Archaeological Projects*. Edinburgh: Historic Scotland,
- HODDER, I., 1989. Writing archaeology: site reports in context. *Antiquity* 63:268–274.
- HOLMEN, J. and E. ULEBERG, 1996. The National Documentation Project of Norway the archaeological sub-project. In H. Kamermans and K. Fennema (eds), *Interfacing the Past Computer Applications and Quantitative Methods in Archaeology*. CAA 95, volume 1, published as *Analecta Praehistorica Leidensia* 28:43–46, Leiden: University of Leiden.
- HOOPES, J.W., 2000. Beyond the World Wide Web: present and future venues for scholarly online publication. *Society for American Archaeology Newsletter* 18(5):20–22
- JONES, S., A. MACSWEEN, S. JEFFREY, R. MORRIS and M. HEYWORTH, 2001. *From the Ground Up. The Publication of Archaeological Projects: A User Needs Survey*. London: Council for British Archaeology. Available online at <http://www.britarch.ac.uk/pubs/puns/index.html>
- KENNY, J., W.G. KILBRIDE and J.D. RICHARDS, forthcoming. Enter the ARENA: preservation and access for Europe's archaeological archives. In *Computer Applications and Quantitative Methods in Archaeology 2002*, Oxford: British Archaeological Reports (International Series).
- KILBRIDE, W., 2000. Digital data and digital archaeology: Needle 1 – Haystack 0, *The Archaeologist* 38:21–22
- KILBRIDE, W. and J. WINTERS, 2001. Observing the Game: What can access statistics really tell us? In Z. Stančić and T. Veljanovski (eds), *Computing Archaeology for*

- Understanding the Past: CAA 2000*: 339–345. Oxford: British Archaeological Reports (International Series 931).
- KILBRIDE, W., K. FERNIE, P. MCKINNEY and J.D. RICHARDS, 2002. Contexts of Learning: The PATOIS project and Internet-based teaching and learning in Higher Education. *Internet Archaeology* 12. Available online at [http://intarch.ac.uk/journal/issue12/patois\\_toc.html](http://intarch.ac.uk/journal/issue12/patois_toc.html)
- MCADAM, E., 1999. Archives: why don't we burn the lot? In H. Hansen and G. Quine (eds), *Our Fragile Heritage*: 97–104. Copenhagen: National Museum of Denmark.
- MERRIMAN, N. and H. SWAIN, 1999. Archaeological archives: serving the public interest? *European Journal of Archaeology* 2(2):249–267.
- MILLER, A.P., 1996. Metadata for the masses: what is it, how can it help me, and how can I use it? *Ariadne* 5. Available online at <http://www.ukoln.ac.uk/ariadne/issue5/metadata-masses/>
- MILLER, A.P., 1999. The Importance of Metadata in Archaeology. In Dingwall, L, S. Exon, V. Gaffney, S. Laflin, and M. van Leusen (eds), *Archaeology in the Age of the Internet: Computer Applications and Quantitative Methods in Archaeology. CAA 97*: 133–136. Oxford: British Archaeological Reports (International Series 750).
- MILLER, A.P. and D. GREENSTEIN, eds, 1997. *Discovering Online Resources: A Practical Implementation of Dublin Core*. Bath: UKOLN/AHDS.
- PETRIE, W.M.F. 1904. *Methods and Aims in Archaeology*. London: Macmillan.
- PITT-RIVERS, A.L.F., 1887–1898. *Excavations in Cranborne Chase*. Vols 1–4. Privately printed.
- RICHARDS, J.D., 1997. Preservation and reuse of digital data: the role of the Archaeology Data Service. *Antiquity* 71(274):1057–1059.
- RICHARDS, J.D., 2001. Anglian and Anglo-Scandinavian Cottam: linking digital publication and archive, *Internet Archaeology* 10. Available online at [http://intarch.ac.uk/journal/issue10/richards\\_index.html](http://intarch.ac.uk/journal/issue10/richards_index.html)
- RICHARDS, J.D., forthcoming. Do we believe in the 'data' in the Archaeology Data Service? In S. Roskams (ed.), *Interpreting Stratigraphy: Papers presented at York*. Oxford: British Archaeological Reports (International Series).
- RICHARDS, J.D., P. MILLER and A. WISE, 1999. Digital Archives in Archaeology. In H. Hansen and G. Quine (eds), *Our Fragile Heritage*: 123–130. Copenhagen: National Museum of Denmark.
- RICHARDS, J.D. and D.J. ROBINSON, eds, 2000. *Digital Archives from Excavation and Fieldwork: Guide to Good Practice*, 2nd edn. Oxford: Oxbow Books. Available online at <http://ads.ahds.ac.uk/project/goodguides/excavation/>
- ROSS, S., 2000. *Changing trains at Wigan: digital preservation and the future of scholarship*. London: National Preservation Office (NPO Preservation Guidance Occasional Papers).
- RUSSELL, K., 2000. Digital preservation and the Cedars project experience. *The New Review of Academic Librarianship* 6:139–154.
- SCHLOEN, D., 2001. Archaeological data models and web publication using XML, *Computers and the Humanities* 35(2):123–152. Available online at <http://www.oi.uchicago.edu/e-ANE/e-ANE.html>

- SHANKS, M., and C. TILLEY, 1987. *Social Theory and Archaeology*. Oxford: Polity Press.
- WHEELER, M., 1954. *Archaeology from the Earth*. Oxford: Clarendon Press.
- WISE, A.L. and A.P. MILLER, 1997. Why metadata matters in archaeology. *Internet Archaeology* 2. Available [http://intarch.ac.uk/journal/issue2/wise\\_index.html](http://intarch.ac.uk/journal/issue2/wise_index.html)
- WISE, A.L. and J.D. RICHARDS, 1999. Digital preservation in archaeology. In Dingwall, L, S. Exon, V. Gaffney, S. Laflin, and M. van Leusen (eds), *Archaeology in the Age of the Internet: Computer Applications and Quantitative Methods in Archaeology*. CAA 97: 137–142. Oxford: British Archaeological Reports (International Series 750).

## BIOGRAPHICAL NOTE

Julian D. Richards is Reader in Archaeology in the Department of Archaeology, University of York, where he is Director of the Archaeology Data Service and the e-journal *Internet Archaeology*. His research interests are in the early medieval archaeology of northwestern Europe, particularly Anglo-Saxon and Anglo-Scandinavian settlement patterns and funerary rites, and in computer applications in archaeology. He undertook a PhD using computer processing to study the significance of Anglo-Saxon mortuary behaviour and went on to co-author *Data Processing in Archaeology* in 1985, the first textbook dedicated to the use of computers in archaeology. He has gone on to publish widely on computer applications and early medieval archaeology, including the popular textbook, *Viking Age England* (2000). He has excavated in England at early medieval sites at Wharram Percy, Cottam and Heath Wood, Ingleby.

*Address:* Department of Archaeology, University of York, King's Manor, York, YO1 7EP, UK.

[email: [jdr1@york.ac.uk](mailto:jdr1@york.ac.uk)]

## ABSTRACTS

### Conservation numérique et accès

*Julian D. Richards*

Nous assistons à une crise dans la publication et l'archivage des données archéologiques en Europe. L'informatisation signifie que les données sont plus fragiles que si elles étaient conservées sur papier, mais en même temps plus facilement accessibles par Internet. Dans cet article, on examine le rôle du service des données archéologiques (Archaeology Data Service), on s'occupe de quelques-unes des questions se rapportant à la collection et à la propagation interactive des données numériques, et des conséquences pour l'avenir de la publication archéologique. On discute les différentes façons d'aborder la conservation des données numériques, le développement de critères d'archivage et comment encourager la réutilisation. On étudie l'évolution des catalogues et archives distribués interactivement et le besoin de normes metadata pour cataloguer les documents. Pour terminer, on réfléchit au rôle de XML comme nouvelle technologie et au projet européen ARENA développant la conservation numérique et une infrastructure d'accès.

*Mot-clés:* archives interactives, communication des documents, conservation numérique, gestion du patrimoine, interfonctionnement, Internet, metadata, systèmes d'information archéologique, XML, Z39.50