TWENTY YEARS PRESERVING DATA: A VIEW FROM THE UK

Julian D. Richards

Director, Archaeology Data Service, University of York, The King's Manor, Exhibition Square, York, YO1 7EP, UK (julian.richards@york.ac.uk)

Abstract

In 2016 the Archaeology Data Service (ADS) was 20 years old. Since its birth the ADS has had to respond to rapid changes in technology, as well as major cultural and organizational changes in the external operating environment, from which a sustainable business model for digital preservation has emerged. This paper will take a retrospective look at challenges that have been faced and will review current and future priorities for those seeking to establish digital repositories. Digital preservation and open access to research data are now much higher up the agenda of funding bodies but there is still lack of agreement on what constitutes a core digital archive from a fieldwork project. The paper will review what the significant properties of an archaeological archive are, and how re-use can be supported, linking data and publications. It will consider the challenge of dealing with the grey literature and of avoiding creating further data silos, featuring new initiatives to provide interoperability between digital repositories. It will review the role of data and metadata standards, and consider how the profession needs to address its responsibilities over the next 20 years.

So how did we get here?

The Archaeology Data Service (ADS) is the UK's national digital data archive for archaeology and the historic environment, and the longest serving repository for archaeological data in the world. In 2016 it celebrated its 20th anniversary. The ADS was originally established in 1996 with two members of staff, and a budget of c. £60,000 per annum (Richards 2008; Richards et al. 2013). It now has 14 members of staff and an annual budget of c. £750,000. Having been set up as part of the Art and Humanities Data Service (AHDS), with an annual core grant, the ADS was the only one of the AHDS service providers retained by the Arts and Humanities Research Council (AHRC) beyond 2008, in recognition of the fact that most archaeological data are special and are a primary source that, once lost, cannot be recovered. Whereas most historical sources can be re-digitized, one cannot go back and re-excavate an archaeological site. However, the AHRC's decision also reflected the fact that the ADS had demonstrated that it could generate alternative sources of income. It now receives no external grant aid support and is entirely dependent on generating its operating costs via research and development, consultancy and preservation services. Over 75 per cent of its annual income is derived from project funding and research and development grants, with major contributions from the European Commission, and Historic England. A growing proportion of its income also comes from the commercial contract archaeology sector, and the level of external income generation demonstrates its value for money and cost-effectiveness.

As of October 2016, the ADS Collections Management System recorded 2,143,497 files in our systems, comprising 12 Tb of data. However, the more significant statistic is the number of recorded processes carried out on those files – 21,327 – to ensure their long-term preservation. Unlike a physical archive, digital preservation requires active curation, migrating files formats to current versions to ensure their re-usability in new software. The ADS is compliant with the Open Archival Information Standard (OAIS) ISO 14721 standard for digital repositories (Lavoie 2014).

Since March 2011 the ADS has been accredited with the Data Seal of Approval, an international 'kite-mark' for digital repositories, becoming one of the first UK repositories to gain this recognition, second only to the UK Data Archive (Mitcham and Hardman 2011). In 2012 it was awarded the Digital Preservation Coalition's first Decennial Award for the most outstanding contribution to digital preservation – in all disciplines – in the last decade.

Over the last 20 years the ADS has inevitably witnessed some significant changes in the digital preservation landscape. From the outset ADS set out to preserve all types of digital research data produced by archaeologists. That data has grown both in variety and in size. For example, an early collaborative project looked at the wide variety of data generated by deep underwater archaeology, particularly that automatically logged by remote vehicles (Drap 2009). As well as the complexities of new data formats there are also new forms of communication, such as the widespread use of social media, which bring their own preservation challenges (Jeffrey 2012). 'Big Data' has also become a buzz phrase in computer-based analysis but, whilst archaeological datasets are not generally large by external standards, we have had to address how best we can archive, and disseminate, some of the terabytes of data produced by laser survey (Austin et al. 2008). As bandwidth increases there is also an increasing appetite for online visualization of complex 3D data. A recent project involved collaboration with the CNR-ISTI laboratory in Pisa in order to develop a web-based tool for the visual analysis of 3D stratigraphic layers (Galeazzi et al. 2016). However, even apparently simple data types may require intervention to ensure that they do not rely on proprietary features and ADS staff have undertaken an analysis of the complexities of archiving PDF files, the most common format for the deposit of text reports (Evans and Moore 2014).

This paper will explore some of the lessons learned over the last 20 years. The initial focus will be on the ubiquitous 'grey literature' – the unpublished fieldwork reports – which have come to dominate the archaeological record, and some of the ways of dealing with them. It will then go onto

consider more ambitious forms of dissemination, and the concept of blurring the distinction between publication and archive. It will attempt to define what are the core elements of an adequate fieldwork archive in the 21st century. The next section will consider how ADS has managed to become a self-sustaining digital archive and discuss business models for archiving in the heritage sector. However, there is no point in spending money on preservation if no one uses the archive. This therefore leads into a discussion of re-use and the cost-effectiveness of digital archiving. Finally, before attempting to draw some conclusions, the paper will end with a consideration of data silos and how they can be avoided. With the increase in the number of archaeological digital repositories around the world, it is essential to ensure interoperability between them, and means of achieving that will be discussed.

Fifty Shades of Grey

One of the greatest challenges facing the archaeological profession is how to make available the results of the extensive fieldwork projects undertaken in advance of modern development. The quantity of fieldwork has outstripped the capability of traditional modes of publication to keep pace and it is even difficult to find a comprehensive record of where and how many archaeological projects have taken place (Evans 2013). The outcome in most countries has been a mountain of unpublished fieldwork reports, the so-called 'grey literature'. The problem with grey literature has been the difficulty of knowing it exists and then of tracking it down. Generally produced in just one or two copies – one for the client and one for the regional HER (Historic Environment Record) or SHPO (State Historic Preservation Office) – these reports have been very difficult for researchers to access. In an important paper Richard Bradley (2006) lamented the fact that most teaching and research taking place in universities was way out of date as it failed to take account of new

discoveries locked in the grey literature. Similar problems have been reported in countries across the world.

In the UK one solution has been the OASIS project, now a collaboration between the ADS and the national heritage agencies for England and Scotland: Historic England and Historic Environment Scotland respectively (Hardman and Richards 2003). OASIS is an online data collection form which collects key information about any type of archaeological fieldwork according to national recording standards, and allows the user to upload a copy of their report. Users must enter core metadata, including the location of the project and who undertook the work, as well as standardized period and subject terms, a summary of what was found, and what the archive contains. Access to the form is provided to the regional archaeological or planning office, as well as the appropriate national agency. Once the form has been signed off by the archaeologists in the planning office or, (by agreement), in the national agency, any attached report is given a DOI (Digital Object Identifier) and released within the ADS library of unpublished fieldwork reports. As of February 2017 there were 40,816 reports available online, providing an essential resource for archaeological research, a figure which grows at the rate of c. 200 a month. Completing an OASIS form has become a requirement in Scotland and in most regions in England, but the main driver for usage of the form came from the contract archaeologists themselves, keen to promote their work online. Ironically the grey literature, available Open Access, is now far more accessible than conventional publication in monographs on regional county society journals. Even if libraries subscribe to these they are usually only available in the library, or if they are online they are still behind a subscription pay-wall, accessibly only to the privileged few whose institutions have subscribed. It is a noticeable trend that as the grey literature has gone online, and has largely replaced traditional journal articles, it has also become more professional and represents adequate publication in its own right. Completion of an OASIS record is not yet routine in the voluntary or

community archaeology sector, or by university-based archaeologists, but a major revision, currently underway and due for release in 2018, is designed to change that. The new OASIS will also allow access to museums, forewarning them of physical archives destined for their store.

One of the remaining challenges has been the adequate indexing of such reports. The OASIS form requires the manual entry of keywords, and was liable to people mistyping or inventing terms which were not part of agreed national controlled vocabularies. In the Archaeotools project the ADS explored the use of Natural Language Processing (NLP) to automatically generate index terms for text reports (Jeffrey et al. 2009; Richards et al. 2011; Richards et al. 2015; Vlachidis and Tudhope 2013). Further work, in collaboration with a team from the University of South Wales in the STAR and STELLAR projects, has led to the development of online tools to allow those doing data entry to draw their terms from controlled pick lists (Tudhope et al. 2011a; 2011b). These projects also allowed the integration of grey literature with other forms of fieldwork data via a semantic cross-search, with a mapping to the international high level ontology for the cultural heritage sector, the CIDOC CRM (Doerr 2003). In the next generation of OASIS, NLP will be used to analyze a report as it is uploaded and will suggest suitable keywords, drawn from the controlled vocabularies. In previous experiments NLP has achieved an 80 per cent success rate in identifying the same terms that the archaeologist would enter by hand. The user will then be able to accept or reject suggested terms, and to add alternatives, thereby leading to a significant increase in efficiency, as well as much better quality metadata.

Blurring Boundaries

Despite using online medium for dissemination the grey literature report remains a traditional form of publication. Whilst reports may increasingly be accompanied by other data types, including images or databases, they rarely link to them in a way which allows the reader to drill down to

verify a particular interpretation. Digital media should allow for much more ambitious linking to primary data, which has been the aspiration of our sister e-journal, *Internet Archaeology*, from the outset (Richards 2015).

From its first issue in 1996 the journal has endeavoured to promote links between the publication of research and supporting datasets. The award-winning Linking Electronic Archives and Publications (LEAP) project set out to provide a series of exemplars of linked publications in *Internet Archaeology* with archives held by the ADS, including the major fieldwork projects at Merv, Silchester, Troodos, and Whittlewood (Richards, Winters et al. 2011). Of course this relationship is not exclusive and *Internet Archaeology* has also published articles linked to data sets held in other data archives, including tDAR in the United States (Holmberg 2010). With its recent move to becoming a full Open Access journal, *Internet Archaeology* has become the journal of choice for many archaeologists who wish to promote access to their research and data.

In 2013 *Internet Archaeology* introduced another publication model to encourage researchers to provide access to their data sets: the data paper. The concept of the data paper was developed in the physical sciences, and has been extended to archaeology via the *Journal of Open Archaeological Data*, established at University College London under the auspices of Ubiquity Press. A data paper is generally a short paper which simply describes and summarizes a research data set, and which outlines how it might be re-used. It is generally a condition of publication that the dataset must have been deposited in an archive and have been allocated a DOI. Thus, for example, a paper by Bevan and Conolly on the Antikythera survey project (2012) references a dataset held by the ADS (Bevan and Conolly 2014). *Internet Archaeology* has developed the concept of the data paper further, adding a published review of the dataset, by a named external reviewer (e.g. Williams et al. 2014).

What is an Adequate Archaeological Digital Archive?

Twenty years ago the establishment of ADS and *Internet Archaeology* met with some skepticism from some members of the academic world. Given the current emphasis on Open Data and on Open Access to research publications, stemming from both government and funding agencies, it seems as if the establishment has finally caught up. However, if digital archiving and dissemination is to be a routine part of archaeological practice it becomes even more important to define what we mean by a digital archive. From the 1990s, the AHDS developed a series of Guides to Good Practice (Mitcham et al. 2010) which were published as hard copy handbooks as well as being made freely available online; the ADS editions covered domain-specific data types, such as geophysical survey, aerial photography, and GIS. The purpose of the Guides was not to standardize or even advise how specific techniques should be employed. Instead, their role was to define the file formats and metadata standards to follow, if a particular technique was being used, in order to safeguard long-term preservation, and facilitate data re-use. The Guides were aimed at those preparing data for archival deposit, and those running digital repositories. Over the years the Guides have been enhanced and updated, with a major upgrade to a wiki-based system undertaken jointly with the US-based Digital Antiquity consortium (Niven 2013), and new sections to cover additional data types, most recently Dendrochronology (Brewer and Jansma 2016) and Thermoluminescence (Kazakis and Tsirliganis 2016).

One of the first guides, aimed at fieldworkers, had a slightly different focus as it attempted to define the minimum standard for a digital archive from an excavation (Richards and Robinson 2000). In the commercial archaeological environment, where the primary driver can be to keep costs to a minimum, it is essential that those setting the specifications for archaeological work specify what they believe should be part of the archive. The Fieldwork Guide attempted to define a

sliding scale of digital archive, according to the importance of a project. For example, a small site evaluation or watching brief yielding few archaeological features would not warrant a major investment in an archive, although it would still be important to preserve some record of the negative evidence, via an OASIS record. On the other hand, in 2000 the minimum standard for the digital archive accompanying a major project was defined as comprising several key elements (Table 1). These features were regarded as the minimum needed to allow future re-use of a fieldwork archive, on the basis that the archive should be seen as a standalone resource. In practice, some 15 years later, very few projects meet these minimum standards. In particular, stratigraphic matrices are rarely preserved, and spreadsheets of finds and animal bones (nowadays often developed by independent specialists) are not regarded as part of the core archive. Instead a text report and a selection of digital photographs of trenches are too often regarded as the project archive. One suspects, however, that far more primary data may have been collected in digital format, and may well have been born digital. In 2012 it was estimated there were some 2.2 Gb of undeposited digital data comprising over 1.25 million files languishing in the hands of archaeological contractors in England (Smith and Tindall 2012). Cost is undoubtedly a key factor here, but the profession needs to take its responsibilities seriously, and the deposit of a proper digital archive should be part of the standard workflow for any archaeological project, undertaken online on completion of a project. This need not be time-consuming but data archives and archaeological curators need to agree on what is the minimum standard for the 21st century, and enforce it. This is one of the greatest challenges facing archaeological repositories today.

Covering the costs of digital curation

Digital archiving requires human intervention and that comes at a cost. One of the most significant achievements of ADS has been the development of a business model designed to ensure that it is

self-sufficient and sustainable, the key prerequisite for any archive. Whilst ADS was initially supported by research council annual grants to provide a free archiving service for university-based researchers, it was soon recognized that the majority of archaeological research data was actually produced outside higher education, in the commercial and public sectors – variously described as development control or contract archaeology, rescue archaeology, or in continental Europe, preventative archaeology. Most countries, including both the UK and USA, follow the principle that 'the polluter pays' i.e. those funding the development should also pay for any archaeological work deemed necessary. In the UK this traditionally includes any charges levied by the museum for the long-term deposit of the physical archive – including the artifacts and the documentary, photographic and drawn record of the fieldwork. In both the UK and the USA this model has been extended to include the digital archive. A one-off charge levied at the point of deposit pays for the digital preservation. The charge can be passed on by the archaeologist undertaking the fieldwork to the funding body – whether developer, government body, or the research council – ensuring that the data can remain freely available as Open Data to all users.

In the case of ADS the deposit charge pays for the accessioning of the data; the creation, where necessary, of a preservation version in open formats; and of a dissemination version made available from a project web page, with a DOI. A proportion of the deposit charge is set aside in an endowment fund to cover the cost of future migration, although our experience shows that if the appropriate steps are taken at the point of accession then future costs can be minimized (Richards et al. 2010). This business model works well in a discipline where most research, including fieldwork, is undertaken on a project-funding basis. For large projects, the digital preservation charges will generally be less than one per cent of the overall project budget, although for smaller projects (such as a geophysical survey) they can represent a much higher proportion. Here a subscription model may be more appropriate, whereby geophysics contractors pay an annual subscription fee which

allows them to deposit a restricted number of project archives per year. Whilst a state grant-funding model for digital archiving has been pursued in some countries it would be politically unacceptable in the UK, and whilst it might seem more attractive in terms of ensuring a regular income stream, experience shows that government funding is never guaranteed and is vulnerable to changes in administration and policy, particularly in times of austerity when archaeology and heritage become soft targets.

Experience also shows, however, that many archaeologists still fail to safeguard the future of their data by depositing it in a trusted repository, either because they claim they cannot afford it, or forgot to include the costs of preservation when preparing their grant application, or because they are reluctant to make it available to others (maybe either through fear of intellectual property theft or simply embarrassment), or simply because they leave it to the end of the project or even beyond and never get round to it. The ADS has endeavored to simplify the data deposition process by developing a semi-automated file upload system, ADS-Easy (Moore et al. 2013). This automates many of the time-consuming processes formerly carried out by its digital archivists and ensures the collection of adequate metadata using standard templates, thereby qualifying depositors for discounted deposit charges. However, it is clear that many archaeologists still find the process difficult and prefer ADS to undertake the work on their behalf, following what tDAR describe as their 'full-service model' (McManamon this volume).

Whilst there are many supposed 'carrots' which should encourage archaeologists to make their data available, including professional esteem and citation, and feedback from others, it is clear that 'sticks' are far more effective. Where funding bodies make it a requirement that digital data are archived in a trusted repository and made freely available, and will not give out future funding until compliance is demonstrated, then it provides the most powerful incentive. In the UK, the Engineering and Physical Sciences Research Council (EPSRC) has adopted the most robust

position, requiring research organizations to publish online appropriately structured metadata describing the research data they hold, normally within 12 months of the data being generated, and for the data themselves to be made available without restriction for a minimum of 10 years (EPSRC 2016). This has led to a flurry of universities establishing their own institutional repositories and requiring researchers to develop data management plans. However, the effectiveness of the policy will still depend upon how far compliance is monitored and policed and audited. The Arts and Humanities Research Council (AHRC) – the funding body which funds most university-based archaeological research in the UK – has adopted a similar, but slightly more conservative position. Under AHRC funding rules digital resources must be maintained for "a minimum of three years after the end of project funding… but in many, if not most cases a longer period will be appropriate" (AHRC 2016, 66). Historic England, the lead state agency for heritage protection in England has adopted a robust position to make sure that the digital outputs from the work it funds are adequately archived. Under their funding guidance all projects are asked "to ensure that digital archives are deposited with the Archaeology Data Service (ADS) (http://archaeologydataservice.ac.uk/) or a similar recognised digital archiving organisation approved by Historic England (HE)" (Historic England 2016, 14).

So is it worth it?

It is clear, therefore, that data preservation does not come cheap. However, data collection is much more expensive. Primary fieldwork tends to be a major cost for any archaeological research but even where a project seeks to draw upon and synthesize existing data the cost of collecting and cleaning that data can be prohibitively expensive. The relative inaccessibility of grey literature, now being addressed, is a major contributor to that. In attempting to write a synthesis of recent work in British and Irish prehistory Richard Bradley spent three person years in data collection from HER

offices (Bradley 2007). In my own research on the *Viking and Anglo-Saxon Landscape and Economy (VASLE)*, using Portable Antiquities Scheme data for metal-detected finds, two person years of a three-year project had to be spent in data cleaning (Naylor and Richards 2005). Most recently, in Michael Fulford's *Roman Rural Settlement* project (Allen et al. 2016) six person years have been spent in data collection.

However, if the data produced by such projects are now properly archived and made available for others to use a huge amount of future effort can be avoided, as well as new research questions made possible (see for example Kintigh et al. 2015). Every ADS archive has a web page which shows the number of archive views and downloads, generally in large numbers. It should also be noted that unlike publications and data in many disciplines (particularly the sciences) most archaeological data does not age. A report or data set published last century may be just as important today as it as then. Thus the value of the ADS repository increases with time, particularly as more resources are added, leading to a critical mass of data for some topics. Undoubtedly the highest levels of use of the ADS comes from those simply seeking something to read – whether grey literature reports or back-runs of journal articles (Figure 1). These figures alone justify the resources required to make the reports and articles available Open Access. By comparison, whilst still substantial, the figures for the re-use of data are much lower (Figure 2). Thus should not be surprising; the effort and skills involved in downloading a raw data file, understanding the metadata, and opening it up on one's own computer should not be underestimated. Although we have little qualitative data to set besides the quantitative download statistics it is reasonable to assume that in order to invest this much effort the user must have a specific research project in mind. Hence the research value added from re-using a data file may be significantly greater than that from reading an existing report.

The use of DOIs is not yet well enough established to undertake citation analysis of those referencing ADS archives in their bibliographies so evidence for re-use often still relies on anecdotal evidence. In my own university recent examples include a doctoral study of the Mesolithic in Northern England which made extensive use of unpublished fieldwork reports held by ADS, for example (Blinkhorn 2012); another PhD on Anglo-Saxon monetization re-used the artifact database collected for the *VASLE* project, referred to above (Abramson 2017); an externally-funded research project developing techniques for image recognition re-used ADS image databases of flint tools and Viking brooches (Power et al 2017). Similarly, the data provided by the *Roman Rural Settlement* project has already been re-used by a study of Roman brooches (Cool and Baxter 2016).

Nonetheless, it is clear that archaeologists appreciate the value of discipline-based repositories. A comparative study of data re-use across a range of disciplines commissioned by the JISC (Joint Information Systems Committee) and the Research Information Network (RIN 2011) found that of the archaeologists surveyed about the impact of the ADS on their work:

- 84 per cent said ADS had an impact on data sharing
- 79 per cent said ADS reduced the time required for data access and processing
- 51 per cent said ADS provided new intellectual opportunities
- 56 per cent said ADS permitted new types of research
- 94 per cent said the data held by ADS were very or quite important for their research

When compared with other repositories, covering domains ranging from atmospheric research to the social sciences, the impact of the ADS emerged very favourably (Table 2).

In 2013 JISC commissioned an independent study to analyze and survey perceptions of the value of digital collections held by the ADS, and to measure, assess and quantify the economic impact of those collections. The work was undertaken by Neil Beagrie, a digital preservation specialist, and John Haughton, a strategic economist who specializes in using standard economic

procedures to try to assign a value to things that are not usually measured. Whilst such procedures are routine in the public sector, for example in assessing the cost-benefit of a new road or rail infrastructure for instance, they are rarely employed in the cultural sector, but may become increasingly important in persuading funding bodies of the economic impact of the heritage sector. Beagrie and Houghton explored a range of methods and sources, including data from 1996-2011, on the growth of collections and users at ADS and how return on investment grows with the collections (Beagrie and Houghton 2013). Their quantitative analysis suggests that the economic benefits of ADS substantially exceed the operational costs. However, when users were asked what they would be willing to pay for access to ADS, the total came to only £1.1m per annum, probably reflecting the relative low level of funding in the sector, as well as the attitude that access to data should be free at the point of use. When the question was turned around, and users were asked how much they would need to be compensated by if access to ADS was taken away, the total came instead to £7.4m per annum. In addition, a very significant increase in research efficiency was reported by users as a result of using the ADS, which was calculated to be worth at least £13 million per annum – five times the costs of operation, data deposit and use (Figure 3). They also identified a potential increase in return on investment in data creation/collection resulting from the additional use that was facilitated by ADS that may be worth between £2.4 million and £9.7 million per annum over 30 years in net present value from one-year's investment – a two-fold to eight-fold return on investment. Due to the conservative treatment of use and user statistics, the value estimates presented are likely to be conservative. Although Beagrie and Houghton did not directly measure the wider impacts of ADS on society as a whole, the returns on investment provide a window onto those impacts.

Joining it all up

Many countries have now recognized the value of developing their own digital repositories for archaeological data. In addition to the ADS in the UK and tDAR in the USA, there are national repositories in the Netherlands (Gilissen 2013; Hollander 2013), Sweden (Jakobsson 2013), and Germany (Schäfer and Trognitz 2013), and another being established in Austria. However, there is inevitably a tension in bringing resources together in single repositories. On the one hand trusted digital repositories need to reach a critical mass if they are to maintain the staffing levels and range of skills required to achieve sustainability. They also need to have organizational backing and long-term commitment, and if their resources are also to be 'trusted' in the more conventional sense then some reputable institutional imprint is essential. On the other hand, as Huggett (2016) has identified, there is a risk that we are simply creating a new set of data silos that challenge the founding principles of a distributed Internet. In fact, digital repositories need to do both: they must bring resources together and make it easy for users to interrogate them via shared and user-friendly interfaces; but they must also open data up: via APIs, harvesting protocols such as OAI-PMH, and as Linked Open Data (LOD) so that it can be viewed and manipulated via multiple routes (May et al. 2015). The ADS has undertaken some LOD experiments linking excavation database records to fieldwork text reports as part of the STELLAR project. There is also a growing community participating in LOD in archaeology, including services such as Pleiades (a community-built gazetteer of ancient places), Pelagios (also joining up places in the Classical world), and Open Context (the web-based system for publishing archaeological data), some of which are demonstrating research results. Opening up data in this totally permissive sense can sometime be at odds with the more conventional gatekeeper role of national heritage bodies and can create internal tensions. Yet there are good reasons for managing domain-based repositories at national level. This both gives them an appropriate scale and also tends to coincide with legislative remits and heritage protection policies, as well as the scope of most funding streams.

Nonetheless, archaeological research questions rarely coincide with modern political boundaries, which were irrelevant for the vast majority of the timespan of the human past. As early as 1992 Henrik Jarl Hansen expressed the need to join up digital resources across Europe (Hansen 1992), and the European Commission has been a key funding agency which has facilitated several projects in this area. In 2002-4 the ADS led a consortium of European partners on the EU-funded ARENA project (Kenny and Richards 2005). One of the outcomes of the project was a portal which provided a distributed cross-search of sites and monuments records for six countries (Dam et al. 2010). However this relied upon dated technologies such as Z39.50 which had been developed for cross-searching library catalogues. In 2009-10 the ADS was able to work with DANS to migrate the ARENA portal into a more flexible web services architecture. A similar technological infrastructure was also employed in a collaborative project between the ADS and tDAR to build a Transatlantic Archaeological Gateway (TAG) (Jeffrey et al. 2012). Subsequently the ADS has played a significant role in the ARIADNE e-infrastructure (Niccolucci and Richards 2013; Aloia 2017) which has developed a powerful cross-search portal, as well as experiments in Linked Open Data. To achieve interoperability across different European languages and cultures is challenging and adherence to data standards is essential for any level of semantic interoperability and cross-search. In ARIADNE national subject terms have been mapped to a common core standard (in this case the Getty Art and Architecture Thesaurus) and archaeological Period terms have been defined according to explicit criteria, working with the North American *Period0* initiative (Shaw et al. 2015). As always there are challenges in turning projects into services, but by linking ARIADNE to the ESFRI roadmap (ESFRI 2016) via the preparatory phase for a new infrastructure dedicated to Heritage Science (E-RIHS) it is anticipated that ARIADNE can achieve sustainability.

Summary and conclusions

In summary, this paper has attempted to demonstrate that whilst digital preservation has a cost, data collection – and data loss – is much more expensive. If we make digital data easily available then they are re-used, and a number of studies have shown both a research and an economic return on that investment. Whilst the greatest demand is for simple text reports, especially the grey literature, we should also explore the potential offered by new models for data publication and dissemination. Underpinning our digital preservation work is fundamental work on data standards and for this to facilitate interoperability and Big Data projects it is essential that digital archives collaborate at an international level. The experience of ADS, over the last 20 years has demonstrated that, whilst it takes time, it is possible to develop a sustainable business model for a self-sufficient national digital repository for archaeology.

References cited

Abramson, Tony

2017 *'Where there's muck there's brass!' Coinage in the Northumbrian landscape and economy, c.575 – c.867.* Unpublished PhD thesis, University of York.

Allen, Martyn, Nathan Blick, Tom Brindle, Tim Evans, Michael Fulford, Neil Holbrook, Julian D. Richards, and Alex Smith

2015 *The Rural Settlement of Roman Britain: an online resource*. Archaeology Data Service, York. doi:10.5284/1030449.

Aloia, Nicolo, Ceri Binding, Sebastian Cuy, Martin Doerr, Bruno Fanini, Achille Felicetti, Johan Fihn, Dimitris Gavrilis, Guntram Geser, Hella Hollander, Carlo Meghini, Franco Niccolucci, Federico Nurra, Chirstos Papatheodorou, Julian Richards, Paola Ronzino, Roberto Scopigno, Maria Theodoridou, Doug Tudhope, Andreas Vlachidis and Holly Wright

2017 Supporting European Archaeological Research: The ARIADNE E-Infrastructure. *Internet Archaeology,* in press.

Arts and Humanities Funding Council (AHRC)

2016 *Research Funding Guide v.2.8.* Electronic document, http://www.ahrc.ac.uk/documents/guides/research-funding-guide/, accessed December 30, 2016.

Austin, Tony, Jenny Mitcham, and Julian D. Richards

2008 From questions to answers: outcomes from the 'Big Data project'. In *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA) Berlin, Germany, April 2-6, 2007,* edited by Axel Posluschny,

Kartsen Lambers and Irma Herzog, pp. 194-199. Kolloquien zur Vor- und Frahgeschichte Band 10, Bonn.

Beagrie, Neil and John Houghton

2013 *The Value and Impact of the Archaeology Data Service. A study and methods for enhancing sustainability.* http://repository.jisc.ac.uk/id/eprint/5509

Bevan, Andrew and James Conolly

2012 Intensive Survey Data from Antikythera, Greece. *Journal of Open Archaeology Data* 1(1). doi:10.5334/4f3bcb3f7f21d.

2014 *The Antikythera Survey Project*. Archaeology Data Service, York. doi:10.5284/1024569.

Blinkhorn, Edward

2012 *The Mesolithic and the Planning Process*. Unpublished PhD thesis, University of York.

Bradley, Richard

2006 Bridging the Two Cultures – Commercial Archaeology and the Study of Prehistoric Britain. *Antiquaries Journal* 86: 1-13.

2007 *The Prehistory of Britain and Ireland.* Cambridge University Press, Cambridge.

Brewer, Peter and Esther Jansma

2016 *Dendrochronological Data in Archaeology: A Guide to Good Practice.* Archaeology Data Service/ Digital Antiquity Guides to Good Practice. Electronic document, http://guides.archaeologydataservice.ac.uk/g2gp/Dendro_Toc, accessed December 30, 2016

Cool, Hilary E.M. and Mike J. Baxter

2016 Brooches and Britannia. *Britannia* 47: 71-98.

Dam, Claus, Tony Austin and Jon Kenny

2010 Breaking Down National Barriers: ARENA - A Portal to European Heritage Information. In *Beyond the Artifact. Digital Interpretation of the Past. Proceedings of CAA2004, Prato,*

13-17 April 2004, edited by Franco Niccolucci and Sorin Hermon, pp. 94-98. Archaeolingua, Budapest.

Doerr, Martin

2003 The CIDOC Conceptual Reference Model: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* 24 (3): 75–92.

Drap, Pierre

2009 VENUS: Virtual ExploratioN of Underwater Sites: Port-Miou C wreck, Marseille. Archaeology Data Service, York. doi:10.5284/1000004.

Engineering and Physical Sciences Research Council (EPSRC)

2016 *EPSRC Policy framework on research data*. Electronic document, https://www.epsrc.ac.uk/about/standards/researchdata/, accessed December 30, 2016

European Strategy Forum on Research Infrastructures (ESFRI)

2016 Strategy Report on Research Infrastructures. Roadmap 2016. Electronic document, https://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/esfri_roadmap_2016_adopted.pdf, accessed December 30, 2016

Evans, Tim N.L.

2013 Holes in the Archaeological Record? A Comparison of National Event Databases for the Historic Environment in England. *The Historic Environment* 4: 19-34.

Evans, Tim N.L. and Ray H. Moore

2014 The Use of PDF/A in Digital Archives: A Case Study from Archaeology. *International Journal of Digital Curation* 9: 123-138.

Galeazzi, Fabrizio, Marco Callieri, Matteo Dellepiane, Michael Charno, Julian D. Richards and Roberto Scopigno

2016 Web-based Visualization for 3D Data in Archaeology: The ADS 3D viewer. *Journal of Archaeological Science: Reports* 9: 1-11. doi:10.1016/j.jasrep.2016.06.045.

Gilissen, Valentijn

2013 Past the Opening: Building Towards the Present, On-going Dissemination of Dutch Archaeological Data as part of the DANS Archive. In *Opening the Past 2013: Archaeology of the Future,* pp. 42-43, MapPapers 1-III. doi:10.4456/MAPPA.2013.17.

Hansen, Henrik Jarl

1992 European Archaeological Databases: Problems and Prospects. In *Computing the Past. Computer Applications and Quantitative Methods in Archaeology*, edited by Jens Andresen, Torsten Madsen and Irwin Scollar, pp. 229-237. Aarhus University Press, Aarhus.

Hardman, Catherine and Julian D. Richards

2003 OASIS: Dealing with the Digital Revolution. In *CAA2002: The Digital Heritage of Archaeology. Computer Applications and Quantitative Methods in Archaeology 2002*, edited by Martin Doerr and Apostolis Sarris, pp. 325-328. Archive of Monuments and Publications Hellenic Ministry of Culture, Athens.

Historic England

2016 *Guidance for Grants Projects.* Electronic document, http://www.english-heritage.org.uk/publications/guidance-grants-projects/grants-guidance.pdf, accessed December 30, 2016.

Hollander, Hella

2013 The Netherlands: The e-Depot for Dutch Archaeology: Archiving and Publication of Archaeological Data. In *Proceedings of the 18th International Conference on Cultural Heritage and New Technologies, Vienna, Austria, November 2013.* Electronic document, http://www.chnt.at/wp-content/uploads/Hollander_2014.pdf, accessed December 30, 2016.

Holmberg, Karen

2010 Placing Immateriality: Situating the Material of Highland Chiriquí. *Internet Archaeology* 28. doi:10.11141/ia.28.5.

Huggett, Jeremy

2016 Digital Data Realities. Electronic document, https://introspectivedigitalarchaeology.wordpress.com/2016/06/29/digital-data realities/#more-502, accessed December 30, 2016.

Jakobsson, Ulf

2013 Sweden: Swedish National Data (SND) Services, the OAIS Reference Model and Archaeological Data, In *Proceedings of the 18th International Conference on Cultural Heritage and New Technologies, Vienna, Austria, November 2013*. Electronic document, http://www.chnt.at/wp-content/uploads/Jakobsson_2014.pdf, accessed December 30, 2016.

Jeffrey, Stuart

2012 A New Digital Dark Age? Collaborative Web Tools, Social Media and Long-term Preservation. *World Archaeology* 44: 553-570.

Jeffrey, Stuart, Julian D. Richards, Fabio Ciravegna, Stewart Waller, Sam Chapman and Ziqi Zhang

2009 The Archaeotools Project: Faceted Classification and Natural Language Processing in an Archaeological Context. In *Crossing Boundaries: Computational Science, E-Science and Global E-Infrastructures*, edited by Peter Coveney, Special Themed Issue of the *Philosophical Transactions of the Royal Society A*, 367: 2507-2519.

Jeffrey, Stuart, Lei Xia, Julian Richards, Jon Bateman, Keith Kintigh, Francis Pierce-McManamon and Adam Brin

2012 The Transatlantic Archaeology Gateway: Bridging the Digital Ocean. In *CAA2011- Revive the Past: Proceedings of the 39th Conference in Computer Applications and Quantitative*

Methods in Archaeology, Beijing, China, 12-16 April 2011, edited by Mingquan Zhou, Iza Romanowska, Zhongke Wu, Pengfei Xu, and Philip Verhagen, pp. 198-208. Pallas Publications, Amsterdam.

Kazakis, Nikolaos A. and Nestor C. Tsirliganis

2016 Thermol*uminescence Dating: A Guide to Good Practice*. Archaeology Data Service/ Digital Antiquity Guides to Good Practice. Electronic document, http://guides.archaeologydataservice.ac.uk/g2gp/TL_Toc, accessed December 30, 2016

Kenny, Jon and Julian D. Richards

2005 Pathways to a Shared European Information Infrastructure for Cultural Heritage. *Internet Archaeology* 18. doi:10.11141/ia.18.6.

Kintigh, Keith, Jeffrey Altschul, Ann Kinzig, Fredrick Limp, William Michener, Jeremy Sabloff, Edward Hackett, Timothy Kohler, Bertram Ludäscher and Clifford Lynch

2015 Cultural Dynamics, Deep Time, and Data: Planning Cyberinfrastructure Investments for Archaeology. *Advances in Archaeological Practice* 3(1): 1-15.

Lavoie, Brian

2014 The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition). *DPC Technology Watch Report 14-02*. doi:10.7207/twr14-02.

May, Keith, Ceri Binding and Doug Tudhope

2015 Barriers and Opportunities for Linked Open Data use in Archaeology and Cultural Heritage. *Archäologische Informationen* 38: 173-184.

Mitcham, Jen and Catherine Hardman

2011 ADS and the Data Seal of Approval – Case Study for the DCC. *Digital Curation Centre - Case-Studies*. Electronic document, http://www.dcc.ac.uk/resources/case-studies/ads-dsa, accessed December 30, 2016.

Mitcham, Jen, Kieron Niven and Julian D. Richards

2010 Archiving Archaeology: Introducing the Guides to Good Practice. In *Proceedings of the 7th International Conference on Preservation of Digital Objects. iPRES 2010*, edited by Andreas Rauber, Max Kaiser, Rebecca Guenthier, and Panos Constantopoulos, pp. 183-187. Books@ocg.net Band 262.

Moore, Ray, Catherine Hardman, Lei Xia and Julian D. Richards

2013 ADS Easy: an Automated E-archiving System for Archaeology. In *Archaeology in the Digital Era*. *Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, 26-29 March 2012*, edited by Graeme Earl, Tim Sly, Angeliki Chrysanthi, Patricia Murrieta-Flores, Constantinos Papadopoulos, Iza Romanowska and David Wheatley, pp. 299-306. Amsterdam University Press, Amsterdam.

Naylor, John and Julian D. Richards

2005 Third Party Data for First Class Research. *Archeologia e Calcolatori* 16: 83-91.

Niccolucci, Franco and Julian D. Richards

2013 ARIADNE: Advanced Research Infrastructures for Archaeological Dataset Networking in Europe. *International Journal of Humanities and Arts Computing* 7: 70-88.

Niven, Kieron (editor)

2013 *Caring for Digital Data in Archaeology.* ADS/ Digital Antiquity Guides to Good Practice, Oxbow Books, Oxford.

Power, Christopher, Andrew Lewis, Helen Petrie, Katie Green, Julian Richards, Mark Eramian, Brittany Chan, Ekta Walia, Isaac Sijaranamual and Maarten De Rijke

2017 Improving Archaeologists' Online Archive Experiences Through User-Centred Design. *ACM Journal on Computing and Cultural Heritage* 10. http://dx.doi.org/10.1145/2983917

Richards, Julian D.

2008 Managing Digital Preservation and Access: The Archaeology Data Service. In *Managing Archaeological Resources: Global context, national programs, local actions*, edited by Francis P. McManamon, Andrew Stout and Jodi A. Barnes, pp. 173-194. One World Archaeology 58, Left Coast Press, Walnut Creek.

2015 Ahead of the Curve: Adventures in E-publishing in *Internet Archaeology*. *Archäologische Informationen* 38: 63-71.

Richards, Julian, Tony F. Austin and Catherine Hardman

2010 Covering the Costs of Digital Curation. *Heritage Managemen*t 3: 255-263.

Richards, Julian D., Stuart Jeffrey, Stewart Waller, Fabio Ciravegna, Sam Chapman and Ziqi Zhang

2011 The Archaeology Data Service and the Archaeotools Project: Faceted Classification and Natural Language Processing. In *Archaeology 2.0 and Beyond: New Tools for Collaboration and Communication*, edited by Sarah Whitcher Kansa, Eric C. Kansa and Ethan Watrall, pp. 31-56. Cotsen Institute of Archaeology Press, Los Angeles.

Richards, Julian D., Kieron Niven and Stuart Jeffrey

2013 Preserving our Digital Heritage: Information Systems for Data Management and Preservation. In *Visual Heritage in the Digital Age*, edited by Eugene Ch'ng and Vince Gaffney, pp. 311-326. Springer-Verlag, London.

Richards, Julian D. and Damian Robinson (editors)

2000 *Guide to Good Practice for Excavation and Fieldwork*. Arts and Humanities Data Service, Oxbow Books, Oxford.

Richards, Julian D., Doug Tudhope and Andreas Vlachidis

2015 Text Mining in Archaeology: Extracting Information from Archaeological Reports. In *Mathematics in Archaeology*, edited by Juan A. Barceló and Igor Bogdanovic, pp. 240-254. Science Publishers, Boca Raton, Florida.

Richards, Julian D., Judith Winters and Michael Charno

2011 Making the LEAP: Linking Electronic Archives and Publications. In *On the Road to Reconstructing the Past, Proceedings of the 36th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA), Budapest, April 2-6. 2008*, edited by Erzsébet Jerem, Ferenc Redö and Vajk Szeverényi, pp. 141-146, Archaeolingua, Budapest.

RIN (Research Information Network)

2011 *Data Centres: their Use, Value and Impact.* The Research Information Network, London. Electronic document, http://www.rin.ac.uk/system/files/attachments/Data_Centres_Report.pdf, accessed December 30, 2016.

Schäfer, Felix and Martina Trognitz

2013 Developing a Research Data Centre for Germany: IANUS and its IT-guidelines. In *Abstracts of the 18th International Conference on Cultural Heritage and New Technologies, Vienna, Austria, November 2013.* Electronic document, http://www.chnt.at/ianus/, accessed December 30, 2016

Shaw, Ryan, Adam Rabinowitz, Patrick Golden, and Eric Kansa

2015 A Sharing-oriented Design Strategy for Networked Knowledge Organization Systems, *International Journal on Digital Libraries*. doi:10.1007/s00799-015-0164-0.

Smith, Roland and Adrian Tindall

2012 *A Survey of Archaeological Archives Held by Archaeological Practices in England,*

Scotland and Wales, Federation of Archaeological Managers and Employers. Electronic document, http://www.famearchaeology.co.uk/wp-content/uploads/2012/12/FAME-Report-on-Archaeological-Archives-Final.pdf, accessed December 30, 2016.

Tudhope, Doug, Ceri Binding, Stuart Jeffrey, Keith May and Andreas Vlachidis

2011 A STELLAR Role for Knowledge Organization Systems in Digital Archaeology. *Bulletin of the American Society for Information Science and Technology* 37 (4): 15-18.

Tudhope, Douglas, Keith May, Ceri Binding and Andreas Vlachidis

2011 Connecting Archaeological Data and Grey Literature via Semantic Cross Search. *Internet Archaeology* 30. doi:10.11141/ia.30.5.

Vlachidis, Andreas and Douglas Tudhope

2013 The Semantics of Negation Detection in Archaeological Grey Literature. In *Metadata and Semantics Research*, edited by Emmanuoel Garoufallou and Jane Greenberg. *Communications in Computer and Information Science* 390: 188-200.

Williams, Alan N., Sean Ulm, Mike Smith and Jill Reid

2014 AustArch: A Database of ^{14}C and Non-^{14}C Ages from Archaeological Sites in Australia - Composition, Compilation and Review (Data Paper). *Internet Archaeology* 36. doi:10.11141/ia.36.6.