ARCHAEOLOGY DATA SERVICE

Five Year Plan

April 2008 - March 2013

Document History
Version 1.3 (final internal draft; last revised 25 June 2008)
Version 1.4 (revised following Oct 1 Management Committee meeting)

Julian Richards
Director, ADS

Summary

Archaeological research creates large quantities of digital data, in an ever-increasing variety of formats. These data are often born digital and may represent the primary record of unrepeatable fieldwork interventions or laboratory analyses. It is essential that such data are preserved as they provide the basis for all existing and new interpretations. Yet digital data are fragile. Complex data formats and the maintenance of the functionality of data sets requires disciplinary-based knowledge and hard-won expertise. New technologies continue to create new challenges for the digital archivist.

This document was first written at the request of AHRC, as part of their review of funding of digital preservation services, and in support of a bid from ADS for continued funding. It outlines the strategic aims of ADS for the next 5 years and also provides a 5-year business plan for how these can be achieved. It will be reviewed at meetings of the ADS Management Committee to assess progress against these objectives.

Background

Digital data are fragile. They require active curation, underpinned by migration, refreshment and back-up. When Newham Museum Archaeological Service was closed down in 1996 the archaeologists had the foresight to send their orphaned digital data to the Archaeology Data Service (ADS). The archive comprised 220 floppy discs and 6432 individual files, representing 1500 excavation reports, or parts of reports. ADS discovered that 5% of the disks were already irrecoverable, having suffered physical degradation of the magnetic media. A further 900 files were unreadable as they were in obsolete and redundant file formats, including old versions of CAD word-processing programs for which no migration path now existed. The main problem, however, was lack of adequate documentation. Without the codes used to record the data, complete spreadsheets and databases of artefactual and skeletal evidence were reduced to meaningless lists of numbers and letters. The significance of the Newham case study is that it demonstrates that researchers should plan for preservation and re-use, and that forward planning is much cheaper.

The ADS has come a long way in the 10 years since it was established. It has developed metadata standards and clear guidelines for the documentation of data sets, including a library of Guides to Good Practice. It has built up a rich archive of primary digital data, managed in a Collections Management System compliant with the Open Archival Information System (OAIS) ISO Standard for digital repositories.

However, the creation of digital data does not stand still. As well as the ongoing task of migrating and preserving existing resources ADS has to embrace emerging techniques which create new file formats and larger and larger data sets. The ADS has addressed some of these issues in the ‘Big Data’ project (funded by English Heritage). In the VENUS project (with European funding) it is developing guidelines for the archiving of data types produced by remote operated vehicles (ROVs) used in the exploration of deep underwater sites. However, this is an ongoing and constantly changing environment in which institutional repositories are incapable of maintaining complex digital objects which frequently demand disciplinary knowledge to ensure the preservation of their significant properties. Other bodies have emerged which can provide advice on digital preservation, such as the Digital Curation Centre and Digital Preservation Coalition. The ADS is unique, however, in that it has built up active expertise in the day-to-day research and practice of digital preservation.

Research Context

Archaeologists routinely make use of ICT in support of their research. Although other arts and humanities disciplines are making increasing use of ICT methods there remains a qualitative difference with archaeology, where digital data are primary and greater in scale and where there is a greater interest in cutting edge techniques and a more mature usage of established technologies. This is partly because the quantities of data have encouraged archaeological take up of electronic data processing since its invention, and partly because many aspects of archaeological research overlap with the biological and physical sciences, encouraging archaeologists to adopt scientific methods. Archaeological usage includes primary digital data collection during fieldwork or in the laboratory, where data are increasingly 'born digital'. They may be captured by geophysics survey equipment, electronic data loggers, and hand-held PDAs, and by increasingly sophisticated techniques, such as LiDAR, 3-D laser scanning and 3-D photogrammetry, and echo-sounding. Additional data may be transferred from paper-based pro forma into complex digital databases and GIS, which are then treated as primary for the purposes of analysis and re-use. These data generally represent the primary digital record of a precious and finite resource which has been destroyed during the process of excavation and recording. The data are fundamental to support higher level observation and theories and are essential in order to allow the evaluation and testing of interpretations. They also represent an invaluable resource to support new research, obviating the need for new and expensive data collection. Therefore there is a strong professional ethic (going back to the nineteenth century and reflected in the current Codes of Conducts of the professional associations – the Institute of Field Archaeologists and the Society for American Archaeology) which requires the preparation and deposit of an archive record. In the twenty-first century this obligates those bodies which fund archaeological research to ensure the provision of appropriate digital archival facilities.

There has also been an increasing trend to complement traditional forms of publication with electronic dissemination of supporting data. This goes back to a series of working parties, starting with the Frere Report (1975) and leading most recently to the recommendations of the Publication User Needs Survey conducted by the Council for British Archaeology. This trend is partly driven by economics. Full hard copy publication of archaeological excavation has become prohibitively expensive and would lead to a major increase in the cost of awards in archaeology. However, electronic dissemination is also often more appropriate and leads to enhanced value from access to numeric and other structured data. It permits a seamless interface between electronic publication of results and interpretation, and access to supporting archival data.

History

The ADS was established in 1996 as one of the discipline-based services under the AHDS umbrella. Unlike the other AHDS services, it has established a role which bridges the academic and professional archaeological sectors, reflecting the fact that there is an equivalent need for digital preservation support outside academia, and that much data which has tremendous value for archaeological research in universities is created outside the academic sector. Core funding, provided (via AHDS) by JISC and AHRC until 2008, has allowed the free provision of advice and preservation services for those employed in Higher Education Institutions, whilst ADS

has developed a charging policy to cover the costs of preservation of research data funded by other bodies (such as quasi-governmental bodies such as English Heritage, or commercial developers like Rail Link Engineering for the Channel Tunnel Rail Link). With the withdrawal of AHRC and JISC funding from AHDS, AHRC support for ADS has been maintained, in recognition of the importance of the primary data sets it preserves, and its ability to generate other revenue streams.

Current situation

From 1 April 2008 AHRC withdrew funding of AHDS but agreed to continue to fund advice and preservation services for Archaeology for five years, but with a view to shifting to responsive funding from April 2013. This was partly on the basis that ADS had developed a charging policy and had demonstrated the potential of income generation from other sectors. However, a responsive funding model poses a number of questions in terms of providing a stable business model for ADS:

1. How can a small and variable number of grant awards sustain advice and technical appendix review services for a larger number of applications (typically five times larger than the number of awards)?
2. Why should a small number of new awards bear the costs of preservation of legacy data acquired before the introduction of the charging policy?
3. How can responsive mode funding cover the maintenance of core background functions (such as standards development and Guides to Good Practice)?
4. How can the development of new business (R&D projects) be factored into responsive mode funding?
5. If the case for continued funding is to be based on usage, how can responsive mode funding be used to encourage re-use or investigate usage figures?

These questions will be faced by a growing number of digital archives and institutional repositories so their investigation may be of wider interest. The ADS bid to AHRC proposed that:

> "Over the course of the five year project it is intended to investigate the extension of charging models developed for the public and professional sectors of archaeology to the academic sector, and to evaluate at the end of three and five years how far responsive mode funding can provide a sustainable business model for future preservation services for archaeology."

As background to this Five Year Plan, it is also worth quoting our response to one of the AHRC grant assessors:

> "Our charging policy provides the basis for a responsive mode operation, but it is currently predicated on charging the additional costs of each deposit, assuming that core infrastructure costs have been covered by core funding. The revised model will involve shifting infrastructure costs from core funding to responsive mode. We will need to reconfigure the charges applied to other funders and model the effects. The 'hard edge or modelling' requested will be an outcome of this project. In reality the transitional elements can only relate to 2011-12 and onwards rather than the whole five year period, because of the time lag in new awards coming to completion. We already share both assessors concerns about the reliability of this income stream given the uncertainties associated with project funding, and have concerns about its impact on individual AHRC grant applicants. Given the small volume of AHRC

awards within our sector each year, we wonder if it may always be necessary to guarantee some minimum threshold core payment to allow us to maintain continuity of staffing. However, this would become clearer at the three year review point.

It is also important to note that whether ADS is financed via core grant, or by charging an overhead on individual responsive mode awards, funding will ultimately come from the public purse, as is the case with any other library or archive service throughout the world. Responsive mode funding does not replace public subsidy; it just changes the route by which it is paid, and loads it onto individual grant applications. Assessor 1 is correct in saying that this removes security, which for an archive is worrying. One of the objectives of the five year funding period is to explore ways to offset this."

The aim of the Five Year Plan is to develop ADS as a sustainable digital preservation service for Archaeology by May 2013, and to investigate how far costs can be transferred to responsive mode funding. There will be an interim report after three years - by May 2011 - on progress towards meeting this aim.

Strategic Aims and objectives

(1) To provide preservation services for primary data and digital resources created in the course of research in archaeology and related disciplines, following the Open Archival Information System (OAIS) ISO standard.

(2) To provide advice to researchers and funding bodies on the creation, dissemination, documentation, and preservation of digital resources created in the course of research in archaeology and related disciplines.

(3) To work with others to develop and implement agreed standards to ensure appropriate documentation and preservation of digital resources created in the course of research in archaeology and related disciplines.

(4) To provide open and easy online access to primary data and digital resources created in the course of research in archaeology and related disciplines, either as downloadable datasets or through online interfaces which allow users to interrogate key data sets online.

(5) To provide encouragement and support for the re-use of primary data and digital resources created in the course of research in archaeology and related disciplines.

(6) To provide appropriate finding aids and resource discovery mechanisms to allow users to discover primary data and digital resources created in the course of research in archaeology and related disciplines.

(7) To undertake research and development into preservation, access and interoperability, with a view to enhancing all the above services.

(8) To maintain effective service management and administration in pursuit of the above objectives, and develop and implement cost models appropriate to higher education, public and commercial archaeology sectors.

These eight objectives reflect the areas of activity which provide core functions for ADS, and can be cross-referenced to the funding matrix diagram and annual workplans and reports. Underpinning all is the need to provide financial sustainability for each area, and to demonstrate the interplay of different sources of funding and how the support of funding in one area may provide added value for another funding body.

1. Preservation services
2. Providing advice to data creators
3. Developing standards
4. Providing access to data
5. Providing support for re-use
6. Resource discovery / interoperability
7. Undertaking research & development
8. Service management & administration

Each of the objectives can also be defined in terms of core minimum staffing levels, and the resources required in order to achieve them. In the table below the second and third columns indicate the staffing and approx costs required to deliver a core service. The fourth and fifth columns indicate the current/desirable staffing levels. These assume provision for research and development and associated project management costs, as well as an increase in staffing levels associated with volume.

Objective	Core Minimum staffing	Approx cost	Current/ desirable staffing	Approx additnl cost
1. Preservation services	Systems Manager	£45,000	Additional curatorial officers according to volume +2	£64,000
2. Providing advice to data creators	Collections Development Manager (0.5 FTE)	£22,500	Collections Development Manager (0.5)	£22,500
3. Developing standards			Standards editor	£35,000
4. Providing access to data	Curatorial Officer	£32,000	Additional curatorial officers according to volume +2	£64,000
5. Providing support for re-use			User Services Manager (c.0.5)	£22,500
6. Resource discovery / interoperability			Applications Developer (c.0.5)	£20,000
7. Undertaking research & development			Applications Developer (c.0.5)	£20,000
			User Services Manager - project mgt (c.0.5)	£22,500
8. Service management & administration	Director (0.1 FTE)	£10,000	Director (0.1 FTE)	£10,000
	Administrator (0.5 FTE)	£15,000	Administrator (0.5 FTE)	£15,000
	Running costs	£25,000		£15,000
Sub-total		**£149,500**		**£310,500**
	Overheads/ FEC @ 46%	£68,700		£142,800
	Preservation legacy	£10,000		£10,000
Totals		**£228,200**		**£463,300**

NB Salary costs are based on average salaries plus on-costs for the current grade

Funded	
Currently unfunded	

Work programme

Objective	Action	Sector/ Funder	Funding	Owner/ Action	Start date	Target	Completed/ Progress
1. Preservation services	1.1 Revise charging policy			CSH	Nov 07	Apr 08	Nov 07
	1.2 Apply charging policy	AHRC	Agreed	CSH	May 08	Jun 08	Jun 08
		British Academy	Under investigation	JDR / CSH	Apr 08	Apr 09	Meeting held July 08
		NERC	Agreed	JDR / CSH	Sep 07	Oct 08	Meeting Sep 07 Awaiting contract
		English Heritage	Agreed	CSH	Nov 07	Apr 08	Jan 08
		Developer-funded	Agreed	CSH	Sep 07	Apr 08	Sep 07
		Leverhulme	ACTION required	CSH	Jan 09	Apr 11	
		EU-funded projects	ACTION required	CSH	Jan 09	Apr 11	
		Heritage Lottery fund	Under investigation	JDR / CSH	Apr 08	Apr 09	Tender awarded Oct 08
	1.3 Model charging policy income			CSH	Jan 11	Apr 11	
	1.4 Develop offline storage in UKDA	AHRC	Agreed	AFA	Nov 07	Apr 09	Server installed
	1.5 Develop and maintain technical infrastructure	ALL	Agreed – covered by charging policy	AFA	Nov 07	Mar 13	Ongoing
	1.6 Develop Fedora based repository	?	ACTION required	AFA / SJW	Jul 10	Oct 11	Need to identify means of funding
	1.7 Develop preservation policy and	?	ACTION required	AFA / JLM	Oct 08	Oct 11	Dependent upon overhead income,

Objective	Action	Sector/ Funder	Funding	Owner/ Action	Start date	Target	Completed/ Progress
	undertake TRAC certification						and external certification framework
2. Providing advice to data creators	2.1 Technical Appendix review	AHRC	Funded until Apr 2103; ACTION required	CSH	Apr 11	Mar 13	
	2.2 Advice to applicants etc	AHRC	Funded until Apr 2013; ACTION required	CSH	Apr 11	Mar 13	
		British Academy	Under investigation	JDR / CSH	Apr 08	Apr 11	Meeting planned
		NERC	Under investigation	JDR / CSH	Apr 08	Oct 08	Costs supplied Jun 08
		English Heritage (ALSF)	Agreed for specific programmes	CSH	Apr 08	Apr 11	In place for ALSF
		Heritage Lottery fund	Under investigation	JDR / CSH	Jun 08	Apr 09	Monitor status awarded Oct 08
3. Developing Standards	3.1 Big Data G2GP	English Heritage	Agreed	JDR	Jan 08	Oct 08	Proposal agreed Feb 09
	3.2 Underwater G2GP	EU (VENUS)	Agreed	SJ	Nov 07	Apr 09	Workshop held Nov 08
	3.3 G2GP review & 2nd editions	Mellon Foundation	Agreed	JDR	Jan 08	Apr 09	2 year funding agreed
	3.4 Standards review	EU (ACE)	Agreed	SJ	Apr 08	Oct 10	
	3.5 Committee attendance	Use UoY overhead to subsidise	Agreed	JDR	Apr 08	Apr 08	Apr 08
4. Providing access to data	4.1 ArchSearch III – facetted browsing	AHRC Archaeotools	Agreed	SJ / SJW	Sep 07	Apr 09	
	4.2 Deeper access – data mining	AHRC Archaeotools pilot	Agreed	SJ	May 08	Oct 09	

Objective	Action	Sector/ Funder	Funding	Owner/ Action	Start date	Target	Completed/ Progress
		Service implementation	ACTION required	SJ / JDR	Oct 09	Oct 11	Need to identify funding source
	4.3 Grey literature library	?	ACTION required	CSH / JDR	Apr 08	Apr 11	Need to identify funder
	4.4 Archaeology image bank	HEA	ACTION required	SJ / JDR	Apr 08	Apr 09	Need to identify funding source
	4.5 Journals	Publisher		CSH / JDR	Apr 08	Apr 11	Self funding basis
	4.6 PhD theses	?	ACTION required	CSH / JDR	Apr 08	Apr 11	Need to identify funder
5. Providing support for re-use	5.1 Help desk 5.2 Outreach 5.3 Visits 5.4 Newsletters	?	Unfunded; ACTION required	SJ / JDR	Apr 08	Apr 11	Need to identify business model/ funding
6. Resource Discovery/ Interoperability	6.1 OASIS	English Heritage/ Historic Scotland	Agreed	CSH	Jan 08	Jun 08	Apr 08
	6.2 'Wyvern' – integrated forms	English Heritage/ Historic Scotland/ CADW/ DoENI	Under investigation	CSH	May 08	Apr 10	Scoping & pilot study proposed
	6.3 Collection level descriptions	Intute	Agreed	SJ / KJN	Oct 07	Oct 07	Oct 07
	6.4 European web services registry and gateway	EU ACE/ EU DARIAH	Agreed	SJ / SJW	Sep 08	Oct 10	
	6.5 Transatlantic gateway	Mellon Foundation	Under investigation	JDR	Apr 09	Apr 11	
7. Undertaking Research & Development	7.1 Build in resource to allow development of R&D bids	Retain proportion of UoY overhead to invest in R&D	Agreed	JDR	Apr 08	Oct 08	Approved UoY Planning Committee Jun 08
8. Service management & administration	8.1 Year One Annual report	AHRC	Agreed	JDR	Aug 07	Sep 08	
	8.2 Year Two	AHRC	Agreed	JDR	Aug 08	Sep 09	

Objective	Action	Sector/ Funder	Funding	Owner/ Action	Start date	Target	Completed/ Progress
	Annual report						
	8.3 Year Three Annual report	AHRC	Agreed	JDR	Aug 09	Sep 10	
	8.4 Year Three Evaluation	AHRC	Agreed	JDR	Jan 11	Apr 11	
	8.5 Year Four Annual report	AHRC	Agreed	JDR	Aug 10	Sep 11	
	8.6 Year Five Final report	AHRC	Agreed	JDR	Jan 13	Mar 13	
	8.7 Staff training – half day a week	Funded from UoY overhead	Agreed	CSH SJW	Jan 08	Apr 08	Jan 08
	8.8 Staff skills audit	N/A	ACTION required	JDR	Jan 11	Apr 11	
	8.9 Review business model	AHRC	Agreed	JDR	Apr 11	Sep 11	
	8.10 Revise business model as necessary			JDR	Sep 11	Mar 13	