Archaeology Data Service:

Preservation Policy

Authors: Tony Austin (ADS Systems Manager)
Julian Richards (ADS Director)
Maintained by: Systems Manager

Version: 1.3 final
Date: 30 September 2009
Last updated: 27 January 2011
Review Due: September 2011 (unless significant change)

Update history:

Date	Version	Description
27 Jan 2011	1.3.1	Web links updated

1. Principal Statement[1]

'The Archaeology Data Service (ADS) supports research, learning and teaching with high quality and dependable digital resources. It does this by preserving digital data in the long term, and by promoting and disseminating a broad range of data in archaeology. The ADS promotes good practice in the use of digital data in archaeology, it provides technical advice to the research community, and supports the deployment of digital technologies'[2]

The long term preservation and reuse (reuse value in itself aids preservation) of digital data is then core to ADS activities in providing 'high quality and dependable digital resources' to its user community. The latter has broadened over time from a largely academic focus to encompass a range of groups with an interest in Archaeology including commercial archaeology, heritage organisations, museums, Further and Secondary Education, community archaeology and the interested public in general.

The ADS actively follows preservation and management strategies based on this policy with the aim of ensuring the authenticity, reliability and logical integrity of all resources entrusted to its care. It further endeavours to provide its user community with usable versions for research, teaching or learning, in perpetuity.

2. Contextual Links

[1] Beagrie, N., Semple, N., Williams, P. & Wright, R. 2008. Digital Preservation Policies Study Part 1: Final Report for the JISC. provides the structure of this document http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf

[2] Mission Statement http://archaeologydataservice.ac.uk/about

This document systematizes an overview of archival practice developed by the ADS since its inception in 1998. It does not exist in isolation but as part of a suite of documents guiding good governance and practice by the ADS. Policy and strategy documents include

- ADS Five Year Plan: April 2008 - March 2013[3] (strategy document)
- ADS Risk Register[4]
- ADS Collections Policy (4th Edition)[5]
- ADS Preservation Strategy
- ADS Disaster Recovery Plan[6]
- ADS Access Policy (in prep)

The ADS is further governed by the policy and strategy of its host institution; the University of York. Documents include

- University of York Records Management Policy 2004[7]
- University of York Information Access and Security Policy[8]
- University of York Legal Statements and linked policy and strategy documents therein[9]

As noted in the Collections Policy the ADS has agreements with a number of funding agencies that support archaeological research, to encourage funding recipients to offer their datasets for deposit[10]

---

[3] http://ads.ahds.ac.uk/manage/agendas/management/011008/ADSFiveYearPlan.pdf (internal)

[4] http://ads.ahds.ac.uk/manage/agendas/management/011008/RiskRegister.pdf (internal)

[5] http://archaeologydataservice.ac.uk/advice/collectionsPolicy

[6] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=DisasterRecoveryPlan (internal)

[7] http://www.york.ac.uk/recordsmanagement/rm/policy.htm

[8] http://www.york.ac.uk/admin/po/cmte/information/information%20access%20and%20security%20policy.rtf

[9] http://www.york.ac.uk/docs/disclaimer/disclaimer.htm

[10] http://archaeologydataservice.ac.uk/advice/collectionsPolicy#section-collectionsPolicy-2.5.AcquisitionStrategies

- Arts and Humanities Research Council (AHRC)
- Natural Environment Research Council (NERC) , for science-based archaeology

The ADS has Service Level Agreements (SLA) with a number of organisations including

- The UK Data Archive (UKDA) [11] for provision of a remote deep storage facility
- To host and provide an image preservation service to the Parks and Gardens Data Service (PGDS)[12]
- To host and provide a preservation service to the online journal Internet Archaeology[13]

Further, the ADS has Memoranda of Understanding (MoU) with a number of external organisations concerned with preservation and reuse of data[14] including

- The Association of Local Government Archaeological Officers (ALGAO)
- The Council for British Archaeology (CBA)
- The Royal Commission on the Ancient and Historical Monuments of Scotland (RCAHMS)
- The Royal Commission on the Ancient and Historical Monuments of Wales (RCAHMW)
- The Royal Commission on the Historical Monuments of England (RCHME now part of English Heritage)
- The mda (formerly the Museums Documentation Association)
- The National Trust

3. Preservation Objectives

[11] http://www.data-archive.ac.uk/

[12] http://www.parksandgardens.ac.uk/

[13] http://intarch.ac.uk/

[14] http://archaeologydataservice.ac.uk/about/memorandaOfUnderstanding

The core objective of the long term preservation of digital data for reuse by a broad archaeological community has been described above.

The ADS endeavours to undertake long term preservation working within a framework conforming to the ISO (14721:2003) specification of a reference model for an Open Archival Information System (OAIS) as defined by a recommendation of the Consultative Committee for Space Data Systems[15].

OAIS provides a conceptual framework in which to discuss and compare archives through developing a common language. It describes the responsibilities and interactions of Producers, Managers and Consumers of digital and paper records. It defines processes necessary for the ingest, long-term preservation and dissemination of information objects.

Specifically the model describes a series of 'transformations, both logical and physical, of the Information Package and its associated objects as they follow a lifecycle from the Producer to the OAIS and from the OAIS to the Consumer'. These packages comprise

- Submission Information Package (SIP): Supplied by a data Producer (creator or depositor) including documentation to facilitate archiving and reuse

- Archival Information Package (AIP): Generated from the SIP and the long term preservation package managed within the OAIS including administrative, technical and reuse documentation

- Dissemination Information Package (DIP): Generated from the SIP/AIP and made available to Consumers (users) including documentation to facilitate reuse.

Clearly OAIS influences archival policy and strategy significantly. OAIS does not proscribe preservation strategies but the active management and lifecycle approaches tend toward migration in various forms rather than other techniques like emulation or technology preservation. The ADS uses a number of migration types for ongoing preservation

- Normalisation: Data may exist natively or is migrated to widely supported open international standards such as ASCII (text) or TIFF (images).

- Version migration: Data is migrated through successive versions of a format. For example, TIFF 5.0 is migrated to TIFF 6.0. Version migration may be the only option for preserving proprietary formats that don't migrate to open standards. This is only practical where the software using proprietary formats is widely used within a community

[15] http://public.ccsds.org/publications/archive/650x0b1.pdf

and accessible (affordable) to an archive. It is not practical for an archive to maintain a suite of limited use proprietary software.

- Format migration: As well as normalisation data may be migrated to other formats for a number of reasons including dissemination. For example, a spatial dataset may be preserved as GML but disseminated as an ESRI Shape file. ESRI software sees wide usage amongst the archaeological community.

- Refreshment: Migration between media which leave data (the bit stream) totally unchanged. For example, from one system to another.

Data that cannot be normalised and/or migrated between versions is not suited to long term preservation within the framework described.

As well as the physical process of preservation OAIS describes Preservation Description Information (PDI) as the 'information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context information' which is preserved with an AIP

- Provenance information: Concerned with 'history' and records, for example, 'the principal investigator'.

- Reference information: Concerned with unambiguously identifying content information through, for example, the provision of an ISBN number for a publication.

- Fixity Information: A fixity value or checksum provides a simple way to protect the integrity of data by detecting errors in data. The MD5 (Message-Digest algorithm 5) and the SHA (Secure Hash Algorithm) are widely used cryptographic hash functions. Applying these algorithms to a file produces an (almost certainly) unique hash or checksum value and will consistently produce this value if a file is unchanged. Thus it provides a mechanism for validating and auditing data.

- Context information: In terms of OAIS is concerned with environment. Examples include 'why the Content Information was created and how it relates to other Content Information objects'.

Documentation including metadata concerned with resource discovery and reuse is then an equally important part of an archival package.

The above defines two of the cornerstones for a successful archival strategy within an OAIS framework

- Use of software (by Producers) supporting formats with clear migration paths for both preservation and reuse.

- The existence of adequate documentation to facilitate ongoing preservation and reuse.

The other cornerstones are

- Ongoing access to adequate hardware systems by skilled staff.
- That robust backup/recovery strategies are in place.

It is widely recognised that there are inherent weaknesses associated with these last two points; equipment fails or needs replacing, skilled staff leave or are difficult to recruit, digital media are notoriously frail to name some. These weaknesses can be quantified through risk assessment[16] and lessened through forward planning including disaster recovery[17] and systems budgets[18].

In terms of reuse the ADS currently supports open access to its holdings (some data may be subject to a time limited embargo at the behest of a Producer or for legal reasons). The contents of most collections are available online. Because of bandwidth concerns larger files may only be available on request either as a specifically organised download or on portable media for which charges at cost may be made. The ADS is actively investigating various network technologies such as Point of Access (PoA) optical networks and Grid Computing seeking better mechanisms for disseminating 'big data'[19]

In order to quantify and qualify success in reaching these stated objectives the ADS actively seeks compliance with the certification document Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist[20] authored by the US based Research Libraries Group (RLG) part of the Online Computer Library Center (OCLC), the Center for Research Libraries (CRL) and the National Archives and Records Administration (NARA). The purpose of the checklist is identifying repositories capable of reliably managing digital collections. The audit checklist is closely tied to the OAIS reference model in terms of a conceptual framework and terminology and considers organisational suitability, repository workflows, user communities and usability of data, and the underlying technical infrastructure including security. All of these areas must be openly documented. Organisations that can demonstrate that they meet the criteria within the checklist will be identified as Trusted Digital Repositories.

---

[16] http://ads.ahds.ac.uk/manage/agendas/management/011008/RiskRegister.pdf (internal)

[17] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=DisasterRecoveryPlan (internal)

[18] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=SystemsBudgetsThreeYearPlan (internal)

[19] http://archaeologydataservice.ac.uk/research/bigData

[20] http://catalog.crl.edu/search~S1?/Xtrac&searchscope=1&SORT=R/Xtrac&searchscope=1&SORT=R&SUBKEY=trac/1,6,6,B/l856~b2212602&FF=Xtrac&searchscope=1&SORT=R&3,3,,1,0

The ADS is currently undergoing self certification[21]. Although mechanisms for formal certification are as yet not in place the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) self-audit toolkit developed by the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE)[22] is one possible direction. It should be noted that DRAMBORA is not OAIS specific. DRAMBORA is also seen as being much more quantitative 'Most existing methods are too static, with too much reference on the OAIS model and too little emphasis on evidence in the auditing process'[23]

4. Identification of Content

Clearly content is driven by community; what the community is producing and what it wants to reuse. Also, as described above, the ADS uses migration in various forms as a long term preservation strategy. This influences which formats the ADS accept. Current practice with regard to content is set out in detail in the ADS Collections Policy (4th Edition)[24].

The cost of long term preservation also influences content. The ADS follows a costing model where Producers are expected to pay for the long term preservation of their data. Where the ADS receives core funding from organisations such as the AHRC basic charges are waived for projects funded by the same organisation. Other projects are subject to the ADS Charging Policy[25]. Thus projects may need to build long term preservation costs into funding applications.

.

5. Procedural Accountability

ADS staff have established job descriptions which define roles and responsibilities. These are formalised following review by the University of York using the Higher Education Role Analysis (HERA) job evaluation methodology[26]. Accountabilities pertaining to preservation and reuse are

- Director: Overall responsibility for financial management and for policy including compliance with legislation affecting digital preservation and its management.

[21] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=TRACcompliance (internal)

[22] http://www.repositoryaudit.eu/

[23] http://www.jisc.ac.uk/events/2007/06/repositories_conference/repos_pres_session34_notes.aspx

[24] http://archaeologydataservice.ac.uk/advice/collectionsPolicy

[25] http://archaeologydataservice.ac.uk/advice/chargingPolicy

[26] http://www.york.ac.uk/univ/mis/cfm/herarecs/

- Collections Manager: Responsible for approaching grant holders, negotiating with depositors and acquiring access to collections; managing collection services for the ADS; first point of contact for information about data deposition, joint cataloguing, or data access and re-use.

- Systems Manager: Responsibilities include monitoring and developing management and preservation strategies for digital data; planning, selecting, purchasing and commissioning new computer equipment; evaluating, purchasing and the installation of software packages; overseeing system and network security of all ADS systems; actively managing digital data deposited with the ADS.

- User Services Manager: responsible for promoting the ADS and usage of its online catalogue and digital resources amongst user communities; undertaking research and project management.

- Administrator: Responsible for essential administrative and financial management.

- Application developer: Responsible for the development of software applications and user interfaces; Accessioning, mounting, cataloguing, validation, conversion, migration and curation of data sets; undertaking data audits and discussion with clients (Producers); and answering user queries.

- Curatorial staff: Responsible for accessioning, mounting, cataloguing, validation, conversion, migration and curation of data sets; development of user interfaces; undertaking data audits and discussion with clients (Producers); and answering user queries.

- All staff: Accountable to their line managers for compliance with this policy and with related policies, strategies, standards and guidelines.

The ADS also has recourse to its Management Committee though it should be noted that this group acts in a purely advisory capacity and without legal liability[27].

6. Guidance and Implementation

The ADS came into being in 1998 as one of the data services grouped under an Arts and Humanities Data Service (AHDS – no longer extant) umbrella. As such it was and still is very much involved in the lifecycle approach to long term preservation as, for example, defined by Neil Beagrie and Dan Greenstein then of the AHDS in their 1998 publication A Strategic Policy Framework for Creating and Preserving Digital Collections[28].

[27] http://archaeologydataservice.ac.uk/about/management

[28] http://www.ukoln.ac.uk/services/papers/bl/framework/framework.html

The generally recognised categories of the lifecycle of digital assets are (equivalent OAIS functional entities in brackets)

- Data creation (Administration)
- Acquisition, retention or disposal (Ingest, Administration)
- Preservation and management (Archival Storage, Data Management, Administration)
- Access and use (Access, Administration)

The ADS maintain a purpose built Collections Management System (CMS)[29] that is used to track and document potential and actual collections of data throughout this lifecycle. The CMS is modular and broadly follows the above flow with People, Tracking, Accessions and User Services modules. Additionally there are Assist (help) and Admin (input controls and security) modules.

6.1 Data creation

Lead role: Collections Manager
Policy document: Collections Policy

The pre-ingest period of a resource or potential resource is of major importance from the time a project is conceptualised. Whereas a well formed SIP aids repository processes a poorly formed one may well preclude ingest (see 6.2). For a SIP to be well formed it must conform to a repository's requirements

The ADS is active in a number of ways in providing guidance to potential depositors during this period including

- Collections Policy[30]
- Guides to Good Practice[31]
- Advisory services[32]
- Guidelines for depositors[33]

---

[29] http://muninn.york.ac.uk/cms/ (internal)

[30] http://archaeologydataservice.ac.uk/advice/collectionsPolicy

[31] http://archaeologydataservice.ac.uk/advice/g2gp

[32] http://archaeologydataservice.ac.uk/advice/adviceForAHRCGrantApplicants

[33] http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors

- Research initiatives such as the 'Big Data' and VENUS projects[34]

The Guides to Good Practice are currently being revised and expanded as an online Wiki[35] with funding primarily from the Mellon Foundation and English Heritage.

6.2 Acquisition, retention or disposal

Lead role: Systems Manager
Policy document: Preservation Policy

A number of documents guide the process of ingesting a SIP including

- Repository Operations[36]
- Ingest Procedures (Ingest Manual)[37]
- Data Procedures (dealing with specific data types and file formats)[38]
- Procedure Checklists[39]
- File formats table - delivery, preservation and presentation[40]. This is further extended by a formats review during the 'Big Data' project for some of the more unusual technologies used by archaeologists[41]
- Security Overview[42]
- Software Inventory (in prep)

---

[34] http://archaeologydataservice.ac.uk/research

[35] http://guides.archaeologydataservice.ac.uk/

[36] http://archaeologydataservice.ac.uk/attach/preservation/ADS_Repository_Operations_V2.pdf

[37] http://archaeologydataservice.ac.uk/attach/preservation/ADS_ingest_manual_V2.pdf

[38] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=DataProcedures (internal)

[39] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=ProcedureChecklists (internal)

[40] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=FileFormats (internal)

[41] http://archaeologydataservice.ac.uk/research/bigDataFormats

[42] http://muninn.york.ac.uk/wiki/Edit.jsp?page=SecurityOverview (internal)

The existence of a SIP and a signed deposit licence pertaining to it triggers accessioning. The licence grants a non-exclusive right to the ADS to distribute supplied data. Copyright is not transferred[43]. As a first step in the process of ingest supplied data is virus checked.

The ADS uses the concept of a collection of digital objects to describe a discrete resource. Thus a collection may be related to a distinct project. Necessarily any number of accessions (SIPs) of related objects may be made into a collection as a project may be ongoing either submitting data in stages or providing reloads (sometimes known as editions). A producer may also deposit multiple collections pertaining to different projects. Collections and accessions are assigned identifiers which are unique within ADS systems.

As already described the ADS migrates files from a producer supplied SIP into its systems in various formats as part of a corresponding AIP (for preservation) and DIP (for dissemination). The retention of the significant properties of files is a primary concern during any migration as detailed in ADS Data Procedures. Copies of supplied files are also maintained in the same systems which are known within the ADS as the original files. These reflect files as delivered in terms of format and content but they may have been processed to, for example, remove spaces from file names (Unix based systems cannot process file names containing spaces).

A formalised directory structure is built under folders reflecting collections and accessions identifiers. These comprise

- original (contains files supplied in the SIP which may have seen some processing as described above)
- preservation (contains the AIP data – see also admin)
- dissemination (contains the DIP)
- previous (contains data that has been updated by a depositor including previous editions of a resource)
- admin (contains data concerned with the administration of a resource including licence information, collection level metadata, preservation metadata; in OAIS terms the Preservation Description Information noted earlier)

As noted the DIP is created as part of the accessioning process; however, the concept of dissemination on demand is becoming an increasingly attractive solution in that less files need to be stored and managed and multiple software applications can be supported. For example, GIS files can be stored as GML and, using GDAL libraries, files in formats used by ESRI or MAPINFO

[43] http://archaeologydataservice.ac.uk/attach/guidelinesForDepositors/ads_licence_form.pdf

software generated on request. This scenario is being actively investigated by the ADS[44].

All processing is recorded in the Collections Management System (CMS) already noted. This is done at the batch level. The ADS is in the process of implementing Fedora Commons for the file level management of resources[45]. This will record metadata; attributes such as physical location, filename, size, format and fixity value of individual files which will allow for greater automation and fluidity in management processes concerned with integrity, authenticity and version control.

Occasionally files are included in a SIP that are not suitable for ingest either by accident, through the lack of a clear preservation path or inadequate documentation. These files are deleted following consultation with the producer. A similar scenario might occasionally apply to an entire SIP.

Currently delivery media of fully accessioned SIPs are retained for a minimum of six months (from time of accessioning) at which point they may be disposed of. In certain circumstances media may be returned to a supplier, for example, where data has been provided on a portable hard drive.

6.3 Preservation and management

Lead role: Systems Manager
Policy document: Preservation Policy

A number of documents on the ADS Wiki[46] inform the ongoing preservation and management of data. These together provide the detail and hence the preservation strategy sitting under this policy.

6.3.1 Storage and resilience

The ADS maintain multiple copies of data in order to facilitate disaster recovery (i.e. to provide resilience). Core data in the form of AIPs and original (the slightly altered SIP) files is regularly synchronised from a local copy in the University of York to a dedicated off site store maintained in the machine room of the UK Data Archive at the University of Essex[47]. The server is running a RAID 5 disk configuration which allows rapid recovery from disk

[44] This is discussed more fully in the Big Data report (p. 15) http://archaeologydataservice.ac.uk/attach/bigDataDeliverables/bigdata_final_report_1.3.pdf

[45] http://www.fedora-commons.org/about/

[46] http://muninn.york.ac.uk/wiki/ (internal)

[47] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=OffsiteStorageFacility (internal)

failure. In the interests of security outside access to this server is via an encrypted SSH tunnel[48] from nominated IP addresses. Data is further backed up to tape by the UKDA (see UKDA Preservation Policy[49]).

Dissemination data or DIPs are maintained on the main ADS production server in the machine room of the Computing Service at the University of York. The Computing Service further back up this data to tape and maintain off site copies of the tapes. Currently the backup system uses Legato Networker and an Adic Scalar tape library. The system involves daily (over-night), weekly and monthly backups to a fixed number of media so tapes are recycled.

Record based data is currently maintained in an Oracle 10 database which is similarly backed up to tape. It is also stored as delimited text and synchronised off site as part of a collection as described above.

Data compression is generally seen as something to be avoided for the preservation copies of files[50]. There are good reasons for this for even lossless compression techniques where bit or byte losses can cause much more damage to files in formats such as jpg or zip that use compression techniques than uncompressed files[51]. This informs on the formats preferred for archiving and on storage. In short preservation data is not compressed for storage by the ADS even though the saving on storage would be significant.

6.3.2 Data management

As already noted the ADS maintain a custom built Collections Management System within which the Accessions Module pertains to data management. Beyond detailing the accessioning of data into a collection it holds 'Generic Preservation MetaData'. The GPMD metadata schema, developed by the AHDS, holds data describing the processing of files such as the conversion of supplied files into versions for the AIP and DIP or later migrations to different versions and formats.

Whilst this processing documentation provides a useful record it is also problematic in that recording can be at either a file or batch level whereas fixity or checksum (MD5, SHA-1, etc) information which documents the outcome of the process and the ongoing integrity of the file is necessarily recorded at the file level. This dichotomy hinders management in terms of automating processes such as auditing and versioning. Consequently, the

[48] http://en.wikipedia.org/wiki/SSH_tunneling

[49] http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf

[50] http://www.erpanet.org/advisory/list.php?start=5&end=10

[51] http://old.hki.uni-koeln.de/people/herrmann/forschung/heydegger_archiving2008_40.pdf (presented at the Archiving 2008 conference)

ADS is moving towards using the open source Fedora Commons repository software[52] for file level management including process history.

Data refreshment is an ongoing process. It is undertaken regularly (minimally on a weekly basis) during the already noted synchronisation of locally held data to an off site data repository within the UKDA. This one way synchronisation compares checksum values at source and destination to detect change and acts accordingly. Refreshment also occurs during periodic system upgrades where data is moved between media. Again synchronisation and comparison software is used to monitor the movement onto new media.

As already described file migration between formats is a common activity during the accessioning process but can also occur throughout the lifecycle of a file. It may become necessary for a number of reasons including

- version change (many formats change or evolve over time)
- format obsolescence (a format is or is becoming deprecated)
- another format becomes a more attractive preservation option

An ongoing Technology Watch is maintained by ADS Curatorial and Technical staff and acted upon as and when necessary. As with migrations during accessioning it is important that the significant properties of a file are retained. However, it should be noted that in some cases significant properties may be altered in order to ensure ongoing preservation and usability (document formatting might be such a case). As such migrations are likely to be complex involving the DIPs or AIPs of multiple resources and multiple systems a migration plan is drawn up before commencing operations[53].

These processes then carry on throughout the lifecycle of data held by the ADS. It was noted in the Principal Statement (1) at the beginning of this document that the avowed intention of the ADS is preservation 'in perpetuity'. However, all lifecycles have a beginning and an end and that some are shorter than others. Thus the reality is that we can only talk about the foreseeable future and there are a number of reasons why a resource or part thereof might have a limited lifecycle including

- There is a breach of the agreement detailed in the deposit licence that cannot be resolved (deposit licence[54] clause 8.9.1)
- A depositor (producer) no longer wishes to make a resource available (deposit licence [ibid] clause 8.9.2)

---

[52] http://www.fedora-commons.org/

[53] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=FileMigration (internal - including example plan)

[54] http://archaeologydataservice.ac.uk/attach/guidelinesForDepositors/ads_licence_form.pdf

- A resource was deposited with a formally agreed lifespan
- A resource is completely subsumed within a new resource
- A resource or part thereof no longer has a suitable migration path for ongoing preservation

In all such cases the ADS will endeavour to contact depositors (or their organisations) to discuss the situation. The data in question may be removed from ADS systems following discussion. It may be returnable to a depositor in certain circumstances (this service may be chargeable). End of life events will be detailed in the CMS.

The ADS maintains, and adds to when circumstances allow, a Preservation Legacy Fund. Should the ADS cease to be a viable organisation the Fund will be used to provide an exit strategy that ensures the ongoing preservation of the data in its care.

6.4 Access and use

Lead role: User Services Manager
Policy document: Collections Policy (part), Preservation Policy (part), an Access Policy is 'in prep'

This section is concerned with the access and use of the Dissemination Information Package or DIP; finding a resource, rights management and receiving a data collection or part thereof. It is also concerned with the availability, reliability and security of delivery systems. As already noted reuse of data can aid preservation.

A dedicated post of User Services Manager has responsibility for investigating ways of aiding and encouraging the use of its collections.

6.4.1 Prerequisites

Access and use of resources held by the ADS is governed by a legal and regulatory framework

- A Deposit Licence for each resource[55]
- Copyright and Liability Statements[56]
- A Common Access Agreement[57]

[55] http://archaeologydataservice.ac.uk/attach/guidelinesForDepositors/ads_licence_form.pdf

[56] http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess

6.4.2 Resource discovery

It should be noted that the ADS holds two distinct types of dissemination data

- DIPs representing a discrete archive which contain files in various formats

- Record level datasets or collections. These may be available as standalone searchable datasets or as part of the ADS union catalogue the contents of which range from national reference collections to single records describing the accessible part of a resource; the DIP.

The ADS uses a qualified Dublin Core metadata schema for describing the collections it holds which reflects its roots as a onetime AHDS Service Provider[58]. Where practical various thesauri are used in order to standardise the terminology used to describe collections[59]. This record level data is currently stored in an Oracle 10 database and is available online through ArchSearch; the ADS union catalogue[60].

Other organisations also consume (use) ADS resource discovery metadata which aids the exposure of resources. For example,

- The intute subject portal[61] for study and research holds collections level metadata about many ADS resources

- The ADS provides record level data mapped to the MIDAS XML schema to the Heritage Gateway[62] as Web Services

- The ADS similarly provides record level data to the Z39.50 based Archaeological Records of Europe - Networked Access (ARENA) portal[63]

- An Open Archive Initiative repository implementation OAI-PMH v2.0 is also maintained by the ADS[64]

---

[57] http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess

[58] http://ahds.ac.uk/public/metadata/discovery.html

[59] http://archaeologydataservice.ac.uk/advice/depositCreate3#section-depositCreate3-2.3.Part3DocumentingTheProject

[60] http://archaeologydataservice.ac.uk/archsearch/

[61] http://www.intute.ac.uk/

[62] http://www.heritagegateway.org.uk/gateway/

[63] http://ads.ahds.ac.uk/arena/

[64] http://ads.ahds.ac.uk/oaicat/, http://ads.ahds.ac.uk/oaicat2/

As a further aid to resource discovery the ADS maintains a list of permanent URLs for its resources. When requested by a user these are resolved using a look up table to the current location of the resource. In being permanent this service must be maintained but the ADS is also actively investigating the use of a formal methodology for persistent identification in the form of Digital Object Identifiers[65] either using a Handles system[66] directly or as part of another system or project.

6.4.3 Rights management[67]

Access to the holdings of the ADS is free at the point of use to users for research and educational purposes. All users are required to accept the terms and conditions of the ADS Copyright and Liability statements and to the AHDS Common Access Agreement before they can use ArchSearch or access any of our archived data (see 6.4.1).

The ADS reserves the right to control the downloading of some or all resources by a system of user authentication at some point in the future.

6.4.4 Receiving data

ADS data is largely available online. Because of possible bandwidth issues some larger datasets may only be made available on request for a dedicated download. Some large datasets may be deemed as too big to deliver via a network but may be supplied on portable media. There may be charges for these services. Note charges would be for staff time in setting up deliveries and not for the data itself.

6.4.5 Security of delivery systems

A number of documents have relevance here

- Systems Overview[68]
- Risk Assessment[69]

---

[65] http://en.wikipedia.org/wiki/Digital_object_identifier

[66] http://www.handle.net/

[67] http://archaeologydataservice.ac.uk/advice/collectionsPolicy#section-collectionsPolicy-5.2.RightsManagement (Collections Policy 5.2)

[68] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=SystemsOverview (internal)

[69] http://ads.ahds.ac.uk/manage/agendas/management/011008/RiskRegister.pdf (internal)

- Disaster Recovery Plan[70]
- Systems Budget[71]
- Software Inventory (in prep)

ADS delivery systems sit behind the University of York firewall within the Computing Service machine room. This provides a relatively safe environment and basic level of security. All production servers are either under warranty or on maintenance contracts with a next business day service. Delivery systems are backed up to tape (as 6.3.1) and external hard drives.

Core applications are installed on mirrored disks with the mirror taking over without loss of service should one disk fail. Thus in most cases of disk failure the mirror would operate until the problem disk is replaced next working day. In the case of multiple disk or other hardware failures recovery from backup media would be facilitated next working day following replacements or repairs.

Application upgrades and migrations between applications are planned and documented unless these constitute a minor operation.

6.4.6 Consumer access analysis

Analytics inform on consumer activity. They can be used to feed back into dissemination systems. The ADS uses tools or services including

- Web access statistics are generated[72] using the Analog log file analyser package[73]
- The use of Google Analytics[74] is currently being trialed on selected resources. The University of York Legal Statements[75] covers this usage

6.4.7 Outage

[70] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=DisasterRecoveryPlan (internal)

[71] http://muninn.york.ac.uk/wiki/Wiki.jsp?page=SystemsBudgetsThreeYearPlan (internal)

[72] http://archaeologydataservice.ac.uk/about/accessStatistics

[73] http://www.analog.cx/

[74] http://www.google.com/analytics/

[75] http://www.york.ac.uk/docs/disclaimer/disclaimer.htm

Records are kept wherever possible of service downtime both organisational (ADS) and institutional (University of York). There is a scheduled maintenance period of Tuesdays 8-9am (UK time). Services may be unavailable during this period.

7. Glossary

Accession: A deposit into a Collection (ADS)

ADS: Archaeology Data Service

AHDS: Arts and Humanities Data Service (now defunct)

AHRC: Arts and Humanities Research Council

AIP: Archival Information Package (OAIS)

ALGAO: Association of Local Government Archaeological Officers

Big Data: An EH funded research project looking at preservation and management strategies for large datasets. See also VENUS

CBA: Council for British Archaeology

CCSDS: Consultative Committee for Space Data Systems (OAIS)

Checksum: see fixity metadata

CMS: Collections Management System (here an ADS system)

Collection: A collection consists of one or more Accessions

Consumer: A user of data (OAIS)

Context information: Concerned with environment. Examples include ‘why the Content Information was created and how it relates to other Content Information objects’ (OAIS)

Data integrity: ensuring data is whole or complete and continues in this state (see fixity metadata)

Digital preservation: Ongoing managed activity to ensure continued access to authentic versions of content

DIP: Dissemination Information Package (OAIS)

DOI: Digital Object Identifier. Managed system for persistent identification of content-related entities on digital networks

DRAMBORA: Digital Repository Audit Method Based on Risk Assessment – self audit toolkit

EH: English Heritage

Fedora Commons: Flexible Extensible Digital Object Repository Architecture – a digital asset management system

Fixity information: Part of a PDI. A fixity value or checksum provides a simple way to protect the integrity of data by detecting errors in data. The MD5 (Message-Digest algorithm 5) and the SHA (Secure Hash Algorithm) are widely used cryptographic hash functions. Applying these algorithms to a file produces an (almost certainly) unique hash or checksum value and will consistently produce this value if a file is unchanged. Thus it provides a mechanism for validating and auditing data

Format migration: Moving data from one format to another. Particular attention should be paid here to maintaining the significant properties of files

GDAL: Geospatial Data Abstraction Library

GIS: Geographic Information System

GML: Geography Markup Language

GPMD: Generic Preservation MetaData (AHDS)

Handles: A DOI technology specification for assigning, managing, and resolving persistent identifiers

MD5: Message-Digest algorithm 5 is a widely used cryptographic hash function. Used to generate checksum or fixity values. See also SHA-1

Metadata: Data about other data (e.g. fixity, PDI and resource discovery information)

MoU: Memoranda of Understanding

OAI: Open Archives Initiative

OAIS: Open Archival Information System

Outage: System downtime whether planned or unplanned

NERC: Natural Environment Research Council

Normalisation: process of migrating files into widely supported open international standards

PDI: Preservation Description Information (OAIS)

PGDS: Parks and Gardens Data Service

Producer: A creator of data (OAIS)

Provenance information: Part of a PDI. Concerned with ‘history’ and records, for example, ‘the principal investigator’ (OAIS)

RAID: Redundant Array of Inexpensive Disks - a technology that provides high levels of storage reliability

RCAHMS: Royal Commission on the Ancient and Historical Monuments of Scotland

RCAHMW: Royal Commission on the Ancient and Historical Monuments of Wales

RCHME: Royal Commission on the Historical Monuments of England (now part of EH)

Reference information: Part of a PDI. Concerned with unambiguously identifying content information through, for example, the provision of an ISBN number for a publication (OAIS)

Refreshment: Migration between media which leave data (the bit stream) totally unchanged. For example, from one system to another

SHA-1: Secure Hash Algorithm is a widely used cryptographic hash function. Used to generate checksum or fixity values. See also MD5

SIP: Submission Information Package (OAIS)

Significant properties: The essential characteristics of a digital object which must be preserved over time for the digital object to remain accessible and meaningful. Proper understanding of the significant properties of digital objects is critical to establish best practices and helps answer the fundamental question related to digital preservation (DPC)

SLA: Strategic Level Agreement

TRAC: Trustworthy Repositories Audit & Certification

UKDA: UK Data Archive

University of York: ADS host organisation

VENUS: Virtual ExploratioN of Underwater Sites; a research project funded under the EU Sixth Framework Programme. As part of the project team the ADS investigated archival strategies, as well as the developing a Guide To Good Practice on managing marine archaeological data. See also Big Data

Version migration: Migrating data through successive versions of a format

Web Services: A software system designed to support interoperable machine-to-machine interaction over a network.

Z39.50: A client/server protocol for the simultaneous searching and retrieving of information from a number of geographically remote databases. Being superseded by Web Services