



Preservation Policy

Version 1.5

Created date:	30 September 2009
Last updated:	07 March 2018
Review Due:	August 2018 (unless significant change)
Authors:	Tim Evans and Ray Moore
Maintained by:	Digital Archivists
Previous version:	Preservation Policy 1.4

1. Principal Statement ¹

'The Archaeology Data Service (ADS) is an accredited digital repository for heritage data that supports research, learning and teaching with free available, high quality and dependable digital resources by preserving and disseminating digital data in the long term. The ADS also promotes good practice in the use of digital data, provides technical advice to the heritage 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. At the same time the ADS adheres to the FAIR data principles³ providing findable, accessible, interoperable and reusable datasets to its user community which can be utilised for research, teaching or learning, in perpetuity.

2. Contextual Links

This document systematizes an overview of archival practice developed by the ADS since its inception in 1996. 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

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

² See the ADS' [Mission Statement](#).

³ <https://www.ukdataservice.ac.uk/news-and-events/newsitem/?id=4615>

- *ADS Five Year Plan: October 2016 October 2021*⁴ (strategy document)
- *ADS Risk Register*⁵
- *ADS Collections Policy*⁶
- *ADS Repository Operations*⁷
- *ADS Disaster Recovery Plan*⁸

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 Policy⁹
- University of York Information Access and Security Policy.¹⁰
- University of York Legal Statements and linked policy and strategy documents¹¹

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.¹²

- Arts and Humanities Research Council (AHRC)
- British Academy
- Carnegie Trust
- Natural Environment Research Council (NERC), for science-based archaeology
- Historic England
- Leverhulme Trust
- Wellcome History of Medicine Project

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

- The UK Data Archive¹³ (UKDA) for provision of a remote deep storage facility
- To host and provide a preservation service to the online journal *Internet Archaeology*¹⁴

The ADS has Memoranda of Understanding (MoU) with a number of external organisations concerned with preservation and reuse of data including:¹⁵

- Association of British Geological Survey¹⁶
- Association of Local Government Archaeological Officers (ALGAO)¹⁷
- Council for British Archaeology (CBA)¹⁸

⁴ <http://archaeologydataservice.ac.uk/resources/attach/strategicPlan/ADSFiveYearPlan2016-21.pdf>

⁵ http://archaeologydataservice.ac.uk/resources/attach/ADS_Risk_Register_v1-10.pdf

⁶ <http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>

⁷ <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>

⁸ http://archaeologydataservice.ac.uk/resources/attach/ADS_Disaster_Recovery_Plan_v1-8_redacted.pdf

⁹ <https://www.york.ac.uk/records-management/records/policy/>

¹⁰ <https://www.york.ac.uk/about/departments/support-and-admin/information-services/information-policy/index/>

¹¹ <http://www.york.ac.uk/docs/disclaimer/disclaimer.htm>

¹² <http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml> particularly section 2.5

Acquisition Strategies.

¹³ <http://www.data-archive.ac.uk/>

¹⁴ <http://intarch.ac.uk/>

¹⁵ <http://archaeologydataservice.ac.uk/about/memorandaOfUnderstanding>

¹⁶ <http://archaeologydataservice.ac.uk/about/bgsMOU.xhtml>

¹⁷ <http://archaeologydataservice.ac.uk/about/algaoMOU.xhtml>

- Royal Commission on the Ancient and Historical Monuments of Scotland (now Historic Environment Scotland, HES)¹⁹
- Royal Commission on the Ancient and Historical Monuments of Wales (RCAHMW)²⁰
- Royal Commission on the Historical Monuments of England (RCHME now part of Historic England)²¹
- MDA (now Collections Trust)²²
- National Trust²³
- Bedern Group²⁴

3. Preservation Objectives

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.²⁵

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

¹⁸ <http://archaeologydataservice.ac.uk/about/cbaMOU.xhtml>

¹⁹ <http://archaeologydataservice.ac.uk/about/rcahmsMOU.xhtml>

²⁰ <http://archaeologydataservice.ac.uk/about/rcahmwMOU.xhtml>

²¹ <http://archaeologydataservice.ac.uk/about/rcahmeMOU.xhtml>

²² <http://archaeologydataservice.ac.uk/about/mdaMOU.xhtml>

²³ <http://archaeologydataservice.ac.uk/about/ntMOU.xhtml>

²⁴ <http://archaeologydataservice.ac.uk/about/Bedern.xhtml>

²⁵ Consultative Committee for Space Data Systems (2012) *Reference Model for an Open Archival Information System (OAIS)*. *Magenta Book*. Issue 2. June 2012.
<https://public.ccsds.org/Pubs/650x0m2.pdf>

- Normalisation: Data may exist natively or is normalised to a 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, AutoCAD Release 9 (AC1004) has been migrated to AutoCAD Release 2010/11/12 (AC1024).²⁶ 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 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 3.2 but disseminated as an ESRI Shapefile. 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, typically, 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

²⁶ See Green, K., Niven, K. and Field G. (2016) 'Migrating 2 and 3D Datasets: Preserving AutoCAD at the Archaeology Data Service'. ISPRS *Int. J. Geo-Inf.*, **5(4)**, 44. <https://doi.org/10.3390/ijgi5040044>.

- 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 assessment²⁷ and lessened through forward planning including disaster recovery²⁸ and systems budgets.²⁹ A fuller account of the ADS implementation of the OAIS model can be found in *Repository Operations* document.³⁰

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 and ethical 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'.³¹

In order to quantify and qualify success in reaching these stated objectives the ADS actively seeks compliance with community driven initiatives for best practice:

1. *Trustworthy Repositories Audit and Certification* (TRAC).³² Criteria and Checklist authored by the US Centre for Research Libraries. The purpose of the checklist is identifying repositories capable of reliably managing digital collections. The ADS regularly undertakes self-certification on an annual basis.
2. *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA).³³ Developed jointly by the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE) DRAMBORA provides a method for self-assessment, encouraging organisations to establish a comprehensive self-awareness of objectives, activities and assets and identify, assess and manage risks implicit within their organisation.
3. *The Data Seal of Approval* (DSA). The Data Seal of Approval was established by a number of institutions committed to the long-term archiving of research data. By assigning the seal, the DSA group seeks to guarantee the durability of the data concerned, but also to promote the goal of durable archiving in general. The Data Seal of Approval is granted to repositories that are committed to archiving and providing access to scholarly research data in a sustainable way. It is assigned by the DSA Board and renewed on a regular basis through a modification procedure.³⁴
4. *CoreTrustSeal Data Repository Certification*. Collaboration between the DSA and ICSU World Data System (WDS), under the umbrella of the Research Data Alliance, has seen the merger of these two data repositories certifications into the CoreTrustSeal. This replaces the certification offered by the DSA and was released in November 2017.³⁵

²⁷ http://archaeologydataservice.ac.uk/resources/attach/ADS_Risk_Register_v1-10.pdf

²⁸ http://archaeologydataservice.ac.uk/resources/attach/ADS_Disaster_Recovery_Plan_v1-8_redacted.pdf

²⁹ The ADS has an annual systems budget for renewal of physical hardware. This budget is reviewed and set by the ADS Director, Administrator and Applications Development team.

³⁰ <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>

³¹ <http://archaeologydataservice.ac.uk/research/bigData>

³² <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories>

³³ <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora>

³⁴ http://www.data-archive.ac.uk/media/57322/dsa_overview.pdf

³⁵ <https://www.coretrustseal.org/about/>

5. *NESTOR-Seal*. Based on the DIN 31644 standard “Criteria for trustworthy digital archives” the NESTOR-seal provides ‘extended certification’ of compliance and trustworthiness for digital repositories. The assessment covers both organisational and technical aspects of the digital archive and builds on ‘basic’ assessment offered by the DSA.³⁶

The ADS applied and was awarded the DSA in 2010³⁷ and subsequently in 2013, and currently holds the latest version of the DSA (2014-2017).³⁸ The ADS is hopeful of making an application for the new CoreTrustSeal during 2018 and continues to investigate the appropriateness of the NESTOR-seal with a view to future application.

4. Identification of Content

Content is driven by community; what the community is producing and what it wants to reuse. Also, as described above, the ADS uses normalisation and 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*.³⁹ All projects are subject to the ADS *Charging Policy*.⁴⁰ Thus projects will 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.⁴¹

Accountability pertaining to preservation and reuse falls to:

- Director: Overall responsibility for financial management and for policy including compliance with legislation affecting digital preservation and its management.
- Administrator: Responsible for essential administrative and financial management.
- Collections Development 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.
- Communications and Access Manager: responsible for developing and managing ADS communication and access strategy, promoting the ADS, with overall responsibility for user services and outreach activities.
- European Projects Manager: manages the ADS contribution to a range of major European research projects.
- Applications Developer (Systems Management): 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. Responsible for the development of software applications and user interfaces;

³⁶ Harmsen, H et al. (2013) Explanatory notes on the Nestor Seal for Trustworthy Digital Archives.
<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2013100901>

³⁷ Mitcham, J and Hardman, C (2011) ADS and the Data Seal of Approval – case study for the DCC.
<http://www.dcc.ac.uk/resources/case-studies/ads-dsa>

³⁸ https://assessment.datasealofapproval.org/assessment_96/seal/html/

³⁹ <http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>

⁴⁰ <http://archaeologydataservice.ac.uk/advice/chargingPolicy.xhtml>

⁴¹ <https://www.york.ac.uk/admin/hr/browse/pay-and-grading/role-evaluation/>

- Digital Archivists: 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. Monitoring and developing management and preservation strategies for digital data; ensuring compliance with preservation best practice and certification; ensuring secure offsite backing up of data. Digital archivists also monitor redundancy and integrity checks of individual and complete ADS archives on a regular basis.
- 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.⁴²

6. Guidance and Implementation

The ADS came into being in 1996 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*.⁴³

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) 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. The CMS also provides a portal through which staff can engage with the Object Management System (OMS) which stores file/object specific metadata and documentation (discussed in Section 6.2).

6.1 Data Creation

Lead role: Collections Development 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 section 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*⁴⁴
- *Guides to Good Practice*⁴⁵

⁴² <http://archaeologydataservice.ac.uk/about/governance.xhtml>

⁴³ <http://www.ukoln.ac.uk/services/papers/bl/framework/framework.html>

⁴⁴ <http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>

⁴⁵ <http://guides.archaeologydataservice.ac.uk/>

- Advisory services⁴⁶
- *Guidelines for depositors*⁴⁷

6.2 Acquisition, Retention or Disposal

Lead role: Digital Archivists

Policy document: Preservation Policy

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

- *Repository Operations*⁴⁸
- Ingest Procedures (*Ingest Manual*)⁴⁹
- *Data Procedures* (dealing with specific data types and file formats)⁵⁰
- Procedure checklists⁵¹
- File formats table delivery, preservation and presentation⁵²
- Security Overview⁵³

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.⁵⁴ Copyright, of the original data, is not transferred.

The ADS uses the concept of a collection of files or notional '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 digital objects 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 normalisation or 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

⁴⁶ <http://archaeologydataservice.ac.uk/advice.xhtml>

⁴⁷ <http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>

⁴⁸ <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>

⁴⁹ <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>

⁵⁰ The ADS Data Procedures documents are maintained on the ADS internal wiki, although downloadable versions of these are available (see Appendix 8.1). These documents are reviewed and updated regularly, or in the light of technological developments in, for example, a DPC Technology Watch report.

⁵¹ ADS procedures are ensured via a number of internal checklists covering every aspect of creation the SIP, AIP and DIP which provide internal consistency checks.

⁵² A list of preferred and accepted formats is provided externally (see <http://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml>), with lists of preservation and dissemination formats maintained on the ADS internal wiki. Static versions of these documents are provided in the Appendix.

⁵³ http://archaeologydataservice.ac.uk/resources/attach/ADS_Security_Overview_v1-13.pdf

⁵⁴ http://archaeologydataservice.ac.uk/resources/attach/ads_licence_form.pdf

processed to, for example, remove spaces from file names (Unix based systems cannot process file names containing spaces).⁵⁵

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)
- migration (contains old versions of files that have been migrated into newer formats)
- 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).⁵⁶

All files within a collection are recorded in an extension to the ADS CMS known as the Object Management System (OMS). The OMS records a high level of technical metadata including physical location, filename, size, format (identified using the National Archives (UK) DROID⁵⁷ software and recording MIME type and PRONOM identifier) and fixity value. The OMS also provides a mechanism to link related objects, so, for example, where metadata is stored externally to the OMS a relationship can be established between the data file and the metadata file. Similarly, where a digital object is part of a wider dataset as in the case of database which has been broken up into discrete tables for preservation and dissemination. Where possible the ADS this metadata and the relationships between objects can be mapped to the PREMIS data dictionary schema.⁵⁸

File normalisations and processing, for example the migration of JPG files in the SIP to TIFF files in the AIP, are recorded in the CMS. The relationship between files within separate parts of the AIP, DIP and SIP are recorded in the OMS and documented in the CMS. Thus a link between the various representations of an object, and any processes to achieve this are recorded and maintained.

Fixity values (checksums) are created during ingest and checked during the AIP process to identify any changes since accession, where discrepancies are identified these are documented in the CMS. Once work on a collection has been completed checksums are validated every 3 months. This is achieved using a semi-automated program within the CMS/OMS which runs a new checksum, and compares it to the version stored within the database. Any discrepancies are investigated by digital archivists.

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 highlighted to the data producer and if possible replacement files sort, if this is not possible then files are removed from the SIP following consultation. This impact is mitigated through

⁵⁵ This specifics of this file naming strategy is outlined in *Repository Operations* document - <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>.

⁵⁶ Full details of the form and nature of this structure are outlined in the *Repository Operations* document - <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>

⁵⁷ <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

⁵⁸ <https://www.loc.gov/standards/premis/v3/index.html>, and more specically to the relationships http://id.loc.gov/vocabulary/preservation/relationshipType/collection_PREMIS.html

clear guidance on selection and retention⁵⁹ and communication with the depositor by the Collections Development Manager and Digital Archivists prior to deposition.

Currently delivery media of fully accessioned SIPs are retained indefinitely as a record of the original deposition, but not with any guarantee of longevity. In certain circumstances media may be returned to a supplier, for example, where data has been provided on a portable hard drive. Retention of physical media is periodically reviewed by digital archivists.

6.3 Preservation and Management

Lead role: Digital Archivists

Policy document: Preservation 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). All data (AIP, DIP, SIPs) are maintained on the main ADS production server in the machine room of the IT Service at the University of York. IT Services further back up this data to tape and maintain offsite copies of the tapes. Currently the backup system uses Legato Networker and an Adic Scalar tape library. The system involves daily (overnight), weekly and monthly backups to a fixed number of media so tapes are recycled.

All data (AIP and SIPs) are synchronised once a week from the 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.⁶⁰ This repository takes the form of a standalone server (see SLA) behind the University of Essex firewall. The server is running a RAID 5 disk configuration which allows rapid recovery from disk failure. In the interests of security outside access to this server is via an encrypted SSH tunnel from nominated IP addresses. Data is further backed up to tape by the UKDA (see UKDA Preservation Policy).⁶¹

CMS and OMS based data is currently maintained in an Oracle 10 database which, similarly to the above, are backed up to tape on a daily (overnight), weekly and monthly schedule.

It should be noted that preservation data is not compressed for storage by the ADS even though the saving on storage would be significant. The issue of data compression within digital repositories continues to be hotly debated amongst digital preservationists,⁶² and while some advocate the use of lossless compression as a pragmatic solution to offset the issues of increasing file sizes and storage costs, concerns over the impact of compression remain. Research into the impact of compression on long term preservation strategies has been limited, but CERN have shown that in “a test with 1000 compressed files” there was “a likelihood of 99.8% [that] a SINGLE bit error makes the whole file unreadable, thus the data loss rate would be much higher for compressed files” when compared to compressed files.⁶³ This in itself must be offset by the increased risk of so called ‘bit rot’ caused by the increased footprint of the uncompressed file.

6.3.2 Data Management

⁵⁹ <http://archaeologydataservice.ac.uk/advice/selectionGuidance.xhtml>

⁶⁰ A responsible archive needs to maintain a copy of its data at a remote site. The ADS currently has an agreement with the UK Data Archive (UKDA) based at the University of Essex in Colchester (approximately 200 miles distant) to act as an off site repository. <http://www.data-archive.ac.uk/>

⁶¹ <http://www.data-archive.ac.uk/curate/preservation-policy>

⁶² see Corrado, EM and Sandy, HM (2017) *Digital Preservation for Libraries, Archives, and Museums*. (2nd edition). Rowman & Littlefield: London, pp.283.

⁶³ after Corrado & Sandy (eds).

As already noted the ADS maintain a custom built *Collection Management System* (CMS) which has been developed to act as a data management system. Beyond detailing the accessioning of data into a collection, through the OMS extension it also holds technical metadata (such as file type and location) for files. The CMS schema also records metadata describing the processing of files such as the normalisation of supplied files into versions for the AIP and DIP or later migrations to different versions and formats.

Through the development of the OMS, the technical metadata, process information and fixity or checksum (MD5, SHA-1, etc) for each file can be linked. This level of management enables short and long-term management of data such as auditing and versioning. At the same time the OMS allows the storage of object, data set specific metadata where it can be more effectively managed and queried. Historically, this information was stored in supplementary metadata files within the archive.

Data refreshment is an ongoing process. It is undertaken regularly (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. Data integrity is also ensured through a semi-automated validation of all checksum values this is undertaken every three months via the CMS/OMS.

As already described file normalisation and 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

The ADS has recently (2015) successfully completed a migration of all CAD files to AutoCAD Release 2010/11/12 (AC1024). A report on this process is in the process of being published.⁶⁴

An ongoing Technology Watch is maintained by Digital Archivists 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.

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 life cycles 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 life cycle including

- There is a breach of the agreement detailed in the deposit licence that cannot be resolved (deposit licence clause 8.9.1)⁶⁵

⁶⁴ See Green, K., Niven, K. and Field G. (2016) 'Migrating 2 and 3D Datasets: Preserving AutoCAD at the Archaeology Data Service'. ISPRS Int. J. Geo-Inf., 5(4), 44. <https://doi.org/10.3390/ijgi5040044>.

⁶⁵ http://archaeologydataservice.ac.uk/resources/attach/ads_licence_form.pdf

- A depositor (producer) no longer wishes to make a resource available (deposit licence ibid clause 8.9.2)
- A resource was deposited with a formally agreed lifespan
- 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. A proportion of the cost of each collection contributes to this 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. A Memorandum of Understanding (MoU) signed with the University of York which states that should the ADS cease to exist, the University of York will make all reasonable efforts to maintain the long-term preservation and access to data held within the archive. The deposit licence, signed by the depositor and the ADS Director at ingest, permits this transfer of responsibility to the University of York, as it is the legal entity to which the ADS belongs and is signed on behalf of the university by the ADS Director.

6.4 Access and use

Lead role: Communications and Access Manager

Policy document: Rights Management Framework

This section is concerned with the access and use of the 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 Communications and Access 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 resource⁶⁶
- a Copyright and Liability Statement⁶⁷
- a Common Access Agreement⁶⁸

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 data sets or collections. These may be available as standalone searchable datasets or as part of the ADS catalogues the contents of which range

⁶⁶ http://archaeologydataservice.ac.uk/resources/attach/ads_licence_form.pdf

⁶⁷ <http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess.xhtml>

⁶⁸ <http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess.xhtml>

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 one time AHDS Service Provider. Where practical various thesauri are used in order to standardise the terminology used to describe collections. The ADS uses a variety of traditional and SKOSified linked data thesauri provided by a range of external bodies and organisations.⁶⁹ Collection level data is currently stored in an Oracle 10 database and is available online through the ADS' ArchSearch, Archives and Library catalogues.⁷⁰

The ADS provides metadata to many aggregators and portals via OAI-PMH and SOAP web services, including:

- The Heritage Gateway⁷¹
- Europeana⁷²
- Thomson Reuters Data Citation Index⁷³
- Keepers Registry⁷⁴
- NERC Data Catalogue Service⁷⁵
- MEDIN Data Discovery Portal⁷⁶
- ARIADNE Portal⁷⁷

The ADS also publishes a number of datasets as a Linked Data RDF based triple store.⁷⁸

The ADS uses the *Digital Object Identifier* (DOI) System for uniquely identifying its digital content. The DOI System is an ISO International Standard and managed by an open membership consortium including both commercial and non-commercial partners. At the ADS DOIs are registered through the British Library⁷⁹, part of the DataCite consortium.⁸⁰

DOIs are persistent identifiers which can be used to consistently and accurately reference digital objects and/or content. Within the ADS, DOIs are used to reference digital archives, and in the future selected individual digital files. The DOIs provide a way for the ADS resources to be cited in a similar fashion to traditional scholarly materials. DOIs can be thought of as a combination of a URL and an ISBN number.

Each DOI has metadata associated with it, such as subject, location (URL), publisher, creator, etc. While the metadata can change for a DOI, the actual DOI name will never change. This allows for an archive's DOI to be permanent while the actual location of the

⁶⁹ For example, Historic England, RCAHMS and RCAHMW (see Heritage Data: linked data vocabularies for cultural heritage - <http://www.heritagedata.org/blog/vocabularies-provided/>) and Library of Congress (<https://id.loc.gov/>).

⁷⁰ <http://archaeologydataservice.ac.uk/search.xhtml>

⁷¹ <http://www.heritagegateway.org.uk/gateway/>

⁷² <http://www.europeana.eu/>

⁷³ http://wokinfo.com/products_tools/multidisciplinary/dci/repositories/

⁷⁴ <https://thekeepers.org/>

⁷⁵ <http://data-search.nerc.ac.uk/>

⁷⁶ <http://portal.oceannet.org/portal/start.php>

⁷⁷ <http://www.ariadne-network.eu/>

⁷⁸ <http://data.archaeologydataservice.ac.uk/query/>

⁷⁹ <https://www.bl.uk/datacite/>

⁸⁰ <https://www.datacite.org/>

archive can change. In this sense, citing a DOI is much more robust and permanent than merely citing a URL, since the DOI will always resolve to the current location of the archive.

6.4.3 Rights management⁸¹

Access to the holdings of the ADS is free at the point of use to users for research and educational purposes. By using the ADS catalogue and resources users accept the terms and conditions of the *Copyright and Liability statement*⁸² and *Common Access Agreement*.⁸³

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 be disseminated in a compressed form or 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 Overview⁸⁴
- Risk Register⁸⁵
- Disaster Recovery Plan⁸⁶

ADS delivery systems are split between a number of dedicated virtual servers hosted by the University of York IT Services (ITS). ADS delivery systems sit behind the University of York firewall with a maintenance contract with a next business day service. All delivery systems are backed up to tape (as 6.3.1) and external hard drives.

Application upgrades and migrations between applications are planned and documented with ITS 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. Since 2013 the ADS use Matomo (formerly Piwik - a free and open source web analytics application)⁸⁷ to collect such data; the University of York Legal Statements covers

⁸¹ See sections 2.6 and 5.2 of the *Collections Policy*
<http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>

⁸² <http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess.xhtml#CopyrightAndLiabilityStatement>

⁸³ <http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess.xhtml#CommonAccessAgreement>

⁸⁴ http://archaeologydataservice.ac.uk/resources/attach/ADS_Systems_Overview_v1-102.pdf

⁸⁵ http://archaeologydataservice.ac.uk/resources/attach/ADS_Risk_Register_v1-10.pdf

⁸⁶ http://archaeologydataservice.ac.uk/resources/attach/ADS_Disaster_Recovery_Plan_v1-8_redacted.pdf

⁸⁷ <https://matomo.org/>

this usage.⁸⁸ Prior to this web access statistics were generated using the Analog log file analyser package.

6.4.7 Outage

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

A glossary of abbreviations is provided on ADS website.⁸⁹

8. Appendix

8.1 ADS data procedures documents

Current versions of the procedures documents are available within the ADS internal wiki, but static versions (current at the time of writing) are reproduced here:

- Audio (version 1.52)⁹⁰
- CAD and vector graphics (version 1.134)⁹¹
- Databases (version 1.98)⁹²
- Geophysics (version 1.92)⁹³
- GIS (version 1.148)⁹⁴
- Moving Images (version 1.18)⁹⁵
- Photogrammetry {currently under revision}
- Raster Images (version 1.107)⁹⁶
- Spreadsheets (version 1.118)⁹⁷
- Statistics
- Binary and Plain Text (version 1.130)⁹⁸
- Virtual reality {currently under revision}
- LiDAR {currently under revision}
- Websites

These documents, along with the *Guides to Good Practice*, form the backbone of current ADS practice. Additional procedures are being created for the following data types:

- BIM {in draft}
- Harris Matrices {in draft}

⁸⁸ <http://www.york.ac.uk/docs/disclaimer/disclaimer.htm>

⁸⁹ <http://archaeologydataservice.ac.uk/advice/Glossary.xhtml>

⁹⁰ http://archaeologydataservice.ac.uk/resources/attach/ADS_Audio_Procedures_v1-52.pdf

⁹¹ http://archaeologydataservice.ac.uk/resources/attach/ADS_CAD_Procedures_v1-134.pdf

⁹² http://archaeologydataservice.ac.uk/resources/attach/ADS_Database_Procedures_v1-98.pdf

⁹³ http://archaeologydataservice.ac.uk/resources/attach/ADS_Geophysics_Procedures_v1-92.pdf

⁹⁴ http://archaeologydataservice.ac.uk/resources/attach/ADS_GIS_Procedures_v1-148.pdf

⁹⁵ http://archaeologydataservice.ac.uk/resources/attach/ADS_Moving_Images_Procedures_v1-18.pdf

⁹⁶ http://archaeologydataservice.ac.uk/resources/attach/ADS_Raster_Procedures_v1-107.pdf

⁹⁷ http://archaeologydataservice.ac.uk/resources/attach/ADS_Spreadsheet_Procedures_v1-118.pdf

⁹⁸ http://archaeologydataservice.ac.uk/resources/attach/ADS_Text_Procedures_v1-130.pdf

- Laser Scanning {in draft}
- Medical Imaging: Radiography & CT Scans {in draft}
- PTM & RTI {in draft}
- Scientific Data {in draft}