

Preservation Policy

Version 1.6

| | |
|--------------------------|---|
| Created date: | 2004 |
| Last updated: | 31st July 2019 |
| Review Due: | August 2020 |
| Authors: | Tim Evans, Ray Moore, Digital Archivists |
| Maintained by: | Archives Manager, Digital Archivists |
| Previous version: | Preservation Policy (Version 1.5) |

1. Purpose of this document

This policy provides an overview of the preservation and management strategies carried out to ensure the authenticity, reliability and logical integrity of all the resources entrusted to the Archaeology Data Service. This document should be read in conjunction with the *Repository Operations Policy* and *Ingest Manual*, and the suite of procedural documents referenced.

2. 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

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](#).

to its care. At the same time the ADS adheres to the FAIR (Findability, Accessibility, Interoperability, and Reusability) data principles³ providing findable, accessible, interoperable and reusable datasets to its user community which can be utilised for research, teaching or learning, in perpetuity.

3. 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:

- *Five Year Strategic Plan*: October 2016 October 2021⁴ (strategy document)
- *Risk Register*⁵
- *Collections Policy*⁶
- *Repository Operations*⁷
- *Disaster Recovery Plan*⁸
- *Information Security Risk Assessment*⁹
- *Policy and Guidance on the Deposition of Sensitive Digital Data*¹⁰
- *Security Overview*¹¹
- *Systems Overview*¹²

The University of York (UoY) is the legal entity for all services, consultancy and other contracts carried out by the ADS. Two memoranda of understanding outline the specifics of this relationship (see ADS MoU with *University of York Registrar and Finance 2012* and ADS *MoU with University Information Services & Library 2016*).¹³ The UoY provides computing and technical support for the ADS; hosting its collections, resources and website(s) through a series of virtual machines. It also provides short and medium-term storage for all datasets held within the repository.¹⁴ As part of the UoY the ADS is also subject to the policies and guidelines outlined by its host organisation. Documents include:

- *Records Management Policy*¹⁵

³ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016) doi: <https://doi.org/10.1038/sdata.2016.18>.

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

⁵ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Risk>.

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

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

⁸ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Disaster>

⁹ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#ITRisk>

¹⁰ <http://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy.xhtml>

¹¹ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Security>

¹² <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Systems>

¹³ Due to the sensitive nature of these documents, they are not publically available.

¹⁴ See *Systems Overview* -

<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Systems>

¹⁵ <https://www.york.ac.uk/records-management/records/policy/>

- *Information Policy*¹⁶
- *Information Security Policies*¹⁷
- *Legal Statements* and linked policy and strategy documents (*Accessibility, Copyright, Privacy, Cookies*)¹⁸

As noted in the *Collections Policy*¹⁹ the ADS actively engages with both national funding councils and UK Local Authorities to provide advice and digital preservation services to projects undertaken under a research or commercial remit. As part of this role, the ADS is currently a mandated archive for heritage data undertaken under the auspices of the following national organisations and funding councils:²⁰

- Natural Environment Research Council
- Marine Environmental Data and Information Network
- Nature (specifically for the journal Scientific Data)
- Historic England
- Arts and Humanities Research Council
- Birmingham City Council + Birmingham Museum Trust
- Devon County Council + Devon Museums
- Durham County Council
- Hampshire County Council Arts and Museums Service
- Hertfordshire Museums
- Southampton City Council
- City of York Council
- Museums Worcestershire

The ADS also works with external organisations and agencies to provide guidance for funding applications and project-based digital preservation:²¹

- British Academy
- Heritage Lottery Fund
- Council for British Archaeology
- Society of Antiquaries of London
- The Carnegie Trust
- Leverhulme Trust

¹⁶ <https://www.york.ac.uk/about/departments/support-and-admin/information-services/information-policy/index/>

¹⁷ <https://www.york.ac.uk/about/departments/support-and-admin/information-services/information-policy/index/information-security-policy/#tab-1>

¹⁸ <https://www.york.ac.uk/about/legal-statements/>

¹⁹ <http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>, particularly section 2.5 *Acquisition Strategies*.

²⁰ For further information see the *Partnerships* section of the ADS website - <https://archaeologydataservice.ac.uk/research/partnerships.xhtml>

²¹ Further detail on the nature of these relationships is available from the *Partnerships* section of the ADS website - <https://archaeologydataservice.ac.uk/research/partnerships.xhtml>

- Wellcome History of Medicine Project

The ADS has Service Level Agreements (SLA) with:

- UoY IT Services²²
- Amazon S3 Glacier²³
- Internet Archaeology²⁴

The ADS has Memoranda of Understanding (MoU) with a number of external organisations concerned with preservation and reuse of data including:²⁵

- Association of British Geological Survey²⁶
- Association of Local Government Archaeological Officers (ALGAO)²⁷
- Council for British Archaeology (CBA)²⁸
- Royal Commission on the Ancient and Historical Monuments of Scotland (now Historic Environment Scotland, HES)²⁹
- Royal Commission on the Ancient and Historical Monuments of Wales (RCAHMW)³⁰
- Royal Commission on the Historical Monuments of England (RCHME now part of Historic England)³¹
- MDA (now Collections Trust)³²
- National Trust³³
- Bedern Group³⁴

4. 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*

²² *Service Level Agreement (SLA) for the Provision of Information Technology Services*. Restricted access.

²³ <https://aws.amazon.com/s3/sla/>, for provision of a remote deep storage facility.

²⁴ <http://intarch.ac.uk/>, to host and provide a preservation service to the online journal

²⁵ <http://archaeologydataservice.ac.uk/about/memorandaOfUnderstanding>

²⁶ <http://archaeologydataservice.ac.uk/about/bgsMOU.xhtml>

²⁷ <http://archaeologydataservice.ac.uk/about/algaoMOU.xhtml>

²⁸ <http://archaeologydataservice.ac.uk/about/cbaMOU.xhtml>

²⁹ <http://archaeologydataservice.ac.uk/about/rcahmsMOU.xhtml>

³⁰ <http://archaeologydataservice.ac.uk/about/rcahmwMOU.xhtml>

³¹ <http://archaeologydataservice.ac.uk/about/rcahmeMOU.xhtml>

³² <http://archaeologydataservice.ac.uk/about/mdaMOU.xhtml>

³³ <http://archaeologydataservice.ac.uk/about/ntMOU.xhtml>

³⁴ <http://archaeologydataservice.ac.uk/about/Bedern.xhtml>

Information System (OAIS) as defined by a recommendation of the Consultative Committee for Space Data Systems.³⁵

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 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.

OAIS influences archival policy and strategy significantly. OAIS does not proscribe preservation strategies but promotes active management, particularly migration or normalisation, throughout the data lifecycle to facilitate preservation. This approach stands in contrast to other preservation methodologies involving emulation or technology preservation. The ADS uses a number of migration types for ongoing preservation

- Normalisation: Data may exist natively, or be normalised, to a widely supported open international standard with properties more conducive to preservation. The preservation of image files, for example, involves normalisation to 'Uncompressed Baseline TIFF v.6' format.
- Version migration: Data may be normalised through successive versions of a format. Historically vector data was normalised to AutoCAD Release 9 (AC1004), as the most stable version of the format. Subsequent changes have seen the most stable version change and consequently data must be migrated to the most stable version, in this case AutoCAD Release 2010/11/12 (AC1024).³⁶ Version migration may be the only option for preserving proprietary formats that do not 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.

³⁵ 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>

³⁶ 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>.

- Format migration: As well as normalisation data may be migrated to other formats for reasons other than preservation, particularly dissemination. The use of the GML 3.2 format to preserve spatial datasets, for example, with dissemination in the ESRI Shapefile format. The choice of format for dissemination may be a consequence of variety of factors, but may include access to software, file size, support within the community, etc.
- Replication: Migration between media that leaves data (the bit stream) totally unchanged. For example, from one system to another.

Data that cannot be normalised, or migrated between versions, are unsuitable for long-term preservation within the framework described. Documentation of all preservation actions and processes within the ADS' *Collections Management System (CMS)*³⁷ and *Object Metadata Store (OMS)*³⁸ ensure a clear data trail for all files and digital objects. To achieve this the ADS utilises terminologies derived from the PREMIS schema³⁹ to document all preservation events.⁴⁰

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: Concerned with 'history' and records, for example, 'the principal investigator'.⁴¹
- Reference: Concerned with unambiguously identifying content information through, for example, the provision of an ISBN number for a publication.⁴²
- Fixity: 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

³⁷ Internal access only.

³⁸ Internal access only.

³⁹ <https://www.loc.gov/standards/premis/>

⁴⁰ PREMIS Editorial Committee (2015) *PREMIS Data Dictionary for Preservation Metadata. Version 3.0* - <https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>. PREMIS terminologies are also utilised to document relationships between digital objects (see *Ingest Manual*, Section 4.8 and specifically Appendix 3 - <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>)

⁴¹ The ADS ensures that all datasets include appropriate collection and file-level metadata, outlined in the *Guidelines for Depositors* -

<https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>. This metadata is stored and managed within the *Collections Management System (CMS)* and *Object Metadata Store (OMS)* as outlined in the *Repository Operations Manual* -

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

⁴² The ADS utilises DOIs (*Digital Object Identifier*) to allow sustainable referencing of resources and collections (see *Ingest Manual*, Section 4.13

<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>).

checksum value and will consistently produce this value if a file is unchanged. The checksum thus provides a mechanism for validating and auditing data.⁴³

- Context: 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'.⁴⁴

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 reuse⁴⁵
- the existence of adequate documentation to facilitate ongoing preservation and reuse⁴⁶

The other cornerstones are

- ongoing access to adequate hardware systems by skilled staff⁴⁷
- that robust backup/recovery strategies are in place⁴⁸

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

⁴³ Fixity checking forms part of the accession process. A 'deposit receipt' is issued at ingest ensuring the successful transmission of data to the repository (see *Ingest Manual*, Section 3.5 - <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>) while fixity values also allow the repository to monitor files throughout the data lifecycle (see *Repository Operations Manual* – Section 6 <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>).

⁴⁴ The ADS requires the submission of contextual information, in the form of collection metadata, for all collections (see the *Guidelines for Depositors* - <https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>).

⁴⁵ Clear normalisation and migration pathways are outlined in the ADS' *Data Procedures* (internal access only, although static versions are provided through the website - <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures>). These are subject to regular review.

⁴⁶ Documentation of all collections and datasets uses the ADS' *Collection Management System* (CMS) and *Object Metadata Store* (OMS). Internal access only.

⁴⁷ Maintenance and development of systems and software fall under the purview of the *Systems Manager*, with the support of other repository staff, in conjunction with UoY IT Services (<https://www.york.ac.uk/it-services/>). The *Service Level Agreement (SLA) for the Provision of Information Technology Services* outlines, more formerly, the specifics of this arrangement. Restricted access.

⁴⁸ Discussed below.

media are notoriously frail to name some. Risk assessment⁴⁹, appropriate planning⁵⁰ and systems budgets⁵¹ allow the ADS to mitigate for such weaknesses. The *Repository Operations* provides a fuller account of the ADS implementation of the OAIS model.⁵²

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 the vast majority of collections are available online. Issues associated with the dissemination of large files/datasets over limited bandwidth connections has proved problematic in the past, necessitating the sharing of data through the exchange of physical media, but as technology has improved this has become much less of a concern. There are still issues associated with the dissemination of very large datasets, but the ADS has actively investigated various network technologies to improve access.⁵³

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).⁵⁴ 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 undertakes self-certification on an annual basis.
2. *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA).⁵⁵ 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.

⁴⁹ The *Risk Register* (<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Risk>) allows the ADS to monitor risk generally. The monitoring of information security through the *Information Security Risk Assessment - Information Security Risk Assessment* (<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#ITRisk>), with the *Security Overview* outlining steps to mitigate risk (<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Security>). These are subject to regular review.

⁵⁰ A redacted version of the *Disaster Recovery Plan* is available (<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Disaster>), with a fuller version made available to all repository staff. This is subject to regular review.

⁵¹ The *Systems Overview* provides details on hardware and infrastructure (<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Systems>), with an annual systems budget for renewal of physical hardware. This budget is reviewed and set by the ADS Director, Administrator and Applications Development team.

⁵² <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>

⁵³ See Austin, T and J Mitcham (2007) *Preservation and Management Strategies for Exceptionally Large Data Formats: 'Big Data'*. Archaeology Data Service/English Heritage: York (<https://archaeologydataservice.ac.uk/research/bigData.xhtml>).

⁵⁴ <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories>

⁵⁵ <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora>

3. *The Data Seal of Approval* (DSA) (2008-2018). Established by a number of institutions committed to the long-term archiving of research data, through formal application and assessment. The DSA seek to guarantee the durability of the data concerned, but also to promote the goal of durable archiving in general. The seal also endeavours to promote those repositories committed to archiving and providing access to scholarly research data in a sustainable way. This accreditation requires renewal on a regular basis ensuring continued adherence to the requirements.⁵⁶
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.⁵⁷
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.⁵⁸

The ADS achieved DSA certification in 2010, reapplying in 2013, and held the latest version of the DSA (2014-2017).⁵⁹ The ADS has made an application for the CoreTrustSeal (in 2019), and continues to investigate the appropriateness of the NESTOR-seal with a view to future application.

5. Identification of Content

Content is driven by community, and is dependent on what the community is producing and what it wants to reuse. The repository maintains a dedicated *Collections Policy* that outlines the scope and priorities for the repository, driven, in part, by community served.⁶⁰ All data resources offered for deposition are evaluated according the policy document and, where necessary, decision deferred to the ‘Collections Evaluation Working Group’, drawn from the ADS’ Management Committee drawn from key stakeholders and user communities, to establish the appropriateness of the dataset for inclusion in the archive.⁶¹

⁵⁶ *CoreTrustSeal* certification has now replaced the DSA (<https://www.coretrustseal.org/about/history/data-seal-of-approval/>).

⁵⁷ <https://www.coretrustseal.org/about/>.

⁵⁸

https://www.langzeitarchivierung.de/Webs/nestor/EN/Services/nestor_Siegel/nestor_siegel_node.html

⁵⁹ Mitcham, J and Hardman, C (2011) *ADS and the Data Seal of Approval – case study for the DCC*. Digital Curation Coalition: Edinburgh. <http://www.dcc.ac.uk/resources/case-studies/ads-dsa>. See also <https://www.coretrustseal.org/about/history/data-seal-of-approval/>.

⁶⁰ <http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>

⁶¹ Details of the Management Committee members are available - <https://archaeologydataservice.ac.uk/about/managementCommittee.xhtml>.

The *Guides to Good Practices*, a collaborative work by the ADS and Digital Antiquity, has sort to raise the profile of good data standards within the profession and authored by professionals and experts from within the sector. These standards, alongside more direct engagement with the repositories own designated community, form the basis of the ADS' own *Data Procedures*⁶² and *Guidelines for Depositors*.⁶³ As outlined above the ADS uses normalisation and migration, in various forms, as part of its long-term preservation strategy, consequently the repository works with its community to select appropriate formats for accession, preservation and dissemination.⁶⁴ Through direct and indirect engagement the ADS ensures that data adequate documentation and metadata forms part of the dataset.

The ADS also actively advocates for best practice within its designated community,⁶⁵ whilst providing information and tools that allow data producers to plan and legislate for the costs associated data preservation at the outset of their research and fieldwork.⁶⁶

6. Procedural Accountability

The ADS considers procedural accountability to be of the utmost significance for its works. With this in mind the ADS regularly conducts both internal and external reviews to ensure that its current practices and workflows are 'fit for purpose' and up to industry standards. As discussed above, the ADS seeks compliance to community driven guidelines for best practice, specifically TRAC and DRAMBORA, but also more formal accreditation standards, including DSA and CoreTrustSeal.⁶⁷ The ADS actively engages with other published guidelines and certification standards as it actively tries to reflect on current activities and approaches, for example, NESTOR-Seal and ISO standards.

In the interest of transparency the ADS makes most/all procedural documentation available through its website,⁶⁸ allowing depositors and users to consider current working practices within the repository.⁶⁹ The ADS publishes an annual report each year and this details the activities and practices carried out in the preceding year.⁷⁰

⁶² See Appendix 9 below.

⁶³ <https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>

⁶⁴ The ADS provides potential depositors with a list of accepted formats - <https://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml>, while the Data Procedures documents provide information on preservation and dissemination formats (see Appendix 9).

⁶⁵ <https://archaeologydataservice.ac.uk/research/presentations.xhtml>

⁶⁶ See

<https://archaeologydataservice.ac.uk/advice/PreparingDatasets.xhtml#Preparing%20Collections%20for%20Deposit>.

⁶⁷ See Section 4.

⁶⁸ See, for example, our *Collection Policy* (<https://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>), *Preservation Policy and Repository Procedures* documentation (<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml>), etc.

⁶⁹ Some procedural documentation is available in a redacted form in the interest of security.

⁷⁰ <https://archaeologydataservice.ac.uk/about/annualReports.xhtml>

The ADS is governed by its own management committee, made up of representatives of key stakeholders, funders, user communities and the ADS management team, monitoring the progress and ensure that it continues to work towards a strategic vision.⁷¹ This committee meets annually and acts in a purely advisory capacity and without legal liability.

ADS staff have established job descriptions that define roles and responsibilities. These are formalised following review by the University of York using the Higher Education Role Analysis (HERA) job evaluation methodology.⁷²

Accountability pertaining to preservation and reuse falls to:⁷³

- Director: Overall responsibility for financial management and for policy including compliance with legislation affecting digital preservation and its management.
- Deputy Director:
- Collections Development Manager: Promoting the role of the ADS for data deposit, providing advice, developing strategies and costings for depositing data, and maintaining and developing key partnerships with a collections focus.
- Archives Manager: Line management of the curatorial team, allocation of curatorial and technical tasks, oversight of ingest operations, maintaining and updating relevant policy documents and standards.
- International Projects Manager: manages the ADS contribution to a range of major international research projects. They are also responsible for maintaining and developing key partnerships with an International and Higher Education focus.
- Systems Manager: Responsible for maintaining the systems stack (with support from UoY IT Services), developing large applications (CMS, ADS-easy etc), and providing oversight of technical standards.
- Applications developer: Responsible for developing large applications, principally OASIS, and providing general maintenance and technical support for all staff where required.
- 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.
- Trainee Digital Archivists: Are responsible for archiving smaller collections (ADS EASY), small maintenance tasks, ADS Helpdesk (including ADS-easy and OASIS), and raising awareness of ADS collections and work through Social Media.

⁷¹ <https://archaeologydataservice.ac.uk/about/governance.xhtml>

⁷² <https://www.york.ac.uk/admin/hr/browse/pay-and-grading/role-evaluation/>

⁷³ More detailed information on roles and responsibilities are available within the internal wiki {restricted access}.

- Administrator: Responsible for essential administrative and financial management.
- All staff: Accountable to their line managers for compliance with this policy and with related policies, strategies, standards and guidelines.

7. 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*.⁷⁴

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 collections and datasets throughout this data lifecycle. The CMS also allows repository staff to record future depositions. The CMS takes a modular approach divided along the following lines:⁷⁵

- Tracking: used to record information about current and future collections
- Collections: provides more comprehensive documentation and metadata about current collections and datasets. This section also provides information on the collection management including documentation of files, processes and web interfaces.
- People: documents individuals and organisations who contact the ADS
- Admin: used for the administration of the CMS
- Advice: used for record informal communications with both users and potential depositors

The CMS also provides a portal through which staff can engage with the *Object Management Store* (OMS) that stores file/object specific metadata and documentation (discussed in Section 7.2).

⁷⁴ <http://www.ukoln.ac.uk/services/papers/bl/framework/framework.html>

⁷⁵ Where personal information is collected the ADS adheres to its own *Privacy Policy* (<https://archaeologydataservice.ac.uk/advice/Privacy.xhtml>).

7.1 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 assists the repository in processing the dataset, a poorly formed one may well preclude ingest entirely (see section 7.2). For a SIP to be well formed it must conform to a repository's requirements in terms of formats, metadata and documentation which are expressed in the *Guidelines for Depositors*. The ADS is active in a number of ways in providing guidance to potential depositors during this period including:

- *Collections Policy*⁷⁶
- *Guidelines for depositors*⁷⁷
- *Preferred and accepted file formats*⁷⁸
- *Guidance on the selection of material for deposit and archive*⁷⁹
- *Policy and Guidance on the deposition of personal, confidential and sensitive data*⁸⁰
- Advisory services

The ADS also provides more generic information and guidance about formats and metadata through the *Guides to Good Practice*.⁸¹

7.2 Acquisition, Retention or Disposal

Lead role: Archives Manager/Digital Archivists

Policy document: Preservation Policy

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

- *Repository Operations*⁸²
- *Ingest Manual*⁸³
- *Data Procedures*⁸⁴

⁷⁶ <http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>

⁷⁷ <http://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>

⁷⁸ <https://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml>

⁷⁹ <https://archaeologydataservice.ac.uk/advice/selectionGuidance.xhtml>

⁸⁰ <https://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy.xhtml>

⁸¹ <http://guides.archaeologydataservice.ac.uk/>

⁸² <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>

⁸³ <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>

⁸⁴ The *Data Procedures* provide detailed information on the processes associated with specific data types and file formats. These are documented internally using the ADS wiki, although downloadable versions of these are available (see Appendix 9.1). These data procedures are regularly reviewed, and/or updated in the light of technological developments.

- Procedure checklists⁸⁵

And are carried out in accordance with the

- *Information Security Risk Assessment*⁸⁶
- *Security Overview*⁸⁷

The existence of a SIP, and a signed deposit licence, triggers accessioning. The licence grants a non-exclusive right to archive and distribute the supplied data.⁸⁸ The depositor, or original copyright holder, retain their intellectual property rights; and the ADS makes no claim over these.

The ADS places no restriction on the number of accessions (SIPs) for a project, and all collections remain 'open' to the submission of new or replacement data; each treated in accordance with the standard procedures and practices for accession. Such submissions, to existing collections, may involve the addition of data in stages, or the submission of an entire 'edition' of the complete dataset. Unique identifiers facilitate the ongoing management of collections.

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 original files is a primary concern during normalisation or migration, as detailed in ADS Data Procedures. All 'original' files included in the deposition are retained becoming part of the AIP; these may be subject to minor changes (e.g. filenames so that they align with the file naming policy)⁸⁹ but remain otherwise unchanged in terms of format and content.

The ADS uses a formalised directory structure to store datasets as outlined in the *Repository Operations*.⁹⁰ Where possible, beneath this structure, the repository will maintain the original data structure.

⁸⁵ A series of checklists ensure adherence to current practice and procedures. These cover aspects of accession, procedure and the AIP process. Detailed versions are available internally through the ADS wiki, with static versions provided through the website

(<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Checklists>).

⁸⁶ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#ITRisk>

⁸⁷ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Security>

⁸⁸ A sample deposit licence is available -

<https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>

⁸⁹ The repository uses Unix-based systems that makes processing of files with names that include spaces and non-standard characters difficult. The repository requests that depositors follow the guidance outlined in the *Guidelines for Depositors*

(<https://archaeologydataservice.ac.uk/advice/PreparingDatasets.xhtml#File%20Management>), but in some circumstances deposits may require additional changes. These are outlined in *Repository Operations* document (<http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>).

⁹⁰ <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>

An extension of the CMS, known as the *Object Metadata Store* (OMS), supports the documentation of all individual files submitted to the repository. The OMS records a high level of technical metadata, including location, filename, size, format, MIME type, PRONOM identifier and fixity value using the National Archives (UK) DROID software.⁹¹ A more detailed account of the process is available from the *Ingest Manual*.⁹²

Digital archivists use the 'match objects' functionality, initiated from within the CMS, to group all related files into notional 'objects' within the OMS. A digital 'object' will, typically, include the original file alongside the normalised versions for preservation and dissemination. The OMS tables allow archivists to record relationships between discrete objects, with the nature of that relationship described using the necessary PREMIS concept.⁹³ The *Ingest Manual* provides a fuller discussion of the OMS.⁹⁴

The CMS also allows the documentation and recording of all processes carried out by repository staff.⁹⁵

In rare instances an accession may include files that are not suitable for ingest and fall beyond the prescribed formats outline in the list of *Preferred and Accepted File Formats*.⁹⁶ In these instances, repository staff take appropriate actions to ensure the submission of new, or replacement, files.⁹⁷ Similarly, where documentation is inadequate ADS staff will negotiate with depositors about these requirements. The provision of guidance on selection and retention⁹⁸ and communication with the depositor by the *Collections Development Manager* and *Digital Archivists* ensures mitigation prior to deposition.⁹⁹

The *Ingest Manual* outlines policies concerning the retention of physical media.¹⁰⁰

7.3 Preservation and Management

Lead role: Archives Manager/Digital Archivists

Policy document: Preservation Policy/Repository Operations Manual

⁹¹ <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

⁹² <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>, see Section 4.8.

⁹³ See *Ingest Manual* <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>, Appendix 3.

⁹⁴ <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>

⁹⁵ These include (data) capture, compression, creation, deletion, migration, renaming, restructure, etc.

⁹⁶ <https://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml>

⁹⁷ See *Ingest Manual*, Section 3.3 -

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

⁹⁸ <http://archaeologydataservice.ac.uk/advice/selectionGuidance.xhtml>

⁹⁹ Additional information on this can be found in the *Ingest Manual*, see Section 3.3 and 3.6,

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

¹⁰⁰ See Section 3.13 - <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>

7.3.1 Storage and Resilience

The ADS maintain multiple copies of data in order to facilitate disaster recovery and provide resilience. The University of York provides short and medium-term storage of all datasets held by the repository, including all SIPs, AIPs and DIPs. While data submitted via one of the ADS submission portals, ADS-easy and OASIS Images, is 'stored', in the short term, on a separate dedicated virtual machine until formal accession can be initiated. In each instance access is strictly controlled and limited to authorized users only. Backups of all data on these virtual machines are made via hourly snapshots (retained for 30 days), with further tape backups of data retained for 90 days.

All data is stored on a pair of Dell Compellent enterprise storage arrays (current capacity ~1Pb), located in two different data centres. Each data centre is dedicated and purpose built, and has full UPS, fire suppression, generators and is 'lights out' and alarmed. Data is protected by being spread redundantly across multiple disks ('RAID') in each location. The storage arrays are automatically monitored, with logs and alerts generated that report failed disks, storage capacity warnings and other hardware and software issues. All logs are emailed to several members of the UoY ITS team for immediate action.

The UoY ITS use Linear Tape-Open (LTO-6) for 90-day backups. UoY ITS plan to continue to migrate to newer LTO versions (with greater durability and storage capacity) as a matter of course; migrating to newer LTO versions will help to ensure against media deterioration. The LTO media is stored in II UPS, fire suppression, alarmed and secured rooms. If a tape error is reported (via a Storage Manager server), the relevant data is migrated to another tape and the tape with the error is removed from circulation. Daily logs are produced by the Storage Manager servers, which alert UoY ITS administrators of any errors or warnings.

The ADS has recently implemented cloud-based storage via Amazon Web Services (AWS) for long-term/deep storage of the off-site backup only.¹⁰¹ The SLA for Amazon S3 Glacier SLA is available.¹⁰² All AIPs and SIPs are synchronised, from the local copy in the UoY to AWS. This process intends to mitigate against data loss caused by hardware degradation and failure of local storage and any associated physical threat to storage device, and build resilience into the preservation of the datasets the repository curates. The initiation of data synchronization of the AIP, from local virtual servers to deep storage, follows completion of preservation activities or after an update. Synchronization follows a semi-automated, prescribed process, documented in the ADS wiki, and utilises file fixity and time stamps to mitigate for error and corruption during transfer. The ADS has intimated that all data stored using AWS should be stored within the European Union, specifically Ireland.

All associated documentation and metadata stored within the CMS and OMS is stored and backed up locally in accordance with the policies and guidelines outlined by the UoY.

¹⁰¹ <https://aws.amazon.com/>

¹⁰² <https://aws.amazon.com/s3/sla/>

Repository staff make additional back-ups, stored outside of the local network, of these databases.

In order to mitigate for the impact of data degradation of files and datasets, the ADS does not use data compression within any of its storage systems even though the saving on storage would be significant. The intention in such an approach is to mitigate for the impact of 'bit rot' on compressed files and effects of 'lossy' compression.¹⁰³

7.3.2 Data Management

As already noted the ADS maintain a custom-built *Collection Management System* (CMS) developed to act as a data management system for all collections from accession onwards.¹⁰⁴ The CMS facilitates the administration of datasets submitted to the repository, but also affords the administration of enquiries about future submissions. As well as internal data management tool, the CMS stores also collection and resource discovery metadata and controls elements archive interface published on the ADS website

Detailed information on specific files and digital objects is stored in an extension of the CMS called the *Object Metadata Store* (OMS). The OMS stores technical metadata,¹⁰⁵ used for the maintenance and management of digital objects, alongside contextual and data type specific metadata to facilitate use and reuse. Historically, this information was stored in supplementary metadata files within the archive.

Together the CMS and OMS ensure the maintenance of a clear data trail for all collections, files and digital objects throughout data lifecycle.

The ADS uses fixity values (checksums) throughout the data lifecycle to monitor changes and ensure the documentation of all actions carried out by repository staff.¹⁰⁶ This also allows the recovery of data should discrepancies between checksums be identified whilst work is being carried out on the archive and on completion of preservation activities. The ADS also operates a quarterly check on all data held by the repository. This process involves a comparison of the checksum (as recorded with the OMS) and a current fixity sum from file held by the archive. These guard against bit-rot and other issues associated with

¹⁰³ Discussion of the impact of compression remains contentious (see, for example, Corrado, EM and Sandy, HM (2017) *Digital Preservation for Libraries, Archives, and Museums*. (2nd edition). Rowman & Littlefield: London), and the ADS keeps a 'watching brief' over this debate.

¹⁰⁴ See section 7.2 above, and discussion within the *Ingest Manual* (<http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>) and *Repository Operations* (<http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>).

¹⁰⁵ See section 7.2 above, and discussion within the *Ingest Manual* (<http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>) and *Repository Operations* (<http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp>).

¹⁰⁶ As outlined in the *Ingest Manual*, see Section 4.6 - <http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest>

data degradation. Repository staff can then take appropriate action to ensure the correct preservation of that files and datasets.

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).¹⁰⁷ Repository staff maintain an ongoing technology watch to ensure the consideration of format changes and software updates and any impacts upon preservation workflows.¹⁰⁸ As is the case with accession, any format/version migration should maintain the significant properties of the original data. Extensive research and good planning prior to commencement ensure that all format and version migrations are correctly and effectively. The documentation of all actions carried out during the migration within the CMS and OMS ensures the maintenance of a record of all preservation actions throughout the data lifecycle.

While the repository works hard to ensure the preservation of datasets 'in perpetuity', in some circumstances this level of preservation may not be possible or required. The ADS recognises that all data lifecycles have a beginning and an end; that, sometimes, digital objects, files or even entire collections may have a shorter data lifecycle than others. The repository takes appropriate steps to ensure the maintenance of a clear preservation pathway and, typically, problems that arise during the data lifecycle are not technical issues or problems with the preservation processes but are often outwit of this. Examples of these may include:

- a breach of the agreement detailed in the deposit licence¹⁰⁹
- a content objection by a data consumer/user¹¹⁰
- a depositor (producer) no longer wishes to make a resource available

¹⁰⁷ For a discussion of this process 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>.

¹⁰⁸ Where necessary the repository will update its list of accepted formats (<https://archaeologydataservice.ac.uk/advice/FileFormatTable.xhtml>) and metadata requirements (<https://archaeologydataservice.ac.uk/advice/FilelevelMetadata.xhtml#File-level%20Metadata%20Requirements>) in light of these developments

¹⁰⁹ A sample deposit licence is provided through the *Guidelines for Depositor* - <https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>.

¹¹⁰ An outline of the content objection process is given in the 'Website Terms and Conditions' - <https://archaeologydataservice.ac.uk/advice/WebsiteTerms.xhtml>.

- a resource was deposited with a formally agreed lifespan¹¹¹
- data was submitted in formats outwith of the list of accepted formats, but on the request of the depositor have been retained and preserved on a 'best efforts' basis¹¹²

When problems do arise with a dataset then, in the first instance, the repository will contact the depositor(s)/rights holder(s) to let them know about any problems/issues in an effort to seek immediate resolution. Whilst resolution is sort the affected file/part of the dataset will be removed the ADS website; this is particularly the case if the issue relates to an infringement of an individual/organisations role as a rights holder. In instances where resolution cannot be agreed, the ADS provides a framework of policies and procedures put in place to retrospectively appraise datasets, deaccession data and support the sustainable management of repositories collections.¹¹³ Documentation of all 'end of life' events and processes, within the CMS, ensures the maintenance of a clear record of the process and its outcomes.

In terms of succession planning, should the ADS cease to function as viable repository, responsibility for its existing collections and datasets would transfer to the University of York (UoY), as the legal entity for all services, consultancy and research carried out by the ADS. A formal Memorandum of Understanding, between the UoY and the ADS, outlines the specifics of this relationship and the responsibilities of each party.¹¹⁴ The deposit licence, agreed with depositors at ingest, permits the transfer of responsibility to the UoY, as it is the legal entity to which the ADS belongs; this licence is signed by the ADS Director in behalf of the UoY. The ADS maintains a *Preservation Legacy Fund*, with a proportion of the cost of each accession/collection added to the fund each year, to enable the enactment of the succession plan and transfer the repository holdings to the UoY.

7.4 Access and use

Lead role: Archives Manager/Digital Archivists

Policy document: Rights Management Framework

This section is concerned with the access and use of the DIP; and includes information on finding a resource, rights management and receiving a data collection or part thereof. As an advocate for the FAIR principles of data stewardship¹¹⁵ the ADS is committed to making its

¹¹¹ Agreement that a dataset/collection should be retained for a finite period is usually agreed prior to ingest.

¹¹² In such circumstances the depositor will be notified at ingest that the repository cannot guarantee preservation of the datasets

¹¹³ *Appraisal and Deaccession Policy* -

<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DeacPol>

¹¹⁴ See *ADS MoU with University Information Services & Library 2016* -

<https://archaeologydataservice.ac.uk/manPages/mou.xhtml> (internal access only).

¹¹⁵ That is making data *Findable, Accessible, Interoperable, and Reusable*, see Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and

holdings findable, accessible, interoperable and reusable allowing its user community to maximise resources for research, teaching or learning. As such the repository recognises that preservation AND dissemination are of equal significance in terms of data preservation and active curation. As such, the repository is working closely with partners in ARIADNE, ARIADNEplus and E-RIHS projects to promote FAIR and expose the collections it curates.

7.4.1 Prerequisites

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

- a deposit licence for each resource¹¹⁶
- a copyright and liability statement¹¹⁷
- a common access agreement¹¹⁸
- website terms and conditions¹¹⁹

7.4.2 Resource Discovery¹²⁰

The repository 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 other catalogues/resources available through the ADS website (e.g. ArchSearch,¹²¹ ADS Library,¹²² etc.

The ADS uses a qualified Dublin Core metadata schema for describing holdings, collected from depositor, using a standardised collection-level metadata template or online form, as part of the deposition process.¹²³ Alongside detailed information about projects and the collections, depositors are encouraged to use terminologies from recognised thesauri to facilitate the cataloguing and categorization of datasets. Where practical the ADS encourages depositors to use terminologies from recognised national, cultural heritage

stewardship. *Sci Data* **3**, 160018 (2016) doi: <https://doi.org/10.1038/sdata.2016.18> and discussion in Section 2, above.

¹¹⁶ See the *Guidelines for Depositors* for a sample deposit licence -

<https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>

¹¹⁷ <http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess.xhtml>

¹¹⁸ <http://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess.xhtml>

¹¹⁹ <https://archaeologydataservice.ac.uk/advice/WebsiteTerms.xhtml>

¹²⁰ See also the *Collections Policy*, Section 5.3,

<https://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>

¹²¹ <https://archaeologydataservice.ac.uk/archsearch/basic.xhtml>

¹²² <https://archaeologydataservice.ac.uk/library/>

¹²³ The template, alongside descriptions of the metadata requirements, are available from the ADS website - <https://archaeologydataservice.ac.uk/advice/DatasetlevelMetadata.xhtml#Collection-level%20Metadata%20Requirements>

thesauri, particularly those published as structured Linked Open Data (LOD).¹²⁴ These linked terms facilitate the creation of semantic searches and published within the repositories own triplestore. Any geographical or locational terms utilise the *Getty Thesaurus of Geographic Names* and to *Ordnance Survey Open Names*.¹²⁵ Where appropriate datasets use geographic coordinates (decimal latitude/longitude) as part of GEMINI compliance. The repository uses thesauri of temporal terms to qualify data and collections chronologically.¹²⁶

The repository uses a number of search facilities (*ArchSearch*,¹²⁷ an *Archive* search¹²⁸ and ADS Library¹²⁹) to promote and improve resource discovery. To facilitate machine harvesting of its metadata via three *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) targets for its archives, journals and OASIS resources.¹³⁰ It also provides access to its linked data repository through a SPARQL endpoint that allows the interrogation of its triplestore through both a query interface, or via a specialised client.¹³¹

ADS surfaces collection and file-level metadata through a number of external aggregators and portals, both within the UK and Europe.

- *Heritage Gateway*¹³²
- *Marine Environmental Data and Information Network* (MEDIN) data portal¹³³
- *Natural Environment Research Council* (NERC) data discovery portal¹³⁴

¹²⁴ *Historic England* (HE), *Historic Environment Scotland* (HES) and the *Royal Commission on Ancient & Historical Monuments of Wales* (RCAHMW) provide thesauri - <https://www.heritagedata.org/blog/>. Alongside these ADS also utilises the Library of Congress Subject Headings, published as Linked data - <http://id.loc.gov/authorities/subjects/sh2013002090.html>.

¹²⁵ See TGN <https://www.getty.edu/research/tools/vocabularies/tgn/> alongside OS terms derived from Open Names <https://www.ordnancesurvey.co.uk/business-government/products/open-map-names>.

¹²⁶ Supplied by *Historic England* (HE), *Historic Environment Scotland* (HES) and the *Royal Commission on Ancient & Historical Monuments of Wales* (RCAHMW) - <https://www.heritagedata.org/blog/>.

¹²⁷ *ArchSearch* is an integrated catalogue of 'thin' metadata records, including ADS collections and resources alongside metadata derived from archaeological and historic environment inventories from national and regional agencies within the UK. A series of search options are provided including simple keyword, a facet browse, map and a more complex query search functionality to provide access <https://archaeologydataservice.ac.uk/archsearch/basic.xhtml>

¹²⁸ A dedicated *Archive* search provides direct access to 'rich' datasets and collections held by the repository, with both a faceted browse and keyword search functionality - <https://archaeologydataservice.ac.uk/archive/>

¹²⁹ The ADS Library provides access to bibliographic references, alongside direct access to textual content, particularly unpublished field reports and journal holdings, held by the repository <https://archaeologydataservice.ac.uk/library/>

¹³⁰ <https://archaeologydataservice.ac.uk/about/endpoints.xhtml>

¹³¹ <http://data.archaeologydataservice.ac.uk/page/>

¹³² <https://www.heritagegateway.org.uk/gateway/default.aspx>

¹³³ The repository exposes metadata for those holdings associated with the historic environment within the marine environment - <http://portal.oceannet.org/portal/start.php>

¹³⁴ Used for science-based archaeological research - <https://csw-nerc.ceda.ac.uk/geonetwork/srv/eng/catalog.search#/home>

- the *Keepers Registry*¹³⁵
- *Europeana*¹³⁶
- ARIADNE Portal¹³⁷

The ADS utilises *Digital Object Identifier* (DOI) handles to facilitate the citation of its resources. These are provided the *British Library*¹³⁸, and part of the *DataCite* consortium.¹³⁹ The DOI also provides a mechanism for the repository to track and measure citations referencing its resources.

7.4.3 Rights management¹⁴⁰

Access to the holdings of the ADS is free at the point of use to all users wishing to use ADS resources for research and educational purposes.¹⁴¹ The deposit licence, 'signed' by both the depositor and the repository, outlines the rights of individuals and organisational rights holders. In each instance, the depositor outlines the terms of access and reuse are explicitly stated.¹⁴² Each archive interface provides details of these terms for data consumers and users.¹⁴³ All depositors must agree to the terms of the deposit licence as part of the deposition process as outlined in the *Guidelines for Depositors*.¹⁴⁴

The repository respects the rights of all its data providers and users and endeavours to highlight problems and concerns about collections during the ingestion data and whilst works are carried out to preserve datasets. When archives have been published depositors and

¹³⁵ Through which the repository surfaces metadata for its journal holdings - <https://thekeepers.org/>

¹³⁶ <https://www.europeana.eu/portal/en>

¹³⁷ <https://ariadne-infrastructure.eu/portal/>

¹³⁸ <https://www.bl.uk/datacite/>

¹³⁹ <https://www.datacite.org/>

¹⁴⁰ See sections 2.6 and 5.2 of the *Collections Policy* - <http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>

¹⁴¹ In a rare instances access may be restricted, typically this a consequence of technical issues to do with the size of the dataset.

¹⁴² See *Collection Policy*, sections 2.6 - <http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml>

¹⁴³ This is expressed in the vertical bar to the left of each piece of archive (i.e. '*This work is licensed under the...*'), for example, Historic England (2018) HE7238 - *Stonehenge Southern World Heritage Site Survey* [data-set]. York: Archaeology Data Service [distributor] <https://doi.org/10.5284/1047612>. Generally, dissemination is under the terms of a *Creative Commons Attribution 4.0 International License* (<http://creativecommons.org/licenses/by/4.0/>), but depositors are free to nominate other terms of access. Historically, however, dissemination was under the ADS' own terms and conditions, copyright and liability statement and common access agreement (<https://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess.xhtml>).

¹⁴⁴ <https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml>

users alike are encouraged to report infringements of rights, or to other content that they might object too. The repository maintains a formal procedure put into effect on receipt of reports any rights breach.¹⁴⁵

7.4.4 Receiving data

As advocates of the FAIR data stewardship principles¹⁴⁶ the ADS regards accessibility as an essential part of the preservation process; consequently all datasets and collections curated by the repository are made accessible according to the terms of the deposit licence.¹⁴⁷ The primary mechanism for the dissemination of data is through the ADS website, but in circumstances where the size of the file or dataset may make sharing difficult the repository will use compression to facilitate downloading.¹⁴⁸ In those instances where compression may be necessary, the repository ensures the provision of sufficient guidance on the use of compressed data.¹⁴⁹ Dissemination of particularly large datasets may require a formal request to the repository and the use of external file sharing services, or even the exchange of physical media.

The repository is keen to ensure the dissemination of data in open and accessible formats that maximise the use and reuse potential of data. With this in mind, repository staff follow the guidance provided in the internal *Data Procedures* in order to create consistent DIPs utilising the accepted dissemination formats.¹⁵⁰ Collection and file-level metadata accompanies all DIPS to facilitate interoperability and reuse.

7.4.5 Security of delivery systems

The repository has created policies and practices then ensure the delivery of its systems and resources and outlined in a series of documents:

- *Information Security Risk Assessment*¹⁵¹
- *Risk Register*¹⁵²

¹⁴⁵ <https://archaeologydataservice.ac.uk/advice/WebsiteTerms.xhtml>

¹⁴⁶ That is making data *Findable, Accessible, Interoperable, and Reusable*, see Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016) doi: <https://doi.org/10.1038/sdata.2016.18> and discussion in Section 2, above.

¹⁴⁷ Only the *Dissemination Information Package* (DIP) is disseminated.

¹⁴⁸ Typically, 'ZIP' or 'multi-ZIP' formats.

¹⁴⁹ See, for example, Nigel Nayling, Garry Momber, Miguel San Claudio, Ana Crespo Solana (2018) ForSEADiscovery [data-set]. York: Archaeology Data Service [distributor] <https://doi.org/10.5284/1048329>, particularly https://archaeologydataservice.ac.uk/archives/view/forseadiscovery_2018/video.cfm?videofile=A-BAY01-00-06-2015-0-XX-V.

¹⁵⁰ These are available for internal use only, but static versions of these procedures are available - <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures>.

¹⁵¹ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#ITRisk>.

¹⁵² <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Risk>

- *Disaster Recovery Plan*¹⁵³
- *Systems Overview*¹⁵⁴
- *Security Overview*¹⁵⁵

The ADS has undertaken a high-level *Information Security Risk Assessment*, which, in conjunction with a low-level *Security Overview*, mitigate risk associated with the management of repository systems and services. These assessments, used in conjunction with the policies and guidelines published by the ADS' host institution, the University of York, form a comprehensive and detailed suite of policies and guidance that ensure the secure and effective management of repository systems. A high level *Risk Register* provides an overarching assessment of risk at an organisational level. Regular review of these policies and guidance ensures their currency. The repository also carries out a self-assessed risk analysis using DRAMBORA,¹⁵⁶ and regularly evaluates the effectiveness of this assessment and associated risk management implications.

The spread of ADS' systems and resources over a number of virtual machines provides mitigation for any risk.¹⁵⁷ The granting of permissions to repository staff on a needs basis, with appropriate levels of access relevant to their working practices assigned to each user, safeguards systems and resources. Access is also restricted using IP address and encrypted passwords. A centralised password administration system facilitates the management of access, with updates to access credentials carried out on regular basis, in line with the *Security Policy* and in accordance with the *Information Security Risk Assessment*. The repository also adheres to the policies, procedures and guidance outlined by its parent organization, the *University of York*.¹⁵⁸ The synchronization of all repository holdings and dataset to an off-site, deep storage facility ensures its ongoing security.

The *Disaster Recovery Plan* summarises the processes and procedures in place to protect, recover and mitigate problems and issues that might affect resources and data storage.

7.4.6 Consumer access analysis

The ADS uses *Matomo Web Analytics*¹⁵⁹ to monitor user experience and collect, non-personally-identifying, statistical information on access to repository resources and datasets.

¹⁵³ A full version of this plan is available internally, but for security reasons only a redacted version is publicly available - <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Disaster>

¹⁵⁴ A full version of this plan is available internally, but for security reasons only a redacted version is publicly available - <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Systems>

¹⁵⁵ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Security>

¹⁵⁶ <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora>

¹⁵⁷ These systems are summarised in the *Systems Overview* -

<https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Systems>

¹⁵⁸ <https://www.york.ac.uk/it-services/security/>

¹⁵⁹ <https://matomo.org/>

More specific access statistics, displayed alongside each archive¹⁶⁰, allows data creators to monitor resource use on a general level. Formal citation, using the DOI, allows the repository to monitor and analyse the referencing of datasets and collections.¹⁶¹

7.4.7 Outage

The ADS monitors all planned and unplanned service downtime. At the same time, the repository's parent organisation, the University of York, carries out its own monitoring of service downtime. There is a scheduled maintenance period of Tuesdays 8-9am (UK time). Services may be unavailable during this period.

8. Glossary

A glossary of abbreviations is provided on ADS website.¹⁶²

¹⁶⁰ See the 'Usage Statistics' for each dataset/resource, see, for example, Chapman et. al. 2019, https://archaeologydataservice.ac.uk/archives/view/trypillia_ahrc_2018/stats.cfm.

¹⁶¹ The use of DOI statistics is still in development but the ADS hope to make these accessible in the near future.

¹⁶² <http://archaeologydataservice.ac.uk/advice/Glossary.xhtml>

9. Appendix

9.1 ADS data procedures documents

Current versions of the procedures documents are available within the ADS internal wiki, but static versions are accessible:¹⁶³

- 3D data (version 1.49)
- Audio (version 1.62)
- BIM (version 1.3)
- CAD and vector graphics (version 1.143)
- Databases (version 1.103)
- Geophysics (version 1.99)
- GIS (version 1.173)
- Harris Matrices (version 1.14)
- LiDAR (version 1.7)
- Medical Imaging (version 1.16)
- Moving Images (version 1.18)
- PTM and RTI (version 1.78)
- Raster Images (version 1.115)
- Scientific Data (version 1.2)
- Spreadsheets (version 1.131)
- Statistics (version 1.38)
- Binary and Plain Text (version 1.141)
- Websites

Some of these are available in draft only, or are under revision. These documents, along with the *Guides to Good Practice*, form the backbone of current ADS practice.

¹⁶³ <https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures>