PRESERVATION POLICY (VERSION 1.7)

DEPUTY DIRECTOR, ARCHIVES MANAGER AND DIGITAL ARCHIVISTS
ARCHAEOLOGY DATA SERVICE
https://archaeologydataservice.ac.uk/

Created date:	2004
Last updated:	02 July 2020
Review Due:	31 July 2021
Authors:	Tim Evans, Ray Moore, Digital Archivists
Maintained by:	Deputy Directory, Archives Manager and Digital Archivists
Required Action:	
Status:	Live
Location:	https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml
Previous version	http://archaeologydataservice.ac.uk/resources/attach/ADS_Preservation_Policy_v1-6.pdf

1. Purpose of this document

1.0.1 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 (ADS). This document should be read in conjunction with the *Repository Operations Manual*[1] and *Ingest Manual*[2], and the suite of procedural documents referenced.[3]

2. Principal Statement[4]

2.0.1 '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.' [5]

2.0.2 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 and cultural heritage including commercial archaeology, heritage organisations, museums, further and secondary education, community archaeology and the interested public in general.

2.0.3 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.

2.0.4 At the same time the ADS adheres to the FAIR (Findability, Accessibility, Interoperability, and Reusability) data principles[6] providing findable, accessible, interoperable and reusable datasets to its user community in order that they can be utilised for research, teaching or learning, in perpetuity.

---

[1] *Repository Operations Manual* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp, accessed 02 July 2020.
[2] *Ingest Manual* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest, accessed 02 July 2020.
[3] Many of these are publically accessible from the *Preservation Policy and Repository Procedures* section of the ADS website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml, accessed 02 July 2020.
[4] 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, accessed 02 July 2020.
[5] See the ADS' *Mission Statement*.
[6] 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.

2.0.5 Likewise, the ADS supports the TRUST framework (Transparency, Responsibility, User Focus, Sustainability and Technology) for digital repositories, as a mechanism to facilitate discussion and implementation of best practice in digital preservation by all stakeholders.[7]

2.0.6 More formerly, the ADS seeks to quantify and qualify its activities and successes in reaching its stated objectives by compliance with community driven initiatives for best practice. Formal accreditation in the form of the *CoreTrustSeal*[8] and membership of the *World Data System* (WDS)[9] ensure that current processes and practices meet recognised standards. At the same time, less formal methods of review and self-regulation, alongside wider involvement in the preservation community, ensure the validity and currency of repository workflows and processes.

3. Contextual Links

3.0.1 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 2021[10] (strategy document)
- *Collections Policy*[11]
- *Preservation Policy*
- *Repository Operations Manual*[12]
- *Ingest Manual*[13]
- *Appraisal and Deaccession Policy*[14]
- *Risk Register*[15]
- *Information Security Risk Assessment*[16]

[7] Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. *Sci Data* **7**, 144 (2020). Doi: https://doi.org/10.1038/s41597-020-0486-7, accessed 02 July 2020.
[8] *CoreTrustSeal* - https://www.coretrustseal.org/, accessed 02 July 2020. The ADS gained CoreTrustSeal accreditation in May 2020 - https://www.coretrustseal.org/wp-content/uploads/2020/04/Archaeology-Data-Service.pdf, accessed 02 July 2020.
[9] *World Data System* (WDS), is an interdisciplinary body of the International Science Council (ISC; formerly ICSU) - https://www.icsu-wds.org/organization, accessed 02 July 2020. The WDS accepted the ADS for membership as a certified 'trusted data service' in June 2020 - https://www.icsu-wds.org/community/membership/regular-members?fid=archaeology-data-service, accessed 02 July 2020.
[10] https://archaeologydataservice.ac.uk/resources/attach/strategicPlan/ADSFiveYearPlan2016-21.pdf, accessed 02 July 2020.
[11] https://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml, accessed 02 July 2002.
[12] *Repository Operations Manual* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp, accessed 02 July 2020.
[13] *Ingest Manual* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest, accessed 02 July 2020.
[14] *Appraisal and Deaccession Policy* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DeacPol, accessed 02 July 2020.
[15] *Risk Register* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Risk, accessed 02 July 2020.
[16] *Information Security Risk Assessment* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#ITRisk, accessed 02 July 2020.

- *Systems Overview*[17]
- *Security Overview*[18]
- *Disaster Recovery Plan*[19]
- *Policy and Guidance on the Deposition of Sensitive Digital Data*[20]

3.0.2 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*).[21] 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.[17] As part of the UoY the ADS is also subject to the policies and guidelines outlined by its host organisation. Documents include:

- *Information and Records Management Policy*[22]
- *University Information Policy*[23]
- *Information Security Policy*[24]
- *Legal Statements* and linked policy and strategy documents (*Accessibility*, *Copyright*, *Privacy*, *Cookies*)[25]

3.0.3 As noted in the *Collections Policy*[26] 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:[27]

- *Natural Environment Research Council* (NERC)
- *Marine Environmental Data and Information Network* (MEDIN)
- *Nature* (specifically for the journal *Scientific Data*)
- *Historic England*
- *Arts and Humanities Research Council*

---

[17] A redacted version of the *Systems Overview* is available - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Systems, accessed 02 July 2020, with a fuller version made available to all repository staff.
[18] *Security Overview* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Security, accessed 02 July 2020.
[19] A redacted version of the *Disaster Recovery Plan* is available (https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Disaster, accessed 02 July 2020), with a fuller version made available to all repository staff. This is subject to regular review.
[20] http://archaeologydataservice.ac.uk/advice/sensitiveDataPolicy.xhtml, accessed 02 July 2020.
[21] Due to the sensitive nature of these documents, they are not publically available.
[22] https://www.york.ac.uk/records-management/records/policy/, accessed 02 July 2020.
[23] https://www.york.ac.uk/about/departments/support-and-admin/information-services/information-policy/index/, accessed 02 July 2020.
[24] https://www.york.ac.uk/about/departments/support-and-admin/information-services/information-policy/index/information-security-policy/#tab-1, accessed 02 July 2020.
[25] https://www.york.ac.uk/about/legal-statements/, accessed 02 July 2020.
[26] http://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml, particularly section 2.5 *Acquisition Strategies*, accessed 02 July 2020.
[27] See *Partnerships: Mandated + Recommended Archive Status* - https://archaeologydataservice.ac.uk/research/partnerships.xhtml, accessed 02 July 2020.

3.0.4 The ADS also acts as a Trusted Digital Repository for digital data generated from projects undertaken within the historic environment, on behalf of the following Local Authorities and Museums:[28]

- *Birmingham City Council + Birmingham Museum Trust*
- *Cambridgeshire County Council*
- *Derbyshire County Council*
- *Devon County Council + Devon Museums*
- *Durham County Council and Darlington Borough Council*
- *Gloucestershire County Council, Gloucester City Council and Gloucester Museums Group*
- *Great North Museum: Hancock*
- *Hampshire County Council Arts and Museums Service*
- *Hertfordshire Museums Association*
- *Lincolnshire County Council*
- *Oxfordshire County Council and Oxfordshire Museum Service*
- *The Potteries Museum and Art Gallery*
- *Southampton City Council*
- *South West Heritage Trust*
- *City of York Council*
- *Museums Worcestershire*

3.0.5 The ADS also works with external organisations and agencies to provide guidance for funding applications and project-based digital preservation:[29]

- *British Academy*
- *Heritage Lottery Fund*
- *Council for British Archaeology*
- *Society of Antiquaries of London*

3.0.6 The ADS has Service Level Agreements (SLA) with:

- UoY IT Services[30]
- Amazon S3 Glacier[31]
- Internet Archaeology[32]

3.0.7 The ADS has *Memoranda of Understanding* (MoU) with a number of external organisations concerned with preservation and reuse of data including:[33]

- *Association of British Geological Survey*

[28] See *Partnerships: Recommended Trusted Digital Repository for digital data, on behalf of the following Local Authorities and Museums* - https://archaeologydataservice.ac.uk/research/partnerships.xhtml, accessed 02 July 2020.
[29] See *Partnerships: Other Partnerships* - https://archaeologydataservice.ac.uk/research/partnerships.xhtml, accessed 02 July 2020.
[30] *Service Level Agreement (SLA) for the Provision of Information Technology Services.* Restricted access.
[31] https://aws.amazon.com/s3/sla/, for provision of a remote deep storage facility.
[32] http://intarch.ac.uk/, to host and provide a preservation service to the online journal
[33] http://archaeologydataservice.ac.uk/about/memorandaOfUnderstanding , accessed 02 July 2020.

- *Association of Local Government Archaeological Officers* (ALGAO)
- *Council for British Archaeology* (CBA)
- *Royal Commission on the Ancient and Historical Monuments of Scotland* (now Historic Environment Scotland, HES)
- *Royal Commission on the Ancient and Historical Monuments of Wales* (RCAHMW)
- *Royal Commission on the Historical Monuments of England* (RCHME now part of Historic England)
- *MDA* (now Collections Trust)
- *National Trust*
- *Bedern Group*

4. Preservation Objectives

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

4.0.2 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.[34]

4.0.3 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.

4.0.4 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.

4.0.5 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:

[34] 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, accessed 02 July 2020.

- 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 (AC1024).[35] 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.
- 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.

4.0.6 Data that cannot be normalised, or migrated between versions, are unsuitable for long-term preservation within the framework described.

4.0.7 The ADS ensures documentation of all preservation actions and processes through its *Collections Management System* (CMS)[36] and *Object Metadata System* (OMS)[37] ensuring that a clear data trail for all files and digital objects. To build interoperability into this process the ADS utilises the PREMIS schema to document all preservation events.[38]

4.0.7 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'.[39]

---

[35] 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.
[36] Internal access only.
[37] Internal access only.
[38] 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*, specifically Appendix 3 - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest), accessed 02 July 2020. More generally see https://www.loc.gov/standards/premis/, accessed 02 July 2020.
[39] 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

- Reference: Concerned with unambiguously identifying content information through, for example, the provision of an ISBN number for a publication.[40]
- 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 checksum value and will consistently produce this value if a file is unchanged. The checksum thus provides a mechanism for validating and auditing data.[41]
- 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'.[42]

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

4.0.9 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[43]
- the existence of adequate documentation to facilitate ongoing preservation and reuse[44]

4.0.10 The other cornerstones are

- ongoing access to adequate hardware systems by skilled staff[45]
- that robust backup/recovery strategies are in place[46]

4.0.11 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,

---

and managed within the *Collections Management System* (CMS) and *Object Metadata Store* (OMS) as outlined in the *Repository Operations Manual*.

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

[41] 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) while fixity values also allow the repository to monitor files throughout the data lifecycle (see *Repository Operations Manual* – Section 6).

[42] 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, accessed 02 July 2020).

[43] Detail of the preservation pathways through normalisation and migration are available in the ADS' *Data Procedures* (see Appendix 1). These procedures are subject to regular review.

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

[45] 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.

[46] Discussed below.

digital media are notoriously frail to name some. Risk assessment[47], appropriate planning[19] and systems budgets[48] allow the ADS to mitigate for such weaknesses. The *Repository Operations* provides a fuller account of the ADS implementation of the OAIS model.[12]

4.0.12 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).[49] 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.[50]

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

1. *CoreTrustSeal Data Repository Certification*. An international, community based, non-governmental, and non-profit organization, developed under the umbrella of the *Research Data Alliance*, which promotes sustainable and trustworthy data infrastructures.[51] The ADS achieved CoreTrustSeal certification in May 2020.[8]
2. *World Data System* (WDS). WDS is an interdisciplinary body of *the International Science Council* (ISC; formerly ICSU) providing certification 'trusted data service'. The ADS achieved membership of the WDS in June 2020.[9]
3. *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 'core' assessment offered by the CoreTrustSeal. Although the ADS does not hold the NESTOR-seal, it continues to investigate the accreditation standard with a view to a future application.[52]

[47] The *Risk Register* allows the ADS to monitor risk generally. The monitoring of information security through the *Information Security Risk Assessment*, with the *Security Overview* outlining steps to mitigate risk. Each is subject to regular review.
[48] The *Systems Overview* provides details on hardware and infrastructure, with an annual systems budget for renewal of physical hardware. Budgetary commitments are subject to regular review.
[49] See the *Collection Policy*, section *2.9 Embargo Periods* - https://archaeologydataservice.ac.uk/advice/collectionsPolicy.xhtml, accessed 15 July 2020.
[50] 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).
[51] This replaces the certification offered by the *Data Seal of Approval* (DSA). The ADS achieved DSA certification in 2010, reapplying in 2013, and held the latest version of the DSA (2014-2017). See 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, accessed 02 July 2020.
[52] *NESTOR-Seal* - https://www.langzeitarchivierung.de/, accessed 02 July 2020.

4.0.14 The ADS also seeks to evaluate infrastructure and workflows through recognised self-assessment standards, conducting regular reviews of processes and procedures based upon these tools and specifications.

1. *Trustworthy Repositories Audit and Certification* (TRAC).[53]
2. *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA).[54]
3. *DPC Rapid Assessment Model* (DPC RAM).[55]

5. Identification of Content

5.0.1 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.[11] 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.[56]

5.0.2 The *Guides to Good Practice*, a collaborative work by the ADS and Digital Antiquity, have sort to raise the profile of good data standards within the profession and authored by professionals and experts from within the sector.[57] These standards, alongside more direct engagement with the repositories own designated community, form the basis of the ADS' own *Data Procedures*[58] and *Guidelines for Depositors*.[59] 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.[60] Through direct and indirect engagement the ADS ensures that data adequate documentation and metadata forms part of the dataset.

---

[53] *Trustworthy Repositories Audit and Certification* - http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories, accessed 02 July 2020.
[54] *Digital Repository Audit Method Based on Risk Assessment* - https://www.dcc.ac.uk/tools/drambora, accessed 02 July 2020.
[55] *DPC Rapid Assessment Model* - https://www.dpconline.org/our-work/dpc-ram, accessed 02 July 2020.
[56] Details of the Management Committee members are available - https://archaeologydataservice.ac.uk/about/managementCommittee.xhtml, accessed 02 July 2020.
[57] *Guides to Good Practice* - https://guides.archaeologydataservice.ac.uk/g2gpwiki/, accessed 02 July 2020.
[58] See *Appendix 9* below.
[59] *Guidelines for Depositors* - https://archaeologydataservice.ac.uk/advice/guidelinesForDepositors.xhtml, accessed 02 July 2020.
[60] The ADS provides potential depositors with a list of preferred and accepted formats, intended to mitigate for long-term risks for preservation (https://archaeologydataservice.ac.uk/advice/Downloads.xhtml, accessed 02 July 2020). The *Data Procedures* also document preservation and dissemination formats used by the ADS (see Appendix 1).

5.0.3 The ADS also actively advocates for best practice within its designated community,[61] 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.[62]

6. Procedural Accountability

6.0.1 The ADS considers procedural accountability to be of the utmost significance for its works. With this is 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.[63] 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*[52] and *ISO standards*.

6.0.2 In the interest of transparency the ADS makes vast majority of its procedural documentation available through its website,[64] allowing depositors and users to consider current working practices within the repository.[65] The ADS publishes an annual report each year and this details the activities and practices carried out in the preceding year.[66]

6.0.3 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.[67] This committee meets annually and acts in a purely advisory capacity and without legal liability.

6.0.4 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.[68]

6.0.5 Accountability pertaining to preservation and reuse falls to repository staff:[69]

- Director

[61] See, for example, the list of presentations (https://archaeologydataservice.ac.uk/research/presentations.xhtml, accessed 02 July 2020) and publications (https://archaeologydataservice.ac.uk/about/aBibliographyOfTheADS.xhtml, accessed 02 July 2020).
[62] See, for example, *Guidelines for Depositors: Preparing Datasets* https://archaeologydataservice.ac.uk/advice/PreparingDatasets.xhtml, accessed 02 July 2020.
[63] See Section 4.
[64] See, for example, *Preservation Policy and Repository Procedures* - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml, accessed 02 July 2020.
[65] Some procedural documentation is available in a redacted form in the interest of security.
[66] *Annual Reposts* - https://archaeologydataservice.ac.uk/about/annualReports.xhtml, accessed 03 July 2020.
[67] Management Committee - https://archaeologydataservice.ac.uk/about/governance.xhtml, accessed 03 July 2020.
[68] *University of York Pay and Grading* - https://www.york.ac.uk/admin/hr/browse/pay-and-grading/role-evaluation/, accessed 06 July 2020.
[69] An overview of roles and responsibilities is available publically - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Roles, accessed 02 July 2020, with more detailed information available internally.

- Deputy Director
- Collections Development Manager
- Archives Manager
- International Projects Manager
- Systems Manager
- Applications developer
- Digital Archivists
- Trainee Digital Archivists
- Administrator

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

7. Guidance and Implementation

7.0.1 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*.[70]

7.0.2 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)

7.0.3 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:[71]

- 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

[70] http://www.ukoln.ac.uk/services/papers/bl/framework/framework.html, accessed 03 July 2020.
[71] Where personal information is collected, the ADS adheres to its own *Privacy Policy* (https://archaeologydataservice.ac.uk/advice/Privacy.xhtml), *Website Terms & Conditions* (https://archaeologydataservice.ac.uk/advice/WebsiteTerms.xhtml) and *Cookies Policy* (https://archaeologydataservice.ac.uk/about/Cookies.xhtml), alongside the wider policies of its parent organisation, the University of York (http://www.ukoln.ac.uk/services/papers/bl/framework/framework.html) accessed 03 July 2020.

- Admin: used for the administration of the CMS
- Project Documentation: used to document research and technical projects undertaken by the repository.
- Advice: used for record informal communications with both users and potential depositors

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

7.0.5 Repository staff can initiate tools and processes through the CMS interface.[72] At the same time, the integration of a suite of tools within the CMS allows repository staff to automate parts of ingest, preservation and AIP checking workflow.[73]

7.0.6 The CMS allows repository staff to track archives and datasets as they progress through repository.[74]

7.1 Data Creation

Lead role: Collections Development Manager
Policy document: Collections Policy

7.1.1 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*[11]
- *Guidelines for depositors*[59]
- *Preferred and accepted file formats*[60]
- *Guidance on the selection of material for deposit and archive*[75]
- *Policy and Guidance on the deposition of personal, confidential and sensitive data*[20]
- Advisory services

[72] DROID (discussed below in Section 7.2: Acquisition, Retention or Disposal), for example, can initiated through the CMS. The CMS also provides a mechanism to call external services, for example, allowing DOI creation, the use of external thesauri for facilitate metadata generation, to metadata import metadata from templates, etc.
[73] These 'CATS tools' allow repository staff to create 'deposit receipts', check for file duplication, export file metadata (for inclusion in archive interfaces), etc.
[74] A 'simple' checklist allows repository staff to see at what stage a dataset is within the preservation workflow and which member of staff has/is working on the data.
[75] *Guidance on the selection of material for deposit and archive* - https://archaeologydataservice.ac.uk/advice/selectionGuidance.xhtml, accessed 02 July 2020.

7.1.2 The ADS also provides more generic information and guidance about formats and metadata through the *Guides to Good Practice*.[76]

7.2 Acquisition, Retention or Disposal

Lead role: Archives Manager/Digital Archivists
Policy document: Preservation Policy

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

- *Repository Operations*[12]
- *Ingest Manual*[13]
- *Data Procedures*[58]
- Procedure checklists[77]

7.2.2 And are carried out in accordance with the

- *Information Security Risk Assessment*[16]
- *Security Overview*[18]

7.2.3 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.[78] The depositor, or original copyright holder, retain their intellectual property rights; and the ADS makes no claim over these. Within the licence depositors are can express the dissemination terms under which the data made available.[79]

7.2.4 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. The *Repository Operations Manual*[12] and *Ingest Manual*[13] provide a detailed explanation of processes and management of such collections. The repository also uses versioning and unique identifiers (DOIs) to facilitate the ongoing usage of such datasets.[80]

7.2.5 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

[76] http://guides.archaeologydataservice.ac.uk/
[77] 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, with static versions provided through the website - https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Checklists, accessed 03 July 2020.
[78] A sample licence is available - https://archaeologydataservice.ac.uk/advice/Downloads.xhtml, accessed 03 July 2020.
[79] See Appendix B. The ADS encourages depositors to dissemination under a CC-BY-4.0 licence, but depositors can request the sharing of data under whatever licencing terms they deem appropriate. Historically, however, dissemination was under the ADS' own terms and conditions, copyright and liability statement and common access agreement.
[80] See Section *7.4.2 Resource Discovery*, for a wider discussion of persistent identifiers and DOIs.

dissemination). The retention of the significant properties of original files is a primary concern during normalisation or migration, as detailed in ADS' *Data Procedures*.[58] 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) [81] but remain otherwise unchanged in terms of format and content.

7.2.6 The ADS uses a formalised directory structure to store datasets as outlined in the *Repository Operations Manual*.[12] Where possible, beneath this structure, the repository will maintain the original data structure.

7.2.7 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.[82] The *Ingest Manual* provides a detailed account of implementation and usage of DROID within the preservation workflow.[13]

7.2.8 Repository staff 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.[83] The OMS tables allow archivists to record relationships between discrete objects, with the nature of that relationship described using the necessary PREMIS concept.[84] The *Ingest Manual* provides a fuller discussion of the OMS.[13]

7.2.9 The CMS also allows the documentation and recording of all processes carried out by repository staff.[85]

7.2.10 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*.[60] As outlined in the *Ingest Manual* in such instances, repository staff take appropriate actions to ensure the submission of new, or replacement, files.[13] Similarly, where documentation is inadequate ADS staff will negotiate with depositors about these

[81] 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*, but additional changes may be required. The *Repository Operations Manual* provides a more detailed discussion, see specifically *Appendix 4: Reserved File Names* and *Appendix 5: File naming policy*, http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#RepOp, accessed 03 July 2020.
[82] Developed by the National Archives (UK) the DROID is a file characterisation tool allowing the identification of files and the creation of technical metadata - http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/, accessed 03 July 2020.
[83] A detailed discussion on 'Match Objects' can be found in the *Ingest Manual*, specifically *Appendix 8: Match Objects*, http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest, accessed 03 July 2020.
[84] See *Ingest Manual,* specifically *Appendix 3: Relationship types*, http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest, accessed 03 July 2020.
[85] These include (data) capture, compression, creation, deletion, migration, renaming, restructure, etc.

requirements. The provision of guidance on selection and retention[75] and communication with the depositor by repository staff ensures mitigation prior to deposition.[86]

7.2.11 The *Ingest Manual* outlines policies concerning the retention of physical media.[87]

7.3 Preservation and Management

Lead role: Archives Manager/Digital Archivists
Policy document: Preservation Policy/Repository Operations Manual

7.3.1 Storage and Resilience

7.3.1.1 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*[88] and *OASIS Images*, is 'stored', in the short term, on a separate dedicated virtual machine until formal accession can been 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.

7.3.1.2 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.

7.3.1.3 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 ll 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.

7.3.1.4 The ADS has implemented cloud-based storage via Amazon Web Services (AWS) for long-term/deep storage of the off-site backup only.[89] The SLA for Amazon S3 Glacier

[86] Additional information on this can be found in *the Ingest Manual*, see *Section 3.3 Check file formats are suitable for deposition* and *3.6 Documentation and integrity check*, http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest, accessed 03 July 2020.
[87] See *Ingest Manual*, specifically *Section 3.16 Store original media* - http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest , accessed 03 July 2020.
[88] ADS-easy - https://archaeologydataservice.ac.uk/easy/, accessed 03 July 2020.
[89] Amazon Web Services - https://aws.amazon.com/, accessed 03 July 2020.

SLA is available.[90] 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, failure and physical threats to local storage and, therefore, build resilience into the preservation of the datasets that 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 internally, and utilises file fixity and time stamps to mitigate for error and corruption during data transfer. The ADS has intimated that all data stored using AWS should be stored within the European Union, specifically Ireland.

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

7.3.1.6 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.[91]

7.3.2 Data Management

7.3.2.1 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.[92] 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 also stores collection and resource discovery metadata and controls elements archive interface published on the ADS website

7.3.2.2 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,[93] 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.

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

[90] Amazon Web Services SLA - https://aws.amazon.com/s3/sla/, accessed 03 July 2020.
[91] 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.
[92] Internal access only.
[93] See *Ingest Manual*[13] and *Repository Operations Manual.*[12]

7.3.2.4 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.[94] 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 data degradation. Repository staff can then take appropriate action to ensure the correct preservation of that files and datasets.

7.3.2.5 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

7.3.2.6 The ADS has successfully completed a migration of all CAD files to AutoCAD Release 2010/11/12 (AC1024).[35] Repository staff maintain an ongoing technology watch to ensure the consideration of format changes and software updates and any impacts upon preservation workflows.[95] 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.

7.3.2.7 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 licence[96]

[94] As outlined in the *Ingest Manual*, see specifically *Section 3.11 Create checksums and technical metadata* and *Section 4.4 Technical metadata creation and metadata checks*, http://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#Ingest, accessed 03 July 2020.
[95] Where necessary the repository will update its list of *Preferred and accepted file formats*[60] and metadata requirements outlined in the *Guidelines for Depositors*[39] and make changes to its *Data Procedures*[58] in light of any developments.
[96] The *ADS Terms of use and Access to Data* provide users with detailed information on the appropriate use and citation of digital resources within the repository - https://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess.xhtml, accessed 03 July 2020.

- a content objection by a data consumer/user[97]
- a depositor (producer) no longer wishes to make a resource available[14]
- a resource was deposited with a formally agreed lifespan[98]
- data was submitted in formats outwit of the list of accepted formats, but on the request of the depositor have been retained and preserved on a 'best efforts' basis[99]

7.3.2.8 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.[14] 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.

7.3.2.9 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.[100] 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

7.4.0.1 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 advocates for the *FAIR data principles* of data stewardship and the *TRUST* framework for digital repositories the ADS is committed to making its holdings findable, accessible, interoperable and reusable allowing its user community to maximise resources for research, teaching or learning. [101] As such the repository recognises that preservation

[97] The *Website Terms and Conditions* provides an outline of the content objection process - https://archaeologydataservice.ac.uk/advice/WebsiteTerms.xhtml, accessed 03 July 2020.
[98] A detailed agreement on the 'life' of the deposition within the repository should be sort, with agreement from repository and depositor, prior to ingest.
[99] In such circumstances the depositor will be notified at ingest that the repository cannot guarantee preservation of the datasets
[100] See *ADS MoU with University Information Services & Library 2016* - https://archaeologydataservice.ac.uk/manPages/mou.xhtml (internal access only).
[101] See discussion in Section 2, above.

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

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

- a deposit licence for each resource[78]
- a copyright and liability statement[102]
- a common access agreement[102]
- website terms and conditions[97]

7.4.2 Resource Discovery[103]

7.4.2.1 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. ADS Archives,[104] ArchSearch,[105] ADS Library,[106] etc.

7.4.2.2 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.[107] 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 thesauri, particularly those published as structured Linked Open Data (LOD).[108] 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*.[109] Where appropriate datasets use

[102] The *ADS Terms of use and Access to Data* - https://archaeologydataservice.ac.uk/advice/termsOfUseAndAccess.xhtml, accessed 03 July 2020.
[103] See *Section 5.3. Data Delivery Mechanisms and Tools* of the *Collections Policy*.[11]
[104] *Archives* - https://archaeologydataservice.ac.uk/archive/, accessed 03 July 2020.
[105] *ArchSearch* - https://archaeologydataservice.ac.uk/archsearch/basic.xhtml , accessed 03 July 2020.
[106] *ADS Library* - https://archaeologydataservice.ac.uk/library/, accessed 03 July 2020.
[107] The template, alongside descriptions of the metadata requirements, are available from *Guidelines for Depositors* - https://archaeologydataservice.ac.uk/advice/Downloads.xhtml, accessed 03 July 2020.
[108] *Historic England* (HE), *Historic Environment Scotland* (HES) and the *Royal Commission on Ancient & Historical Monuments of Wales* (RCAHMW) provide discreet regional thesauri (https://www.heritagedata.org/blog/). Alongside these ADS also utilises the *Library of Congress Subject Headings* (http://id.loc.gov/authorities/subjects/sh2013002090.html). Accessed 03 July 2020.
[109] 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). Accessed 03 July 2020.

geographic coordinates (decimal latitude/longitude) as part of GEMINI compliance. The repository also uses thesauri of temporal terms to qualify data and collections chronologically.[110]

7.4.2.3 The repository uses discrete search facilities for each of its catalogues/resources to promote and improve resource discovery.[111] 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.[112] It also provides access to its linked data repository through a SPARQL endpoint that allows the interrogation of its triple-store through both a query interface, or via a specialised client.[112]

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

- *Heritage Gateway*[113]
- *Marine Environmental Data and Information Network* (MEDIN) data portal[114]
- *Natural Environment Research Council* (NERC) data discovery portal[115]
- the *Keepers Registry*[116]
- *Europeana*[117]
- ARIADNEPlus Portal[118]

7.4.2.5 The ADS use *Digital Object Identifier* (DOI) to facilitate the citation of its resources. The ADS, through an ongoing partnership with the British Library, mint DOIs with DataCite. DOIs also provides a mechanism for the repository to track and measure citations referencing its resources.[119] The DOI also provides a mechanism for the repository to track and measure citations referencing its resources.

[110] For UK projects the ADS use thesauri curated by the *Forum for Information Standards in Heritage* (FISH) (http://www.heritage-standards.org.uk/fish-vocabularies/). This encompasses the terminologies supplied by *Historic England* (HE), *Historic Environment Scotland* (HES) and the *Royal Commission on Ancient & Historical Monuments of Wales* (RCAHMW).For non-UK projects, the ADS use a generic list of archaeological periods first defined in the 1998 version of the *Manual and Data Standard for Monument Inventories* (MIDAS) temporal terms (http://www.heritage-standards.org.uk/midas-heritage/). Accessed 03 July 2020.
[111] An Apache Solr architecture (https://lucene.apache.org/solr/, accessed 03 July 2020) form the basis of series of discrete catalogues/indexes.
[112] *Endpoints* - https://archaeologydataservice.ac.uk/about/endpoints.xhtml, accessed 03 July 2020.
[113] *Heritage Gateway* - https://www.heritagegateway.org.uk/gateway/default.aspx, accessed 03 July 2020
[114] *MEDIN portal* - http://portal.oceannet.org/portal/start.php, accessed 03 July 2020.
[115] *NERC data catalogue service* - https://data-search.nerc.ac.uk/geonetwork/srv/eng/catalog.search#/home, accessed 03 July 2020.
[116] *The Keeper's Registry* - https://thekeepers.org/, accessed 03 July 2020.
[117] *Europeana* - https://www.europeana.eu/portal/en, accessed 03 July 2020.
[118] *ARIADNEPlus* - https://ariadne-infrastructure.eu/portal/, accessed 03 July 2020.
[119] DataCite - https://www.datacite.org/, accessed 03 July 2020.

7.4.3 Rights management[120]

7.4.3.1 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.[121] 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.[122] Each archive interface provides details of these terms for data consumers and users.[123] All depositors must agree to the terms of the deposit licence as part of the deposition process as outlined in the *Guidelines for Depositors*.[59]

7.4.3.2 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 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.[97]

7.4.4 Receiving data

7.4.4.1 As advocates of the FAIR data stewardship principles[124] 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.[125] 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.[126] In those instances where compression may be necessary, the repository ensures the provision of sufficient guidance on the use of compressed data. 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.

7.4.4.2 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.[43] Collection and file-level metadata accompanies all DIPS to facilitate interoperability and reuse.

---

[120] See *Section 2.6. Rights Management* and *Section 5.2. Rights Management* of the *Collections Policy*.[11]
[121] In a rare instances access may be restricted, typically this a consequence of technical issues to do with the size of the dataset, but also come as the result of embargo. See the *Ingest Manual*, *Section 5.14 Archive release*, for a more detailed discussion of the associated process.[13]
[122] See *Collection Policy, Sections 2.5. Acquisition Strategies*.[11]
[123] This is expressed in the vertical bar on the left side of each archive interface (i.e. '*This work is licensed under a …*). See also *Section 7.2 Acquisition, Retention or Disposal* above.
[124] See discussion in *Section 2 Principle Statement* and *7.4 Access and use* above.
[125] With access to the *Dissemination Information Package* (DIP) only.
[126] Typically, 'ZIP' archive, or in rare instances, the supply of data may be enacted 'on request'.

7.4.5 Security of delivery systems

7.4.5.1 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*[16]
- *Risk Register*[15]
- *Disaster Recovery Plan*[19]
- *Systems Overview*[17]
- *Security Overview*[18]

7.4.5.2 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,[54] and regularly evaluates the effectiveness of this assessment and associated risk management implications.

7.4.5.3 The spread of ADS' systems and resources over a number of virtual machines provides mitigation for any risk.[17] 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 Overview*[18] and in accordance with the *Information Security Risk Assessment*.[16] The repository also adheres to the policies, procedures and guidance outlined by its parent organization, the *University of York*.[18] The synchronization of all repository holdings and dataset to an off-site, deep storage facility ensures its ongoing security.

7.4.5.4 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

7.4.6.1 The ADS uses *Matomo Web Analytics*[127] to monitor user experience and collect, non-personally-identifying, statistical information on access to repository resources and datasets. More specific access statistics, displayed alongside each archive[128], allows data

[127] *Matomo Web Analytics* - https://matomo.org/, accessed 03 July 2020.
[128] See the 'Usage Statistics' for each dataset/resource, see, for example, Kevin Camidge, CISMAS (2019*) Isles of Scilly Designated Wrecks Interpretation* [data-set]. York: Archaeology Data Service [distributor] https://doi.org/10.5284/1051619.

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.[129]

7.4.7 Outage

7.4.7.1 The ADS monitors all planned and unplanned service downtime. At the same time, the repositories 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.[130]

[129] The use of DOI statistics is still in development but the ADS hope to make these accessible in the near future.
[130] http://archaeologydataservice.ac.uk/advice/Glossary.xhtml

Appendix 1

A1.1 ADS Data Procedures documents

A1.1.1 Current versions of the procedures documents are available within the ADS internal wiki, but static versions are accessible:[131]

- 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

A1.1.2 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.

[131] https://archaeologydataservice.ac.uk/advice/PolicyDocuments.xhtml#DataProcedures