# ADS Preservation Policy

*Tim Evans*

***July 2022***

## Document Control Grid

| | |
|---|---|
| Title: | ADS Preservation Policy |
| File name: | ADS_preservation_policy_v1-8 |
| Status: | LIVE |
| Version: | 1.8 |
| Last updated: | 20th July 2022 |
| Created date: | 30 July 2004 |
| Review due: | July 2023 |
| Authors: | Tim Evans |
| Maintained by: | Deputy Director |
| Required Action: | None |

Archaeology Data Service

# Glossary

| | |
|---|---|
| Accession | The formal act of accepting a dataset or object into the ADS repository system |
| ADS | Archaeology Data Service |
| ADS-easy | A web-based application which allows Depositors to upload data to the ADS |
| AIP | Archival Information Package |
| AWS | Amazon Web Services |
| DIP | Dissemination Information Package |
| Depositor | The person or persons responsible for submitting data to the ADS |
| Metadata | Descriptive information about data |
| NFS | Networked File Storage |
| Object | A package of information that is defined by a single set of metadata, for example a report or a database. |
| SIP | Submission Information Package |
| UoY | University of York |
| User | An individual that uses the ADS website to access data |
| VM | Virtual Machine |

# 1. Introduction

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).

# 2. Principal statement

The Archaeology Data Service (ADS) is a digital repository for archaeology and historic environment data generated by UK-based fieldwork and research. Founded in 1996, the core activity of the ADS is the long-term digital preservation of the data entrusted to our care. The ADS follows a policy of active data management and curation to ensure the integrity, reliability and accessibility in perpetuity of all data. All resources archived with the ADS are Open Access, and delivered through our website to facilitate re-use by the heritage sector and wider community. The ADS actively follows preservation and management strategies with the aim of ensuring the authenticity, reliability and integrity of all resources entrusted to its care.

The ADS champions and encourages re-use of the data it preserves to our user community. The latter has broadened over time from a largely UK-based academic focus to an international audience encompassing all 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.

The ADS adheres to the FAIR data principles (see Wilkinson et al 2016) 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.

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 (see Lin et al 2020).

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 via CoreTrustSeal and membership of the World Data System (WDS) ensure that current processes and practices meet recognised standards. At the same time, less formal methods of review and self-regulation, including the DPC RAM, alongside wider involvement in the preservation community, ensure the validity and currency of repository workflows and processes.

# 3. Supporting documents

This document outlines the basis for the archival practice developed by the ADS. It does not exist in isolation but as part of a suite of documents guiding good governance and practice by the ADS. Policy and procedural documents include:

- Five Year Strategic Plan (October 2021 - October 2026)
- Collections Policy
- Repository Operations
- Ingest Manual
- Assessment and Appraisal Policy
- Deaccession Policy
- Sensitive Data Policy
- Risk Register and the Information Security Risk Assessment

- Systems Overview
- Security Overview
- Disaster Recovery Plan
- Cataloguing Policy
- Roles and Responsibilities Overview

The University of York (UoY) is the legal entity for all services, consultancy and other contracts carried out by the ADS. Memoranda of understanding outline the specifics of this relationship:

- ADS MoU with University of York Registrar and Finance (2012). This document contains the agreement on the hosting of ADS by the Department of Archaeology, University of York.
- ADS MoU with The University of York Library and Archives and The University of York Department of Archaeology (May 2022). The purpose of this agreement is to ensure that should ADS be unable to continue its digital preservation, that all objects and related metadata will be transferred to the UoY digital repository and digital archive.
- ADS MoU with University of York IT Services (May 2020). The UoY Information Technology Services (ITS) provides computing and technical support for the ADS; hosting its database(s), applications and website(s) through a series of virtual machines. The Network File Service (NFS) used by the ADS for storage of AIP and DIP objects is also managed by UoY ITS.

A full overview of UoY and third party dependencies is provided in the ADS Outsource Partners document.

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
- University Information Policy
- Information Security Policy
- Legal Statements and linked policy and strategy documents (Accessibility, Copyright, Privacy, Cookies)

# 4. Preservation aims and objectives

The core preservation aims of the ADS are:

- The continued preservation of digital data, so that it can be accessed and reused by our designated community in perpetuity.
- That (meta)data within the ADS archive meets the highest technical and thematic standards, to facilitate access and understanding, and interoperability with other digital preservation repositories.

Encompassed within this broad statement are a number of defined objectives.

- Implement a clear, understandable, and portable **Framework for digital preservation**.
- Implement a clearly defined **Method of preservation**, with procedure and operations that document practice undertaken by ADS staff.
- Ensure a strategy for individual **Object Management**; ensure independent utility for each object, preservation of intra-specific relationships between representations, and inter-specific relationships between objects.
- Ensure that the ADS archive meets the highest appropriate level of peer **Accreditation**, ensuring that ADS operations are well-documented, accessible, transparent, and trusted.

- Maintain a flexible **Repository system** that facilitates best preservation practice, and is adaptable to changing requirements of ADS staff.
- Support preservation work, and ensure capacity through clearly defined **Roles and responsibilities**.
- Maintain a **Standards and preservation watch**, ensuring knowledge of advances or changes in the wider environment are fed back into ADS procedure, including file formats.
- Maintain and implement clear **Data procedures** that reflect the significant properties of specific file types and ensure their preservation throughout the ADS life cycle.
- Ensure that each digital object is described using **Consistent and interoperable metadata** that covers all preservation description information.
- Ensure all necessary steps are taken to ensure **data integrity (fixity)**.
- Facilitate **Reuse and transparency of data** by ensuring all objects are publicly available, of where relevant, covered by a clear embargo or redaction policy.
- Implement a strategy to ensure **Sustainability** of the archive by mitigating risk.

# 5. Meeting preservation objectives

## 5.1. Framework for digital preservation

The ADS conceptual framework is based on 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. OAIS provides a useful conceptual framework and common language with which to define ADS practice, and helps describe the responsibilities and interactions of depositors, ADS staff and users.

Of specific importance is the OAIS model's definition of Information Packages (and associated objects) throughout a typical lifecycle from deposition to public access. Thus the ADS will use the following OAIS definitions within our Repository Operations:

- Submission Information Package (SIP): Supplied by a data producer (creator or depositor) including documentation to facilitate archiving and reuse. It is important to note that in the case of the ADS, the SIP is the actual data formally accessioned.
- Archival Information Package (AIP): Generated from the SIP and the long term preservation package managed within the OAIS including administrative, technical and reuse documentation. The ADS, through use of the term "digital object", concurs with OAIS in defining two 'specialisations' of an AIP: an Archival Information Unit (AIU), and an Archival Information Collection (AIC).
- Dissemination Information Package (DIP): Generated from the SIP/AIP and made available to consumers (users) including documentation to facilitate reuse.

## 5.2. Method of preservation

The ADS practise **preservation through migration or normalisation**. This approach stands in contrast to other preservation methodologies involving emulation or technology preservation.

The ADS uses the following event types for ongoing preservation:

- Normalization: Within the ADS, all digital objects of a particular type (e.g.raster images) are converted into a single chosen file format that is thought to embody the best overall compromise amongst characteristics such as functionality, longevity, and preservability. The formats chosen are set within the individual ADS Data Procedures.

- Migration: As well as normalisation data may be migrated to other formats as part of longer-term, and repository wide, reviews of data types and standards. Data may be migrated through successive *versions* of a format. Version migration may be the only option for preserving proprietary formats that do not migrate to open standards, or where . This is only practical where the software using proprietary formats is widely used within a community and accessible (affordable) to an archive.

- Replication: migration between media that leaves data (the bit stream) totally unchanged.

Data that cannot be replicated, normalized, or migrated between, are unsuitable for long-term preservation within the framework described, and should not be accessioned.

The details for the implementation of this strategy is contained in the ADS Repository Operations document, and the individual Data Procedures documents which detail data type specific interventions.

## 5.3. Object Management

The ADS ensures that each digital object is a clearly defined information package, and uses the PREMIS Data Dictionary definition of file objects, representation

objects, and bitstream objects as a basis for the ADS Object Management Store (OMS).

The OMS, as defined in Repository Operations, ensures the maintenance of relationships between individual representations, and conversely larger groups of objects that make up an intellectual - or otherwise human defined - entity.

## 5.4. Accreditation

The ADS actively seeks compliance with community driven initiatives for best practice. These include:

- CoreTrustSeal Data Repository Certification. An international, community based, non-governmental, and non-profit organisation, developed under the umbrella of the Research Data Alliance, which promotes sustainable and trustworthy data infrastructures. The ADS achieved CoreTrustSeal certification in May 2020.
- 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

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.

- Trustworthy Repositories Audit and Certification (TRAC).
- Digital Repository Audit Method Based on Risk Assessment (DRAMBORA).
- DPC Rapid Assessment Model (DPC RAM).
- DPC Competency Framework.

## 5.5. Repository system

The ADS ensures documentation of all preservation actions and processes through its Collections Management System (CMS) and Object Metadata Store (OMS).

The ADS CMS is used to track and document collections and datasets throughout the data lifecycle. The CMS also allows repository staff to record future depositions. The CMS takes a modular approach divided along the following lines:

- Tracking: used to record information about current and future collections. Information recorded includes Administrative metadata (such as Creator, Primary Contact, Copyright Holder), deposit licences. The Tracking module also records each accession, creating a unique identifier for each event.
- 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 used for delivery.
- People: centrally manages individuals and organisations who contact the ADS. Where personal information is collected, the ADS adheres to its own Privacy Policy.
- 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

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. The Ingest Manual provides a detailed account of implementation and usage of DROID within the preservation workflow.

## 5.6. Roles and responsibilities

The ADS Support preservation work, and ensure capacity, through clearly defined Roles and responsibilities. This includes named responsibilities for:

- Deputy Director: oversight of all Policy and Procedural documents, including implementing annual review.
- Digital Archivist (Standards): lead role in undertaking technology watch and for updating internal procedural documentation.
- Systems Manager: maintaining CMS/OMS systems.
- Digital Archivists: undertaking review and assessment of data, performing preservation actions (e.g. normalization of objects), recording processes and documentation.

It is widely recognised that there are inherent risks associated with this objective, as skilled staff may leave or are difficult to recruit or replace. The ADS Risk Register includes appropriate planning to allow the ADS to mitigate for such weaknesses.

## 5.7. Standards and preservation watch

The ADS will actively monitor current standards for change, including but not limited to:

- Updates to current formats i.e. new versions / new software.
- New software and formats for deposit.

- New data types or techniques.

This will be achieved through:

- Monitoring developments in software and organisations associated with our current suite of formats.
- Monitoring work of organisations associated with digital preservation (e.g. DPC.).
- Monitoring the standards in use by other archiving organisations (e.g. DANS, SND).

Resources that will be routinely consulted include:

- DPC Technology Watch Publications (Reports and Guidance Notes).
- DPC News section (e.g. any other relevant work).
- Federal Agencies Digital Guidelines Initiative (FADGI).
- Data Curation Network.

Archives and Repositories monitored should include, but will not be limited to:

- DANS
- Swedish National Data Service
- IANUS (DAI)
- The National Archives
- Library of Congress
- NARA
- University of Minnesota Digital Conservancy (UDC).

## 5.8. Data procedures

The ADS will maintain a set of documents which act as the internal manuals for how to deal with every specific data type accepted by the repository.
Each document will list:

- Accepted and preferred formats for deposit.
- Formats used for preservation and dissemination versions.
- The required metadata for that data type
- Issues to check in Assessment and Appraisal.
- Definition and explanation of significant properties to be retained.
- Instructions and guidance for post-processing file naming (for example for extracted metadata).
- Instructions on how to convert files for normalization, with a list of preferred tools or software and instructions for post-process check of output format.
- Further notes.

The ADS Data procedures will be reviewed every year or, where the Standards and preservation watch has identified a change, at the earliest opportunity.

## 5.9. Consistent and interoperable metadata

The ADS are committed to maintaining high standards in all facets of metadata

- Administrative: the Metadata used in managing and administering collections and information resources (e.g. licence or rights). This is recorded and managed in the ADS CMS.

- Descriptive: the Metadata used to identify, authenticate, and describe objects. This is recorded in the ADS CMS, and follows the ADS Cataloguing Policy to ensure the recording of relevant and community standards.
- Preservation: the Metadata related to the preservation management of objects, including editing or migration events. This is recorded in the ADS CMS and OMS, and utilises the PREMIS data dictionary, with procedure defined in the ADS Repository Operations.
- Technical: the Metadata that allows us to understand what an object is. File identification is primarily based on The National Archives DROID tool, and is stored in the ADS OMS, as documented in the ADS Repository Operations.
- Use: the Metadata related to the level and type of use of objects and information resources. Is recorded in the Matomo tool, and displayed with each resource.

## 5.10. Data integrity (fixity)

The ADS have a defined procedure for the generation and storage of checksums within the ADS OMS and a local manifest file. Objects are checked against their checksums on a regular (three month) basis. Further details are recorded in Section 7 of this document.

## 5.11. Facilitate reuse and transparency of data

The ADS is committed to an ethos of 'no preservation without reuse'. The ADS supports open access to all its holdings. Each collection or object is accompanied by a rights statement, clear and obvious licence, and guidance on citation.

The ADS allows some data to be subject to a time-limited embargo at the behest of a producer, or for legal and ethical reasons. It is ADS policy that the metadata for embargoed data is made available at the earliest opportunity.

The ADS will create a Persistent Identifier for each:

- Collection.
- Information Object where the data type is a Text document.

## 5.12. Sustainability

The ADS will take all necessary steps to mitigate against future events that may impact upon the repository service. These are managed via the ADS Risk Register and the ADS Disaster Recovery Plan. Particular examples of risk or disaster resolution include:

- Future migrations: given that the ADS method of preservation is via normalization/migration, it is inevitable that large-scale file conversions will occur in the future. The ADS holds a reserves account (defined in the MoU with the UoY), which can be drawn upon to undertake or fund data preservation activities. For example, seconding staff resources to oversee and document a particular object migration.
- End of repository: as above, the ADS reserves account can be drawn upon to assist in the transfer of ADS responsibilities to another organisation.
- Loss of access to UoY data storage: the ADS maintains an off-site (non-UoY) copy of all data.

# 6. Procedural accountability

The ADS considers procedural accountability to be of the utmost significance for its works. To enable this, the ADS will conduct an annual internal and external review of all documentation to ensure that current practices and workflows are 'fit for purpose' and up to community standards. This review will be led by the Deputy Director, with support from key personnel including the Systems Manager and Standards lead. This piece of work will include a review of community driven guidelines for best practice.

The ADS will make all key procedural documentation available through its website (redacted where required), allowing depositors and users to understand current working practices within the repository.

The ADS will publish an annual report each year and this details the activities and practices carried out in the preceding year.

The ADS is governed by a management committee, made up of representatives of key stakeholders, funders, user communities and the ADS management team. This will always include a representative from a digital preservation focussed organisation. The committee will meet annually and will act in a purely advisory capacity and without legal liability.

# 7. Guidance and Implementation

## 7.1. Data lifecycle

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)

## 7.2. Data creation

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. For a dataset to be accepted by the ADS it must conform to requirements in terms of formats, metadata and documentation. These are expressed in the ADS Instructions for Depositors. The dataset content must also adhere to the following Policy and guidelines:

- ADS Collections Policy
- ADS Guidance on the selection of material for deposit and archive
- ADS Sensitive data

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

## 7.3. Acquisition, Retention or Disposal

A number of documents guide the process of creating a SIP:

- ADS Assessment and Appraisal Policy
- Ingest Manual
- Data Procedures

All submissions of data to the ADS are subject to Assessment and Appraisal. Only if the repository requirements have been met is a formal accession triggered, and SIP created. Every new dataset (i.e. not a new accession to an existing collection) must have a signed deposit licence. The ADS deposit licence grants a non-exclusive right to archive and distribute the supplied data, the depositor, or original copyright holder, retain their intellectual property rights. Within the licence depositors can express the dissemination terms under which the data made available including but not limited to CC-BY and OGL.

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 and Ingest Manual1 provide a detailed explanation of processes and management of such 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) 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 Manual. Where possible, beneath this structure, the repository will maintain the original data structure.

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. The Ingest Manual provides a detailed account of implementation and usage of DROID within the preservation workflow.

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. The OMS tables allow archivists to record relationships between discrete objects, with the nature of that relationship described using the necessary PREMIS concept. The Ingest Manual provides a fuller discussion of the OMS.

In instances a deposit 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. As outlined in the Ingest Manual in such instances, repository staff take appropriate actions to ensure the submission of new, or replacement, files. Similarly, where documentation is inadequate ADS staff will negotiate with depositors about these requirements. The provision of guidance on selection and retention and communication with the depositor by repository staff ensures mitigation prior to deposition.

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

## 7.4. Preservation and Management

### Storage and Resilience

The ADS maintains 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. In each instance access is strictly controlled and limited to authorised 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 uses 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 all 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 implemented cloud-based storage via Amazon Web Services (AWS) for long-term/deep storage of the off-site backup only. The SLA for Amazon S3 Glacier 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 the Republic of 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. Repository staff make additional back-ups, stored outside of the local network, of these databases.
In order to mitigate 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 savings on storage would be significant. The intention in such an approach is to mitigate the impact of 'bit rot' on compressed files and effects of 'lossy' compression.

## Data Management

As already noted the ADS maintains a custom-built Collection Management System (CMS) developed to act as a data management system for all collections from accession onwards. The CMS facilitates the administration of datasets submitted to the repository, but also affords the administration of enquiries about future submissions. As

well as an internal data management tool, the CMS also stores 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, 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 the 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. 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 a 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.

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

Repository staff maintain an ongoing technology watch to ensure the consideration of format changes and software updates and any impacts upon preservation workflows. 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.

## Removing Data

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 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
- a content objection by a data consumer/user
- a depositor (producer) no longer wishes to make a resource available

All of these cases are covered in full by the **Deaccession and Data Disposal** policy.

## End of ADS

Should the ADS cease to function as a 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. 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 on 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.

## 7.5. Access and use

### 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 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. 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.

### Prerequisites to access

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

- a deposit licence for each resource
- a copyright and liability statement
- a common access agreement
- website terms and conditions

## Resource Discovery

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.

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). These linked terms facilitate the creation of semantic searches and are published within the repository's own triplestore. Any geographical or locational terms utilise the Getty Thesaurus of Geographic Names and to Ordnance Survey Open Names.

Where appropriate datasets use geographic coordinates (decimal latitude/longitude) as part of GEMINI compliance. The repository also uses thesauri of temporal terms to qualify data and collections chronologically.

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.

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

The ADS uses 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. The DOI also provides a mechanism for the repository to track and measure citations referencing its resources.

## Rights management

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. Each archive interface provides details of these terms for data consumers and users. All depositors must agree to the terms of the deposit licence as part of the deposition process as outlined in the Guidelines for Depositors.

The repository respects the rights of all its data providers and users and endeavours to highlight problems and concerns about collections during the ingestion of 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 to. The repository maintains a formal procedure put into effect on receipt of reports of any rights breach.

## 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
- Risk Register
- Disaster Recovery Plan
- Systems Overview
- Security Overview

The ADS undertakes an annual high-level Information Security Risk Assessment, which, in conjunction with a Security Overview, mitigates 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, 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. 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 a regular basis, in line with the Security Overview and in accordance with the Information Security Risk Assessment.

The repository also adheres to the policies, procedures and guidance outlined by its parent organisation, the University of York. The synchronisation 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.

## References and Bibliography

Lin, D., Crabtree, J., Dillo, I. et al (2020). The TRUST Principles for digital repositories. *Sci Data* 7, 144. https://doi.org/10.1038/s41597-020-0486-7

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