

## Preservation policy template for repositories

As part of its policy work, the DCC has been creating generic policy templates for users to download and customise. Based on the findings of the curation policy report<sup>1</sup> we felt templates for data management and sharing plans, institutional policy statements and preservation policies were most needed. Work has been undertaken to create a Data Management Plan template<sup>2</sup> and online tool<sup>3</sup> and the DCC is involved in parallel activity through the ERIS project<sup>4</sup> to develop an overarching institutional policy framework.

The following template is intended for repositories to assist in the definition of a digital preservation policy. We looked at four resources to define common elements for inclusion:

- AHDS Collections Preservation Policy, v1.0 (2004)  
<http://ahds.ac.uk/documents/colls-policy-preservation-v1.pdf>
- DataShare, Policy-making for Research Data in Repositories, (2009)  
<http://www.disc-uk.org/docs/guide.pdf>
- OpenDOAR Policies Tool  
<http://www.opendoar.org/tools/policytool.php>
- UK Data Archive Preservation Policy, v3.10 (2009)  
<http://www.data-archive.ac.uk/news/publications/preservationpolicy.pdf>

There are two parts to this template:

1. A basic overview of seven suggested sections to include, with outline points for consideration of what information to include in each;
2. Example texts, guidance and levels of service provision based on the four resources noted above to help repositories compose their own policy.

The DCC is always keen for feedback on its resources, so if you have suggestions of how this could be improved, or have used the template and have lessons or example policies we could share with others, please let us know. You can email [info@dcc.ac.uk](mailto:info@dcc.ac.uk)

---

<sup>1</sup> Sarah Jones, *A report on the range of policies required for and related to digital curation*, version 1.2, (DCC, Glasgow, March 2009), available at:

[http://www.dcc.ac.uk/sites/default/files/documents/reports/DCC\\_Curation\\_Policies\\_Report.pdf](http://www.dcc.ac.uk/sites/default/files/documents/reports/DCC_Curation_Policies_Report.pdf)

<sup>2</sup> The DCC content checklist covers all the components researchers may be asked for in a Data Management or Sharing Plan, see:

[http://www.dcc.ac.uk/sites/default/files/documents/templates/DMP\\_checklist.pdf](http://www.dcc.ac.uk/sites/default/files/documents/templates/DMP_checklist.pdf)

<sup>3</sup> DMP Online is available at: <http://dmponline.hatii.arts.gla.ac.uk/>

<sup>4</sup> For details of the JISC-funded ERIS project, see: <http://eriscotland.wordpress.com/>

## Preservation policy template for repositories

1. **Aim**  
A clarification of the mission to preserve
2. **Standards**  
What standards, frameworks and models for digital preservation will be used?
3. **Content coverage**  
What type of material can be deposited / will be preserved?
4. **Overview of preservation strategy**  
Explanation of the main preservation approach(es) adopted e.g. normalisation on ingest, migration, emulation, media refreshment.  
Are certain data formats preferred? Will only specified formats be preserved?
5. **Methods / levels of preservation**  
What different types / levels of preservation service will be offered and why?  
Are there some things which the repository does not guarantee to preserve?  
How long will deposited items be retained?
6. **Implementing the strategy (operational details)**
  - a) **Procedures for preservation**  
What preservation procedures will take place - when, how often and by whom?
  - b) **Security, authenticity and integrity**  
How will you demonstrate the integrity and authenticity of items i.e. that they are what they purport to be and have not been inadvertently changed?  
How will items be safeguarded against damage or unauthorised access?
  - c) **Media refreshment**  
How and when will storage media be refreshed?
  - d) **Versioning**  
Can new versions be submitted? If so, how will these be handled?
  - e) **Withdrawal of collections**  
Can deposited items be withdrawn?  
What are the circumstances in which this is permitted?  
What is the procedure for withdrawal?  
Will metadata or references to the item persist?
7. **Sustainability plans**  
What will happen if the repository is closed or funding is reduced?  
Will services be cut or have ongoing costs been planned for?  
Are plans in place to transfer repository content if necessary?

## Preservation policy template for repositories (with examples)

The template below includes colour coded example policy statements. These come from:

- [AHDS Collections Preservation Policy, v1.0 \(2004\)](#)
- [DataShare, Policy-making for Research Data in Repositories, \(2009\)](#)
- [OpenDOAR Policies Tool](#)
- [UK Data Archive Preservation Policy, v3.10 \(2009\)](#)

### **Please note:**

*Only the preservation-specific elements of the OpenDOAR tool and the DataShare guide are covered by this template as it aims to help repositories create preservation policies.*

*The examples provided here will vary in relevance depending on repository content and preservation services. The AHDS, UKDA and DataShare examples focus on data curation, while the OpenDOAR example is more publication specific. The OpenDOAR and DataShare examples provide more generic descriptions or points for consideration, as they are tools to assist in the creation of policies rather than example policies.*

### **1. Aim**

A clarification of the mission to preserve

The AHDS Strategic Plan for 2002-2005 identifies the preservation, in useable form, of digital resources of use for research, learning, and teaching in the UK Higher Education sector as a key role for the AHDS.

The goal of the AHDS is to preserve all data and supporting material indefinitely.

This policy outlines the principles which underpin the main activities of the UK Data Archive (UKDA) – the active preservation of digital resources for use and re-use within its core user community.

The UKDA follows a policy of active preservation with the aim of ensuring the authenticity, reliability and logical integrity of all resources entrusted to its care while providing usable versions for research, teaching or learning, in perpetuity.

## 2. Standards

What standards, frameworks and models for digital preservation will be used?

The AHDS intends to adhere to emerging standards and best practice for digital preservation and curation. In particular, the AHDS will manage its collections in a manner that conforms to the requirements of the Open Archival Information System reference model (ISO 14721:2003).

The Open Archival Information System (OAIS) reference model is an international standard which proposes common terms and concepts and a framework for entities and relationships between entities in digital preservation environments. The UKDA follows the broad guidance given in the OAIS reference model, and this version of the preservation policy is the first to explicitly use OAIS terminology. From a preservation point of view this policy is generally conformant to the OAIS Reference Model, with additions and alterations which are specific to the materials held within the Archive.

In terms of national standards for the management of information security, the UKDA follows:

- BS ISO/IEC 27001: 2005;
- BS ISO/IEC 27002: 2005, BS 7799-1: 2005;
- Cross Government Actions: Mandatory Minimum Measures

The UKDA uses industry standard lossless compression tools, which are outlined in the UKDA Information Security Policy.

The UKDA follows best practice in the storage and housing of magnetic and optical media. In particular, for environmental conditions for storage media (BS 4783, ISO/IEC22051, BS ISO 18921:2002 and BS ISO 18925:2002) and for the storage of archival materials (BS 5454).

### 3. Content coverage

What type of material can be deposited / will be preserved?

Only data that conforms to appropriate standards and best practices, as assessed by the AHDS, will be accepted for indefinite preservation.

The AHDS aims to preserve data-based digital resources. Digital resource types accepted by the AHDS include, but are not limited to:

- Text documents (plain and marked-up)
- Still Image collections
- Datasets (data designed for use in spreadsheets, databases and statistical packages)
- Digital audio recordings
- Digital moving image recordings
- CAD
- GIS
- Virtual reality
- Websites

The AHDS does not preserve:

- Software
- Hardware
- Original deposited media
- Hardcopy material (paper, analogue recordings, physical objects)

Are there language considerations? Will translations be included or required? (Will text within data files, metadata or other documentation in other languages be translated into English, for example?)

What kinds of research data will be included?

- Scientific experiments
- Models and simulations (2 parts - the model and computational data arising from it)
- Observations
- Derived data
- Canonical or reference data
- Accompanying material / supplemental objects

Consider what formats will be accepted for deposit, and which are preferred.

- Will the repository accept only file formats which are well-developed, have been widely adopted and are de facto standards in the marketplace?
- Will the repository guarantee that specific file formats will be converted into data formats that remain readable and usable? (See section 5.c, File Preservation.)
- Will the repository minimise the number of file formats to be managed as far as is feasible/desirable?
- Will the repository accept only 'open' non-proprietary, well-documented file formats wherever possible?
- Will compression formats be accepted? (e.g.7zip, gzip, winzip)

Type of material

- The repository holds all types of materials.
- The repository holds all types of materials except: (please specify)
- The repository is restricted to: (please specify)
- Special materials (please specify)

Deposited items may include:

- working drafts
- submitted versions (as sent to journals for peer-review)
- accepted versions (author's final peer-reviewed drafts)
- published versions (publisher-created files)

Principal languages (please specify)

#### 4. Overview of preservation strategy

Explanation of the main preservation approach(es) adopted e.g. normalisation on ingest, migration, emulation, media refreshment.

Are certain data formats preferred? Will only specified formats be preserved?

The AHDS has adopted a migration based approach to digital preservation. In principal, the AHDS aims to ensure that the significant properties (as defined by the AHDS in consultation with the depositor) of all digital resources deposited with it can be preserved indefinitely, assuming periodic future migrations and on-going media refreshment.

It may not be possible to guarantee the readability of some file formats due to software obsolescence, but the repository may choose to promise to maintain the usability and understandability of the specific file formats over time.

- If the repository promises usability and understandability over time, what specific file formats will be included in this guarantee?

#### Example lists of preferred /accepted data format include:

Arts and Humanities Data Service: <http://www.ahds.ac.uk/depositing/deposit-formats.htm>

Florida Digital Archive: <http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>

UK Data Archive: <http://www.data-archive.ac.uk/sharing/acceptable.asp>

*'Over time, items stored in DSpace will be preserved as is, using a combination of time-honoured techniques for data management and best practices for digital preservation. As for specific formats, however, the proprietary nature of many file types makes it impossible to make guarantees' (MIT Libraries, 2002).*

Normalisation: file format conversion by the repository when data are submitted.

- Will the repository convert digital objects of a particular type (e.g. statistical software system files) to a single chosen file format that is thought to embody the best characteristics of functionality, longevity and preservability?
- Will the repository convert proprietary formats to non-proprietary formats?
- Will the repository create plain text versions of datasets (encoded in either ASCII or Unicode character sets)?
- Will the repository retain the original bitstream (file) with the item, in addition to its converted formats?

#### Functional preservation

- The repository will try to ensure continued readability and accessibility:
- Items will be migrated to new file formats where necessary.
- Where possible, software emulations will be provided to access un-migrated formats.
- It may not be possible to guarantee the readability of some unusual file formats.

The UKDA's preservation policy is based upon open and available file formats, data migration and media refreshment.

The UKDA preservation strategy is predicated on two basic principles: first, that digital storage media are inherently untrustworthy unless stored appropriately; second, that all file formats and physical storage media will ultimately become obsolete.

## 5. Methods / levels of preservation

What different types / levels of preservation service will be offered and why?

Are there some things which the repository does not guarantee to preserve?

How long will deposited items be retained?

The exact level of preservation service offered by the AHDS will depend on the technical quality and completeness of the deposited data.

There are three objectives for digital preservation:

1. Preservation of the bit stream (basic sequences of binary digits) that ultimately represent the information stored in any digital object
2. Preservation of the *information content* (words, images, sounds etc.) stored as bits and defined by a logical data model, embodied in a file or media format
3. Preservation of the *experience* (speed, layout, display device, input device characteristics etc.) of interacting with the information content

The AHDS has adopted a migration based approach to digital preservation. Through a combination of backup, media refreshment and migration to standard formats during ingest (migration at ingest is intended to reduce the frequency and complexity of future migrations), the AHDS aims to provide indefinite bit preservation and indefinite preservation of information content for all deposited data, subject to stated conditions. The AHDS does not guarantee, or actively attempt, to preserve the original experience of using deposited digital resources.

### Retention period

- Items will be retained indefinitely.
- Items will be retained for at least xxx years from the date of deposition.
- Items will be retained for the lifetime of the repository.
- Retention periods may be set for individual items, as required.

### File preservation

- Will varying levels of support be offered for various file formats?
- If not all formats are supported, will some formats be preserved only at the bit level (no migrations or transformation are planned)?
- Will the repository use encryption or compression for archival files?
- Will the repository regularly back up its files according to current best practice?
- Will data files be migrated to new file formats where necessary to preserve access to their intellectual content?

### File preservation

- The repository regularly backs up its files according to current best practice.
- The original bit stream is retained for all items, in addition to any upgraded formats.
- Items are microfilmed for long-term preservation.

The repository is working with external partners to:

- convert or migrate file formats
- develop and implement software emulations for old file formats
- record preservation metadata
- backup items in external archives

Preservation decisions at the Archive must always be made within the context of its Collections Development Policy, balancing the constraints of cost, scholarly and historical value, and user accessibility alongside the requirements of levels of authenticity and legal admissibility. Hence, different ingest processes may be required for material with different levels of quality and significance. Data collections are assigned an ingest activity level as outlined in the Archive's Acquisition Review Process document.

## 6. Implementing the strategy (operational details)

### Procedures for preservation

What preservation procedures will take place - when, how often and by whom?

To support bit preservation of all deposited data, the AHDS will:

1. Maintain at least two copies of each file/other volume, with each copy held at a physically separate (minimum: not the same building) location. Copies will be maintained on hard disk and/or tape and/or optical media. Media will be refreshed according to a refreshment schedule.
2. Create or collect metadata for each file/other volume and for logical sets of files/other volumes as described by the AHDS Guidelines for Preservation Metadata

To support preservation of the information content of deposited data, the AHDS will:

1. Provide bit preservation for the data, as described above.
2. Create 'preservation copies' of deposited data during ingest, which conform to appropriate standards, best practices, and definitions of significant properties as described in the AHDS Preservation Handbooks
3. Periodically migrate these preservation copies as needed.
4. Make use of both commercial and open source software to migrate data between file formats.

The **ingest** function transforms all elements of the deposited files into a valid preservation format for the specified data type. Files for preservation are copied to a different machine, as the ingest and preservation directory structures are created. The ingest function also includes the creation of descriptive metadata for a variety of purposes and the production of multiple DIPs for usability.

In the UKDA the **archival storage** function receives AIPs and DIPs from the ingest function and adds them to the permanent storage facility, oversees the management of this storage, including media refreshment and monitoring. This function is also responsible for ensuring that AIPs can be retrieved. For any data collection there are always at least three versions residing on each of the different preservation systems: the original SIP, the ingested AIP and multiple DIPs. When a new version of a DIP is created it must be created from the AIP (or possibly the SIP) but not from an earlier recension of a DIP.

### Security, authenticity and integrity

How will you demonstrate the integrity and authenticity of items i.e. that they are what they purport to be and have not been inadvertently changed?

How will items be safeguarded against damage or unauthorised access?

The OAIS standard (CCSDS, 2002) defines fixity as information which can be used to validate the authenticity of information extracted from a digital object. Fixity checks such as checksums, message digests, and digital signatures are used to verify that a digital object has not been changed between two points in time or events. Information created by these fixity checks provides evidence for the integrity and authenticity of the digital objects.

When should fixity information be created and verified?

- At point of creation
- At point of accession
- At point of ingest
- At point of transformation
- As part of normal maintenance routines
- At point of dissemination.



To ensure authenticity of the digital objects and the metadata:

- Will the repository establish protocols and audit trails to show who has accessed each dataset, and who has enhanced or annotated it? (RIN, 2008 p.14)
- Will existing relationships between datasets and explicit links be maintained?

#### Integrity measures

The UKDA takes its role as custodian of data collections seriously. To this end the complete chain of custody of all data collections is documented through metadata. All actions are explicit, complete, correct and current. However, only the 'original' version can be said to be an integral copy of the version deposited with the UKDA. The preservation and dissemination versions are considered to be authentic and there is an audit trail of all alterations in the preservation and dissemination versions which relates back to the original deposited version.

#### Security

The UKDA is committed to taking all necessary precautions to ensure the physical safety and security of all data collections that it preserves. The repository rooms are equipped with multiple key entries and a security-protected swipe-card system linked to an on-site alarm system and to the University Security Office. The repository rooms are located outside of the secure working area of the UKDA. All machine room computer systems are locked by a logon password system to prevent unauthorised access in the case of a security breach of the room.

#### Media refreshment

How and when will storage media be refreshed?

LTO-1 tape media: Mean Time Between Failures = 250,000 hours; Estimated archival life = 15-30 years; 100-250 full file passes. Media are likely to be replaced by new technology before they need to be refreshed. Recommend review of tape media after 5 years, with refreshment if necessary.

The UKDA follows a policy of multiple copy resilience. Five versions of the complete preservation system are held. The UKDA operates a media monitoring procedure as part of its AMASS® preservation system. This allows it to check for potential future problems of wear and tear on media and act before the problems become severe. A log is kept of all refreshment results and all storage media are provided with a date stamp indicating the time they were written and the next renewal date.

#### Versioning

Can new versions be submitted? If so, how will these be handled?

Policy considerations for the deposit of multiple versions of a dataset:

- The repository uses explicit version numbers which are reflected in dataset names.
- The repository records version and status e.g. draft, interim, final, internal.
- The repository stores multiple copies of a dataset in different formats.
- The repository keeps the original copies of data and documentation as deposited.
- The repository stores supplemental digital objects with the data file/s.
- The repository records relationships between items, such as 'supercedes' or 'is superceded by'.

#### Version Control

- Changes to deposited items are not permitted
- Errata and corrigenda lists may be included with the original record if required.
- If necessary, an updated version may be deposited.

In the case of new versions being deposited:

- The earlier version may be withdrawn from public view.
- The item's persistent URL will always link to the latest version.
- There will be links between earlier and later versions, with the most recent version clearly identified.
- Items are allocated a checksum to facilitate the detection of alterations.

Ensuring that any alterations to the preserved version of any part of a Data Collection is accurately documented is integral to the authenticity of any Data Collection. The UKDA distinguishes between two forms of alteration post ingest.

- new version (Definition: when there is a change to the preserved metadata);
- new edition (Definition: when there is change to data or documentation).

When there is a new version of a data collection, the relevant descriptive and structural metadata will need revising and the old file will be retained.

When there is a new edition of a data collection, all descriptive and structural metadata will need to be recreated, the old file will be retained and also, the old AIP and DIPs will be retained within the preservation system and identified as not for issue.

### Withdrawal of collections

Can deposited items be withdrawn?

What are the circumstances in which this is permitted?

What is the procedure for withdrawal?

Will metadata or references to the item persist?

### Withdrawal policy

- No withdrawal policy defined.
- Items may not normally be removed from the repository.
- Items may be removed at the request of the author/copyright holder (but this is strongly discouraged.)

Acceptable reasons for withdrawal include:

- Journal publishers' rules
- Proven copyright violation or plagiarism
- Legal requirements and proven violations
- National Security
- Falsified research
- Other - specify

### Withdrawn items

- No deletion method for withdrawn items defined.
- Withdrawn items are deleted entirely from the database.
- Withdrawn items are not deleted per se, but are removed from public view.

### Withdrawn items' identifiers/URLs

- Are retained: Indefinitely; Transiently: Not at all.
- URLs will continue to point to 'tombstone' citations, to avoid broken links and to retain item histories together with:
  - a) a link to a replacement version, where available
  - b) a note explaining the reasons for withdrawal
- The metadata of withdrawn items will not be searchable.

#### Withdrawal policy

- Will items be removed from the repository?
- Will items be removed at the request of the depositor?

#### Under what conditions will the repository choose to remove items?

- copyright violation
- legal requirements and proven violations
- national security
- falsified research
- confidentiality concerns etc.

#### If items are withdrawn, indicate the terms of the withdrawn items, for example:

- Withdrawn items are deleted entirely from the database.
- Withdrawn items are not deleted, but are removed from public view.
- Identifiers/URLs for withdrawn items are retained and for how long?
- URLs will continue to point to 'tombstone' citations, to avoid broken links and to retain item histories together with a link to replacement versions, where available, and with reasons for withdrawal.

#### Metadata for withdrawn items:

- Should a dataset be removed by either the repository or the depositor, the repository reserves the right to retain its metadata record in the repository as trace of the dataset.
- The metadata of withdrawn items will / will not be searchable.

The UKDA operates a multifaceted policy towards Data Collection withdrawal. Like the TNA, the UKDA distinguishes between soft deletion whereby certain references to the withdrawn content are deleted, but not the content itself, and hard deletion whereby the content and all references to it are deleted. The UKDA chooses soft deletion as the default method of withdrawal as it is too expensive to remove data collections, and their physical removal would present unacceptable risks to other parts of the collection.

### **7. Sustainability, closure and succession plans**

What will happen if the repository is closed or funding is reduced?

Will services be cut or have ongoing costs been planned for?

Are plans in place to transfer repository content if necessary?

#### In the event of the repository being closed down:

- will the data be transferred to another appropriate archive?
- will items be returned to their originators?
- are there other institutional or domain-specific repositories that can offer succession arrangements?

No repository closure policy is defined. OR

#### In the event of the repository being closed down:

- the database will be transferred to another appropriate archive
- items will be returned to their originators.

The UKDA is committed to supporting continued funding for all of the operations relating to preservation management. The Archive has established a rolling planning scheme for lifetimes of computer equipment and storage media to facilitate forward planning for the necessary upgrades. The Archive makes every effort to remain up-to-date with any relevant technological advances to ensure continued access to its collections.