

Collections Preservation Policy

Document type: Policy

Author: Hamish James

Draft / Version: Version 1.0

Date of edition completion: 04/08/2004

Contents

1. Summary
2. Audience
3. What Does the AHDS Preserve?
4. OAIS Conformance
5. Types of Preservation Undertaken by the AHDS
6. Levels of Preservation Service Offered by the AHDS
7. Appendix 1: AHDS Definition of Good Practice
8. Appendix 2: Media Refreshment Schedule

Change history

1.0 First version, 4th August 2004

Summary

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 AHDS implements its approach to digital preservation in a manner that conforms to the requirements of the OAIS reference model.

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. Recognising the potential of alternative approaches to digital preservation, such as emulation and migration-on-demand, the AHDS also preserves the original bit stream of deposited digital resources. At present, however, the AHDS does not preserve resources, such as file format documentation and rendering software, that may be needed to make direct or emulated use of the deposited bit stream in the long-term.

The exact level of preservation service offered by the AHDS will depend on the technical quality and completeness of the deposited data. Only data that conforms to appropriate standards and best practices, as assessed by the AHDS, will be accepted for indefinite preservation.

Audience

- AHDS staff responsible for collections development and collections management
- Individuals and groups depositing digital resources with the AHDS

What Does the AHDS Preserve?

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 will preserve resources where they have been determined to be of long-term value to our user communities, in accordance with the AHDS

Collections Development Policy. The AHDS aims to preserve the significant properties, such as formatting, 'look and feel', functionality, and information content, of deposited digital resources. The significant properties of a deposited digital resource will be defined by the AHDS in consultation with the depositor of the resource. **Preservation of all of the significant properties of complex digital resources, notably websites, may not be possible given the current state of digital preservation knowledge and tools.**

The AHDS does not preserve:

- Software
- Hardware
- Original deposited media
- Hardcopy material (paper, analogue recordings, physical objects), but the AHDS may create a digital surrogate of such material which will be preserved

OAIS Conformance

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).

Types of Preservation Undertaken by the AHDS

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

Backup and media refreshment (periodic transfer of data from an existing tape or disk to a new tape or disk) procedures can achieve the first objective, but preservation of the bit stream alone will not ensure that the information content represented by those bits can still be rendered (displayed and used) in the future.

In conjunction with backup and media refreshment, the second objective can be achieved by periodically *migrating* the information content held in an obsolescent file or media format to a newer format, so that the information content remains accessible with each new generation of hardware and software.

Objective 3, preserving the experience of using the information content, is the most difficult to achieve. This could be achieved by archiving all the software and hardware needed to access the information content in its original format,

or the original software needed to access the information content could be *emulated* on current hardware and software.

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 the conditions listed below. The AHDS does not guarantee, or actively attempt, to preserve the original experience of using deposited digital resources.

Bit Preservation

The AHDS will maintain all the bits, in their correct sequence, that makes up the deposited data (that is, the files or other volumes such as DV cassettes) subject to the following considerations:

1. The AHDS does not undertake to preserve deposited media, only the bits recorded on that media and excepting those bits that are used solely to manage the internal organisation or retrieval of data from the media.
2. The AHDS does not undertake to gain access to, acquire or maintain hardware or software that may be required to directly render the information content represented by the deposited data.
3. The AHDS does not undertake to gain access to, acquire or maintain hardware or software that may be needed to transfer data from media not listed in the AHDS Data Transfer Options list (preservation will therefore not be possible, given point 1 above).

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 given in Appendix 2 of this policy.
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

Bit preservation alone does not guarantee that the information content in a file/other volume can be rendered.

Preservation of Information Content

In addition to providing bit preservation for all deposited data, the AHDS will endeavour to preserve the significant properties of the information content represented by the deposited data in a renderable state when the following conditions are met:

1. Preservation of the significant properties of the data, as defined by the AHDS in consultation with the depositor, does not require the exact reproduction of the original look and feel of the rendered data, or the exact reproduction of the experience of using the data with the software and hardware it was originally designed for.
2. Software, hardware and expertise needed to preserve data in a renderable state is available to the AHDS at an affordable cost.
3. The data does not take the form of source code or an executable file. The AHDS will not guarantee to preserve the functionality of any executable code.

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.

Levels of Preservation Service Offered by the AHDS

Preferred Level of Preservation Service

The level of preservation service provided by the AHDS will depend on the technical quality and completeness of the data and its supporting material (metadata and documentation). *The goal of the AHDS is to preserve all data and supporting material indefinitely.* The AHDS will encourage and assist depositors to provide data and supporting material in a form that meets the requirements for indefinite preservation. Where possible, the AHDS will work to improve data so that it can be preserved indefinitely.

Minimum Level: Bit Preservation Only

The AHDS will undertake the bit preservation of deposited data, as discussed above, where the following requirements are met:

- The data adheres to a recognised data format (including formats defined by depositors, providing full documentation is supplied to the AHDS)
- The data can be successfully rendered using software designed to display, playback or provide other forms of access to the specific data format used by the data
- The data has not been deposited in a corrupted state
- The data does not contain malicious executable code
- The AHDS's right to preserve the file is established in a valid deposit licence

- The content of the data meets the requirements of the AHDS Collections Development Policy

The AHDS aims to provide bit preservation services such that a copy of any deposited bit stream will be available as long as the AHDS is funded, and for at least one further year, assuming only that physical media are held in a secure location, in the event of the AHDS not being funded.

Ten Year Level: Bit and Information Content Preservation

Functional preservation, as described above, will be supported for a period of *at least* ten years, where data meets the requirements for the minimum level service and these additional requirements:

- Data is deposited in an 'acceptable' or 'preferred' format included on the AHDS Deposit Formats List

Indefinite Level: Bit and Functional Preservation

The AHDS will aim to maintain indefinitely the bits and content of deposited data where the data meets the requirements for the minimum level, and the data has been constructed using appropriate good practice and standards, as assessed by the AHDS. Good practice includes the provision of adequate documentation and the use of open or de-facto standards that are widely supported in software. The AHDS defines good practice in a number of documents, detailed in Appendix 2. Specific requirements of this policy are:

- Data is deposited using a 'preferred' method from the AHDS Data Transfer Options List
- Data is deposited in an appropriate format given its original purpose and content type (e.g. databases are not provided as HTML)
- Sufficient documentation accompanies the data to assure that the content of the data can be understood.
- The data is not deposited for preservation only

Appendix 1: AHDS Definition of Good Practice

The AHDS definition of good practice is given in a number of documents:

- AHDS Deposit Methods List
- AHDS Guides to Documentation
- AHDS Guides to Good Practice series
- AHDS Information Papers

Appendix 2: Media Refreshment Schedule

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.