Gloucestershire Archives @ The Hub

Digital Preservation Policy

1. Policy statement and purpose

This policy explains why and how Gloucestershire Archives preserves records in its care that are in digital format. Without active curation these records will not survive. Digital records cannot just be put into a secure strongroom and left safely for an indefinite period. Access to digital records may depend on hardware and software that becomes obsolete within a few years. Also, stored digital records can become corrupted and inaccessible without warning.

Our key objectives are to ensure that preserved digital records can be located, accessed and trusted, now and in the future.

2. Scope

Preserved digital records include both born-digital and digitised records.

3. Terminology

Archives are the record of everyday activities of governments, organisations, businesses and individuals. Archives may take many different forms – handwritten, typed, printed, photographic or electronic – and include audio-visual material such as video and sound recordings. As authentic and reliable records, they are preserved permanently because of their evidential value.

Born-digital records are created digitally for example by a word processor or a digital camera.

Digitised records are digital copies made from an analogue original, for example a photograph of a page of a parish register.

Further terminology is explained at the relevant point in the policy.

4. Background

Gloucestershire Archives gathers archive collections and local and family history resources to ensure they are kept secure and made accessible.

We are an accredited archive service recognised by The National Archives as the place of deposit for public records relating to Gloucestershire and South Gloucestershire.

We are a Gloucestershire County Council service. By agreement, we also provide an archive service for South Gloucestershire Council and are the appointed Record Office for the Diocese of Gloucester.

We are a lead partner in the Gloucestershire Heritage Hub, a network of local people and organisations created with a common interest in our historic county's documented heritage.

5. Digital Preservation Principles

This digital preservation policy is informed by, and where relevant, conforms to several international and other standards and particularly makes use of the OAIS reference model and the BagIt specification (see Appendix 1).

Gloucestershire Archives

- adopts a sustainable, modular approach to digital preservation. This approach is flexible in that it allows for the substitution of individual components employed in an overall solution. It also maintains a strict distinction between the storing of preserved digital records and their processing. For example, replacement processing software can then be installed without any change to the way preserved digital records are stored.

- avoids any reliance upon specific single hardware or software products or suppliers

- prefers standards based, open source and cross platform (that is not hardware or software specific) solutions to proprietary or patent encumbered solutions

- entrusts the storage of the digital records' "bits" to an appropriate trusted data storage provider

- prefers format migration as opposed to "emulation", (emulation involves using current hardware/software to mimic possibly obsolete hardware/systems).

- ensures authenticity of digital records by implementing transparent and fully documented preservation processes, and by capturing and providing the metadata required to describe the content, context and provenance of the record.

- ensures integrity of digital records (making sure they are complete and protected against unauthorised or accidental alteration) by keeping the original record (bitstream preservation), fixity monitoring, and capturing an audit trail of preservation actions undertaken.

6. Stages of digital preservation

Digital preservation at Gloucestershire Archives comprises six stages (see Appendix 2 for further details):

6.1   Planning and review
Continuous planning and review of requirements covering

- legal requirements
- scope or the categories of records to be preserved
- capacity planning, especially in respect of storage management
- "technology watch" (that is, monitoring relevant hardware/software in order to forecast end of life)
- "standards watch" (that is, identifying changes to standards and practices)

6.2   Selection and appraisal
The process of selecting digital records for preservation is carried out in accordance with our collecting and appraisal policies. For digital records we aim to be proactive and collect at point of creation (see also Gloucestershire County Council's Digital Continuity Policy). Please see section 10 below for links to these policies.

6.3   Integrity and authenticity
Gloucestershire Archives has adopted the OAIS reference model and the BagIt specification (see Appendix 1) and in accordance with these uses its in-house developed tool, SCAT (Scat is Curation and Trust) to process digital records as Archival Information Packages (AIPs). The AIPs are transferred to trusted storage. These processes mean we can demonstrate integrity, authenticity and our resultant digitally preserved records can be

trusted.

6.4   Trusted Storage

The day-to-day storage of the Archival Information Packages (AIPs) is currently provided by a County Council ICT supplier.

"Preservation redundancy" is a key strategy to support the long term preservation of digital objects. That is at least two independent (organisationally and technologically discrete) storage solutions are operated concurrently. We will work with GCC colleagues to implement this strategy.

Fixity values for each AIP are calculated on creation and held in a non-proprietary database stored by Gloucestershire Archives ('ops-db'). The fixity of each AIP is confirmed on receipt by the storage provider and is thereafter regularly monitored. Fixity is again confirmed on download of AIP from the storage provider.

We use the [Planning Tool for Trusted Electronic Repositories](#) (PLATTER) to specify the requirements of our trusted digital store. This tool "provides a basis for a digital repository to plan the development of its goals, objectives and performance targets over the course of its lifetime in a manner which will contribute to the repository establishing trusted status amongst its stakeholders" (Digital Preservation Europe).

6.5   Access

- Access to digital records is provided onsite as required.

- Gloucestershire County Council's and South Gloucestershire Council's intellectual property rights are asserted.

- The Freedom of Information Act (2000) and current Data Protection legislation apply to digital records.

- Access to digital records may be restricted by the terms of the deposit or donation.

6.6   Maintenance and migration

At Gloucestershire Archives we are able to identify "at risk" records where future access may be no longer possible. In the unlikely event that records are identified as "at risk" the AIP will be repackaged to include records in a more sustainable format. This maintenance activity will be appropriately recorded.

7. Emergency recovery plan

Our emergency recovery plan envisages two loss scenarios. The first is an event affecting trusted storage. Redundant cloud storage mitigates the risk of data loss.

The second is an event affecting anything other than the data centre, for example the destruction of our Heritage Hub site. In these circumstances digital preservation/curation activities will be re-established at a recovery site. This plan is assisted by other elements of this policy, in particular the digital preservation principles set out in section five.

8. Risk management

Exposed risk areas include:

- the narrow skills base in respect of digital records preservation/curation which is focussed on a small number of individuals. This risk is reduced by good documentation
- unique copies of digital records awaiting processing. This risk is reduced by minimising processing delays and also by the submitter (donor or deposit or) retaining a copy of the record until processing is confirmed
- unique hardware/software instances (e.g. a zip drive) that support particular physical format processing. This risk is reduced as we will not accept material we cannot process and if necessary will direct the owner to another appropriate repository or museum

9. Roles and Responsibilities

Gloucestershire Archives' Head of Service and its Collections Development Manager are responsible for strategic development.

Gloucestershire Archives' Collections Team supported by the Collections Development Manager is responsible for implementing the policy and undertaking associated tasks.

10. References

This policy should be read alongside our related policies, and in particular our collecting policy which sets out the statutory framework of our service and summarizes our existing collections, and our appraisal policy. All our policies can be found at www.gloucestershire.gov.uk/archives/policies

In addition, Gloucestershire County Council's digital continuity policy (which relates to the Council's electronic records) can be found at: www.gloucestershire.gov.uk/council-and-

democracy/strategies-plans-and-policies/information-management-and-security-policies/information-and-data-management-policies/

Our advice for owners and depositors of digital records can be downloaded at: www.gloucestershire.gov.uk/archives/preserving-collections/adding-to-our-collections/

Details of our digital curation activities can be found at https://www.gloucestershire.gov.uk/archives/preserving-collections/digital-preservation/

The National Archives' digital strategy is available here: Our Digital Century - Archives sector (nationalarchives.gov.uk)

11. Review and Revision

The policy has been developed and reviewed with reference to current national and international research and best practice in digital preservation. This is a field that continues to develop rapidly and so the policy is continuously monitored. It will be formally reviewed every three years.

Please write to archives@gloucestershire.gov.uk if you wish to give feedback on this policy.

Document control

Author:	Claire Collins, Digital Archivist/Collections Development Manager
Owner:	Heather Forbes, Head of Archives Service
Approval Body	Gloucestershire Archives Management Team (GAMT); Gloucestershire County Council's Director of Policy, Performance and Governance; South Gloucestershire Archives Liaison Group
Date Approved	August 2023
Document Number:	v3.1

Version history

Version	Version date	Summary of Changes
1.0	July 2008	New policy approved by Libraries Senior Management Team
2.0	March 2013	Updated to include progress in digital curation capabilities of the service, link to digital continuity policy, guidance to donors/depositors etc • Include explicit reference to relevant standards • Adopt OASIS terminology • Add 'standards watch' • Add selection and appraisal reference to SIP • Update references to GAip • Remove references to 'preservation format migration' • Update ingest reference to metadata • Clarify ingest reference to fixity • Clarify position in regard to metadata harvesting • Include position in respect to PLATTER • Clarify position in regard to fixity management • Clarify position in regard to format migration • Include position in respect to PRONOM and maintenance • Clarify some details in respect of risk management
2.1	Sept 2014	Brief review and minor re-formatting
2.2	Jan 2018	Revision following the creation of the Heritage Hub partnership. Standards updated
3.0	Oct 2020	Reformatted and refreshed terminology to clarify for a lay audience; principles of authenticity and integrity made more explicit; standards reviewed, job roles and fixity arrangements updated; South Gloucestershire Council's intellectual property rights added.
3.1	August 2023	Brief review and minor revision including updating links. • Explicit reference to BagIt specification • Explicit reference to redundancy • Catalogues updated • Standards updated

Date of next revision: 2026

Appendix 1

Gloucestershire Archives' digital preservation policy is informed by and where relevant conforms to several international and de-facto standards. In particular:

- ISO 14721:2012 Space data and information transfer systems - Open archival information system (OAIS) – reference model,
- ISO 16363: 2012 Space data and information transfer systems – Audit and certification of trustworthy digital repositories,
- ISO 17068:2017 Information and documentation – Trusted third party repository for digital records,
- ISO/TR 18492: 2005 Long term-preservation of electronic document-based information [technical report],
- Extensible markup language (XML) 1.0 (Fifth edition),
- XML schema definition language (XSD) 1.1,
- ISO/IEC 19757-2:2008 Information technology – document schema definition language (DSDL) – Part 2: regular-grammar-based-validation – RELAX NG: amendment 1: compact syntax,
- Metadata encoding and transmission standard (METS), 1.11,
- Preservation metadata: implementation strategies (PREMIS), 3.0,
- Extensible metadata platform (XMP),
- Secure hash algorithm (SHA),
- PRONOM file format identification,
- Internet Engineering Task Force "The BagIt File Packaging Format", 0.97, and
- Copyright, Patents and Design Act (1988).
- Universally Unique Identification (UUID) - ISO/IEC 11578:1996 Information technology – Open Systems Interconnection - Remote Procedure Call
- Javascript object notation (JSON) - The JSON data interchange syntax, Ecma International technical committee 39

Appendix 2

Six Stages of Digital Preservation with technical details

6.1 Planning and review

Continuous planning and review of requirements covering,

- legal requirements for example, how the evidential characteristic of a digital record is maintained
- scope or the categories of records to be preserved for example, to include websites, emails and databases
- capacity planning, especially in respect of storage management,
- "technology watch" that is monitoring relevant hardware/software in order to forecast end of life
- "standards watch" that is identifying changes to standards and practices

6.2 Selection and appraisal

The process of selecting digital records for preservation is carried out in accordance with our collecting, appraisal and digital continuity policies. For digital records we aim to be proactive and collect at point of creation.

We can preserve the following categories of records:

- image files
- word processing files
- simple spreadsheet files
- sound files
- audio-visual files
- website extracts

We are not yet in a position to preserve email in-boxes, podcasts/live streamed content, compound documents and complex spreadsheets and databases.

As outlined in the OAIS reference model wherever possible we accept digital records as a Submission Information Package (SIP) that includes the submitted files in the formats in which the records were created, even if that format is now obsolete or proprietary. Where this is the case we require the depositor to deposit a copy of the original submitted files in an open format. The SIP also includes message digests[1] that validates the fixity[2] of the submitted files.

---

[1] Message digest - that is, a form of digital fingerprint

6.3 Integrity and authenticity

Integrity is demonstrated through a process known as 'Ingest'. Ingest refers to the creation of an Archival Information Package (AIP) and its transfer to trusted storage. The SCAT (Scat is Curation and Trust) tool creates AIPs together with their associated fixity records.

Standard ingest workflow procedures are semi-automated and scalable thus meeting the future need for bulk ingest of large numbers of born-digital records.

Tools and procedures continue to be developed in order to take advantage of improvements in the field of digital preservation.

Intellectual property

We ask depositors/donors to grant permission to copy digital records in order to create Archival Information Packages and to manage their storage. The resultant AIP is the intellectual property of Gloucestershire County Council. Records relating to South Gloucestershire will be deemed to be the intellectual property of South Gloucestershire Council.

Metadata

We embed metadata in the AIP using recognised standards and platforms such as METS, PREMIS, XMP and JSON.

Fixity

The long-term stability of the AIP is monitored by its message digest. We use specific cryptographic hashes (SHA1, SHA256, SHA512 and MD5).

Catalogues

Sufficient metadata to identify the AIP is captured and stored within the AIP. Full catalogue entries are created and stored within the archive management system (catalogue).

When surrogate copy digital records are ingested then the existing catalogue entry for the original record is annotated.

---

[2] Fixity, that is, the continued stability of the pattern of "bits" of submitted files.

6.4 Trusted Storage

The day-to-day storage of the Archival Information Packages is provided by a County Council ICT supplier. We provide routine fixity management of AIPS in order to generate evidence that files remain unaltered.

We will use the Planning Tool for Trusted Electronic Repositories (PLATTER) to specify the requirements of a trusted digital store. This tool "provides a basis for a digital repository to plan the development of its goals, objectives and performance targets over the course of its lifetime in a manner which will contribute to the repository establishing trusted status amongst its stakeholders" (Digital Preservation Europe).

6.5 Access

Dissemination Information Packages (DIPs) are created onsite as necessary when access to records in a particular AIP is required. The original digital formats of records in the AIP are converted to the format (including proprietary formats) best suited to the particular information access requirement.

6.6 Maintenance and migration

Gloucestershire Archives does not routinely migrate the original digital formats of records in AIPs to non-obsolete file formats.

However we are able to identify "at risk" records where the future creation of DIPs may become no longer possible, by using the PRONOM file format identification scheme in combination with technology watch. In the unlikely event that records are identified as "at risk" the AIP will be repackaged to include contingent copy records in a more sustainable format. This maintenance activity will be appropriately recorded in PREMIS.