

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.

2. Scope

Preserved digital records include both born-digital and digitised records, for example surrogate copies.

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. 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 available for research.

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 in 2017 with a common interest in our historic county's documented heritage. The network benefits from Hub facilities at Alvin Street, Gloucester and a Hub website. Its members support each other to gather, keep and share their personal and community archives.

5. Associated technological policy

This digital preservation policy is informed by and where relevant conforms to several international and other standards (see Appendix).

In order to reduce technological dependencies and to manage the risks of hardware/software obsolescence or storage failure, Gloucestershire Archives' digital preservation technology policy is to:

- avoid any reliance upon particular single hardware or software products or suppliers
- prefer standards based, open source and cross platform (that is not hardware/software specific) solutions to proprietary or patent encumbered solutions
- entrust the storage of the digital records' "bits" to an appropriate trusted data storage provider
- prefer format migration, that is, regeneration of the digital record in a format that can be rendered using current hardware/software to provide user access to information, as opposed to "emulation", (emulation involves using current hardware/software to mimic possibly obsolete hardware/software).

- adopt a modular rather than a monolithic approach to digital preservation. The modular approach is more sustainable since it allows for the flexible substitution of individual components employed in an overall solution. The modular approach also maintains a strict distinction between the storing of preserved digital records and their processing. Hence, for example, replacement processing software can be installed without any change to the way preserved digital records are stored.

6. Stages of digital preservation/curation

Digital preservation/curation at Gloucestershire Archives comprises six stages:

- planning and requirements specification
- selection and appraisal
- ingest
- trusted storage
- dissemination, including publication
- maintenance and migration

6.1 Planning and requirements specification

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” (for example, monitoring relevant hardware/software in order to forecast end of life)
- “standards watch” (for example, identifying changes to standards and practices)

6.2 Selection and appraisal

The process of selecting digital records for preservation is the same as that for paper records, that is, it is carried out in accordance with our collecting and appraisal policies.

We can preserve the following categories of records:

- image files
- word processing files
- simple spreadsheet files
- sound files

- audio-visual files

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

Wherever possible we accept digital records for ingest 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 that validate the fixity, that is, the continued stability of the pattern of “bits” of submitted files.

6.3 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 GAip files (Gloucestershire Archives information packages) 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/curation.

Intellectual property

We ask depositors/donors to grant permission to copy digital records in order to create AIPs and to manage their storage. The AIP is the intellectual property of Gloucestershire County Council.

Metadata

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

Fixity

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

Catalogues

The GAip archival information package anticipates the requirement to create a searchable database of preserved digital records by harvesting metadata from the AIP. Until this can be implemented metadata is copied manually in order to enter data in the catalogues (currently CALM).

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

6.4 Trusted storage

The day-to-day storage of the Archival Information Packages, that is the GAip files, is provided by the County Council's ICT supplier.

We are working with them to introduce routine fixity management of AIPs in order to provide an evidence based relationship.

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” (DigitalPreservationEurope).

6.5 Dissemination (access) including publication

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.

Gloucestershire County Council's intellectual property rights will be asserted.

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

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

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.

7. Emergency recovery plan

The emergency plan envisages two loss scenarios. The first is an event affecting the trusted store (currently the data centre operated by the County Council's ICT supplier). The data is protected by the general procedures and recovery processes which safeguard the County Council's data.

Gloucestershire Archives' digital preservation/curation processes will be suspended until after a full restoration of data centre services.

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 associated technological policy 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 but still remains a significant exposure
- unique copies of digital records awaiting ingest. This risk is reduced by minimising ingest delays and also by the submitter (donor or depositor) retaining a copy of the SIP until ingest is confirmed
- unique hardware/software instances (eg a zip drive) that support particular physical format ingest. This risk is reduced as we will not accept such material and if necessary will direct the owner to another appropriate repository or museum

9. Roles and Responsibilities

Gloucestershire Archives' Head of Service, its Collections Leader and its Digital Preservation and Access Officer are responsible for strategic development.

Gloucestershire Archives' Collections Team supported by the Digital Preservation and Access Officer 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. 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-policies/information-management-and-security-policies/

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

Details of our digital curation activities can be found at www.gloucestershire.gov.uk/archives/digital-curation

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/curation. This is a field that continues to develop rapidly and so the policy is continuously monitored. It will be formally reviewed every three years.

Document control

Author:	Claire Collins, Digital Preservation and Access Officer
Owner:	Heather Forbes, Head of Archives Service
Approval Body	Gloucestershire Archives Management Team (GAMT); Gloucestershire County Council's Director of Strategy & Challenge; South Gloucestershire Archives Liaison Group
Date Approved	March 2018
Document Number:	v2.2

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 <ul style="list-style-type: none"> • 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 as necessary

Date of next revision: 2021

Appendix

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).