Digital Preservation Policy

Cambridge University Libraries

V2.0

UNIVERSITY OF CAMBRIDGE

Cambridge University Libraries

Table of Contents

Questions about this policy?
Contact digitalpreservation@lib.cam.ac.uk

1. Policy Overview

Cambridge University Libraries (CUL, the Libraries) include the main University Library (the Library) as well as its faculty and departmental libraries. This policy outlines how the Libraries care for their digital collections to ensure these materials can be accessed by different categories of users, including staff as well as onsite and online readers, now and into the future.

The Libraries' digital collection materials are described as follows:
- Digitised representations of physical and analogue collection objects that are created in-house or externally.
- Born-digital objects that are created by the Libraries, acquired from members of the University community or external donors, or purchased.
- Metadata that describes individual objects or a group of objects.
- Metadata-only objects.

Ensuring access to digital collection materials is central to this policy. Access not only demonstrates that appropriate steps have been taken to ensure longevity of these materials, it also enables users to discover and feed back any errors or missing content so that library staff can take action if necessary.

The Libraries take a lifecycle approach to caring for digital collections. This means that activities that help ensure current and ongoing access do not just happen once or at one stage. They happen when and where needed, and for as many times necessary, from the time materials are created or acquired to when they are accessed, used, and re-used so long as they are held by the Libraries.

Besides taking a lifecycle approach, there are core principles that guide how the Libraries manage their digital materials to support scholarship and enable world-leading research. These are detailed in Section 4.

1.2 Background

This policy follows Version 1.0[1] of the Libraries' Digital Preservation Policy that was written during a period of significant change in how the Libraries care for their digital collections. Digital preservation within the University continues to mature, and in October 2020, the Cambridge University Libraries were awarded funding from the University to establish a digital preservation service.

---

[1] https://doi.org/10.17863/CAM.32927

The Cambridge University Libraries Digital Preservation Service will be implemented through a five-year programme beginning in Q2 2021. This programme will deliver identified capabilities to support digital collection materials using open-source systems, tools, and standards.

The CUL Digital Preservation Strategy (forthcoming Q3 2021) outlines how this policy will be upheld, both by the Digital Preservation Programme and by other work being carried out by the University Library.

1.3 Audience

Due to the range of activities that help care for digital collection materials, this policy has a broad audience, including library and other Cambridge University staff as well as users from across the University.

It is recommended that new and existing library staff consult this policy whenever it is relevant to their work. This policy should also be consulted as part of setting up and delivering any library and University projects or other initiatives that will result in the creation or acquisition of digital collection materials that the University Library will need to preserve. Examples include the Digital Preservation Programme as well as project applications to external funding bodies.

External parties might also be interested, including:
- Onsite and online readers of digital collection materials.
- Research partners.
- Existing and potential donors of material.
- Existing and potential funding bodies.
- Publishers and other content creators.
- Suppliers of infrastructure and services for digital collection materials.
- The digital preservation community as well as related communities.[2]

2. Why digital preservation matters

Digital preservation is an ever-evolving discipline focused on ensuring access to digital materials for as long as necessary.

Similar to physical collection objects, digital collection objects must also be looked after to ensure their longevity. Without the necessary knowledge, skills, workflows, business

---

[2] Web archiving, time-based media conservation, digital humanities, to name a few.

processes, and infrastructure to manage digital materials, the risk of losing access to them will occur and potentially escalate.

These are just a few reasons why risks could occur for digital materials:
- Hardware, software, and storage obsolescence.
- Human error.
- Storage media failure that could result in bit rot or complete loss of data.
- Institutional change (e.g., change in strategic direction, loss of funding, loss of staff).
- Natural disasters (e.g., flood, fire).

2.1 Lifecycle Approach

Activities that enable current and long-term access to digital collection materials are not carried out once or at one stage. They are performed when and where needed and however many times necessary whilst materials are within the Libraries' custody.

For this reason, the Libraries look across a lifecycle when managing digital collection materials to ensure that appropriate capabilities exist. The Libraries use a lifecycle comprised of the following stages:

Create involves creating digitised representations of physical or analogue objects. This work is undertaken by the University Library's Digital Content Unit, staff working on research projects or exhibitions, and external suppliers. This stage also applies to born-digital materials created by library staff, projects, and University bodies (e.g., administrative bodies of the University that create official records of University business).

Acquire is the act of taking in digital material from an external source as well as appraisal and accessioning. For the University Library, this could include digital materials acquired by purchase, donation, transfer, or deposit.

Describe involves describing digital objects using different types of metadata—descriptive, provenance, rights, structural, preservation, and technical—that have either been created by a person or system at the Library or deposited with an object.

Ingest packages and transfers digital objects and metadata for submission to systems that process and manage data.

Store involves moving digital objects to their appropriate storage environment. Backed-up, resilient storage is used for day-to-day working, storing objects prior to ingest, providing access, as well as storing preservation copies of files.

Search and Discovery involves readers using metadata to locate or browse digital collection materials.

Access, Use, and Re-use make digital collection materials available to users, both onsite and online. Re-use means our readers can carry out other activities with digital collection materials beyond simply reading, viewing, listening to, or watching them.

Maintain helps ensure current and future access to digital collection materials.[3]

Although not explicitly mentioned as stages, administrative and data management activities will take place throughout.

3. Scope

3.1 In scope

3.1.1 Collection materials

Digitised versions of physical and analogue works as well as born-digital materials (e.g., research outputs, archival materials) that must be preserved can be found across the Libraries. All of the Libraries' digital collection objects and associated metadata are in-scope for long-term preservation.

The Digital Preservation Programme will develop capabilities[4] for digital collection materials held by the Libraries and prioritise those held by the University Library. The Libraries' collections will continue to grow as new collections are established and materials and formats are added to existing collections. Capabilities will be designed for known use cases, with the understanding that new use cases will be identified and existing use cases will evolve over time.

Prior to ingest into systems, digital collection materials will be kept on backed-up, resilient storage with checksums to ensure the files remain unchanged, and integrity can be later verified. Some digital collection materials will be acquired on handheld media (e.g., hard drives, CDs) that will need to be transferred to this other storage; there are also analogue materials that will need to be digitised. The care of the media and analogue carriers are the responsibility of the collecting or creating area to which the materials belong.

---

[3] One example could be migrating digital objects from one format to another when the current format is identified as having some sort of risk. Another activity could occur when physical objects are re-digitised because digitisation technology has improved, and these resulting digitised objects are preserved alongside their older preservation copies.

[4] The capabilities being developed address the stages within the lifecycle described in Section 2.

3.1.2 Associated materials

Digital objects that are themselves not collection objects but provide additional context about a collection object or group of objects and/or aid in the provision of access, amongst other purposes, are in scope for preservation.[5]

3.1.3 Hosted digital materials

The Library hosts and provides access to digital materials owned by third parties. In some cases, these materials are also in scope for the Library to preserve. This type of arrangement is made on a case-by-case basis and the terms and conditions will be outlined in official documentation (e.g., contract, memorandum of understanding).

3.2 Out of scope

3.2.1 Physical and Analogue collection materials

The collection management of physical and analogue collection materials is out of scope for this policy; however, good practice that enables current and ongoing access of digitised copies must be considered when digitisation is undertaken and throughout lifecycle stages.

3.2.2 UK Non-Print Legal Deposit collection

As one the six UK Legal Deposit Libraries,[6] the University Library shares ownership of and responsibility for a collection of digitally published works that are made publicly available in the UK. University Library staff contribute to decisions about this collection through various strategic and subject-specialist operational groups. The British Library ingests the digital files and maintains the systems, storage infrastructure, workflows, and processes for these materials.[7]

---

[5] One example is software that is required to provide access to a type of file format within the Libraries' collections. Another example is email correspondence between library staff (e.g., an archivist) and a donor that provides information about a donation.

[6] Along with the British Library; the National Library of Scotland; the National Library of Wales; the Bodleian Libraries, University of Oxford; and the Library of Trinity College, Dublin.

[7] At present, the following formats are collected: eBooks, eJournals, websites, maps and geospatial data, and sheet music. The LDLs also carry out ongoing research into digitally published works that fit into one of the above categories but are created in formats that are not currently collected (e.g., an eBook that is created as a mobile app and has both software and hardware dependencies).

3.2.3 Websites

The University Library does not have its own web archive but instead contributes to the UK Web Archive (UKWA), which is a collaborative initiative involving all six UK legal deposit libraries.

Websites with a UK top-level domain (e.g., .uk, .scot, and .wales) are in scope to be crawled by the UKWA along with other websites that are in scope[8] under the UK's Non-Print Legal Deposit Regulations. Library staff contribute websites created by the Libraries, the University, and city of Cambridge, as well as websites relating to relevant teaching and research, to the UK Web Archive. The Library also supports this archive through involvement in the Legal Deposit Libraries Web Archiving Subgroup.

3.2.4 Active records and data

The Library is responsible for the long-term preservation of the University Archives that include official records created by bodies within the University in print and digital formats. Digital records are transferred to the Library only when they are finalised and signed off for transfer; any files that are still in use will remain with their respective body until ready. This approach also applies to other digital materials that are 'active', including research data, which falls under the remit of the University's institutional repository.

3.2.5 Subscriptions to digital materials

Digital materials that are subscribed to and are hosted by external suppliers (e.g., Portico) are not in scope. The Libraries are not responsible for the preservation of these materials since they are not owned by the Libraries nor are part of collections.

3.2.6 Access copies and other derivatives

An access copy is a derivative of a preservation copy that is intended to be used by readers, both on- and offsite, where permissions allow. As technology progresses, and software and hardware dependencies change, it is possible that a new access copy might need to be created. In some cases, the preservation copy and the access copy will be the same file, in which case the object is in scope for current and long-term preservation.

Library staff create derivatives of digital collection materials for use online and onsite (e.g., use on social media or exhibitions). These objects will need to be managed over the course

---

[8] These latter websites use a different domain (.com, for example) and are in scope if they are identified as being hosted in the UK.

of their life until they are no longer needed. Although these objects are not intended to be preserved over the long term, it is likely that they will share systems, workflows, and processes with preserved objects. Digital objects will be deleted according to agreed criteria and an audit trail will be maintained of such activity.

4. Principles

Along with the overarching principle of taking a lifecycle approach to preservation, this policy is guided by the following principles.

4.1 Understanding collections

Understanding the digital materials that exist within collections is essential for setting up and maintaining the capabilities needed to steward them throughout stages within a collection management lifecycle. Decisions around the care of digital collections need to be guided by information about these materials.

This principle will be followed by:
- Documenting existing and new workflows, including:
  - Flow of materials through systems and processes.
  - File types, formats, and standards used.
  - Conditions under which materials are created or acquired.
  - Conditions under which materials can be accessed, used, and re-used.

4.2 Integrity, Authenticity, Security

It is crucial that digital materials remain unchanged or that they change only in a managed and documented way, as well as remain secure, whilst in the Libraries' custody. This principle ensures that digital materials are trustworthy and representative of when they were created or acquired.

This principle will be followed by:
- Ingesting digital collection materials into system(s) that undertake relevant activities that aid in their preservation, access, storage, and ongoing management.
- Ensuring current and ongoing bit-level and content-level integrity.
- Capturing and creating relevant metadata.
- Securely storing digital collection materials in geographically separate areas.
- Identifying and addressing risks and issues to monitor and report on the health of preserved materials.

- Periodic auditing using an appropriate audit framework to first create a baseline and to later measure improvement.

4.3 Work in the open

The Libraries are committed to using open-source tools, systems, and standards to ensure the longevity of digital collection materials. The Libraries also recognise that it is important to contribute to the open-source communities whose outputs they use, as well as to share openly about how they are managing their collections.

This principle will be followed by:
- Using open-source systems, tools, and standards across areas of collection management.
- Supporting open-source communities through publishing code to its own open repository and/or contributing code to other repositories; creating or contributing to documentation; and contributing knowledge more generally speaking (e.g., participating in community events).
- Collaborating with external organisations on open initiatives.
- Communicating openly internally and externally about challenges and successes.

4.4 Library-wide collaboration

Digital Preservation sits within the Library's Digital Initiatives Directorate, but activities that ensure current and ongoing access to digital materials are a library-wide undertaking. Activities do not take place by one individual or team; instead, they are carried out directly by, or in consultation with, library staff that are responsible for the ongoing care of digital collection materials.

This principle will be followed by:
- Identifying activities that need to take place and areas of business change.
- Carrying out in-house and external training when and where needed.
- Documenting and explaining complex technical matters in clear and accessible language.
- Continuous improvement to work in the most efficient ways possible.
- Identifying reader needs to enable staff to better support teaching, learning, and research.

4.5 Access is fundamental to preservation

Access is fundamental to preservation. Not only does access enable readers to carry out their research, it enables them to alert library staff to any potential risks to collection materials (e.g., do files open and render as expected?) and helps them identify whether further action is necessary.

This principle will be followed by:
- Providing as wide access as possible to staff and readers in ways that faithfully represents digital objects when they were created or acquired.
- Periodic manual sampling to ensure that files can open and to visually check for any potential issues.
- Communicating to staff and readers the conditions under which digital collection materials can be accessed, used, and re-used.
- Providing a mechanism by which staff and readers can alert staff to access-related issues so that action can be taken.

5. Governance, Roles, and Responsibilities

5.1 Strategic oversight of policy

Role	Responsibility
Library Syndicate	Endorses the policy and supports its implementation.
Senior Leadership Team	Oversees the policy's implementation.
Deputy Director of Digital Initiatives	Ensures that appropriate infrastructure is provided to support digital collection activities.
Head of Digital Preservation	Responsible for this policy on a day-to-day basis and for leading a consultation with stakeholders responsible for digital collection materials to ensure it reflects the needs of staff from across the Library, the faculty and departmental libraries, and the University. Responsible for communicating progress, as well as risks and issues, to the Deputy Director of Digital Initiatives and the Senior Leadership Team.

5.2 Operational roles and responsibilities

The Library's Digital Initiatives Directorate is responsible for the infrastructure that enables necessary capabilities to support associated activities to be carried out.

The following teams sit within this area:

Team	Role (in the context of this policy)
Digital Preservation	Business owner of digital preservation for the Libraries. Works with colleagues to ensure that activities that support the longevity of collections are embedded and staff are supported. Oversees the ingest of digital files and metadata into systems and care of materials in these systems post-ingest.
Digital Development	Technical owner of the systems. Provides technical resource and expertise for implementing and configuring as well as maintaining tools and systems to carry out activities on digital objects and metadata to ensure their longevity.
Digital Services	Technical supplier of Library infrastructure and storage needed for pre-ingest.
Open Research Systems	Business owner of research publications and data, as well as technical owner of the University's institutional repository.
Digital Content Unit	Oversees workflows for digitising print and physical collection materials, creating metadata about digitised copies, as well as storing files and metadata on secure storage. This area is also responsible for rights management and licensing of digitised materials for access and use.
Digital Library	Business owner of the Cambridge University Digital Library that provides access to digitised collection objects and associated materials.

With the exception of research outputs (publications and data), the Research Collections Directorate is responsible for the collection management of digital materials within the Library's Special Collections.

6. Review

This Policy will be reviewed and updated (where applicable) on a yearly basis. The next review will take place in spring 2022.

Glossary

Access: "Access is assumed to mean continued, ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes the digital material was created and/or acquired for"[9] (Source: Digital Preservation Handbook Glossary).

This policy also recognises that providing access to digital collection materials depends on having the software and/or hardware needed to render these files as well as staff knowledge and skills needed to identify and address access needs. Different types of digital materials will have their own requirements for access, some of which are more complex than others.

Analogue:

- "An adjective describing any signal that varies continuously as opposed to a digital signal, which contains discrete levels.
- A system or device which operates primarily on analog signals"[10] (Source: Columbia University Computer Science)

Bit-level integrity: Ensuring that the bits (the 1's and 0's) that make up a digital file remain unchanged from the point of creation and over time (Source: author).

Born-Digital: "Digital materials which are not intended to have an analogue equivalent, either as the originating source or as a result of conversion to analogue form" (Source: Digital Preservation Handbook Glossary).

Cambridge University Libraries: The University Library as well as thirty-three faculty and departmental libraries within the University affiliated with the UL (Source: author).

Carrier: "The physical package (i.e., disc, film, tape, etc.), in or on which the audiovisual data or signal is fixed or recorded"[11] (Source: IASA). Other types of data, not just audiovisual data, can also be recorded on carriers (Source: author).

---

[9] https://www.dpconline.org/handbook/glossary#M
[10] https://www.cs.columbia.edu/~hgs/rtp/glossary.htm#c
[11] https://www.iasa-web.org/cataloguing-rules/appendix-d-glossary

Content level integrity: Ensuring that the content that makes up a digital file renders and performs as expected when accessed (Source: author).

Digital materials: "A broad term encompassing digital surrogates created as a result of converting analogue materials to digital form (digitisation), and 'born digital' for which there has never been and is never intended to be an analogue equivalent, and digital records" (Source: Digital Preservation Handbook Glossary).

Digital Object: "An object composed of a set of bit sequences"[12] (Source: OAIS Reference model).

Digital preservation: "Refers to the series of managed activities necessary to ensure continued access to digital materials for as long as necessary" (Source: Digital Preservation Handbook Glossary).

Digitisation: "The process of creating digital files by scanning or otherwise converting analogue materials. The resulting digital copy, or digital surrogate, would then be classed as digital material" (Source: Digital Preservation Handbook Glossary).

Metadata: "Information which describes significant aspects of a resource… [that is] required successfully to manage and preserve digital materials over time and which will assist in ensuring essential contextual, historical, and technical information are preserved along with the digital object" (Source: Digital Preservation Handbook Glossary).

Migration: "A means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology" (Source: Digital Preservation Handbook Glossary).

Open source: "Open source is a way of developing and distributing software. The code is often written collaboratively, and it can be downloaded, used and changed by anyone"[13] (source: GOV UK).

The University Library: The main research library of the University of Cambridge (Source: author).

---

[12] https://public.ccsds.org/pubs/650x0m2.pdf
[13] https://www.gov.uk/guidance/be-open-and-use-open-source

Approved by	Senior Leadership Team
Approval date	27 May 2021
Written by	The CUL Digital Preservation Policy and Strategy Group Caylin Smith, Head of Digital Preservation (Chair) Katrina Dean, Keeper of Archives and Modern Manuscripts Peter Lund, Librarian Agustina Martinez-Garcia, Head of Open Research Systems Suzanne Paul, Keeper of Rare Books & Early Manuscripts Maciej Pawlikowski, Head of Digital Content Unit Jill Whitelock, Head of Special Collections
Policy owner	Caylin Smith, Head of Digital Preservation