

THE BRITISH LIBRARY



Sustaining The Value

The British Library Digital Preservation Strategy
2017-2020

Executive Summary

“Our vision for digital preservation in the British Library is that by 2020, end-to-end workflows are in place that deliver and preserve our digital collections in a trusted long term digital repository so that they may be accessed by future users.”

Digital Preservation is the combination of actions and interventions required throughout the digital content lifecycle to ensure continued and reliable access to authentic digital materials. This document presents the British Library’s strategic approach to achieving this over the next three years so that our 2020 Vision may be achieved.

Digital Preservation differs from traditional conservation and practices in that threats to content longevity typically manifest earlier and from multiple different sources. The diversity, scale, and importance of the Library’s digital collection further complicate this challenge. Control and consistency throughout the lifecycle is therefore an essential aspect of large scale, sustainable preservation.

This Digital Preservation Strategy identifies the key activities required to sustain the value of our digital collections over the next few years. A core objective is the replacement of our current digital repository infrastructure and the preservation of our digital collection content within a new, dynamic and large scale digital repository.

In summary, our priorities are to:

- ◆ Replace and enhance the existing technical repository infrastructure
- ◆ Ingest our digital collections into the new infrastructure for long term preservation
- ◆ Implement robust processes and reporting mechanisms to provide evidence of preservation
- ◆ Ensure content can be accessed by the Library’s designated community.

Governance of the strategy sits with the Library’s Collection Management Group, a cross-divisional group that approved the strategy in January 2017. Inter-library collaboration is essential for delivery of the strategy; this is recognised in the implementation plan and activities listed towards the end of the document. These will be further developed with each passing year.

1. Purpose

This document provides the framework for the Library's digital preservation activities between 2017 and 2020. Building upon progress made as a result of the 2013 – 2016 strategy, it contextualises activities within our dynamic operational and strategic environment, clearly identifies our strategic priorities, and identifies the roles and responsibilities of those tasked with delivering the strategy.

2. Vision

Our vision for digital preservation in the British Library is that by 2020, end-to-end workflows are in place that deliver and preserve our digital collections in a trusted long term digital repository so that they may be accessed by future users. This is fundamental if we are to achieve our corporate 2020 Vision and guarantee access to future generations.

3. Context

3.1 The Digital Collection

The Library's digital collection is heterogeneous, comprising content such as eBooks, eJournals and the UK Web Archive, digitised newspapers and manuscripts, digital maps and digital sheet music, electoral registers, patents, and personal digital archives. A significant proportion of the collection is acquired via non print legal deposit and is born digital. Further content is acquired via voluntary deposit mechanisms or digitised from existing analogue resources, either internal or external to the Library.

All content acquired by the Library and destined for long term preservation is in scope of this strategy.

3.2 Challenges

Technological obsolescence is often regarded as the greatest threat to digital material: as technology changes, it is increasingly difficult to reliably access content created on and intended to be accessed on older computing platforms. Yet this is just the long term view: in the shorter term we must also consider everything from media integrity and bit rot to digital rights management and metadata.

Other notable differences between analogue and digital content further add to the challenge:

- ◆ Proactive Lifecycle management: Ongoing technological advances and the fragility of digital content require preservation actions to be taken much earlier in the lifecycle than for traditional collections, and at a much greater frequency. A lifecycle management approach is needed to ensure appropriate actions are taken in good time;
- ◆ Integrity & validation: It is easier to make unnoticed changes to digital content than to traditional objects, changes which may affect the authenticity and integrity of the content. Malicious change must be prevented and appropriate change managed;
- ◆ Fragility of storage media: The integrity of storage media for digital materials diminishes at a more rapid pace than analogue archival storage. Resulting bit rot can prevent files from rendering correctly if at all; this can happen with no notice and within just a few years, sometimes less, of the media being produced.

Digital preservation is thus not simply a technical challenge. It necessitates an ongoing and typically recursive series of actions and interventions throughout the lifecycle to ensure continued & reliable access to authentic digital objects, for as long as they are deemed to be of value.

3.3 Stakeholders & Users

Digital preservation services are provided by the Library for its designated community. The designated community comprises all external users of Library digital collections and metadata. It includes academic researchers, businesses, creative users, and incidental communities.

3.4 Strategic Fit

Effective management of digital collection content and metadata supports the delivery of other key Library strategies, including Collection Management, Collection Metadata, and IT. Delivery of the Library's core purposes is also dependent upon digital preservation, specifically:

- ◆ **Custodianship:** Curation and preservation of the national digital collection requires a long term, implemented and sustainable digital preservation solution;
- ◆ **Research:** Research of all kinds is supported and stimulated via the use of generations of digital content and the technological eras they represent;
- ◆ **Business:** Innovation can be stimulated via the implementation of cutting edge digital preservation solutions and the availability of digital materials for use and re-use;
- ◆ **Culture:** Our digital collections represent our digital cultural heritage; digital preservation enables the persistence of this element of the Library's collection;
- ◆ **Learning:** We collaborate with academic institutions around the world to progress our mutual understanding of digital preservation challenges and solutions;
- ◆ **International:** We work with global partners to enhance and exchange digital preservation best practice and expertise, advance development of preservation standards and develop re-usable preservation solutions.

4. Strategic Focus

The 2017 – 2020 digital preservation strategy is oriented around the digital repository used to manage and preserve the Library's digital collections. The repository is a core system for the Library and the current system will be replaced within the duration of this strategy. The focus is therefore upon ensuring this transition enables delivery of our strategic vision.

4.1 Replace and enhance the technical repository infrastructure

A new digital repository and set of preservation tools will be implemented for preserving and managing our digital collections:

- ◆ A new repository infrastructure will be deployed for long term preservation of collection content and metadata;
- ◆ An Integrated Preservation Suite will provide the functionality needed for preservation of intellectual content, managing known and emergent risks to content longevity;
- ◆ File integrity will be maintained and objects appropriately replicated, both content and metadata.

4.2 Ingest digital collections for long term preservation

Content and metadata destined for long term preservation shall be ingested into the Library's digital repository, regardless of form or type:

- ◆ Processes will be in place for timely ingest of both digitised and born digital material as appropriate, both contemporary and legacy;
- ◆ Metadata will be ingested and linked persistently to the content it describes.

4.3 Management and Reporting

Robust processes will provide assurance and evidence that content is appropriately managed and preserved:

- ◆ Management Information will be generated to demonstrate bit-level integrity;
- ◆ Manual sampling will take place on specific collections to ensure content and processes comply with the Library's digital preservation policy;
- ◆ Staff will have the skills needed to deliver effective technical and curatorial management of the digital collection;
- ◆ Processes will be documented and consistent with the digital preservation policy.

4.4 Access for the designated community

Access to digital collections will be timely and reliable:

- ◆ Content will be delivered to users from the Library's long term repository;
- ◆ Users can have confidence in the integrity of the content and its provenance.

5. Expected outcomes

By 2020 we will have:

- ◆ A trusted digital repository infrastructure and processes for managing and preserving our digital collection;
- ◆ End to end workflows in place for processing, preserving, and providing access to our digital collections;
- ◆ Capability to evidence successful preservation so that content is reliable and accessible;
- ◆ Embedded the skills and resources needed to sustain our approach into the future.

6. Delivering the Strategy

6.1 Governance & Support

Overall governance rests with the Head of Collection Management and the Collection Management Group (CMG). CMG is responsible for monitoring delivery of the Collection Management Strategy, of which the Digital Preservation Strategy is one component.

Support will be provided by the Infrastructure Renewal group. This group has governance over renewal of the Library's technical infrastructure, of which the digital repository is one component.

6.2 Implementation

The strategy will be delivered through a combination of project and business as usual activities. Annual implementation plans describe requisite tasks in more detail in order to meet strategic priorities and future requirements. Activities will be undertaken collaboratively by staff from

departments across the Library alongside the Digital Preservation team. Plans for later years in particular will be developed in more detail as the time approaches.

6.2.1 2017/18

- ◆ **System procurement:** a new digital repository system will be procured and plans initiated for implementation of the system;
- ◆ **Content and metadata migration planning:** existing holdings from within the current repository will be analysed; plans will be developed and tested for migration of existing content and metadata into the new system;
- ◆ **Legacy born-digital collection content:** end to end workflows for preservation of legacy born-digital collection content stored on handheld media will be implemented
- ◆ **Integrated Preservation Suite:** development will commence on state of the art tools for implementing long term preservation plans
- ◆ **Digital Preservation Policy:** an assurance group will coordinate institutional responses to digital preservation policy issues as they arise

6.2.2 2018/19

- ◆ **System implementation:** the new repository system will be implemented;
- ◆ **Content and metadata migration:** existing content and metadata will be migrated into the new system and the migration validated;
- ◆ **New process development:** new workflows will be established to enable ingest, management, and access to new ingests, both content and metadata;
- ◆ **Training:** staff will be trained to understand and perform new roles and responsibilities;

6.2.3 2019/20

- ◆ **System utilisation:** use of the new repository system will begin in earnest;
- ◆ **Preservation planning:** functionality is available to manage and mitigate against the risks associated with long term digital preservation at scale;
- ◆ **Audit and certification:** the library will be recognised as a trustworthy memory organisation preserving digital collections for current and future users.

6.3 International engagement

Digital preservation is an emergent and niche discipline, with practical and academic research around the world contributing the development of best practice. International engagement will continue throughout the duration of the strategy to promote developments within the Library and ensure the Library can benefit from relevant developments elsewhere. This will include working with national memory organisations, academic institutions, and membership organisations such as the Digital Preservation Coalition, the Open Preservation Foundation, and the BitCurator Consortium.

6.4 Metrics

Suitable metrics will be developed and trialed in 2017/18 to enable effectiveness of the strategy to be measured. These will be supplemented by an annual informal assessment using an approved digital preservation audit methodology.

7. Appendix

7.1 Related documents

1. British Library Digital Preservation Policy 2017 – 2020
2. British Library Digital Preservation Strategy 2013 – 2016
3. British Library IT Strategy 2016 – 2020

7.2 Glossary

Term	Definition
Access	The act of making objects available to users
Accessibility	The ability to access intellectual content held in files
Acquisition	The process of acquiring objects from any source
Authenticity	The quality of being genuine and free from tampering or alternation, malicious or otherwise
Designated Community	An identified group of users, defined by the Library, who should be able to understand preserved content.
Digital collections	One or more groups of digital materials with a unifying characteristic and/or source
Digital preservation	The series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value
Digital repository	A technical system for managing, storing, and preserving digital content
End-to-end workflow	A complete and unbroken sequence of connected steps for managing digital content across the lifecycle.
Lifecycle	All of the distinct stages of an object's existence, from conception or creation, through to preservation and access and re-use.
Metadata	Data about data, or information that describes an object. Preservation metadata is a particular class of metadata that supports and documents actions related to preservation, specifically supporting the functions of maintaining the fixity, viability, renderability, understandability, and/or authenticity of digital materials in a preservation context.
Technological obsolescence	The state of being sufficiently technically out of date so as to impede access to digital content