

Digital Preservation Framework

2024-26



STATE LIBRARY®
NEW SOUTH WALES

Table of Contents

Objective	1
Scope	1
Roles and responsibilities	1
Review	2
Framework	2
Access and use	2
Arrangement and description	2
Methods of access	2
Rights and restrictions management	3
Acquisition	3
Characterisation of digital collections	3
Community of practice and collaboration	3
Continuous improvement	4
DPC Rapid Assessment Model	4
Fixity	4
Management	5
Deletion of digital material	5
Digital collection categories	5
Reporting	6
Technical Issues	6
Preferred preservation file formats	6
Migration and normalisation	6
Pre-ingest preservation events	6
Preservation levels	7
Preservation planning	7
Staff skills and training	8
Storage	8
Related documents	9
Library documents	9
Standards, good practice and resources	9
Document control	9

Objective

Digital preservation refers to the ongoing, managed activities necessary to ensure continued access to digital materials for as long as necessary. The Digital Preservation Framework identifies and outlines the processes and operational rules to preserve the digital collections of the State Library of New South Wales to ensure accessibility for future generations. It provides a mechanism to monitor and assess the Library's digital preservation maturity, capability, and infrastructure to prioritise related activities and resources.

The Framework aligns with the Library's Strategic Plan, is guided by the principles outlined in the Collection Preservation Policy, and supported by operational procedures and documentation.

Scope

The Framework applies to all digital collections managed within the Library's digital preservation system, including born-digital and digitised material, and covers all functions relating to digital preservation across the Library.

The Framework excludes digital collections managed by vendors or other organisations, such as National eDeposit (NED) and eResources. Where collections are managed by external service providers, the Library will ensure service agreements include terms that cover digital preservation to ensure collections are protected and managed according to the Library's requirements.

Roles and responsibilities

With distributed management of digital preservation and digital collections at the Library, roles and responsibilities fall across divisions, branches and stakeholder groups as follows:

- The Digital Preservation Framework is approved by the Integrated Library Systems Strategy Group (ILSSG)
- The Framework is reviewed and endorsed by the Digital Preservation and Access Development Group (DPADG)
- The Lead Digital Archivist, Catalogue Standards and Digital Curation branch, Library Services Division (LSD), is responsible for coordinating and facilitating the ongoing review and update of the Framework
- Operational procedures and documentation are created, reviewed, and managed by the Digital Collections Operations Working Group
- The Digital Curation team, Catalogue Standards and Digital Curation branch, LSD, is responsible for managing born-digital material and associated workflows for acquisition, ingestion, preservation, and access
- The Digitisation and Imaging branch, Digital Experience Division (DXD), is responsible for defining standards and specifications for digitised material, and managing the digitisation workflow including creation and quality assurance for digitised material
- The Systems and Applications team, Digital Library Systems and Services branch, DXD, is responsible for:
 - Managing digital material and associated workflows for ingestion, preservation, and access
 - Maintaining, supporting, and developing systems related to all digital preservation activities
- Information, Communication & Technology Services branch, DXD, is responsible for:
 - Management of infrastructure, including storage and disaster recovery for digital collections
 - ICT strategy and policy compliance with the Framework and associated documents

- Communicating and consulting with DPADG on system, infrastructure, and storage changes that may affect digital preservation processes and operational rules
- All staff working with digital collections are responsible for understanding and adhering to the Framework and related operational procedures and documentation

Review

The Framework and related operational procedures and documentation must be reviewed on a regular basis to ensure that they are aligned with good practice, the Collection Preservation Policy, and the Library's Strategic Plan. Individual documents have their own review date and version number in order to track and manage their review.

The Framework will be reviewed every two years and will include the completion of the Digital Preservation Coalition Rapid Assessment Model (DPC RAM). Operational procedures and documentation will be reviewed annually or as required.

Framework

Access and use

The Library creates, collects, and preserves digital collections to provide long-term access and use. Where possible, the Library aims to provide meaningful access online or onsite through the Catalogue and Digital Collections platform to minimise the technical challenge for users. This includes the creation of derivative copies and viewers that are suitable for the web for digital objects including image, audio, text and video.

Arrangement and description

Arrangement and description are the foundation for both preservation and access for digital collections. The Library will adhere to the latest operational procedures, ensuring catalogue records provide access at a level appropriate to the significance and anticipated usage of the resource.

Methods of access

Access strategies vary based on the type of digital object. Some files require specific hardware and software to be rendered and accessible. Methods of access denote the usability of digital collections and are defined as follows:

- **Online access** where a file format (or derivative) can be rendered in viewers within the catalogue or digital collections interface
- **Onsite access** where a file format can be rendered on the digital collections access computer in the reading room
- **Unsupported** where there is currently no method to render a file format within a Library environment

Where online or onsite access is not supported and rights agreements allow, users will have the option to access a copy of files with an indication of software and hardware that may be required to interact with the material. The version and representation of digital collections that is most appropriate to the access request should be delivered, for example a compressed derivative copy or uncompressed preservation copy.

The Library will investigate emulation strategies for access to complex born-digital collections.

Rights and restrictions management

The Library will adhere to designated access conditions for all digital collections and aims to implement the following access levels:

1. Open: unmediated access to materials
2. Conditional: collection material that has restrictions, such as permissions, including material deemed sensitive or under copyright
3. Closed: not available

Acquisition

Collecting is guided by the general principles outlined in the Collection Development Policy and related guidelines. This includes appraisal of digital collections and assessing their suitability for long-term preservation and access.

Transfer of digital materials to the Library will be undertaken following operational procedures and will include validation to ensure they are complete and authentic. This includes:

1. Conducting malware and virus scans
2. Creation of fixity metadata, or validation of existing fixity metadata
3. Quality checks against available guidelines
4. Ensuring appropriate metadata and documentation is provided

Where the creation of content or preservation copies can be controlled or influenced, the Library will encourage the use of preferred preservation file formats and provide guidance to vendors or donors on file and folder naming and metadata requirements.

The Library will aim to resolve issues with digital material that is password protected, encrypted, or contains malware or viruses at the time of acquisition.

Characterisation of digital collections

Characterisation, in digital preservation, is a process that provides an overview of digital collections to inform management and preservation decisions. The Library's digital preservation system conducts this process on ingest and includes file format identification, file format validation, and metadata extraction. The system also reports on information such as number of files, file types, and sizes. The Library aims to implement appropriate tools within the digital preservation system for file format characterisation and periodically review preserved digital material to improve characterisation of digital collections over time.

Community of practice and collaboration

Engagement with the local and global digital preservation community of practice enables sharing of knowledge and expertise and provides opportunities to collaborate on standards and good practice across diverse sectors facing similar challenges.

The Library will share information on its strategies, systems, tools, and experiences by actively contributing to and engaging with the community of practice through involvement with initiatives such as the Digital Preservation Coalition, NSLA Digital Preservation Network, Australasia Preserves and other relevant networks, collaborative opportunities, and events.

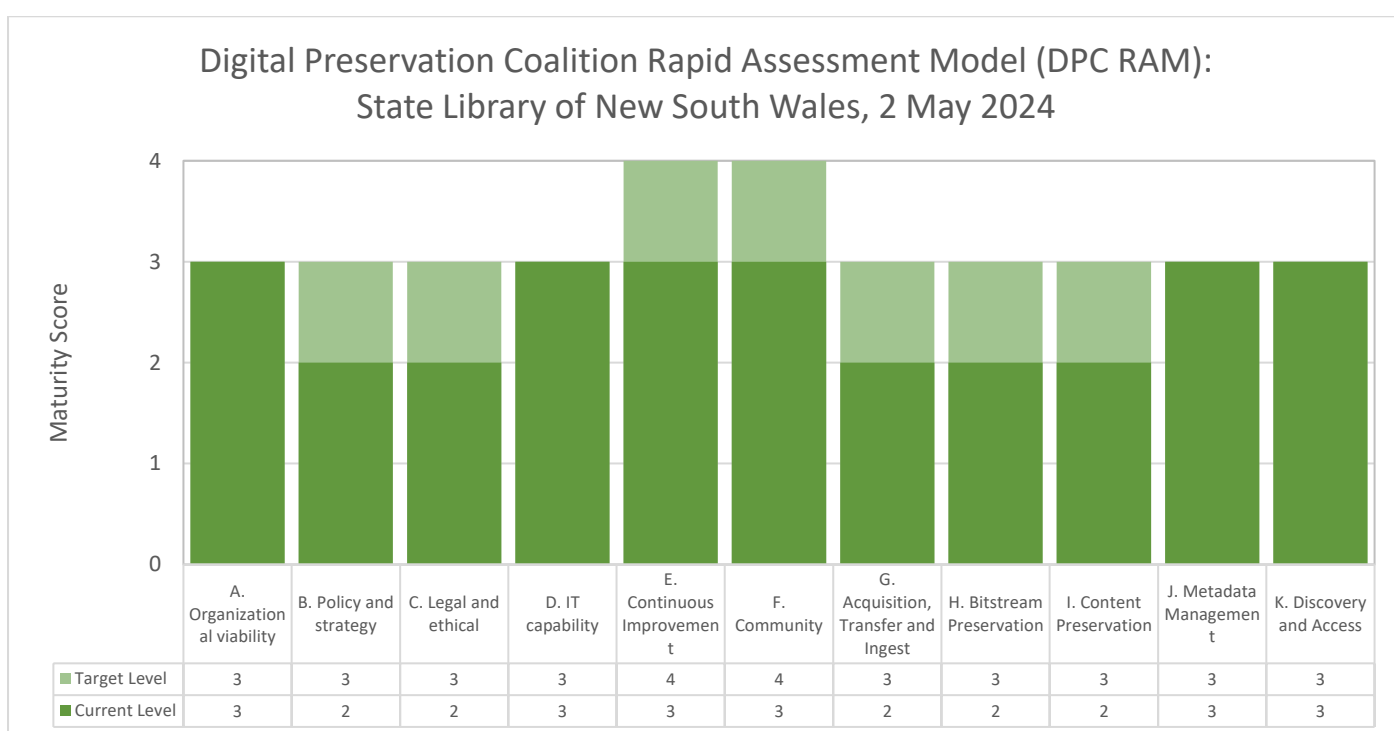
Continuous improvement

DPC Rapid Assessment Model

The DPC RAM is a maturity modelling tool that has been designed to enable rapid benchmarking of digital preservation capability and provides a mechanism to assess where the Library is now and consider where it would like to be in the future. The current and target levels indicated in the DPC RAM align with the Framework and its review period and will be used to reflect and report on progress and continuous improvement. The scope of assessment for the Framework is all digital collections, both born-digital and digitised materials, stored in the Library’s digital preservation system.

The chart below highlights the results for the current review period. The Library is predominantly at Basic (2) and Managed (3) level of maturity across the 11 sections which cover organisational capabilities and service capabilities. The Library aims to reach Managed (3) and Optimised (4) levels by 2024.

A detailed worksheet with analysis of current levels and goals to reach the target levels will be provided as part of internal review and approval of the Framework.



Fixity

Fixity metadata is used to verify that digital content has not changed over time through corruption or unauthorised changes. It is established and monitored through the use of checksums; a unique ‘digital fingerprint’ created with an algorithm that will be modified by any change made to a file.

The Library will adhere to the following fixity practices for digital collections:

1. Fixity will be established at the earliest possible time and will not act as a barrier for acquisition
2. MD5 checksums are preferred for pre-ingest fixity validation, including for send and receipt of digital material between the Library and vendors or donors
3. Fixity validation is conducted during any move across storage mediums

4. During ingest into the digital preservation system, fixity metadata is validated or generated if it does not exist
5. Fixity validation is conducted for digital material in permanent storage within the digital preservation system on a rolling, ongoing basis and in response to specific events
6. Fixity validation is reported to the Digital Preservation and Access Development Group and action will be taken to address any identified issues

Management

The Library's digital preservation system is used to manage digital collections for long-term access, with preservation processes undertaken during ingest as well as on digital material already within the system. This includes characterisation, migration and monitoring of digital materials stored in the repository at scale, with error-checking and evaluation of the outcome of preservation actions an ongoing process over time.

With digital preservation management responsibilities across multiple teams, the Library aims for a consistent approach managing digital collections through the Digital Collections Operations Working Group and use of operational procedures and documentation. This group comprises technical analysts and specialists and has oversight on processes undertaken within the digital preservation system.

Deletion of digital material

Digital material selected and ingested into the digital preservation system for long-term preservation will not be deleted unless one of the following conditions apply:

1. It has been ingested in error or for testing purposes
2. Physical collection material has been re-digitised, and the existing preservation copy is being replaced
3. Deletion of born-digital material has been authorised by a sanctioned authority as outlined in the Collection Retention and Withdrawal Policy

Where deletion has taken place, a record must exist of the process in accordance with recordkeeping requirements.

Digital collection categories

The Library's digital collections are broadly defined in three categories:

1. **Born-digital** – those created, maintained, and acquired in digital format
2. **Digitised primary** – digitised representations of physical material that is not part of the collection, acquired in digitised format or created prior to disposal of physical material
3. **Digitised surrogate** – digitised representations of physical material that is part of the collection, created for access or preservation purposes

These categories do not relate to the condition of physical collection items for digitised material.

Reporting

Regular reporting on the digital preservation repository ensures comprehensive oversight and management of digital preservation activities. A digital preservation report will be provided to the Digital Preservation and Access Development Group at each meeting, including information on the number and size of preserved files, identified file formats, unidentified files, methods of access for preserved collections, categories of digital collections, and preservation levels.

Technical Issues

Technical issues arise where errors occur during processes undertaken within the digital preservation system during ingest or as a result of actions applied to preserved digital material. These issues can be caused by a number of factors such as failures for fixity validation or file format characterisation, failures with tools, or interruption of system processes.

Technical issues are typically identified by the system and solutions should be applied in a consistent and timely manner. The Digital Collections Operating Working Group will document technical issues and solutions for a consistent approach across all digital collections.

Preferred preservation file formats

In determining preferred file formats for acquisition and preservation, the Library will aim for the use of formats with the following characteristics:

1. Open and non-proprietary
2. Widely adopted global use and support
3. Uncompressed or lossless compression
4. Well-documented with publicly available specifications
5. Support for metadata, including technical metadata and other significant properties

The Library may collect digital materials in a variety of forms with emerging, obsolete, and other file formats that may not have clearly defined preservation practices.

Migration and normalisation

For born-digital material the original file is preserved and retained regardless of format, but a migrated or normalised copy in preferred file formats may be created. Migration and normalisation will be undertaken based on defined pathways with established source and target file formats, which will include quality checking with a definition of acceptance criteria of change for digital objects and documentation of actions.

Pre-ingest preservation events

Preservation events record activities that affect the long-term preservation of digital objects and are stored in the digital preservation system utilising the PREMIS Data Dictionary for Preservation Metadata. Preservation events are recorded automatically for any events that take place within the system, but actions may be required pre-ingest to prepare content, resolving any technical issues that may affect files before they can be ingested into the digital preservation system. It is critical that all treatments and events performed on preserved digital materials are accountable through documentation which can be made available to future users.

Where pre-ingest preservation events are significant, they must be clearly documented and stored within the preservation system. Pre-ingest preservation events may include preconditioning, recovery, or other actions agreed to by members of the Digital Collections Operations Working Group.

Pre-ingest preservation events that must be recorded are defined in the Preconditioning Events Controlled Vocabulary and must adhere to the following principles:

1. Pre-ingest preservation actions must be recorded as events as part of a Submission Information Package in addition to a sidecar XLSX file per Intellectual Entity
2. Recorded events must include all metadata elements and values as defined in the Sidecar Provenance Representation template and contain:
 - a. Why the action was required
 - b. What actions were taken on the digital object
3. Changes made to digital content through preconditioning must not change the intellectual content of the object
4. Preconditioning events must be reversible, except when dealing with recovery or modification

Preservation levels

Preservation levels define the extent of confidence and comprehension regarding the sustainability of digital materials over time. They are defined as follows:

- **Level 1** designates a foundational stage where digital material is preserved at the bitstream-level, with the file format not identified
- **Level 2** indicates a file format has been identified and offers a structured approach for preservation planning and actions
- **Level 3** signifies a high degree of preservation confidence, where the file format is not only identified but also validated and/or understood with significant properties defined and extracted

As a minimum, the Library will ensure bitstream preservation (level 1) for all digital collections stored in the digital preservation system. This basic level of preserving the digital object aims to ensure that current and future users can access authentic copies of preserved files and provides options for interpretation and access of content over time, such as emulation or migration.

The Library will aim to achieve level 2 for digital collections so that preservation actions may subsequently be performed based on preservation planning activities. Achieving level 3 demands significant resources and is not a realistic aspiration for every file format.

Where preservation plans clearly identify the need, software applications may be preserved to aid future use and access for born-digital material in emulated environments.

Preservation planning

Preservation planning defines appropriate strategies to mitigate risk and ensure the Library can provide access to digital collections in the long-term. This is an ongoing process that involves monitoring and recommending appropriate revisions to preservation strategies in response to evolving technology changes and good practice. The Library will manage, document, and regularly review preservation plans and actions to ensure that they are valid and in line with good practice.

Staff skills and training

The Library aims to ensure that staff undertaking digital preservation activities have sufficient access to knowledge, training, and professional development opportunities. This is a significant benefit of engagement with community of practice through active participation and membership in relevant networks and organisations.

The Library aims to conduct periodic staff skills audits to ensure expertise and capabilities are developed to provide a workforce with sufficient knowledge in digital preservation to enable the Library to develop and maintain robust digital materials.

Storage

The Library will utilise a multi-copy digital storage strategy for preservation copies managed in the digital preservation system to mitigate risk to digital content and provide a high level of safety and security for digital collections. The Library will store three independent preservation copies utilising the following characteristics:

1. Geographic separation with different disaster threats
2. Use of different storage technologies and vendors
3. Adherence to any legal requirements, agreements, or obligations relating to data and privacy
4. Active monitoring for timely error detection and correction
5. Exit strategies will be considered, staying informed of developments with storage vendors and solutions, and proactive migration of storage where required

Related documents

Library documents

- [Strategic Plan 2019-2023 \(Currently under review\)](#)
- [Collection Preservation Policy](#), 2021
- [Collection Development Policy](#), 2021
- [Collection Retention and Withdrawal Policy](#) (Currently under review)
- [Digital Collecting Strategy](#) (Currently under review)
- ICT Disaster Recovery Plan 2022

Standards, good practice and resources

- [Curation Lifecycle Model](#)
- [Digital Preservation Handbook](#)
- [DPC Rapid Assessment Model \(RAM\)](#)
- [ISO 14721:2012 - Open archival information system](#) (OAIS)
- [Levels of Born-digital Access](#), 2020
- [Library of Congress Recommended Formats Statement](#)
- [Library of Congress Sustainability of Digital Formats](#)
- [NDSA Levels of Digital Preservation](#), 2019
- [PREMIS Data Dictionary for Preservation Metadata](#)

Document control

Version:	2.2		
Endorsed by:	Digital Preservation and Access Development Group	Date of endorsement:	3 June 2024
Approved by:	Integrated Library Systems Strategy Group	Date of approval:	24 June 2024
Contact Officer:	Lead Digital Archivist, Catalogue Standards and Digital Curation Branch	Date of next review:	June 2026