



California State Library Digital Preservation Policy

Vision

The State Library has a long-standing mandate¹ to collect, preserve and provide access to comprehensive materials to enrich cultural and educational needs -- not only for Californians but also for patrons and scholars throughout the world who desire access to the State Library's unique materials.

Because this obligation also encompasses digital collections, the library actively digitizes material and provides open access to it on the Internet via its digital repositories and catalogs. The State Library is committed to meeting U.S. and international standards and "best practices" for preserving and delivering digital resources to its patrons.

This vision supports the Library's mission to be "the state's information hub, preserving California's cultural heritage and connecting people, libraries and government to the resources and tools they need to succeed and to build a strong California."

Purpose

The purpose of this document is to outline the California State Library's approach to digitization and digital preservation and to provide guidance on the preservation of digital materials to ensure their reliability, authenticity and long-term sustainable accessibility.

¹ <https://www.library.ca.gov/publications/2015CaliforniaLibraryLaws.pdf>

It first states general principles and rules regarding specific aspects of digital preservation strategies, then prepares those principles and rules for implementation. Steps to implement the digitization and preservation program must be in accordance with this policy to guarantee their coherence.

This is an interim document, based on current and soon to be implemented practices. This policy will be re-evaluated to reflect the acquisition of any new digital preservation solutions, and thereafter biennially or as needs arise.

Needs Assessment

The California State Library's digital material is part of its broader collection of unique resources and information for public research. As such, digital material requires a policy that establishes a selection, retention and preservation program that safeguards these materials while providing long-term accessibility.

Risk Assessment

The California State Library and its collections are currently at medium risk because the Library does not engage in a full preservation strategy.

(Low risk is assigned to institutions with a full digital preservation process and workflow; high risk to institutions with no preservation process and workflow).

The State Library has digitally converted close to 55,000 items of its analog collection, using sophisticated, high quality equipment and following recommended standards. Most of this has been cataloged and is discoverable online.

Preservation masters of all turned-digital materials are stored on semi-secure, RAID-protected storage that is backed up daily and geo-replicated. Periodic integrity checks are yet to be implemented.

The State Library has digital collections with a wide range of formats, some of which are proprietary and/or becoming obsolete, a challenge the library's digital preservation policy must also address.

Scope and Objectives

Within the State Library's scope of responsibility for digital preservation:

- Digital resources -- born-digital and turned-digital -- that are acquired by the library and which are unlikely to be preserved anywhere else.
- Materials that are in danger of loss due to deterioration or obsolescence.

Out of scope:

- Digital material whose format or lack of metadata make long-term management and accessibility impossible.
- Components of a digital object that may not be necessary to preserve for the long term, such as thumbnail images that are useful for the access system but could easily be generated from an archival master TIFF image and are thus not essential to store permanently.

The main scope of the digital preservation policy is to achieve the following objectives:

1. Preserve and provide continued, sustainable access to digital material, both born-digital and turned-digital.
2. Maintain and develop processes and systems to capture, manage, preserve, find and make accessible digital materials now and into the future.
3. Ensure that the authenticity and integrity of preserved digital material is established and maintained over time.
4. Comply with copyright and other legal rights relating to copying, storage, modification and use of digital resources.
5. Ensure that the highest quality, fullest version of the resource (file) is preserved as the archival master, in a lossless, non-proprietary format whenever possible, with the necessary descriptive, structural and administrative metadata associated with it to make the file usable.
6. Adapt preservation strategies to incorporate capabilities afforded by new and emerging technologies.
7. Minimize the risk associated with technology and personnel changes and be flexible enough to react to or support new technology developments.
8. Migrate and change the format of digital materials to formats suitable and acceptable for long-term preservation and access, when necessary.
9. Undergo periodic risk assessment and audits to ensure that activities are meeting stated commitments, that risks are reduced, and that authenticity and accessibility of digitally preserved content can be verified.

Definitions

Digital preservation: Strategies and actions to guarantee that digital objects remain useable and accessible over time.

Born-digital: Objects that originate in a digital form.

Turned-digital: Objects that are a digital representation of a resource that did not originate in digital form.

Digital Object: Throughout this document, the term digital object is used interchangeably with the terms digital resource, file, and digital material.

OAIS Reference model: The Reference Model for an Open Archival Information System provides a common framework for describing and comparing architectures and operations of digital archives (see “Appendix”).

Authenticity: A digital object (or resource) has clear and known provenance and is what its external metadata purports it to be.

Integrity: A digital object (or resource) has not been corrupted over time or in transit; in other words, it has the same set of sequences of bits that came into existence when the object was created.

PREMIS: The PREMIS Data Dictionary for Preservation Metadata is the international standard for metadata to support preservation of digital objects and ensure their long-term usability.

RPO: The Recovery Point Objective is an industry term used in business continuity planning and is the maximum targeted period in which data might be lost from an IT service due to a major incident or accidental deletion.

Standards and Best Practices

In order to ensure discovery and interoperability among diverse systems, the Digital Preservation Policy will follow standards where they exist. These standards will be employed from the moment of capture through the life cycle of the digital object. Where standards do not exist, the policy will be to share in best practices.

Standards and guidance, which inform digital preservation strategies and processes developed by the State Library, will include the following:

- The Open Archival Information System (OAIS) Reference model (ISO 14721:2012)²
- The Trusted Digital Repository Checklist (ISO 16363)³
- The National Digital Stewardship Alliance’s recommendations on levels of preservation⁴ (also see “Appendix”)
- Digital Library and archival metadata standards, such as: Metadata Object Description Standard (MODS), MARC21/MARCXML, Dublin Core, Metadata Encoding and Transmission Standard (METS), PREMIS

Metadata

The California State Library follows appropriate standards and best practices, which include the adoption of the PREMIS model as the basis for preservation metadata for its digital resources. PREMIS-based metadata is only one part of a larger package of metadata needed to support the use and re-use of digital assets, which additionally includes descriptive, technical, structural and administrative metadata (See “Standards and Best Practices”).

Digital Storage

Until the State Library’s digital preservation capability changes, digital files designated for preservation will be managed with the library’s current digital asset management system, Aleph/ADAM from Ex Libris. Aleph/ADAM provides bit-level protection and security by using the library’s existing digital storage infrastructure and current and augmented workflows, in a manner that is consistent with accepted best practices in the digital preservation community.

Digital files are preserved and managed within structured, secured, dedicated folders on local and remote servers. Servers in this context include the disk media and network connectivity. Full backup routines and processes as well as replica copies of all digital files are stored offline in a geographically dispersed server in a secure room in the Sutro Library, San Francisco, 90 miles away.

Removable media, if used, such as USB-connected drives and tape media, is checked annually for signs of physical deterioration and technical obsolescence. Migration onto new removable

² http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

³ http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

⁴ http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf

media or to new servers is carried out where deemed necessary to ensure continuing access to the digital files.

File formats digitally stored are reviewed annually for risk of obsolescence. If a file is considered to be in danger of obsolescence, appropriate measures are taken to ensure that the file remains in an accessible state. A surrogate copy for public access in an appropriate format is created for each file as part of the digital asset management system.

Commencing with the publication of this policy: Fixity checks on digital objects are run annually or after transit to another system, using SHA-256 or better checksum algorithms, to ensure the ongoing integrity of digital objects and their metadata. Three copies of all archive master files exist, with at least one geographically separate; allowing for the recovery of files in the event data integrity is compromised. Fixity check events are recorded and described using PREMIS metadata.

Access/Use

The digital resources at the State Library must remain accessible into the future. To accomplish this, the library adheres to the following guidelines:

1. Provide web-based access using the most up-to-date and standards-based technology and software available.
2. Inform users of any necessary software needed to render files that they may download, as well as using openly documented formats for which rendering tools are readily available such as PDF and WAVE files.
3. When a resource may only be rendered using the original software application that created the resource, migrate that resource to a newer format in order to support continued access or consider emulation environments to support access. Maintain and preserve the original bit stream of any migrated file.

Roles and Responsibilities

- The State Librarian is responsible for:
 - Approving and promoting this policy and ensuring sufficient resources are allocated to all areas of the workflow and that staff receive appropriate and ongoing access to training.
 - Ensuring biennial review and update of this policy.

- The State Library's staff members involved in digitizing materials shall ensure adherence to relevant standards and smooth integration with other business processes and policies.
- The CIO and IT staff:
 - Ensure smooth integration with technology systems and workflows as well as for business continuance/disaster recovery.
 - Provide physical security to the servers, network and other infrastructure and adherence to the appropriate level of preservation assurances.

Digitization

The California State Library follows relevant standards and best practices for approaches to digitization, including technical specifications, choice of format and metadata, as documented by the Library of Congress, National Archives and Records Administration (NARA), California Digital Library, the Digital Library Federation (DLF), US Federal Agencies Digitization Guidelines Initiative (FADGI), and other professional bodies and institutions.

Digitization Principals

1. Digitization activities encompass all formats.
2. All digitized materials have metadata in a form supported by the California State Library and its catalog and repository systems.
3. Digitized collections are available for discovery by library patrons via the catalog and digital asset management and repository systems at the library and via union catalogs and discovery systems to which the library exports metadata.

Selection for Digitization

Library section heads and reference librarians make selection decisions. Selection decisions must support the library's objective to digitize, preserve and make accessible collections that are unique to the State Library and that have state, national and international interest.