UVic Libraries digital preservation framework

Digital Preservation Working Group

29 March 2017

Purpose

This document formalizes the University of Victoria Libraries' continuing commitment to the long-term preservation of a diverse and extensive range of digital assets under its care. It identifies high-level principles and priorities for implementing a digital preservation strategy in support of the Libraries' overall goals and strategic directions. Specific action plans for content groups capture policies specific to that group, and the processes, workflows, and tools that implement those policies.

What is digital preservation?

Digital preservation is the active management of digital content over time to ensure ongoing access.[1] It is a formal endeavor on the part of an organization to ensure that digital information of continuing value remains accessible and usable. Much more than just long-term storage, it involves a number of key activities, as well as building the organizational capacity to undertake them. These activities include:

1. Identification
2. Selection
3. Storing
4. Protecting
5. Managing
6. Providing access

Objectives

The primary objective of the Libraries' digital preservation efforts is to preserve and make accessible over the long term materials in support of the teaching, research, and community engagement activities of the University, as well as institutional materials with legal, administrative, and historical value. The Libraries will:

- Protect the University's investments in digital materials through preservation activities;
- Identify, through systematic selection, materials to be preserved, consisting of born-digital and digitized materials
- Comply with prevailing community standards for preservation and access; observe legal compliance requirements related to retention, preservation, and access
- Seek, expand, and develop digital preservation methods that are appropriate for the Libraries; and
- Promote inter-institutional and other types of collaboration.

---

[1] Library of Congress. http://www.digitalpreservation.gov/about/

Guiding principles

Foundational principles

- The Libraries' digital preservation strategies are driven by and support the Libraries' collection building mandate. Digital preservation programs and activities complement and share the same mission as the Libraries' analogue preservation programs.
- Long-term access to digital assets is the purpose of digital preservation.
- Digital preservation is not a binary state where something is preserved or not. It is a property of the policies and procedures used to manage digital assets that either increases or diminishes their usefulness over time.
- Risk identification and mitigation is a central part of the digital preservation strategy at the Libraries.
- The Libraries cannot and will not preserve everything, or all materials at the same level.
- Digital preservation must be technologically, organizationally, and financially feasible and sustainable.

Operating principles

- Specific responsibilities for actions related to preservation should be clearly defined and decisions related to all aspect of digital preservation should be made systematically.
- Developing and maintaining thorough documentation is essential to the success of digital preservation effort. Documentation will be updated and reviewed on a regular basis and be accessible to departments and units within the Libraries, where this does not pose a security or privacy risk.
- This document and its associated action plans must be accessible and defensible and decisions around digital preservation and the activities of the Digital Preservation Working Group must be undertaken in as transparent a way as possible.
- This document and its associated action plans must effectively adapt to organizational, environmental, and technological change, and be subject to a meaningful and periodic review process.
- Efforts should align where possible with community-accepted standards. When community-based tools, services, and standards are not employed for specific content groups, the action plan must provide an explicit rationale for this.
- Each specific action plan must define testable and achievable success measures.

Roles and responsibilities

The traditional role of librarians and archivists includes preserving a wide variety of materials in analogue formats, and the same is true for materials in digital form. Digital preservation activities at the Libraries are overseen by the Digital Preservation Working Group. The Group's functions are:

- To ensure that this document and associated action plans are being created, maintained, implemented, and updated as necessary to support the strategic directions and priorities of the Libraries;
- To undertake work and make decisions in reaction to changes in in the wider UVic and community context, and in relation to general trends, activities, and advances in the field of digital preservation;

- To coordinate the Libraries' efforts in the field of digital preservation with those being undertaken at a regional, national, and international level; and,
- To lead regular strategic planning activities related to digital preservation to ensure that the long-term goals of this document and associated action plans are met, including re-prioritization of activities related to specific content groups as needed.

The Digital Preservation Working Group is also responsible for undertaking activities, or engaging with stakeholders across the University that are currently undertaking activities, related to digital preservation. These activities include:

- Selecting material for long-term access based on scholarly value and the Libraries' technical capacity to preserve the materials over time;
- Actively monitoring technology;
- Reviewing licenses for preservation implications;
- Promulgating good practice for preservation of digital records, publications, and other material;
- Ensuring quality creation of digital materials through adherence to digitization best practices and through engagement with content producers where appropriate;
- Responding to technological obsolescence through migration or other strategies;
- Establishing reliable and comprehensive archival storage for born-digital institutional and private archives, as well as other digital material;
- Providing security for content and systems; and,
- Collaborating with other institutions on possible shared services and activities focused on digital preservation.

Scope and prioritizations

The Libraries are responsible for identifying and providing the means to preserve and ensure ongoing access to selected digital assets. Not all of the digital content the Libraries creates or acquires will be preserved. The Libraries commits to these classes of objects with associated preservation priorities and levels of commitment:

- *Priority 1:* Born digital materials: Rigorous effort will be made to ensure preservation of selected material selected, both library resources and institutional records, in all formats, as is reasonable and practicable. This includes, where applicable, not only simple digital objects like text and video files, but complex digital objects, such as websites and research environments.
- *Priority 2:* Digitized materials (available or limited availability of analogue): Every reasonable steps will be taken to preserve materials without an analogue original, or when re-digitizing is not possible, or very limited numbers of analogue versions are available. Also included are digitized materials that have annotations or other value-added features making them difficult or impossible to recreate.
- *Priority 3:* Digitized materials (available analogue): Reasonable measures will be taken to extend the life of digital objects with a readily available analogue original. However, the cost of re-digitizing as needed will be weighed against the cost of preserving the existing digital objects.
- *Priority 4:* Published items and other miscellanea: No preservation steps will be taken for ephemeral materials such as materials scanned for the Libraries' reserves, document delivery, or

related services, odds and ends and portions of text scanned from analogue collections, and content that is deemed unessential to the comprehensiveness of collections.

- *Priority 5:* Widely-available subscription resources: No preservation steps will be taken for commodity materials such as licensed database content and electronic journal packages, although the Libraries may choose to support consortial and other efforts aimed at ensuring post-cancellation access and the long-term preservation of these materials.

Cooperation and collaboration

UVic Libraries is committed to collaborating within the University and with other institutions to advance the development of digital preservation capacity, share lessons learned with other organizations, and extend the breadth of our available expertise. Collaborative opportunities for digital preservation will be prioritized.

Challenges

There are many challenges associated with digital preservation. The Libraries efforts must especially take the following into consideration:

- Change: Technology changes rapidly in relation to content creation and access, but also in relation to preservation-related systems and services. Staying up-to-date and responding to such changes is a huge challenge.
- Sustainability: The Libraries requires sufficient resources, including designated library funding, for ongoing efforts related to the long-term stewardship of digital materials. The scope of preservation activities needs to accurately reflect the availability of resources.
- Flexibility: To respond to evolving technological capabilities and changing user expectations, the Libraries must continually revise requirements for preservation without jeopardizing the ongoing care of existing digital content.
- Content creators and partnerships: The Libraries needs to partner with content creators to facilitate preservation-friendly creation that restricts as little as possible scholarly and creative expression.

Content groups

UVic Libraries will preserve the following content groups[2].

High priority

- Electronic theses and dissertations and other institutional repository content
- High-risk physical holdings requiring immediate reformatting in Special Collections and University Archives
- Born-digital material held in Special Collections and University Archives, e.g. records on loose magnetic and optical media and laptop hard drives
- Institutional records held in University Archives
- Websites
- Research data generated by UVic researchers and curated by the Library

[2] Additional content groups will be added as required, and priorities may change over time. Existing content groups tend to share a certain level of provenance (e.g. electronic theses and dissertations, websites, etc.) or are created as a result of a specific organizational activity (e.g. digitization, research data management, etc.).

Medium priority

- Already digitized collections, e.g. local physical holdings digitized from Special Collections and University Archives and other digital files produced as part of the Libraries' digitization initiatives

Low priority

- Digital content acquired from commercial vendors

Operational details of preservation strategies and workflows as applied to each content group are outlined in action plans specific to that content group.

Preservation levels for content groups

This document defines a set of preservation levels that outline the basic processes and the amount of resources applied to each content group to ensure that it is accessible indefinitely. Preservation levels and access goals for each content group are described in that content group's action plan.

Level[3]	Description	Access goals[4]	Success measures[5]
1 (bit-level)	UVic Libraries ensures bit-level integrity of files over time	Authorized users can access copies of material in the same format it was originally in when it entered the Libraries' preservation workflows. Bit-level preservation does not ensure that files will be accessible in the same software used at time of access.	Checksum at time of original processing is the same as at time of future access
2 (normalization)	At time of capture or ingest, UVic Libraries normalizes content into open formats, which are then preserved along with the original at preservation level 1 (bit-level).	Authorized users can get a copy of the files that make up a resource in an open, non-proprietary format that was current at time of capture or ingest, with significant characteristics[6] of the	The normalized versions of all files that make up a resource have checksums that are identical to ones derived at the time of normalization.

---

[3] At all four preservation levels, metadata (descriptive, structural, rights, technical, preservation) is captured along with content files, as detailed in each content group's action plan. This metadata will support activities required by the various levels of preservation.

[4] "Access goals" describe the type of activity that users of the preserved content are anticipated to be able to perform with that content in the future.

[5] "Success measures" identify activities that UVic Libraries staff or external organizations can perform on a routine basis to verify that access goals as described above can be achieved.

[6] "Significant characteristics" are the characteristics of a digital object that make it accessible and useful to end-users. UVic Libraries will employ the significant characteristics as defined by Archivematica

	Normalization does not imply ongoing format migration (level 3).	original reasonably intact.	
3 (migration)	As formats of files created by normalization become obsolete over time, UVic Libraries migrates content to contemporary, open formats. Migrations to new formats is essentially the same as normalization at time of ingest, except that it happens sometime in the future as required. Original files are preserved at level 1 (bit-level).	Authorized users can access the resource in file formats that are current at the time of access. Files may not correspond one-to-one with the original files, but the significant characteristics of the original resource will be reasonably intact.	The migrated version of the resource retains as many of the significant characteristics of the obsolete version as is practical. Migrated versions of the original are usable in software common at time of access. Migrated versions of all files have future checksums that are identical to the ones derived at the time of migration.
4 (emulation)	UVic Libraries preserves the original files so that they can be viewed, played, or used in an appropriate manner within a computing environment that emulates the one that was current at the time the files were created. Original files are preserved at level 1 (bit-level).	Authorized users can access the original files in a computing environment (via original hardware, a virtual machine, within a browser, etc.) that is similar to the one the files were designed for. Significant characteristics of the original file are unaltered or as intact as is reasonably possible.	All significant characteristics of the original files are reasonably intact, since the computing environment they are tested within in the future is reasonably identical to their original environment. Original files are renderable in a suitable emulation environment.

Storage levels for content groups

This document defines a set of storage levels that outline the basic storage infrastructure applied to each content group to ensure that it is accessible indefinitely. Storage levels for each content group are described in that content group's action plan.

	Description
Level 1	Two complete copies that are not collocated.
Level 2	At least three complete copies. At least one copy in a different geographic location.
Level 3	At least three complete copies. At least one copy in a geographic location with a different

https://wiki.archivematica.org/Significant_characteristics. These represent a well-researched, community-led, operationally tenable set of characteristics.

	disaster threat.
Level 4	At least three copies in geographic locations with different disaster threats.

Review and succession criteria

This document will be reviewed and updated as needed with a full review every two years to ensure timely revisions as technology, preservation strategies, and experience evolve. This review process will be initiated by the Digital Preservation Working.

Last updated: 29 March 2017, Corey Davis

Addenda

Platforms, preservation tools, and storage infrastructure
Digital asset management and related platforms
The following digital asset management and related platforms are being used by UVic Libraries to provide access to locally-produced or locally-stewarded digital materials:

- DSpace
- ContentDM
- OJS
- OCS
- AtoM (local and hosted)
- Archive-it
- Dataverse
- Omeka

Preservation processing and related tools
The following preservation processing and related tools are being used by UVic Libraries locally-produced or locally-stewarded digital materials:

- COPPUL's Archivematica-as-a-Service http://coppul.ca/archivematica. Hosting and digital object storage and back-up are provided through UBC's EduCloud Server Service.
- Digital forensics workstations:
  - New/Old workstation: Native capabilities of a 15 year old PC with 5.25" and 3.5" floppy drives, configured to boot into DOS, Windows 3.1 and Linux (Centos 5).
  - USB Floppy workstation. 5.25" floppy drive attached to an FC 5025 USB floppy controller. Running Windows 7, and comes with point and click software that greatly simplifies the process of imaging 5.25" floppy disks.
  - Kryoflux workstation. This workstation can read damaged 3.5" and 5.25" floppy disks that the other workstations cannot, using low-level magnetic flux transition sampling.

Storage infrastructures
The following storage infrastructures are available to UVic Libraries for locally-produced or locally-stewarded digital materials

- COPPUL LOCKSS. This is the most reliable storage but is very limited in capacity (2TB shared by 10 sites) and it is difficult to get content into it. It has potential though if COPPUL can build capacity and iron out some of the management issues. Cost to the library is $8000/year plus the cost of administering a LOCKSS node.
- University Storage Area Network (aka Netdrive). The Library has two file shares on Netdrive, one for administrative file storage (450G capacity) and one for digital project files (12TB capacity). Cost to the Library is $300/TB/year. Netdrive greatly exceeds consumer grade storage in terms of reliability, but is not sufficient to ensure long term preservation of our assets. The University's storage area network is built to corporate standards, with onsite and offsite backups. It is currently being refreshed and upgraded.

- TSM, the University's tape archiving system. TSM stores data in a tape library with copies both on and offsite. Cost to the Library is free. TSM is less robust than Netdrive. Archived TSM content is migrated periodically to fresh tape.
- UVic enterprise server storage. Content on our servers is stored on networked storage in the UVic data centre. Cost to the Library is $1000/TB/year.
- UBC EduCloud. Our Archivematica instance runs on the EduCloud infrastructure at the University of British Columbia. Similar to Netdrive it is a corporate-grade IT infrastructure with onsite and offsite backups, and an option to send backups to a data centre at UBC Okanagan.

External documents on digital preservation standards, best practices, and recommendations

- **ISO 14721:2012: Space Data and Information Transfer Systems -- Open Archival Information System (OAIS) Reference Model:** "OAIS defines the reference model for an open archival information system (OAIS). An OAIS is an archive, consisting of an organization, which may be part of a larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a designated community. It meets a set of such responsibilities as defined in this International Standard, and this allows an OAIS archive to be distinguished from other uses of the term 'archive'. The term 'open' in OAIS is used to imply that ISO 14721:2012, as well as future related International Standards, are developed in open forums, and it does not imply that access to the archive is unrestricted." (http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284)
- **International Research on Permanent Authentic Records in Electronic Systems (InterPARES):** "a collaborative research initiative led by the University of British Columbia that is focused on addressing issues of long-term preservation of authentic digital records. The research is being conducted by focus groups from various institutions in North America, Europe, Asia, and Australia, with an objective of developing theories and methodologies that provide the basis for strategies, standards, policies, and procedures necessary to ensure the trustworthiness, reliability, and accuracy of digital records over time." (https://en.wikipedia.org/wiki/Digital_preservation#InterPARES)
- **ISO 16363:2012: Audit and Certification of Trustworthy Digital Repositories (TDR):** defines a recommended practice for assessing the trustworthiness of digital repositories. It builds off of Trustworthy Repositories Audit & Certification (TRAC): Criteria & Checklist Version 1.0, 2007. "In 2006 the Center for Research Libraries began certifying repositories of interest to the CRL community. Certification augments CRL's strategic archiving of print and supports the responsible transition by CRL libraries to electronic-only formats where appropriate. (https://www.crl.edu/archiving-preservation/digital-archives/certification-assessment)
- **PREMIS (PREservation Metadata: Implementation Strategies):** "The PREMIS Data Dictionary [for Preservation Metadata] defines a core set of semantic units that repositories should know in order to perform their preservation functions. Preservation functions can vary from one repository to another, but will generally include actions to ensure that digital objects remain viable (i.e., can be read from media) and renderable (i.e., can be displayed, played or otherwise interpreted by application software), as well as to ensure that digital objects in the repository are not inadvertently altered, and that legitimate changes to objects are documented." (http://www.loc.gov/standards/premis/understanding-premis.pdf)

- **METS (Metadata Encoding and Transmission Standard):** "a metadata standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation." (https://en.wikipedia.org/wiki/METS) METS allows for the use of externally developed metadata schemes to be used for descriptive or administrative metadata. PREMIS has been registered as a recognized metadata scheme to be used as administrative metadata with METS. (http://www.loc.gov/standards/mets/mets-extenders.html)