University of British Columbia Library

Persistent Digital Collections Implementation Plan

Final project report
Summary version

May 16, 2012

Prepared by

artefactual systems inc.

1. Introduction

In 2011 Artefactual Systems Inc. was contracted by the University of British Columbia Library to assist in the development of a long-term preservation strategy for the Library's digital collections. This year-long project consisted of a detailed analysis of UBC Library's digital preservation capacity, drafting of a implementation plan and overall system architecture for digital preservation at the Library, and running a series of preservation pilot projects using the open-source Archivematica digital preservation system. The pilots assisted with fine tuning the plan, ensuring that it recommends viable, practical next steps to enable the transition to operational digital preservation services.

The persistent digital collections implementation plan developed for UBC Library consists of using the open-source Archivematica digital preservation system to provide preservation functionality for the Library's digitized and born-digital holdings. The plan identifies the software requirements, existing and new system components, staffing and business processes that can be implemented to establish operational digital preservation systems and processes by the end of 2012. Using open-source software and existing UBC Library systems and staff resources, this can be achieved at minimal cost. Archivematica will be the central tool that interfaces with existing systems (such as DSpace and CONTENTdm) and processes (such as digitization and acquisition of born-digital records) to establish physical control over digital objects, monitor technological obsolescence risks, implement long-term format-specific preservation plans (normalization, migration, and/or emulation), maintain preservation masters and generate and manage preservation metadata for all of the Library's digital materials.

2. A digital preservation implementation plan for UBC Library

During the first few months of this project, Artefactual Systems conducted a digital preservation gap analysis of the workflows and technical architectures which support the cIRcle and CONTENTdm systems, and found that although they support access to and, in the case of cIRcle, submission of digital objects, preservation functionalities are almost entirely absent. In addition, various Library units and projects acquire digital objects on external media but lack the ability to apply any preservation processes to them.

The ISO 14721 Reference Model for An Open Archival Information System (OAIS) is the de facto best practice for defining these preservation processes and entities. Using OAIS as a baseline reference, the core requirements used in the digital preservation gap analysis were:

1. Ability to transfer a Submission Information Package (SIP), consisting of the digital objects to be preserved and descriptive metadata, from Producers to archival storage
2. Verification of successful transfer (i.e. no data corruption/loss during transfer from one system and/or storage media to another)
3. File format identification and validation
4. Extraction of technical metadata
5. Implementation of preservation plans (such as normalization to preservation-friendly file formats)
6. Generation and capture of preservation metadata using best practice standards, namely PREMIS and METS

7. Packaging of the Archival Information Package (AIP) consisting of the original objects, normalized objects and descriptive, preservation and technical metadata
8. Secure storage and backup of the AIP
9. Periodic integrity checking of the AIP and restoration from backup if needed.[1]
10. Generation of a Dissemination Information Package (DIP) consisting of access copies of the digital objects and descriptive metadata, linked back to the stored AIP
11. Upload of the DIP into a web-based access system
12. Management of changes to the descriptive and preservation metadata over time
13. Ongoing monitoring of preservation risks and implementation of updated preservation plans

The Archivematica installation and UBC Library systems' gap analysis work resulted in a draft high-level architecture of integrated digital preservation services for UBC Library systems. A diagram of this architecture is provided in **Appendix 1**.

2.1. Establishing a trusted digital repository (TDR)

Since the goal of a preservation program at UBC Library will be to meet the requirements of a Trusted Digital Repository (TDR), Library staff should be mindful of the requirements for a TDR as laid out in Trustworthy Repositories Audit and Certification Checklist (TRAC).[2] A self-assessment based on TRAC would help to expose any weaknesses or gaps in the program and would assist with planning as UBC Library continues to develop its digital preservation capability. TRAC is divided into three main areas, as follows:

1. **Organizational infrastructure**: governance and organizational viability; organizational structure and staffing; procedural accountability and policy framework; financial sustainability; and contracts, licenses and liabilities.

2. **Digital object management**: as per OAIS requirements: ingest, including acquisition of content and creation of the archivable package; preservation planning; archival storage and preservation/maintenance of AIPs; information management; and access management.

3. **Technologies, technical infrastructure and security**: system infrastructure; appropriate technologies; and security.

TRAC is laid out as a series of questions requiring a summary of analysis along with documentary evidence and comments. A full TRAC audit intended for an external audience (such as a certification body) would be a lengthy and resource-intensive exercise; however, an informal self-assessment in 2012 based on TRAC for internal planning purposes will provide valuable information for continued

[1] UBC Library submits a limited number of e-journals to COPPUL, a LOCKSS-based collaborative storage network managed by a consortium of Canadian University Libraries, and is also considering submitting cIRcle and CONTENTdm content This would provide geo-remote storage and automatic restoration of backups in the event of data corruption, but not the full suite of digital preservation functions required by OAIS.

[2] Available at http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories.

development of UBC Library's digital preservation program. Once the program is in production, preparations could be made for an external TRAC audit in 2013 or later if desired.[3]

3. UBC Library and Archivematica

The open-source Archivematica digital preservation system is one of the key components of UBC Library's persistent digital collections plan. Vancouver-based Artefactual Systems is leading the development of this system in direct collaboration with UBC Library, City of Vancouver Archives, Simon Fraser University Archives, Rockefeller Archive Center, International Monetary Fund Archives and UNESCO Memory of the World. As well, a growing number of early adopters worldwide are integrating and testing the system on their own initiative.[4]

Archivematica is a free and open-source tool providing back-end long-term preservation of digital objects which can be linked to dissemination copies in an access system. The system provides an integrated suite of software tools that allows users to process digital objects from ingest to access in compliance with the OAIS functional model. Archivematica implements a micro-services approach to digital preservation, in which the OAIS Submission, Archival and Dissemination information packages (SIPs, AIPs, DIPs) move through a series of services using a Unix pipeline. The services, which are provided by a combination of Archivematica application scripts and one or more of the free and open-source tools bundled into the system, can be distributed across multiple processing clusters for highly scalable configurations. The system also implements normalization plans to convert files to preferred preservation formats for those cases where a reliable, tested open-source conversion tool is available. For access, Archivematica comes bundled with an open-source web-based access system called ICA-AtoM, also developed by Artefactual Systems.

The Archivematica project implements an agile software development methodology which is focused on rapid, iterative release cycles, each of which improves upon the system's requirements, software, documentation and development infrastructure. This is accomplished by facilitating close collaboration between the developers and end users of the system. As an Artefactual client project, the UBC Library persistent digital collections plan is now playing a key role in defining and prioritizing system requirements and software features for the Archivematica community as a whole. Thus far, the UBC Library has contributed to the development of the 0.7.1-alpha and 0.8-alpha releases of the Archivematica system.

The Archivematica improvements and digital preservation knowledge being generated by the UBC Library project can be freely re-used by other institutions under the terms of the AGPL3 and Creative Commons open-source licenses. Therefore, the financial and technical resources invested in this project are being leveraged well beyond UBC Library's internal strategy to the benefit of the library and archives community at large, which is helping to build UBC Library's reputation as a community leader in this domain.[5]

---

[3] An external audit would likely be carried out by the Center for Library Research or a similar organization. See http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying.

[4] See http://archivematica.org.

[5] UBC Library's contribution to the Archivematica project is recognized on the Archivematica homepage at http://www.archivematica.org (under "About", bottom of page) and is also highlighted at public dissemination events such as Access 2011 (http://access2011.library.ubc.ca/) and the upcoming UNESCO Memory of the World Conference in Vancouver (http://www.unesco.org/new/en/communication-and-information/events/calendar-of-events/events-websites/the-memory-of-the-world-in-the-digital-age-digitization-and-preservation/).

4. UBC Library pilot projects

In order to refine requirements for a digital preservation program, UBC Library conducted a number of pilot projects using both Archivematica and ICA-AtoM. The pilot projects involved private records held by Rare Books and Special Collections, faculty and student materials in the cIRcle institutional repository, digitized Library holdings managed in CONTENTdm, UBC records held by the University Archives, and UBC websites. The pilot projects allowed the project team to identify feature gaps in Archivematica and ICA-AtoM, which are being addressed during the 2012 software development cycles. Of particular note are integration with both DSpace and CONTENTdm and the development of new workflows and functionalities for handling acquisitions of large bodies of digital objects.

5. Steps for 2012 persistent digital collections project

The work completed in 2011 laid the foundation for building an integrated production-ready digital preservation program at UBC Library by the end of 2012. Artefactual Systems will be working with UBC Library over the course of the year to accomplish the following:

1. Complete Archivematica/DSpace integration

2. Complete Archivematica/CONTENTdm integration

3. Finalize digital object acquisition processes, including external media processing tools and procedures

4. Complete access system integration and a unified discovery interface for UBC Library digital collections

5. Conduct LOCKSS/COPPUL pilot, in which objects ingested into Archivematica are uploaded to COPPUL for storage

6. Conduct eScience research data pilot project

7. Conduct ISO 16363 TRAC self-audit & remediation as preparation for possible external audit in 2013

Appendix 1: UBC Library digital preservation architecture