UBC Library's Digital Preservation Strategy

Bronwen Sprout, Digital Initiatives Coordinator
Sarah Romkey, Rare Books and Special Collections Archivist
with input from Artefactual Systems

UBC Library
April 28, 2014
overview

• What is digital preservation?
• Our approach to digital preservation
• Working with Artefactual
• Archivematica
• Born digital archives
• Archivematica and integration with other systems (AtoM, CONTENTdm, cIRcle)
what is digital preservation?

• active management of digital information over time to ensure its accessibility
• consists of policies, strategies, and actions

<table>
<thead>
<tr>
<th>Digital preservation requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>accessibility</td>
</tr>
<tr>
<td>usability</td>
</tr>
<tr>
<td>authenticity</td>
</tr>
</tbody>
</table>
our approach

• Develop a comprehensive preservation system for UBC Library
• Work with Artefactual
• Use Archivematica as a tool to apply OAIS-compliant preservation processes AND integrate it with existing systems used to manage digital objects
• Build internal technical and staff capacity
working with artefactual

• 2011 - present:
  – gap analysis
  – identification of pilot projects
  – early testing of archivematica
  – legacy media from RBSC
  – procedures for born-digital material
  – website archiving
  – more archivematica testing
  – CONTENTdm and DSpace integration development and testing
archivematica

• “a free and open source digital preservation system that is designed to maintain standards-based, long term access to collections of digital objects” http://www.archivematica.org

• micro-services provide integrated suite of software tools in compliance with ISO-OAIS model
## Archivematica Dashboard

### Submission Information Package

<table>
<thead>
<tr>
<th>UUID</th>
<th>Ingest start time</th>
</tr>
</thead>
<tbody>
<tr>
<td>bbb6c198-12e3-439e-9c8b-0e99823354ed</td>
<td>2012-03-19 10:08</td>
</tr>
<tr>
<td>mba931be-ed2a-4a31-92e5-8a43c0c776403f</td>
<td>2012-03-17 15:26</td>
</tr>
</tbody>
</table>

### Jobs

- **Micro-service:** Uploaded DP
- **Micro-service:** Store DP
- **Micro-service:** Prepare DP
- **Micro-service:** Process submission documentation
- **Micro-service:** Normalize submission documentation to presentation format
- **Micro-service:** Identify Files ByExtension
- **Micro-service:** Characterize and extract metadata on submission documentation
- **Micro-service:** Scan for viruses in submission documentation
- **Micro-service:** Sanitize file and directory names in submission documentation
- **Micro-service:** Extract packages in submission documentation
- **Micro-service:** Assign checksums and file sizes to submissionDocumentation
- **Micro-service:** Assign file UUIDs to submission documentation
- **Micro-service:** Move submission documentation into objects directory
- **Micro-service:** Check for submission documentation
- **Micro-service:** Copy transfer submission documentation
- **Micro-service:** Move to processing directory
- **Micro-service:** Normalize
- **Micro-service:** Clean up names
- **Micro-service:** Remove cache files
- **Micro-service:** Include default SIP processingMCP.xml
- **Micro-service:** Rename SIP directory with SIP UUID
- **Micro-service:** Verify transfer compliance
- **Micro-service:** Verify SIP compliance

### Dependencies

- **Digitized**
- **Ingest**
what is “born digital?”

• Quick survey:
  • Do you take digital photographs?
  • Do you print some of your digital photographs?
  • Do you print ALL of your digital photographs?
RBSC/UA born-digital acquisition workflow
legacy media

- Floppy and optical disks acquired with primarily analogue collections.
- Hardware set-up and technological challenges
- Intellectual challenges:
  - Arrangement and description
  - Selection of material
  - Rights
a new acquisition

- "First" born-digital acquisition
- Donor questionnaire
- Physical transfer of files
  - Hardware and software
  - Workflow documentation
- Donor gift agreement
- First AIPs stored April 2013; still in processing
Browse archival storage
Total size: 149.76 MB  Total files: 61 Indexed

<table>
<thead>
<tr>
<th>AIP</th>
<th>Size</th>
<th>UUID</th>
<th>Date stored</th>
</tr>
</thead>
</table>
# Preservation Levels

<table>
<thead>
<tr>
<th>Level of Preservation</th>
<th>Use Case</th>
<th>Minimum Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Imaging files from working hard drives,</td>
<td>Write blocker, file synchronization software (e.g.</td>
</tr>
<tr>
<td></td>
<td>directly from creator</td>
<td>Grsync)</td>
</tr>
<tr>
<td>Level 2</td>
<td>Imaging/copying files from external media</td>
<td>File synchronization software (e.g. Grsync), write</td>
</tr>
<tr>
<td></td>
<td></td>
<td>blocker if possible</td>
</tr>
<tr>
<td>Level 3</td>
<td>Preserving external media only</td>
<td>Hardware necessary to open media (floppy drive,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>optical drive, etc)</td>
</tr>
</tbody>
</table>
AtoM

• Master files and PDFs uploaded to Archivematica for display in AtoM
CONTENTdm

- Master files uploaded to Archivematica
- Archivematica produces access versions and pushes to CONTENTdm
clRcle (DSpace)

- Archivematica receives submissions from DSpace
- Also have Archivematica to DSpace workflow
# BIRS math videos

## Browse archival storage

Total size: 17241.27 MB  Total files: 27 indexed

<table>
<thead>
<tr>
<th>AIP</th>
<th>Size</th>
<th>UUID</th>
<th>Date stored</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIRS-VIDEO-12ss131-2636</td>
<td>3703.80 MB</td>
<td>72e7e87c-cdb3-97e4-83ab-96e57555261e</td>
<td>2014-04-24 T11:59</td>
<td>Stored</td>
</tr>
<tr>
<td>BIRS-VIDEO-12ss131-2637</td>
<td>4262.58 MB</td>
<td>94140d17-624f-f7ea-37ba-5c0980f2d120c</td>
<td>2014-04-25 T11:19</td>
<td>Stored</td>
</tr>
<tr>
<td>BIRS-VIDEO-12ss131-2650</td>
<td>3979.05 MB</td>
<td>09b1b9b2-8a99-4c3e-a9aa-848f0e8b3b0d</td>
<td>2014-04-24 T14:36</td>
<td>Stored</td>
</tr>
<tr>
<td>BIRS-VIDEO-13w856-6655</td>
<td>5295.85 MB</td>
<td>573c70f9-6016-8bfb-8c72-af9c04b98c83</td>
<td>2014-04-22 T18:59</td>
<td>Stored</td>
</tr>
</tbody>
</table>
Public access provided through a unified discovery interface integrating Archivematica index and UBC Library document viewer application.

- UBC faculty & students
- CHRP contributors
- Other

DSpace
- SIP (automated OAI-PMH updates)

CONTENTdm
- DIP (reconfigured for upload to CONTENTdm)

ICAAtom
- DIP

Archivematica
- External media acquisitions
- Born-digital network transfers
- Web archiving harvests

Archivematica processing workstations
- Digitization workstations
- UBC SAN & backup
- LOCKSS server
Next steps and future work

• TRAC (Trusted Repository Audit Checklist) self-audit
• Policy development
• Research data
• Website archiving