Preservation of Digital Theses at SFU

Jordan, Mark

http://hdl.handle.net/1880/49063

Downloaded from PRISM Repository, University of Calgary
Preservation of Digital Theses at SFU

Mark Jordan, Head, Library Systems, SFU

COPPUL Digital Preservation Workshop, April 20, 2012
• Policy
  – Theses at SFU go all digital January 2013
  – Digital preservation and data curation are part of the Library's current Strategic Plan

• Technical
  – Theses Submission and Management System
    • BagIt Drupal module
  – Summit Institutional Repository
    • Automated ingestion
  – LOCKSS

Context
• Create a robust, standards-based, and sustainable digital preservation program for SFU's theses and associated content
• Integrate preservation services into existing workflows
• Review platform options
• Use theses as a test case for more general preservation activities
  – Monitor required resources, costs, SWOTs
  – Refine policies and technology
• Develop comprehensive and integrated digital preservation services for selected assets
• Migrate to standardized digital preservation platforms and services (e.g., Archivematica)
  – i.e., within current Strategic Planning timeline
<table>
<thead>
<tr>
<th>Pros</th>
<th>Archivematica</th>
<th>Locally developed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete digital preservation platform</td>
<td>Services tuned specifically to theses</td>
</tr>
<tr>
<td></td>
<td>Robust support and user community</td>
<td>Can be integrated into existing workflows and tools</td>
</tr>
<tr>
<td></td>
<td>Can be ready prior to Jan. 1/13</td>
<td></td>
</tr>
<tr>
<td>Cons</td>
<td>End-user interaction required</td>
<td>Will require development work</td>
</tr>
<tr>
<td></td>
<td>Will require development work to integrate fully with existing workflows</td>
<td>Is not a complete digital preservation platform</td>
</tr>
<tr>
<td></td>
<td>Lack of support for LOCKSS to store AIPs</td>
<td>Lack of support and user community</td>
</tr>
</tbody>
</table>
Technical overview

Preservation of Digital Theses at SFU
• 1 representation (thesis as intellectual work) per Bag
• 1+ files per Bag
  – Thesis PDF
  – Supplemental files, e.g. data, video, other PDFs
  – METS file containing
    • ETDMS metadata file
    • Technical metadata (FIDO, FITS output) for each thesis and supplemental file
    • structMap indicating relationships of supplemental files

• 1+ versions per Bag

PREMIS Object model
• Dark archive
  – Store it and forget it
  – AU status reports
• PREMIS events
  – Will record capture, ingest, normalization, validation
  – LOCKSS makes fixity check, replication redundant
• Embargos and other restrictions
  – Distributed preservation raises access issues
• Build, borrow, buy, collaborate?
• What constitutes “handing off the baton”?
  – 15% of first 500 files submitted to the Theses Management System failed JHOVE validation
• Policy and technical decisions around supplemental files
  – Video, raw data, etc.
  – Offline versions of these, e.g. video on DVD
• Optimizing workflow integration

Issues to resolve