

2012-04-20

Preservation of Digital Theses at SFU

Jordan, Mark

Jordan, M. "Preservation of Digital Theses at SFU". Presented at Digitization and Digital Preservation: Strategies for COPPUL Libraries, Friday April 20, 2012, University of Calgary Taylor Family Digital Library.

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

Downloaded from PRISM Repository, University of Calgary

The SFU logo consists of the letters 'SFU' in a white, bold, sans-serif font, centered within a solid red square.

SIMON FRASER UNIVERSITY
ENGAGING THE WORLD



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

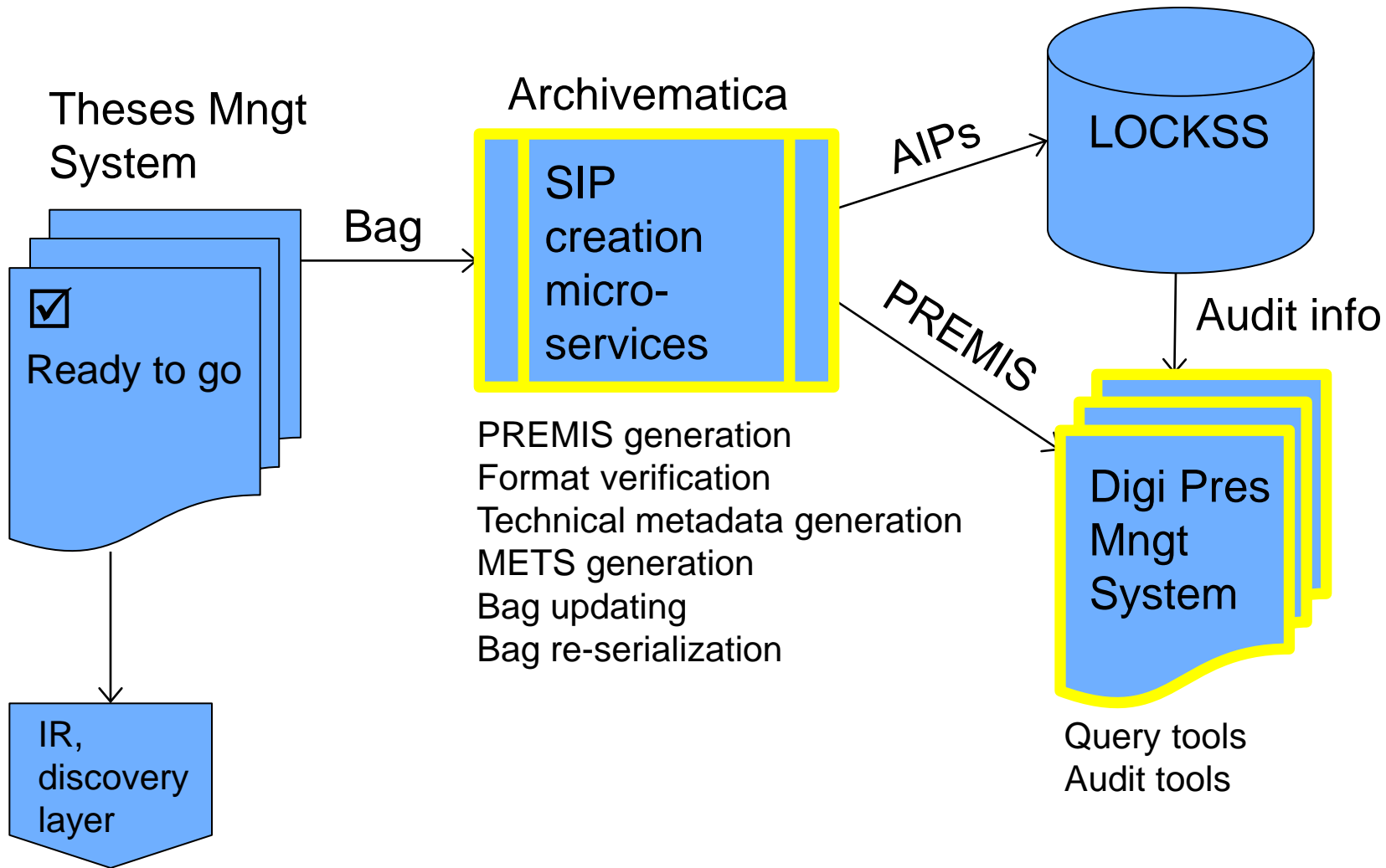
Short-term goals

- 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

Long-term goals

	Archivematica	Locally developed
Pros	Complete digital preservation platform	Services tuned specifically to theses
	Robust support and user community	Can be integrated into existing workflows and tools
		Can be ready prior to Jan. 1/13
Cons	End-user interaction required	Will require development work
	Will require development work to integrate fully with existing workflows	Is not a complete digital preservation platform
	Lack of support for LOCKSS to store AIPs	Lack of support and user community

Platform choices



Technical overview

- 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

LOCKSS

- 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