

A Persistent Digital Preservation Strategy for UBC Library

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

UBC Library
September 26, 2012



a place of mind

THE UNIVERSITY OF BRITISH COLUMBIA

introduction

- Pilot and early stages of digital preservation program at UBC Library
- Supports Library's strategic focus
- New digital initiatives unit
- Opportunity to develop comprehensive approach



overview

- Working with Artefactual Systems
- 2011: pilot projects and testing of Archivemata
- 2012: full implementation of Archivemata and production services



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

The screenshot shows the Archivematica Dashboard in a Mozilla Firefox browser window. The address bar displays the URL `the.wood.library.abc.ca/ingest/`. The dashboard header includes the Archivematica logo and navigation tabs: Transfer, Ingest, Archival storage, Preservation planning, Access, Administration, and DiAdmin01. The user is logged in as 'Connected'.

The main content area displays a table of Submission Information Packages (SIPs). The table has three columns: Submission Information Package, UUID, and Ingest start time. The first SIP is for 'Evelyn' with UUID `b3bb6c196-12e3-4384-8cb2-0e9982336e6d` and an ingest start time of 2012-09-19 10:08. The second SIP is for 'Digitized' with UUID `eba931be-edac-4a31-92a5-843cc776403f` and an ingest start time of 2012-09-17 15:25.

Below the table, a detailed log of jobs and micro-services is shown for the 'Evelyn' SIP. The log includes the following entries:

- Micro-service: Upload DIP
- Micro-service: Store AIP
- Job: Store AIP (review) [?] - Awaiting decision
- Job: Move to the store AIP approval directory - Completed successfully
- Micro-service: Prepare AIP
- Micro-service: Prepare DIP
- Micro-service: Process submission documentation
 - Job: Normalize submission documentation to preservation format - Completed successfully
 - Job: Normalize submission documentation to preservation format - Completed successfully
 - Job: Identify Files ByExtension - Completed successfully
 - Job: Characterize and extract metadata on submission documentation - Completed successfully
 - Job: Scan for viruses in submission documentation - Completed successfully
 - Job: Sanitize file and directory names in submission documentation - Completed successfully
 - Job: Extract packages in submission documentation - Completed successfully
 - Job: Assign checksums and file sizes to submissionDocumentation - Completed successfully
 - Job: Assign file UUIDs to submission documentation - Completed successfully
 - Job: Move submission documentation into objects directory - Completed successfully
 - Job: Check for submission documentation - Completed successfully
 - Job: Copy transfer submission documentation - Completed successfully
 - Job: Move to processing directory - Completed successfully
- 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

The 'Digitized' SIP log shows:

- Micro-service: Store AIP
- Job: Store AIP (review) [?] - Awaiting decision
- Job: Move to the store AIP approval directory - Completed successfully
- Micro-service: Prepare AIP

gap analysis

- Born digital material in Rare Books and Special Collections and University Archives
- Locally created digital collections
 - cIRcle digital repository (DSpace)
 - Digital collections (CONTENTdm)
- Among other things...



a place of mind

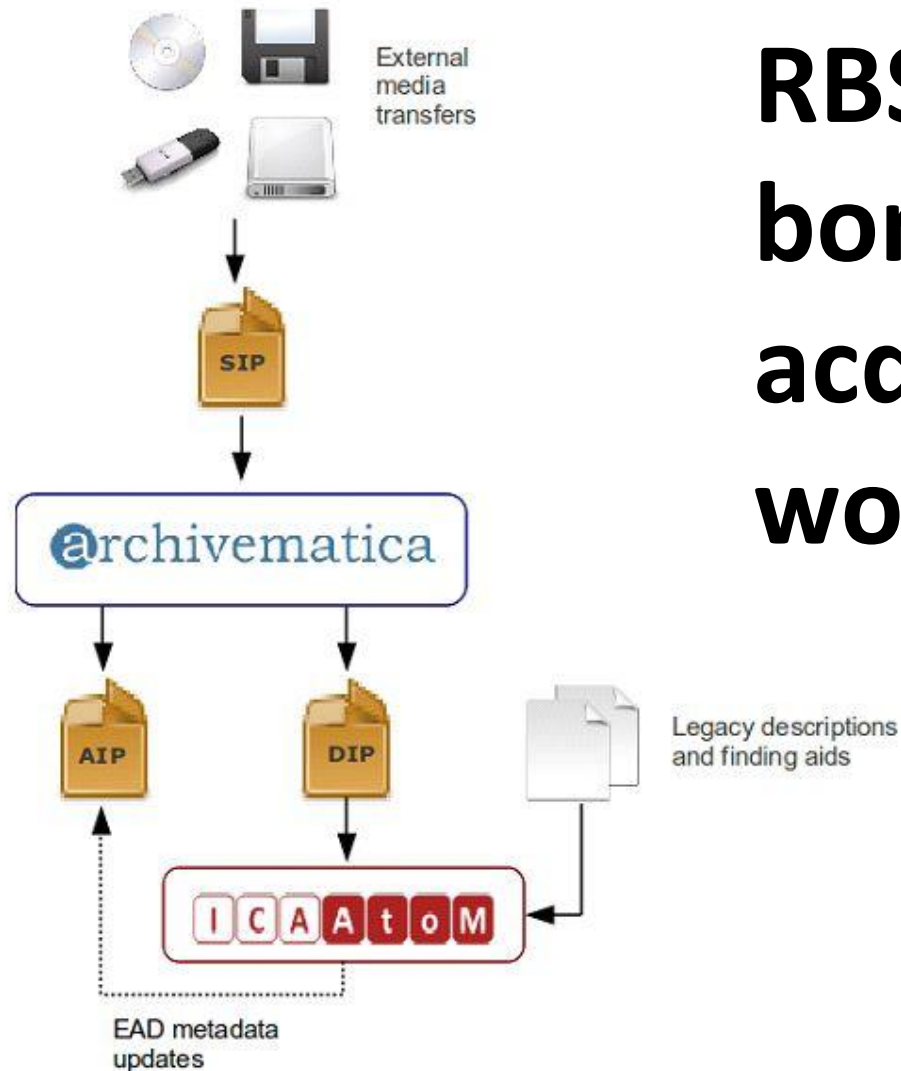
THE UNIVERSITY OF BRITISH COLUMBIA

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



RBSC/UA born-digital acquisition workflow





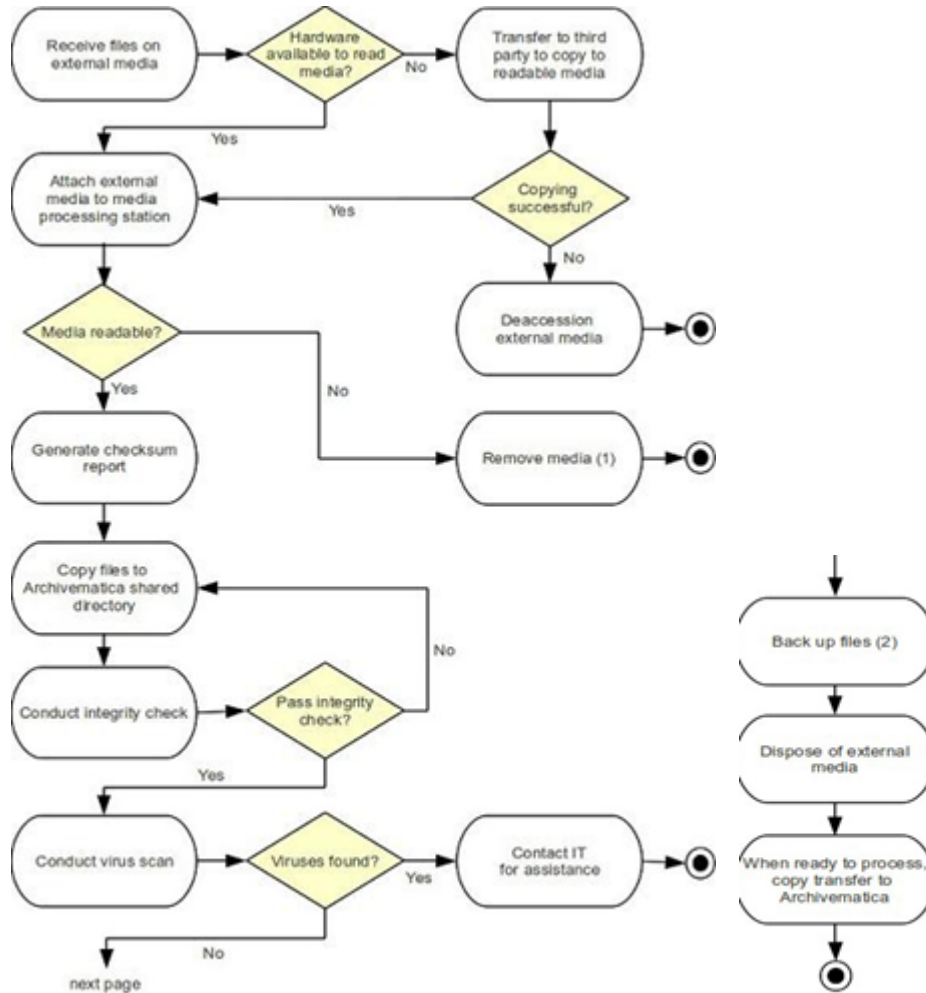
The Library



a place of mind

THE UNIVERSITY OF BRITISH COLUMBIA

external media station workflow



a new acquisition

- "First" born-digital acquisition
- Donor questionnaire
- Physical transfer of files
 - Hardware and software
 - Levels of preservation
 - Workflow documentation
- Donor gift agreement



preservation levels

Level of Preservation	Use case	Minimum tools
Level 1	Imaging files from working hard drives, directly from creator	Write blocker, file synchronization software (e.g. Grsync)
Level 2	Imaging/copying files from external media	File synchronization software (e.g. Grsync), write blocker if possible
Level 3	Preserving external media only	Hardware necessary to open media (floppy drive, optical drive, etc)



a place of mind

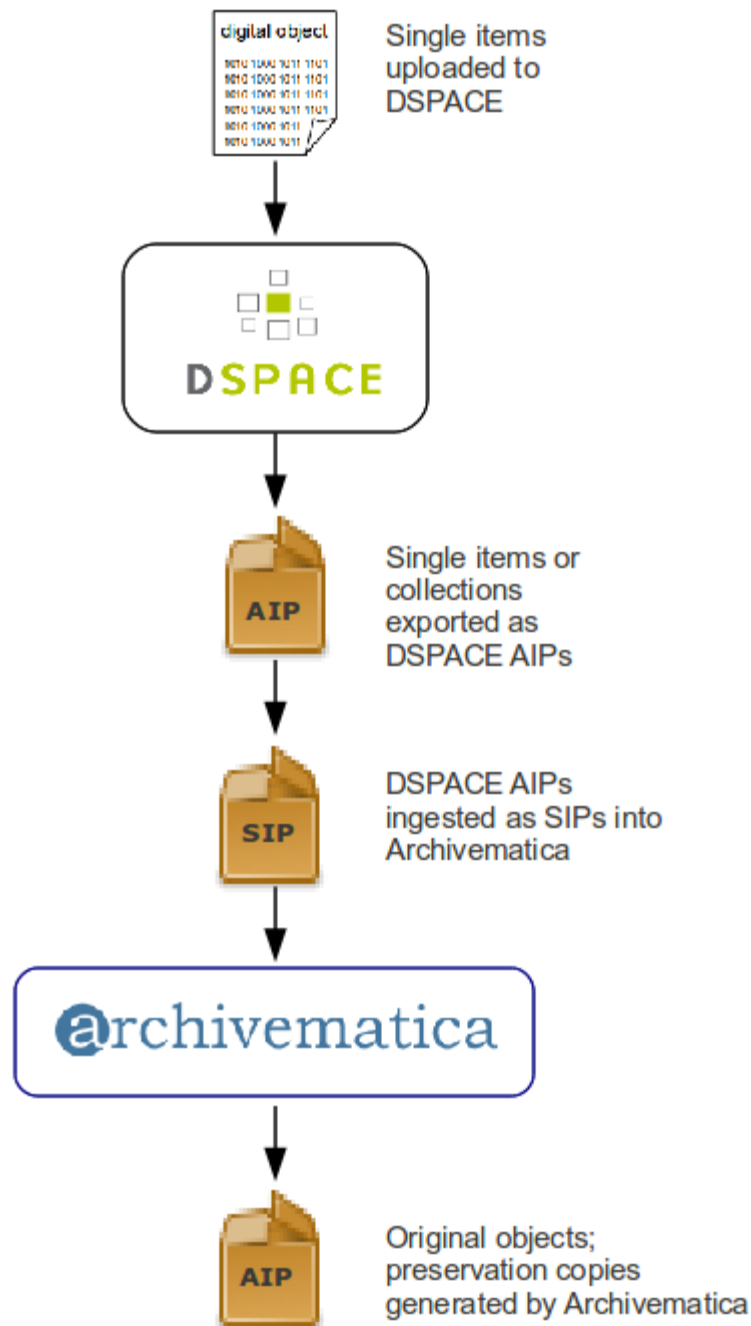
THE UNIVERSITY OF BRITISH COLUMBIA

some additional
programming required



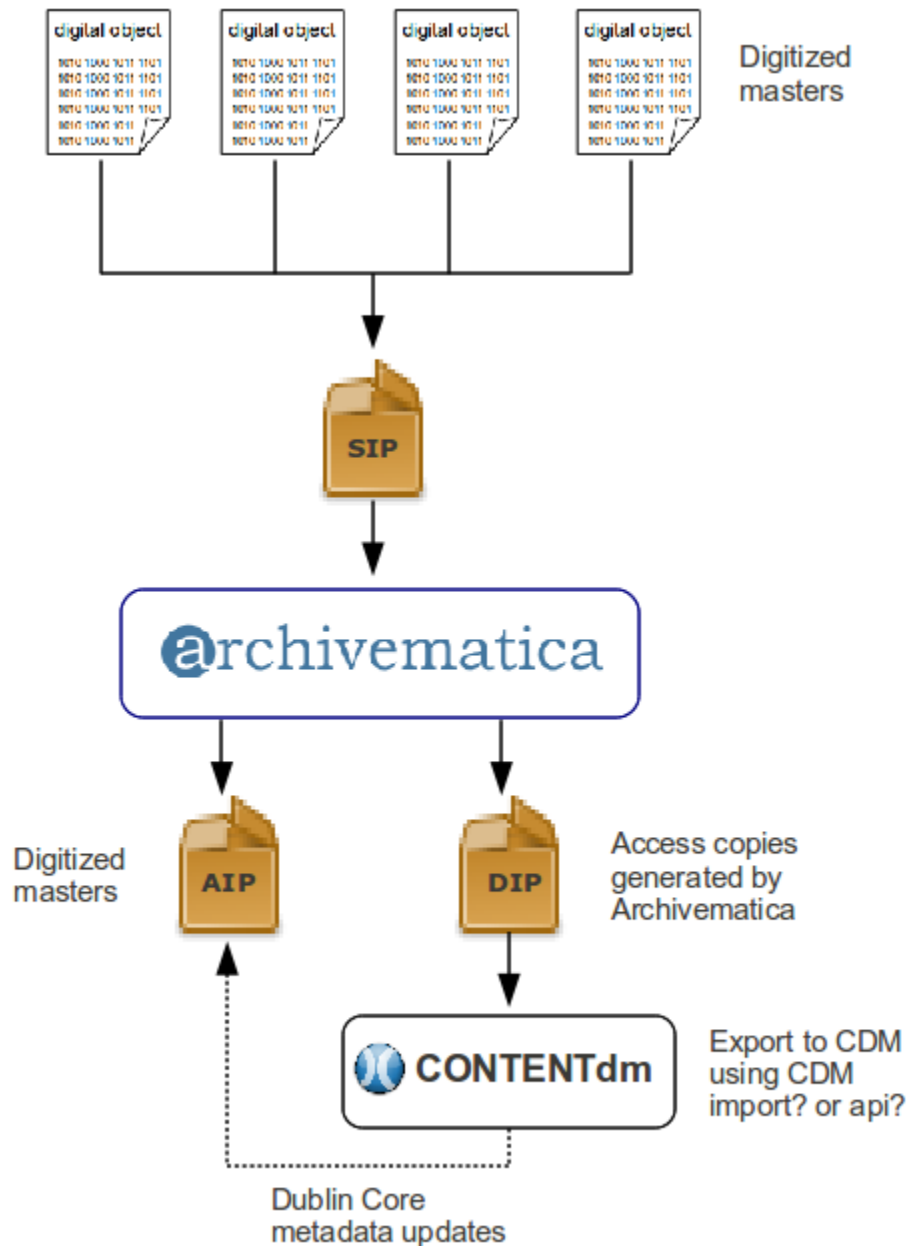
a place of mind

THE UNIVERSITY OF BRITISH COLUMBIA



cIRcle

- Archivematica receives submissions from DSpace



CONTENTdm

- Master files uploaded to Archivematica
- Archivematica produces access versions

next steps

- Production services by end of 2012
 - Complete the Archivematica/DSpace integration.
 - Complete the Archivematica/CONTENTdm integration.
 - Complete Archivematica/ICA-AtoM integration and ingest new acquisition.
 - Web-archiving
- Research data
- LOCKSS PLN pilot
- Develop a unified discovery interface for UBC Library digital collections (Archivematica index)
- Prepare to undertake TDR self-audit



conclusion

- Establishing comprehensive digital preservation program is complex - many moving parts
- CONTENTdm and DSpace integrations show promise - for us, for other institutions, for future integrations with Archivematica

