

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



a place of mind

THE UNIVERSITY OF BRITISH COLUMBIA

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

Digital preservation requirements	
accessibility	discover retrieve
usability	render contextualize re-use
authenticity	identify trust

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

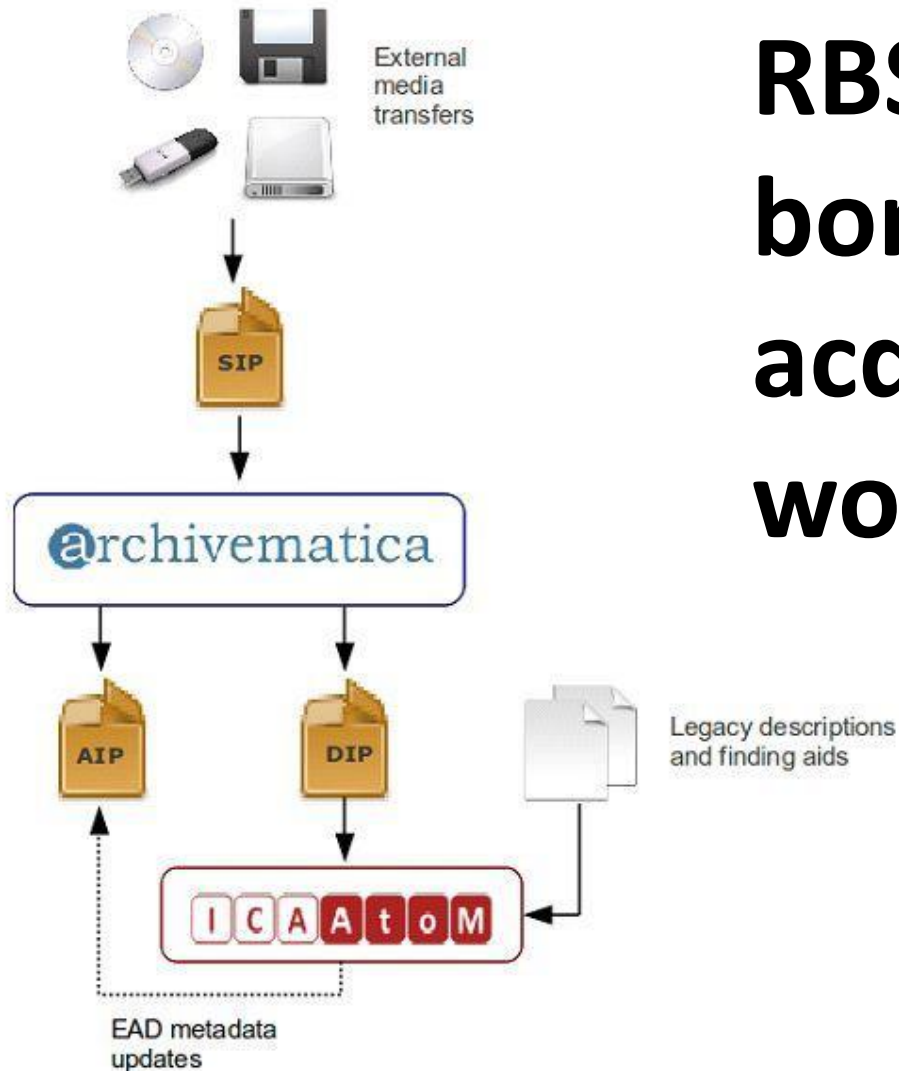
The screenshot shows the Archivematica Dashboard in a Mozilla Firefox browser window. The dashboard displays submission information for two packages: 'Evelyn' and 'Digitized'. The 'Evelyn' package is currently 'Awaiting decision', while the 'Digitized' package is 'Completed successfully'.

Submission Information Package	UUID	Ingest start time
Evelyn	bbb6c196-12e3-4304-8cb2-0e9982336e6d	2012-09-19 10:08
+ 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		
Digitized	ebs931be-edac-4a31-92a5-843cc776403f	2012-09-17 15:25
+ Micro-service: Store AIP		
Job: Store AIP (review) [?]		Awaiting decision
Job: Move to the store AIP approval directory		Completed successfully
+ Micro-service: Prepare AIP		

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



Any Keyword Search archival storage Show files?

Add New

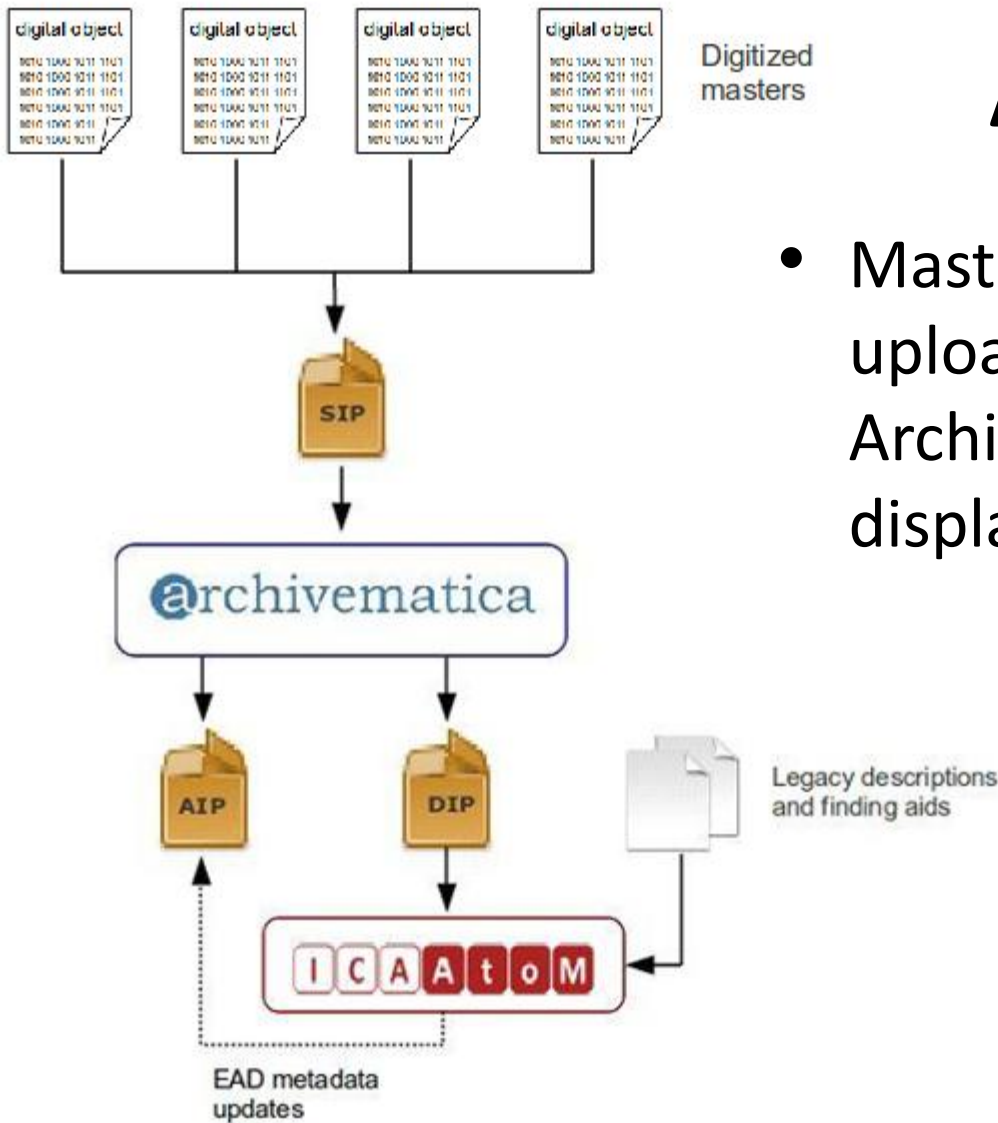
Browse archival storage

Total size: 149.76 MB Total files: 61 indexed

AIP	Size	UUID	Date stored
051207AHB_mvGuemes	149.76 MB	6da2d555-6710-4b58-8641-046101970970	2013-05-30 19:28

preservation levels

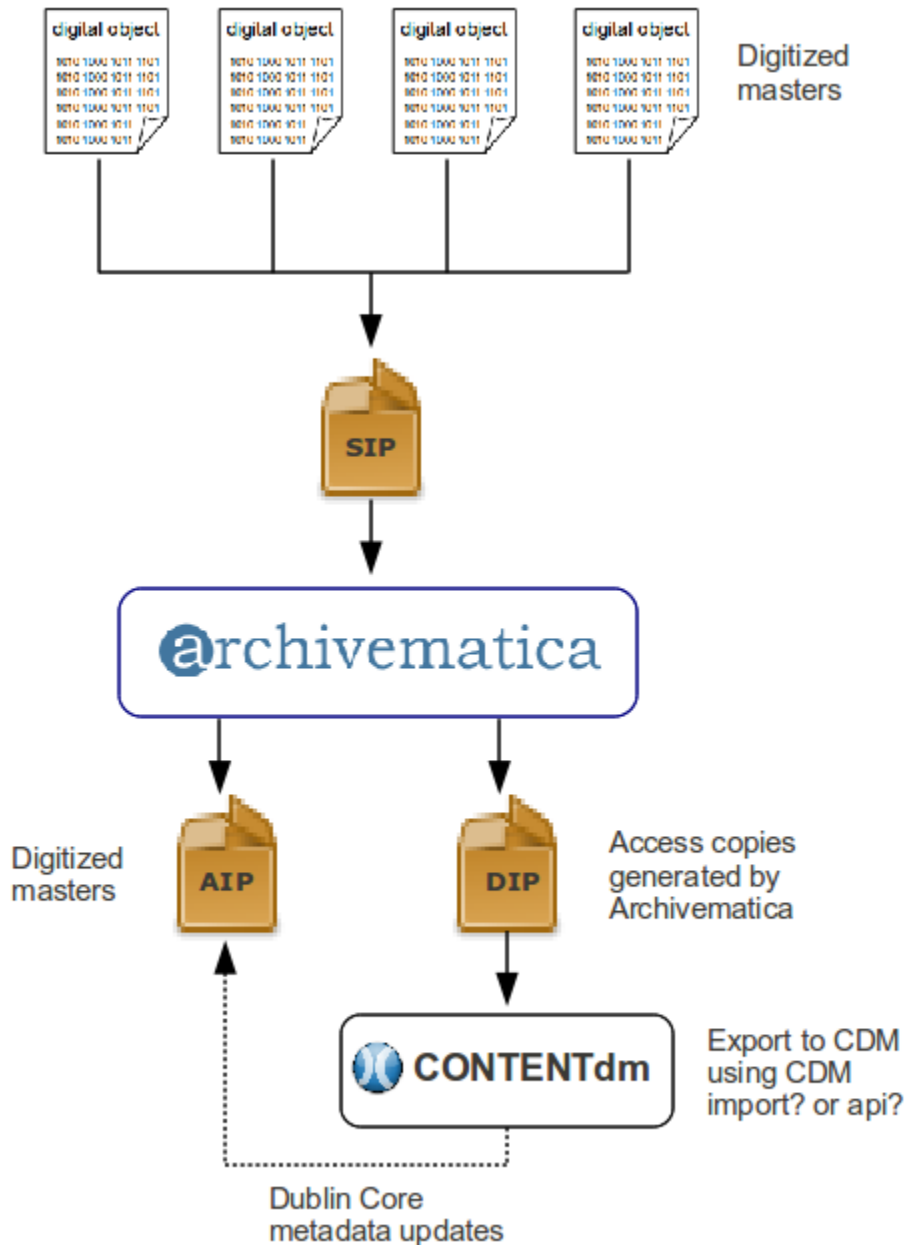
Level of Preservation	Use case	Minimum tools
Level 1	Imaging files from working hard drives, directly from creator	Write blocker, file synchronization software (e.g. Grsync)
Level 2	Imaging/copying files from external media	File synchronization software (e.g. Grsync), write blocker if possible
Level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



Digitized masters

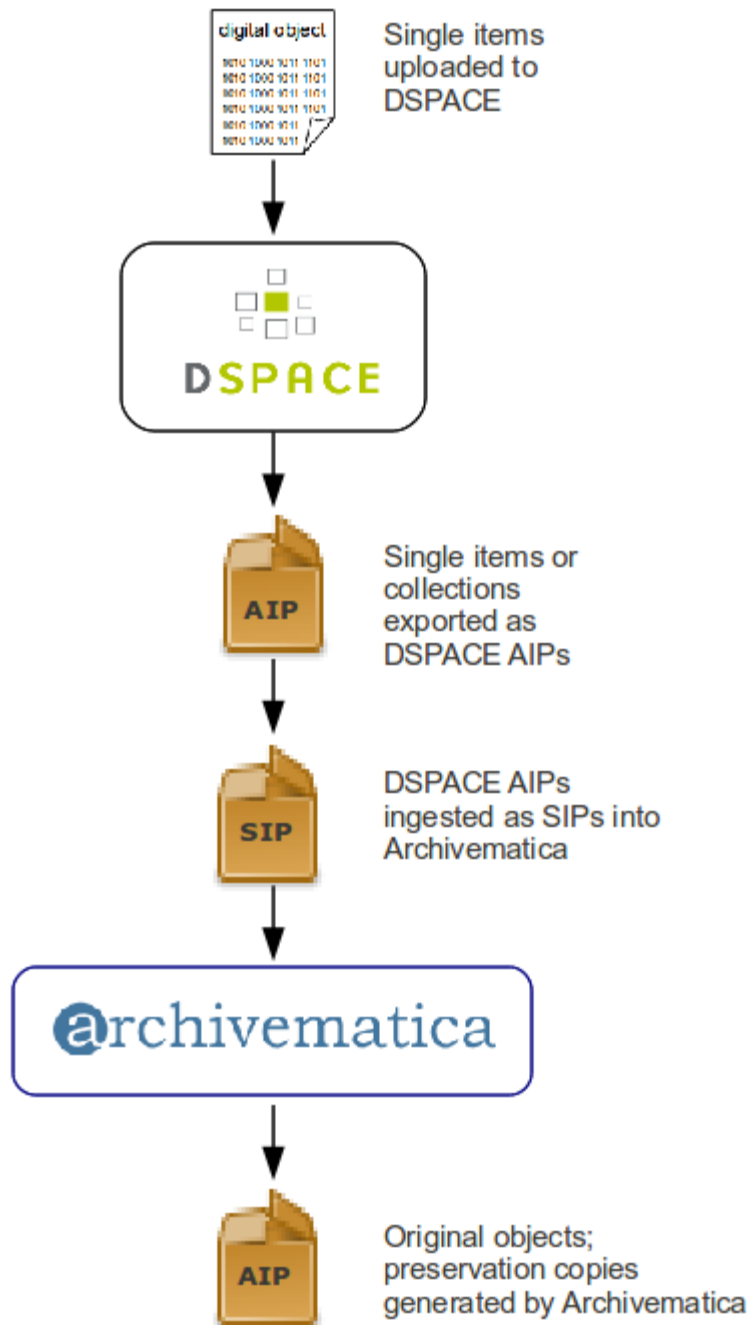
AtoM

- 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



cIRcle (DSpace)

- Archivematica receives submissions from DSpace
- Also have Archivematica to DSpace workflow

BIRS math videos

Any ▾ Keyword ▾ Search archival storage Show files? Show AICs?

[Add New](#)

Browse archival storage

Total size: 17241.27 MB Total files: 27 indexed

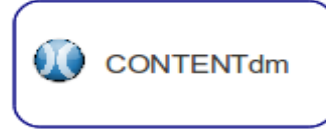
AIP	Size	UUID	Date stored	Status	
BIRS-VIDEO-12ss131-2636	3703.80 MB	72e7e87c-ccb3-47a4-83ab-99e87555261e	2014-04-24 11:59	Stored	
BIRS-VIDEO-12ss131-2637	4262.58 MB	94140d17-624f-4f7a-87ba-5c0480fd120e	2014-04-25 11:19	Stored	
BIRS-VIDEO-12ss131-2650	3979.05 MB	09b1b9b2-ba69-4c3c-a9aa-c848f0ab3b0d	2014-04-24 14:36	Stored	
BIRS-VIDEO-13w5058-6655	5295.85 MB	574c70f9-4046-4bfb-8e72-af3e04b9fe82	2014-04-22 18:59	Stored	



Public access provided through a unified discovery interface integrating Archivemática index and UBC Library document viewer application



- UBC faculty & students
- CHRP contributors
- Other



automated OAI-PMH updates



DIP reconfigured for upload to CONTENTdm



UBC librarians & archivists
Digital Initiatives,
RBSC, Archives



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



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