

ORGANIZATIONAL CONSIDERATIONS FOR IMPLEMENTING ARCHIVEMATICA

Kari R. Smith, Digital Archivist
Institute Archives and Special Collections
Massachusetts Institute of Technology

 **archivematica.**[®]

**MIT**Libraries

... BUT CAN WE DO IT?

MIT Libraries has been an evaluation partner with Archivematica since 2012

- Institute Archives and Special Collections
- Digital Preservation Services
- Digital Sustainability Lab

Join me in Bellflower AB

- Overview of *Archivematica* functions and components
 - *Storage Service*
 - *FPR* (Format Policy Repository)
 - Base application
 - Pipelines
 - Dependencies
 - Workflows
- Using the *Archivematica* Google group (listserv)
- Having an SLA with Artefactual

- **Lessons learned at MIT Libraries**
 - System and storage requirements
 - Developing Pipelines
 - Dependencies
 - Developing Workflows

- **Understanding our content**
 - FPR
 - Packaging
 - Deciding what a SIP should contain

- **Managing our AIPs and DIPs**
 - Location pointers
 - Fixity
 - Integrating metadata into collections management system (Archivists' Toolkit)

Archivematica Storage Service

Welcome to the Archivematica Storage Service! Here you can configure the storage accessible to your Archivematica pipelines.

Pipelines

Each pipeline is an Archivematica installation, and has a unique ID that identifies the server and all associated clients. Locations are assigned to pipelines and can only be accessed by that pipeline.

Spaces

Spaces contain Locations as well as information on how to access them. All information related to the access protocol, such as remote hostname and location of the export are stored in the Space.

Locations

Locations have a purpose, such as AIP storage or transfer source, and are where packages are stored. Each Location is associated with a pipeline and can only be accessed by that pipeline.

Packages

Packages are Transfers, SIPs and AIPs uploaded to a Location managed by the storage service.

Administration

Configure default settings and manage users.

**STORAGE
SERVICE**
FPR
BASE
APPLICATION
PIPELINES
DEPENDENCIES
WORKFLOWS
GOOGLE GROUP
SLA

PRESERVATION PLANNING

- ✓ Identification
- ✓ FPR
- ✓ Characterization
- ✓ Event Detail
- ✓ Extraction
- ✓ Normalization
- ✓ Transcription
- ✓ Validation
- ✓ Verification

The screenshot shows the Archivematica web interface. At the top, the logo 'archivematica' is followed by navigation tabs: 'Transfer' (with a red notification badge '3'), 'Ingest', 'Archival storage', 'Preservation planning' (the active tab), 'Access', and 'Administration'. Below the navigation, the 'Formats' section is displayed. It includes a 'Groups' button, a 'Show' dropdown menu set to '25 entries', and a search box. The main content area is a table with two columns: 'Description' and an empty column. The table lists various file formats, each with a blue background row. The categories on the left are: Identification (Tools, Rules, Commands), Format policy registry (Tools), Characterization (Rules, Commands), Event Detail (Rules, Commands), and Extraction (Rules).

	Description	
Identification	Audio	
Tools	Binary (Data)	
Rules	Binary (Executable)	
Commands	CAD	
Format policy registry	Database	
Tools	Desktop Publishing	
Characterization	Disk Image	
Rules	Email	
Commands	Font	
Event Detail	GIS	
Rules	Image (Raster)	
Commands	Image (Vector)	
Extraction	Package	
Rules	Portable Document Format	
Commands	Presentation	
Extraction	Spreadsheet	
Rules	Statistics	
Extraction	Text (Markup)	
Rules	Text (Plain)	

Standard ▾ /home ▾

Type Transfer name Accession no.

Transfer	UUID	Transfer start time	
Christer ▶ Micro-service: Create SIP from Transfer	9d824842-46fd-41b5-b206-ae296aeb78c	2016-01-12 10:32	
Job: Create SIP(s) [?]		Awaiting decision	Actions ▾
Job: Load options to create SIPs		Completed successfully	
Job: Check transfer directory for objects		Completed successfully	
▶ Micro-service: Complete transfer			
Job: Move to SIP creation directory for completed transfers		Completed successfully	
Job: Create transfer metadata XML		Completed successfully	
Job: Index transfer contents		Completed successfully	
▶ Micro-service: Examine contents			
▶ Micro-service: Validation			
▶ Micro-service: Characterize and extract metadata			
▶ Micro-service: Update METS.xml document			

STORAGE
SERVICE
FPR
BASE
APPLICATION
PIPELINES
DEPENDENCIES
WORKFLOWS
GOOGLE GROUP
SLA

- Pipelines are developed within the Storage Service
- One pipeline per instance of Archivemata
- Pipelines have their own workflows

STORAGE
SERVICE
FPR
BASE
APPLICATION
PIPELINES
DEPENDENCIES
WORKFLOWS
GOOGLE GROUP
SLA

Dependencies

- **GhostScript can cause issues!**

WORKFLOW

archivematica. Transfer ³ Ingest Archival storage Preservation planning Access Administration demo@exam

Processing configuration

General
Failures
Transfer source locations
AIP storage locations
Processing storage usage
AtoM DIP upload
Archivists Toolkit DIP upload
PREMIS agent
REST API
Version

Processing configuration

<input checked="" type="checkbox"/> Send transfer to quarantine	No
<input checked="" type="checkbox"/> Approve normalization	Yes
<input checked="" type="checkbox"/> Store AIP	Yes
<input checked="" type="checkbox"/> Transcribe files (OCR)	Yes
<input checked="" type="checkbox"/> Generate transfer structure report	Yes
<input type="checkbox"/> Remove from quarantine after	<input type="text"/> days
<input type="checkbox"/> Create SIP(s)	--Actions--
<input checked="" type="checkbox"/> Extract packages	Yes
<input type="checkbox"/> Normalize	--Actions--
<input type="checkbox"/> Reminder: add metadata if desired	--Actions--
<input checked="" type="checkbox"/> Examine contents	Skip examine contents
<input checked="" type="checkbox"/> Select file format identification command (Transfer)	Fido version 1 PUID runs Identif
<input checked="" type="checkbox"/> Select file format identification command (Ingest)	Use existing data
<input checked="" type="checkbox"/> Select file format identification command (Submission documentation &	Fido version 1 PUID runs Identif

<http://sandbox.archivematica.org/administration/processing/>

CHARACTERIZATION OF CONTENT

- Understanding our content in order to work with Digital Preservation services
- DP Services determines if the rules and results are appropriate for MIT Libraries preservation strategies
 - Preservation formats
 - Preservation metadata

Formats + number records (Z:\Accessioned, Dec 2015)

Format Name	Puid	
Null	Null	43,701
7Zip format	fmt/484	5
Acrobat PDF 1.0 - Portable Document Format	fmt/14	14
Acrobat PDF 1.2 - Portable Document Format	fmt/16	1,497
Acrobat PDF 1.3 - Portable Document Format	fmt/17	967
Acrobat PDF 1.4 - Portable Document Format	fmt/18	1,188
Acrobat PDF 1.5 - Portable Document Format	fmt/19	259
Acrobat PDF 1.6 - Portable Document Format	fmt/20	352
Acrobat PDF 1.7 - Portable Document Format	fmt/276	13
Acrobat PDF/A - Portable Document Format	fmt/354	1
Adobe Flash	fmt/505	42
	fmt/506	49

GOOGLE GROUP - LISTSERV

	Sarah Romkey	Re: [archivematica] Empty directories removed between transfer and AIP	Wed 1/13/2016 1:3...	37 KB
	Justin Simpson	Re: [archivematica] installation assistance - getting errors when installing dashboard	Mon 1/11/2016 8:5...	33 KB
	Jonathan Wilson	[archivematica] installation assistance - getting errors when installing dashboard	Mon 1/11/2016 6:3...	26 KB
	Sarah Romkey	Re: [archivematica] Re: SIP upload / Custom admin metadata / Data Management / Loc...	Mon 1/11/2016 1:0...	35 KB
	natkeeran@gmail...	[archivematica] Re: SIP upload / Custom admin metadata / Data Management / Localiza...	Mon 1/11/2016 10:...	29 KB
	Sarah Romkey	Re: [archivematica] SIP upload / Custom admin metadata / Data Management / Localiza...	Fri 1/8/2016 11:19 ...	36 KB
	Sarah Romkey	Re: [archivematica] Directory and Object Metadata Import in 1.4	Fri 1/8/2016 11:19 ...	32 KB
	lindsay.simons@a...	Re: [archivematica] Empty directories removed between transfer and AIP	Fri 1/8/2016 9:22 A...	32 KB
	DrewK	[archivematica] Directory and Object Metadata Import in 1.4	Thu 1/7/2016 3:24 ...	25 KB
	natkeeran@gmail...	[archivematica] SIP upload / Custom admin metadata / Data Management / Localization	Tue 1/5/2016 4:22 ...	26 KB
	Kari R Smith	RE: [archivematica] Processing PDFs, stalling at "Normalize for preservation"; "End Time...	Tue 12/29/2015 10:...	52 KB
	Courtney Whitmo...	Re: [archivematica] Processing PDFs, stalling at "Normalize for preservation"; "End Time...	Tue 12/29/2015 9:0...	2 MB
	Sarah Romkey	Re: [archivematica] Converting and normalization rules	Thu 12/24/2015 4:2...	34 KB
	Courtney Whitmo...	Re: [archivematica] Processing PDFs, stalling at "Normalize for preservation"; "End Time...	Tue 12/22/2015 1:2...	959 KB
	Sarah Romkey	Re: [archivematica] Converting and normalization rules	Tue 12/22/2015 1:0...	33 KB
	Mattias Djupdahl	Re: [archivematica] Adding METS metadata to validate against specific package schema	Tue 12/22/2015 2:5...	30 KB
	Sarah Romkey	Re: [archivematica] Temporarily turning off Format Policy Rules/Commands	Mon 12/21/2015 5:...	33 KB
	Sarah Romkey	Re: [archivematica] Processing PDFs, stalling at "Normalize for preservation"; "End Time...	Mon 12/21/2015 4:...	34 KB
	Pawel	Re: [archivematica] Adding / Changing Locations within Default Space	Fri 12/18/2015 11:5...	42 KB
	Courtney Whitmo...	[archivematica] Processing PDFs, stalling at "Normalize for preservation"; "End Time: Inv...	Thu 12/17/2015 9:5...	28 KB
	Sarah Romkey	Re: [archivematica] Time commitment to install, maintain, and upgrade	Tue 12/15/2015 6:3...	36 KB
	Shira Peltzman	[archivematica] Time commitment to install, maintain, and upgrade	Mon 12/14/2015 6:...	27 KB
	Sarah Romkey	Re: [archivematica] Adding / Changing Locations within Default Space	Mon 12/14/2015 4:...	41 KB
	Pawel	[archivematica] Adding / Changing Locations within Default Space	Mon 12/14/2015 1:...	32 KB
	Lauren Goodley	Re: [archivematica] Bagit basics	Fri 12/11/2015 1:36...	38 KB
	Mark Jordan	Re: [archivematica] Bagit basics	Fri 12/11/2015 1:32...	40 KB
	Mark Jordan	Re: [archivematica] Bagit basics	Fri 12/11/2015 1:29...	36 KB
	Lauren Goodley	Re: [archivematica] Bagit basics	Fri 12/11/2015 1:08...	31 KB
	Mark Jordan	Re: [archivematica] Bagit basics	Fri 12/11/2015 12:5...	30 KB
	Lauren Goodley	[archivematica] Bagit basics	Fri 12/11/2015 12:3...	25 KB
	Genevieve HK	Re: [archivematica] Support for Geospatial Data and Metadata Standards	Fri 12/11/2015 11:4...	32 KB

KRS tasks

STORAGE CONSIDERATIONS

- Transfer = content needs to be cleared out periodically
 - Backlog = accumulates until run as SIPs
 - Processing = 3 or 4 times the space of the transfers
 - AIP storage = you choose!
-
- <http://sandbox.archivematica.org/archival-storage/>
 - <http://sandbox.archivematica.org/administration/usage/>