

A Place Where You Process: An Introduction to Archivematica Workflows



Ethan Gates
Talking Tech for Advanced Topics in Preservation Studies
10/19/2017

Itinerary

1. Overview: Why Are We Here? (15 min.)
2. Explore Storage Service and Dashboard interfaces (30 min.)
3. Sample Ingests (45 min.)
4. Q&A/Mess Around Activities (30 min.)
5. Resources and Documentation

What Is Archivematica?

- a networked application/platform
- a suite of integrated open-source preservation tools
- an extensible, flexible, customizable series of micro-services to aid the ingest and transfer of digital objects to and from storage
- a practical example of OAIS in action

List of MicroServices and Tools used by Archivematica

○ Receive SIP <ul style="list-style-type: none">• verifyChecksum	
○ Review SIP <ul style="list-style-type: none">• extractPackage• assignIdentifier• parseManifest• clean Filename	EXT3, Thunar, incron, flock
○ Quarantine SIP <ul style="list-style-type: none">• lockAccess• virusCheck	UUID, Detox, Easy Extract, ClamAV
○ Appraise SIP <ul style="list-style-type: none">• identifyFormat• validateFormat• extractMetadata• decidePreservationAction	FITS, JHove, DROID, NLNZ Extractor
○ Prepare AIP <ul style="list-style-type: none">• gatherMetadata• normalizeFiles• createPackage	FFident, Unoconv, Ffmpeg, OpenOffice
○ Review AIP <ul style="list-style-type: none">• decideStorageAction	ImageMagick, Inkscape, Xena
○ Store AIP <ul style="list-style-type: none">• writePackage• replicatePackage• auditfixity• readPackage• updatePackage	Bagit, SAMBA, NFS-common, Poster
○ Provide DIP <ul style="list-style-type: none">• uploadPackage• updateMetadata	ICA-AtoM, DCB Dashboard
○ Monitor Preservation <ul style="list-style-type: none">• checkFormatRegistry• migrateFormat• synchronizeAIPsandDIPs	

What It's Not



A repository



A catalog

Installation Requirements

Operating System:

- Ubuntu 14.04 or 16.04 (beta)
- CentOS 7

Hardware minimums:

- Testing: dual-core CPU, 2 GB RAM, 7 GB disk space (+ 2-3x the size of collection materials)
- Full-scale processing: dual-core CPU, 4 GB RAM, 20 GB disk space (+ 3-4x size of collection materials)
 - likely need more power if working with A/V

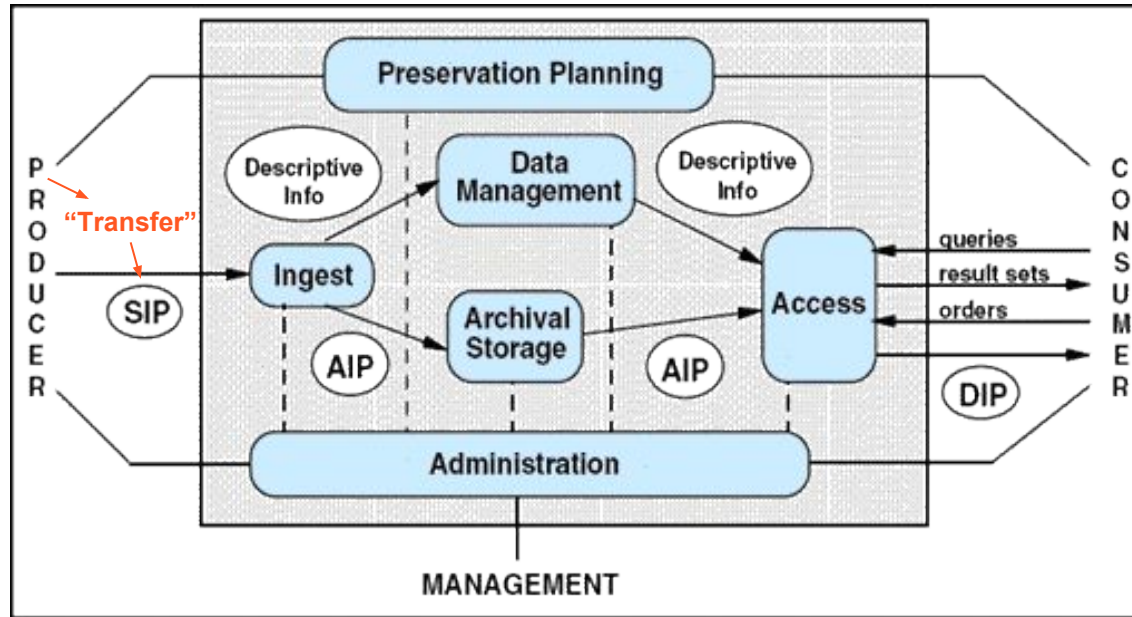
See:

<https://www.archivematica.org/en/docs/archivematica-1.6/admin-manual/installation/installation/> (last accessed October 2017)

When/Where/Why Would You Use It?

- Lots and lots of files...
- ...in a digital repository/on a server (or you want to set up one)
- Regular/frequent acquisition
- Already using integrated front-facing content management/access software:
 - Archivist's Toolkit/ArchivesSpace
 - AtoM
 - DSpace

Let's Look At This Diagram Again



Some advice:

- Take your time (play in a Vagrant test machine or the [Sandbox](#))
- Think micro, not macro
- Don't get too caught up in SIP/AIP/DIP



Dave Rice
@dericed

 Follow

I'm officially announcing my defection from OAIS to pioneer the OHAI model with 3 electronic-resource-packages: HERP, DERP & BERP.

RETWEETS

11

LIKES

29



2:52 PM - 10 Dec 2015



11



29



Before We Start!

Please run the following:

- Open Terminal
- `$ cd archivematica-vagrant`
- `$ vagrant up`

Archivematica Storage Service

- Navigate in web browser to “<http://10.10.10.20:8000>”
- Log in with “test/test”

Archivematica Storage Service Home Pipelines Spaces Locations Packages Administration Logout

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, DIPs and AIPs uploaded to a Location managed by the storage service.

Administration

Configure default settings and manage users.

Archivemata Storage Service

Pipeline: a single installation of the Archivemata dashboard software

UUID	Description	Enabled	Edit
7f20eb03-a0de-435e-a34b-59e212bb3d91	Archivemata on miap-VirtualBox	Enabled	Edit Disable Delete
Showing 1 to 1 of 1 entries			Previous Next



Archivematica Storage Service

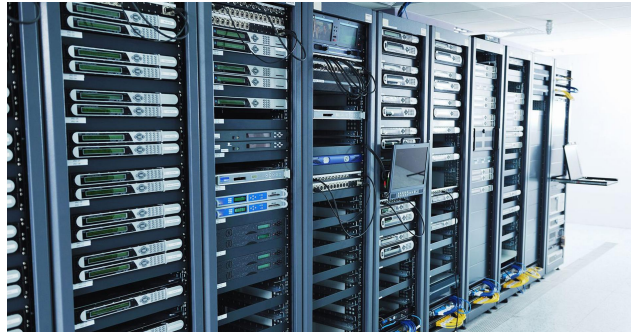
Space: a specific storage device (local hard drive, server accessed via network or remote protocol)

Archivematica Storage Service [Home](#) [Pipelines](#) [Spaces](#) [Locations](#) [Packages](#) [Administration](#) [Logout](#)

All Spaces



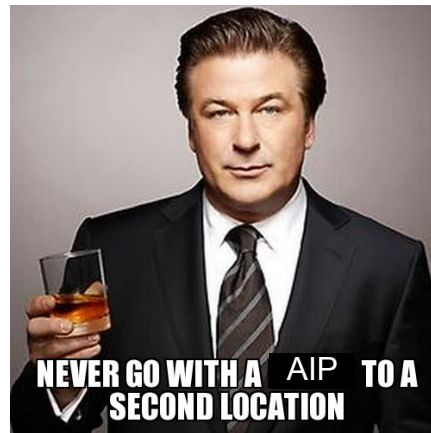
All information related to the access protocol, such as remote hostname and location of the export are stored in



Archivematica Storage Service

Location: sub-division of a Space, used for particular purpose during an Archivematica ingest (AIP Storage, DIP Storage, etc.)

Archivematica Storage Service								
Home Pipelines Spaces Locations Packages Administration Logout								
All 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.								
Show 10 entries		Search:						
Purpose	Pipeline	Path	Description	Space	UUID	Usage	Enabled	Actions
AIP Recovery	Archivematica on miap-VirtualBox,	/var/archivematica/storage_service/recover	Default AIP recovery	e9d100a8...	4e945324-0508-4d94-a097-0c0270f1a821	0B / unlimited	Enabled	Edit Disable Delete
AIP Storage	Archivematica on miap-VirtualBox,	/var/archivematica/sharedDirectory/www/AIPsStore	Store AIP in standard Archivematica Directory	e9d100a8...	6d8481d3-db0d-4c30-bb69-1ad03e2fa50a	0B / unlimited	Enabled	Edit Disable Delete
Currently Processing	Archivematica on miap-VirtualBox,	/var/archivematica/sharedDirectory	None	e9d100a8...	4cb31e80-c620-4392-b21f-a66bd9b69906	0B / unlimited	Enabled	Edit Disable Delete
DIP Storage	Archivematica on miap-VirtualBox,	/var/archivematica/sharedDirectory/www/DIPsStore	Store DIP in standard Archivematica Directory	e9d100a8...	f1789a3f-a9df-4652-83b1-b39d143ee8d4	0B / unlimited	Enabled	Edit Disable Delete
Storage Service Internal Processing	No pipelines	/var/archivematica/storage_service	For storage service internal usage.	e9d100a8...	4194da42-12c7-40d4-a14e-b769a5f8c4b4	0B / unlimited	Enabled	Edit Disable Delete
Transfer Backlog	Archivematica on miap-VirtualBox,	/var/archivematica/sharedDirectory/www/AIPsStore/transferBacklog	Default transfer backlog	e9d100a8...	fd7121aa-4898-476e-9e98-21368b6de473	0B / unlimited	Enabled	Edit Disable Delete
Transfer Source	Archivematica on miap-VirtualBox,	/home	None	e9d100a8...	b0dd6325-4b56-43e6-83a7-a09d566b8d53	0B / unlimited	Enabled	Edit Disable Delete
Showing 1 to 7 of 7 entries				Previous Next				



Archivematica Storage Service

Package: a bundle of files, stored in a Location

All Packages

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

[View recovery requests](#) | [View delete requests](#)

Show 10 entries

Search

UUID	Description	Originating Pipeline	Current Location	Size	Type	Pointer File	Status
34cdac17-1da2-45dc-b70c-37761a8a3dde	None	Archivematica on miap-VirtualBox (7f20eb03-a0de-435e-a34b-59e212bb3d91)	/var/archivematica/sharedDirectory/www/DIPsStore/34cdac17/1da2/45dc/b70c/3776/1a8a3dde/workshop_02-7cf36f79-22e7-4818-9cbc-85d7f9f5319f	457.3 KB	DIP	None	Uploaded (Update Status)
54bb64a0-c092-4365-a8a5-b913b42c0016	None	Archivematica on miap-VirtualBox (7f20eb03-a0de-435e-a34b-59e212bb3d91)	/var/archivematica/sharedDirectory/www/DIPsStore/54bb64a0/c092/4365/a8a5/b913b42c/0016/workshop_03-c5151fc6-357f-4c2b-ba4b-8d2bc346f27c	8.0 MB	DIP	None	Uploaded (Update Status)
7cf36f79-22e7-4818-9cbc-85d7f9f5319f	None	Archivematica on miap-VirtualBox (7f20eb03-a0de-435e-a34b-59e212bb3d91)	/var/archivematica/sharedDirectory/www/AIPsStore/7cf36f79/22e7/4818/9cbc/85d7f9f5319f/workshop_02-7cf36f79-22e7-4818-9cbc-85d7f9f5319f.7z	200.1 KB	AIP	Pointer File	Deleted
c5151fc6-357f-4c2b-ba4b-8d2bc346f27c	None	Archivematica on miap-VirtualBox (7f20eb03-a0de-435e-a34b-59e212bb3d91)	/var/archivematica/sharedDirectory/www/AIPsStore/c5151fc6/357f/4c2b/ba4b/8d2bc346f27c/workshop_03-c5151fc6-357f-4c2b-ba4b-8d2bc346f27c.7z	11.3 MB	AIP	Pointer File	Deleted
d9e8a4f4-214c-42d4-96c8-011d6f7a7605	None	Archivematica on miap-VirtualBox (7f20eb03-a0de-435e-a34b-59e212bb3d91)	/var/archivematica/sharedDirectory/www/AIPsStore/d9e8a4f4/214c/42d4/96c8/011d6f7a/7605/workshop_01-d9e8a4f4-214c-42d4-96c8-011d6f7a7605.7z	651.9 MB	AIP	Pointer File	Deleted





Translate

Turn off instant translation



Russian

English

Spanish

Archivematica



English

Russian

Spanish



Translate

I used a local pipeline to transfer and ingest a package to a location on a space

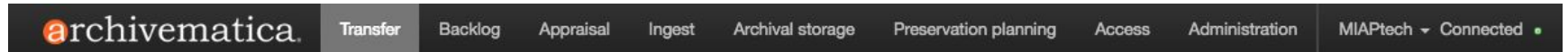


I used Archivematica, installed on my laptop, to move files from an external hard drive to a directory on a server

Type text or a website address or [translate a document](#).

Archivematica Dashboard

- Open new tab, navigate in browser to “<http://10.10.10.20>”
- Log in with “miap/miap”



Exercises to Try

- Using SampleTransfers > Multimedia, create a DIP that normalizes .wmv (Windows Media Video) into .mkv (Matroska)
- Activate the micro-service to transcribe incoming files using OCR, then use on SampleTransfers > OCRImage to create an AIP and/or DIP that contains a transcribed .txt file
- Maximize automation as much as possible - what's the fewest steps you can make it through a full transfer/ingest?

Further Resources

Archivematica Wiki: <https://wiki.archivematica.org>

User Forum (Google Group): <https://groups.google.com/d/forum/archivematica>

Sarah Romkey ([@ArchivesSarah](#)) - Program Manager for Archivematica at Artefactual

“Adding Metadata and Ingesting Large Born-Digital Archives with Archivematica”,
Dina Sokolova and Jane Gorjevsky, presentation for SAA Forum, 2014
<https://academiccommons.columbia.edu/catalog/ac%3A180720>