archivematica

Brought to you by
Anne Kofmehl, Kate Neptune, Claire Cella, Nicole Feldman and Colleen Hobbs

because 76% of us agree that archiving matters

Presentation Outline

- **Introduction** — The Who, What, When, Where, Why and How of Archivematica
- **Installation** — Stories, Challenges and Tips
- **Software** — Archivematica's Dashboard, Standards and Microservices
- **Demo** — An Archivematica Walk-Through
- **Evaluation** — Should You Choose Archivematica?

Introduction to Archivematica

- Archivematica is an **integrated suite** of **software tools**, not a single technology stack.
- It's **ISO-OAIS** compliant (International Organization for Standardization-Open Archival Information System).
- It's an **open-source digital preservation system** developed through Artefactual Systems, a for-profit software developer.
- It's designed to be a **system incorporating people, procedures, and software** that connects existing tools and accommodates changing technology.

Who is Developing and Using Archivematica?

Archivematica was developed in Canada.

Its early research and development partners include —

The InterPARES Project	**1999**
UNESCO	**2007**
Artefactual Systems	**2008**
The City of Vancouver	**2009**
International Monetary Fund Archives	**2010**

Early Digital Preservation Research: InterPARES

1999

- The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) develops strategies for long-term preservation of authentic records created and/or maintained in digital form. It provides the basis for standards, policies, strategies to ensure the longevity of such materials and the ability of users to trust its authenticity.
- It was conducted at University of British Columbia's School of Library, Archival and Information Studies.
- Artefactual founder, **Peter Van Garderen**, was a project coordinator for InterPARES Part 1 in 1999-2001, called "Establishing and Maintaining Trust in Electronic Records."
- InterPARES' Project Director is **Dr. Luciana Duranti.**

UNESCO Memory of the World Programme Sub-Committee on Technology

2007

The **proposal** suggests a holistic approach to digital preservation, considering all aspects of digital repositories.

It argues that **solutions** for digital preservation are understood, and what is needed are affordable tools, technology, and training in those systems.

The White Paper —

"Towards an Open Source Repository and Preservation System: Recommendations on the Implementation of an Open Source Digital Archival and Preservation System and on Related Software Development"

Kevin Bradley, *National Library of Australia*, with Junran Lei and Chris Blackall, *Australian Partnership for Sustainable Repositories*

http://portal.unesco.org/ci/en/ev.php- URL_ID=24700&URL_DO=DO_TOPIC&URL_SECTION=201.html

UNESCO Proposal Recommendations

2007

The UNESCO Memory of the World Programme takeaway —

- Archivists should build a **sustainable** system; archivists *shouldn't* expect a permanent storage media to solve the digital preservation problem.
- A digital archival and preservation system should consider **all elements** of the preservation process: Ingest, Access, Administration, Data Management, Preservation Planning, Archival Storage.

Artefactual Systems

2008

- Artefactual Systems was founded by Peter Van Garderen.
- It is a business plan based on selling microservices to support open-source software.
- It is located in Vancouver, and the City of Vancouver was an early client.

Artefactual develops open-source software,
but it is a ***business***, not a non-profit organization.

The City of Vancouver and the 2010 Olympic Archives

2009

- The City of Vancouver had a legal obligation to collect, maintain, and provide public access to Olympic records.
- In 2009, the City of Vancouver hired **Courtney Mumma** to acquire and process the incoming Olympic collection.
- The City of Vancouver partnered with **Artefactual Systems** to develop **Archivematica.**
- The Olympic Reserve Legacy Fund helped underwrite and fund Archivematica’s development.

The City of Vancouver and the 2010 Olympic Archives

2010

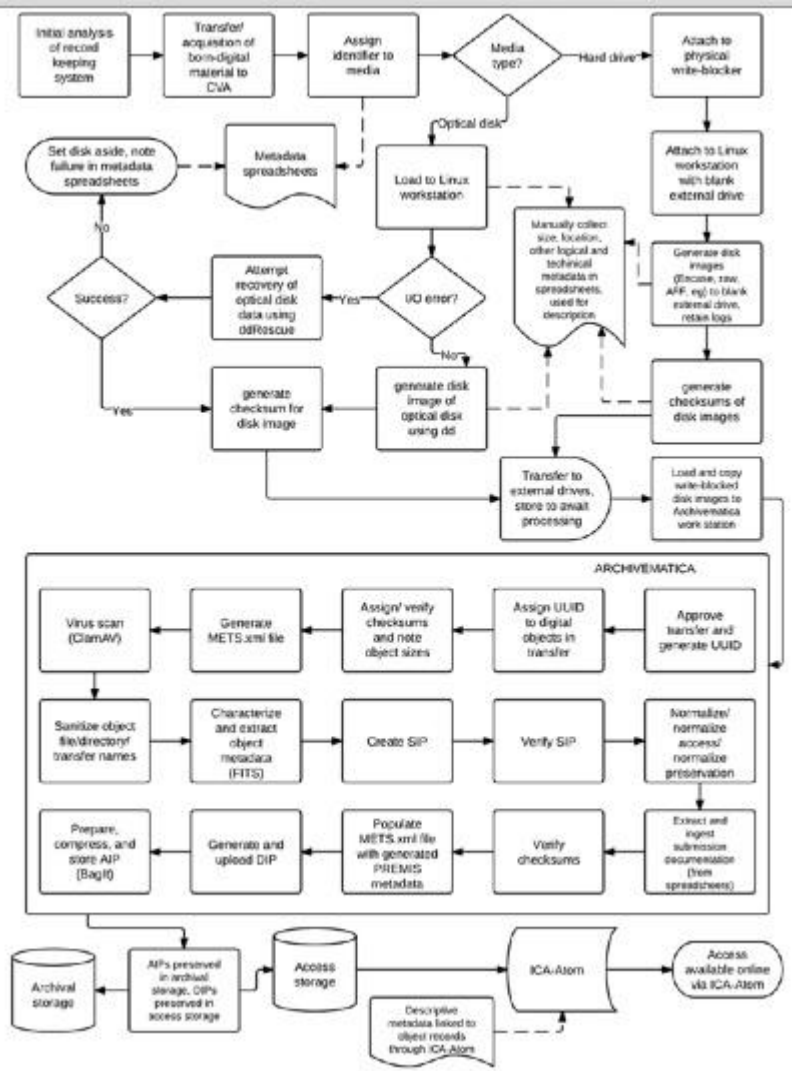
For the 2010 Olympics, Artefactual developed a system that —

- complies with the Open Archival Information System (ISO-OAIS) standard
- ingests a variety of born-digital objects (compared to existing systems that were largely concerned with only digitized files)
- stores objects securely with preservation metadata
- addresses preservation planning
- provides logging to demonstrate what has been done to objects
- is entirely free and open source
- is flexible and able to change or add features as digital curation best practices develop
- is scalable, to accommodate the ingest of large acquisitions

(http://public.ccsds.org/publications/archive/650x0b1.pdf)

City of Vancouver's Born-Digital Workflow

from Martin J. Gengenbach's "The Way We Do it Here": Mapping Digital Forensics Workflows in Collecting Institutions. http://digitalcurationexchange.org/system/files/gengenbach-forensic-workflows-2012.pdf



Archivematica's place in workflow

AtoM's place in workflow

Clients Currently Using Archivematica

About **30** major clients, including Yale University Library, MoMa, and Rockefeller Archive Center

Archivematica works by...

making many *commercial* and *open source* products *compatible*.

Archivematica works with existing collections management and storage tools and architectures.

Accordingly, it is designed to be compatible with systems such as DSpace, ContentDM, ICA-AtoM, and Archivists' Toolkit.

Archivematica provides strategies for...

implementing *emulation, migration,* and *normalization* as strategies for risk management.

Emulation: attempts to recreate or virtualize an underlying technology environment to render a file format that has already gone obsolete.

Migration: attempts to transform a current file format that may be in danger of going obsolete, into a closely related and supported format (proprietary or nonproprietary) with more long-lived potential.

Normalization: converts current file formats into file types that are nonproprietary and rely on open standards and specifications. Normalization ensures the best chance for digital content to surviving technology obsolescence.

Archivematica creates...

processing *workarounds*.

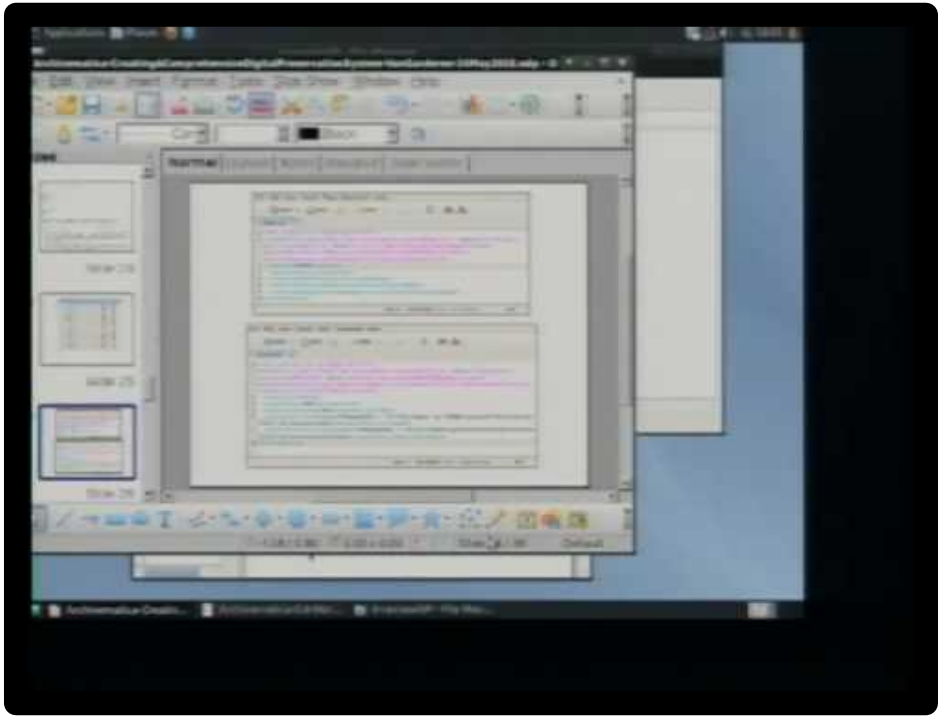
Processing batches of digital materials creates bottlenecks because of limited computing resources.

The Archivematica server can create processing clusters that route tasks to other servers (for example, a virtual machine) and reports back when the task is done.

This workflow allows users to work with the server and process other objects while a task is being completed.

Why is Archivematica being developed?

According to Archivematica's founder, Van Garderen, the software is being developed to provide **practical**, **comprehensive** archival **solutions** that even small repositories can afford.



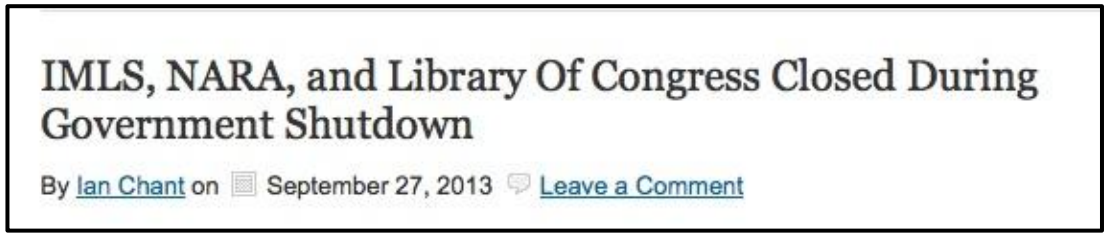
https://www.youtube.com/watch?v=czQx4sCO88k

A Budget-Conscious Archival Approach

Archivematica's designers think that —

- repeatedly moving data from proprietary standards and software is cost-prohibitive and will threaten data security.
- members of the digital preservation community should combine resources to create open-source software rather than independently purchasing commercial licenses.
- users need to contribute time and resources to the software's design. The project's designers argue that archivists need to feel empowered to use digital technology to code and debug open-source software.

Archives' budget constraints



IMLS, NARA, and Library Of Congress Closed During Government Shutdown

By Ian Chant on September 27, 2013 Leave a Comment

Timeframe for Development

"Release early, release often."

Archivematica's "agile development method" launches upgrades on scheduled dates.

This software development plan promotes adaptive planning and rapid and flexible response to change.

The developers' goal is not to perfect an application, but to continue to upgrade and enhance performance.

Feb 2009: Release 0.1-alpha * May 2010: Release 0.6-alpha

February 2011: Release 0.7-alpha * December 2011: Release 0.8-beta

September 2012: Release 0.9-beta

First Production Release

Archivematica 1.0, which was released in January 2014, includes numerous upgrades and fixes —

•Upgrade file identification used as the basis to trigger format policy actions (aka 'preservation plans'). •Include a manual normalization workflow. •Improve email handling. •Add ability to edit format policies from preservation tab in the dashboard. •Add ability to add/change format policies from FPR updates. •Add a workflow for applying updated format policies to pre-existing AIPs. •include advanced search screens for searching AIP contents in the dashboard. • Generate DIPs from the access tab in the dashboard. •Include visualization of transfers. • Include file-level Dublin Core and rights metadata entry. • Include field validation in rights templates. •Index transfers and identify/flag personal information. Evaluate Bit Curator tool to determine how much functionality/data can be integrated/re-used prior to Archivematica ingest. •Customize statistical reporting

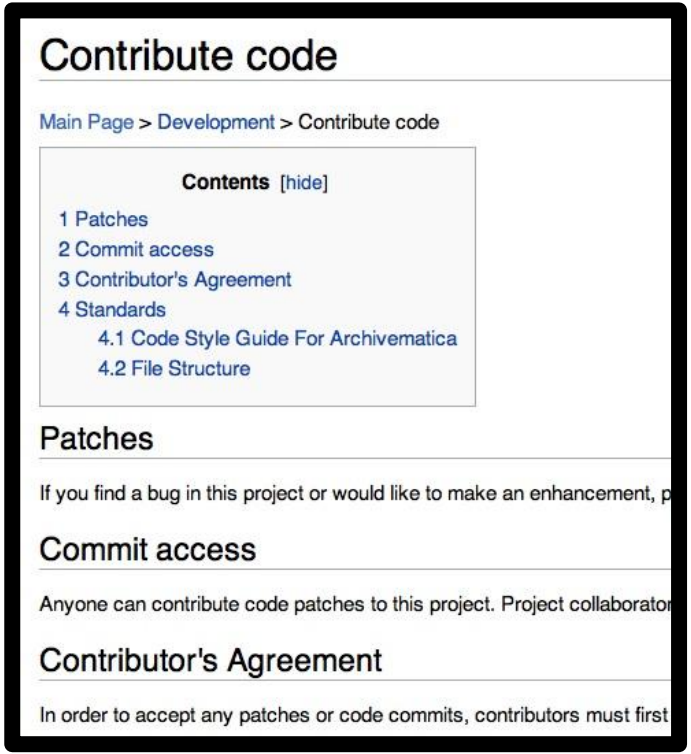
How is Archivematica Being Developed and Made Available?

Archivematica's microservices are developed through collaboration. Examples include —

- code contributions
  - bug reports
- wiki documentation updates
  - chat rooms
- Google Groups
- discussion lists

Archivematica's source code is freely available at —

https://github.com/artefactual/archivematica



Contribute code

Main Page > Development > Contribute code

Contents [hide]

1 Patches
2 Commit access
3 Contributor's Agreement
4 Standards
4.1 Code Style Guide For Archivematica
4.2 File Structure

Patches

If you find a bug in this project or would like to make an enhancement, p

Commit access

Anyone can contribute code patches to this project. Project collaborator

Contributor's Agreement

In order to accept any patches or code commits, contributors must first

Examples of Archivematica's Collaborative Efforts

"We have a new feature that has been sponsored by the University of Alberta which will allow for adding PREMIS event MD after doing manual normalization. You'll find it on our roadmap for the 1.1 release and in our issues list: https://projects.artefactual.com/issues/5216. You may have to experiment with the feature to see how it can be used with your own local workflow."

>> I'm meeting with the creator of the Kakadu JPEG2000 library next week
>> to see if there is a way to incorporate a J2K in FFMEG specifically
>> licenced for AV archives.

We would love to get some kind of standard MJ2K implementation into Archivematica. This would be a major contribution and much appreciated. How did your meeting go? Do you foresee any licensing issues? Obviously we would need some type of free software license for Kakadu to bundle it into the system. Anything like MIT, BSD, Apache, LGPL, GPL2 or AGPL3 would be compatible with our AGPL3 license.

Paul James - ARCW / NLW Digital Preservation 10/24/13

Hello UK users

The National Library of Wales are looking into arranging a UK gathering of Archivematica/AtoM users and other interested parties. We think some networking could prove to be very valuable. So if you have any experiences with Archivematica or are looking into using it in the future we would be delighted to hear from you.

We are also looking into the viability of moving over to using ICA AtoM so if you also have any experiences with this then we would also like to hear from you too. (I will be posting a similar invitation on the AtoM forum so I apologise if you receive this twice)

Kind regards

Paul James
National Library of Wales

Development collaboration from Archivematica Google group:

https://groups.google.com/forum/#!forum/archivematica

How does Archivematica relate to competitive software?

It's more comprehensive and compatible.

According to Liso Spiro's CLIR Report in 2009 —

- "Others caution...that ***importing existing finding aids*** into **Archon** *can be* ***difficult***, given the variability of EAD." (Archon is another open-source system.)

- "Archivists noted that it can be ***difficult to import existing finding aids*** and make **AT [Archivists' Toolkit]** accommodate existing workflows." (Archivist's Toolkit is another open-source system.)

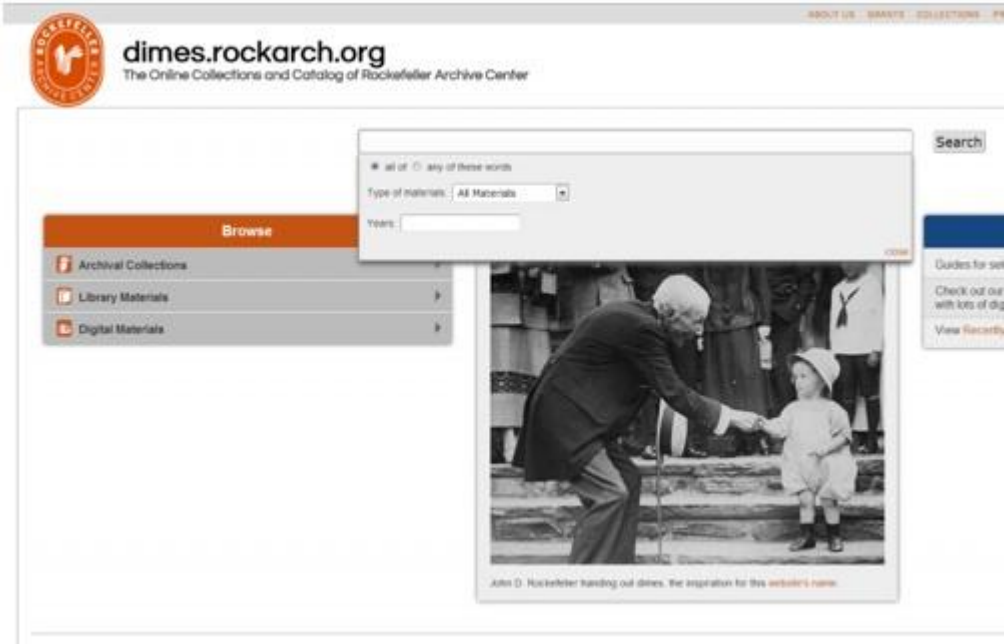
- "**Calm for Archives,** developed by DS, bills itself as 'the leading archival solution in the UK.' It has a client/server architecture and ***requires Windows***." (Calm for Archives is a commercial product.)

How does Archivematica work collaboratively?

- Archivematica can receive transfers from **Bit Curator**, **ContentDM**, and **DSpace**. It can also receive bags in the Library of Congress **Bagit** format. Each of these types of transfer is slightly different than transferring from a drive. The transfer source is specified at the start of processing.
- AIPs from Archivematica are in **Bagit** format.
- DIPs from Archivematica can be uploaded to **AtoM**, **ContentDM**, or **DSpace**.
- When Archivematica is used with **DSpace**, it functions as dark archive while **DSpace** is used for access.

Archivematica in Use

At the Rockefeller Archive Center, Archivematica is used to view digitized material within a finding aid.



"The PDFs you see online are access copies of high-resolution TIFFs that are stored in and managed by Archivematica. We've been working with Artefactual over the past few months to come up with a way that Archivematica can **connect to the Archivists' Toolkit database** to insert metadata about these access copies and link them to the correct components in a resource record."

http://rockarch.org/programs/digital/bitsandbytes/?p=767#more-767

Archivematica in Use

In 2012, the library at Simon Fraser University (SFU) initiated a set of microservices to transfer electronic theses and dissertations theses from its Theses Registration System (TRS) to its institutional repository, Summit, without human intervention (apart from a sign off by library staff that the thesis has become ready for publication).

Shortly after the initiation of this automated workflow, the Library started moving theses from the TRS into the Archivematica digital preservation platform, a process which is also fully automated.

http://purl.pt/24107/1/iPres2013_PDF/Automating%20the%20Preservation%20of%20Electronic%20Theses%20and%20Dissertations%20with%20Archivematica.pdf

Future Plans

Because Artefactual Systems and Archivematica are **businesses**, they require a sustainable profit margin.

Archivematica's business model builds in **constant, incremental change** to keep pace with changing technologies. Innovations take place over several release cycles; when work is sponsored by a partnering institution, it can be completed more quickly.

Per Courtney Mumma, future plans are to —

"Innovate, innovate, innovate. And make the product awesome. If it's awesome, more people will like it and use it, and there will be more demand for our stuff."

Installing Archivematica: What You Need

Archivematica is PC and Mac Compatible.

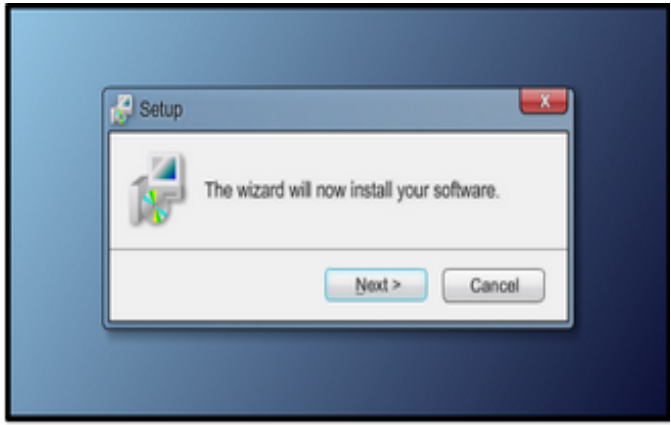
Archivematica can also be installed using Ubuntu Linux and the command line (instructions for 1.0 version only).

System Requirements, according to Archivematica's wiki —

- Processor: Intel core 2 or AMD Opteron
- Memory: 2 GB for the virtual appliance ('guest') operating system, **Depending on the operating system, machines with less than 2 GB total memory will likely have trouble.** Note that the default allocation setting in Archivematica is 512 MB; however, the more allocated the better the system will run. The setting can be changed once Archivematica is running.
- Hard Drive Space: a minimum of 3 GB to test the system on a small scale (i.e., use the available test files or import a small set of test files); 12 GB or more for larger implementations

https://www.archivematica.org/wiki/Virtual_appliance_instructions

Since we aren't installing this in an institutional setting, **we opted to allocate 512 MB of memory** to make it easier on our smaller machines.



Installing Archivematica: What is VirtualBox?

VirtualBox is a type of **virtual machine (VM).**

VMs are software-based emulations of a physical machine (i.e., a computer) that can run various ‘virtual appliances’ (i.e., Archivematica, Bit Curator etc.)

PROS —

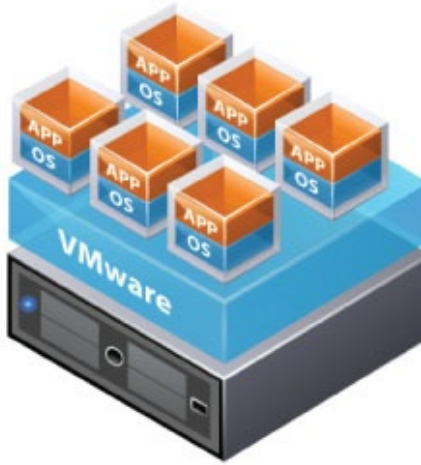
- Faster system speeds, due to freeing up memory and space on the host machine
- Reduces cost and saves energy
- Allows you to run multiple OS at once
- Virtual files are easier to back-up

CONS —

- Time-consuming and laborious for some
- Security and firewall issues, require extra monitoring
- When multiple machines are running at once, it can affect performance

http://computer.howstuffworks.com/how-virtual-computing-works2.htm

Traditional Architecture



Virtual Architecture

Installing Archivematica: Our Story, Challenges & Tips

Our main concerns were **SPACE** and **SPEED.**

If you want to install this software you need...

something **bigger** than a Macbook Air,

at least **2 GB** of space on your hard-drive,

AND

faster internet connections make all the difference.

We opted to install the **0.9 beta version** (for the reasons above) and at least one of our group members had an issue unzipping the file for v. 0.10 (v. 1.0 wasn't yet available)

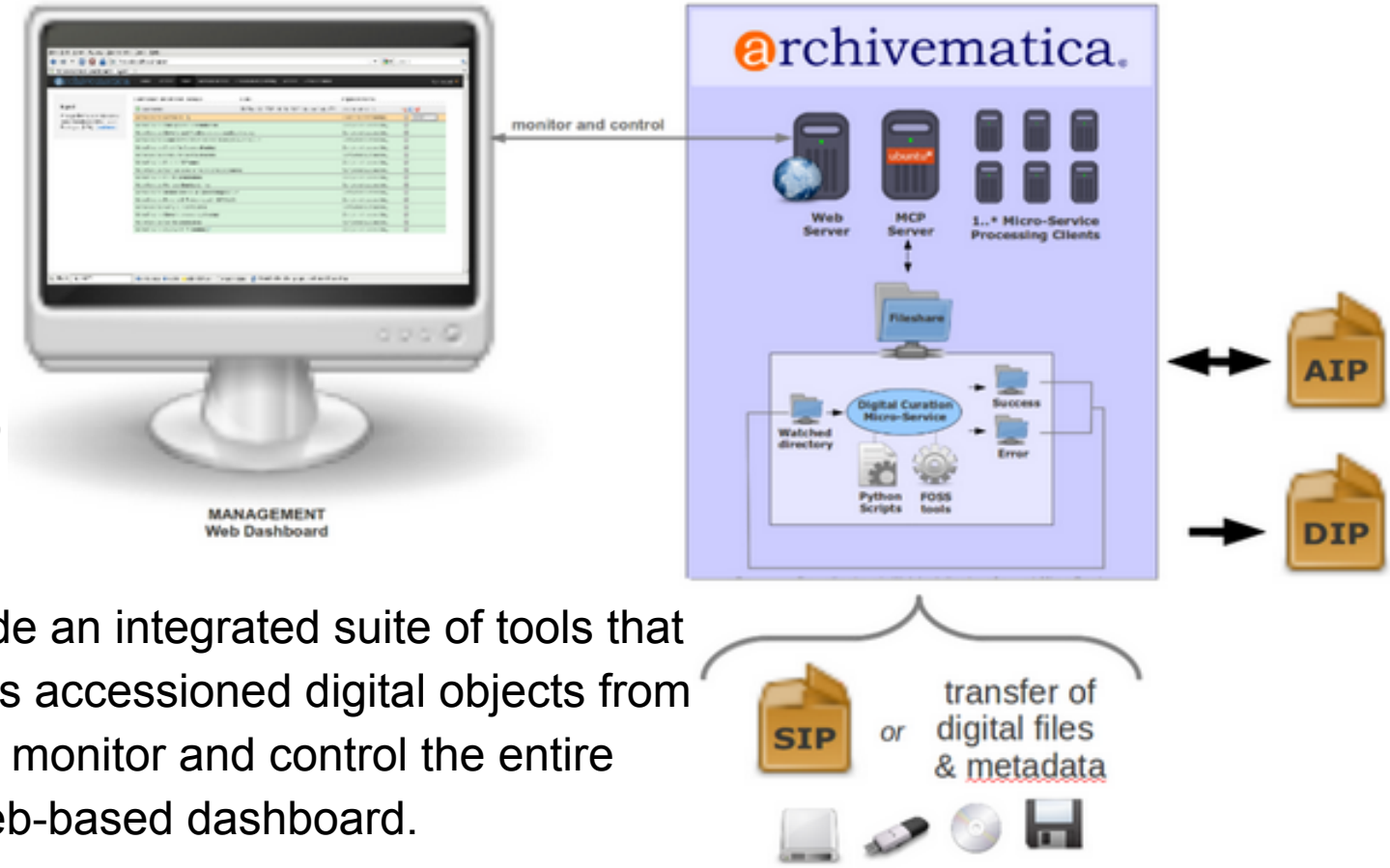
We also failed to perform step 12 of the wiki's installation instructions, which would've allowed us to run Archivematica from a web browser.

Installing Archivematica: Your Story (Audience Participation Portion)

QUESTIONS? THOUGHTS? COMMENTS?

The Software

Archivematica is a free, open source, digital preservation system designed to maintain **standards-based**, **long-term** access to digital objects.



Microservices provide an integrated suite of tools that allow users to process accessioned digital objects from ingest to access, and monitor and control the entire process through a web-based dashboard.

Standards Supported

METS (Metadata Encoding and Transmission Standard) — the standard for encoding descriptive, administrative, and structural metadata to manage objects in a digital library or exchange them between repositories, through the creation of an XML document. A METS document consists of seven sections: a header describing the document itself, descriptive metadata, administrative metadata, a file section that lists contents, a structural map that outlines and links each object, structural links that record the hyperlinks from the structural map and a behavior section that describes the executable elements.

PREMIS (PREservation Metadata: Implementation Strategies) — an international working group sponsored by OCLC and RLG that produced the PREMIS Data Dictionary for Preservation Metadata, an XML schema, and supporting documentation. The Data Dictionary is the international standard for metadata to support preservation of digital objects and ensure their long-term usability and viability. The Data Dictionary defines a core set of units that are important in digital preservation activities and archiving: Intellectual Entities, Objects, Events, Rights, and Agents.

Dublin Core — a vocabulary of fifteen (or more) properties to use in description of objects, using a list of concepts with natural-language definitions, intended to be used in combination with terms from other, compatible vocabularies (i.e., developed by the archival institution) to ensure interoperability.

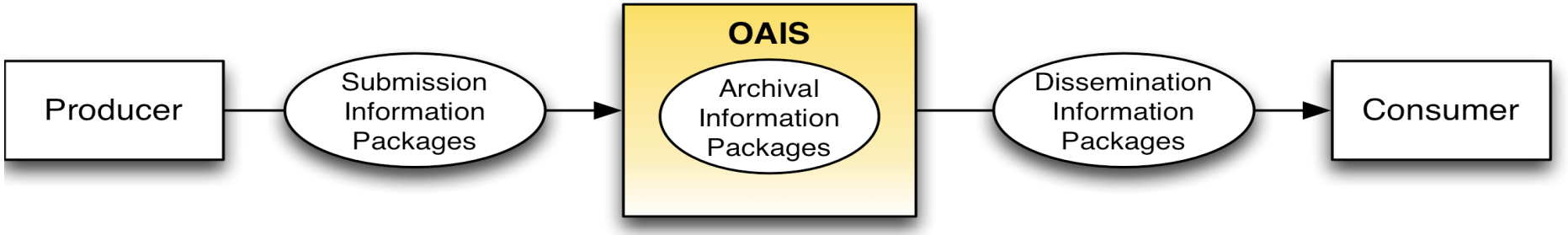
Standards Supported

The OAIS Reference Model — defines those involved in long-term preservation of digital objects and provides a model to manage objects through a system. A significant component is the Information Package (IP), which consists of the digital object(s) to be preserved, the metadata required at that point in the system, and its packaging information.

The **Submission Information Package (SIP)** is the metadata supplied by the producer or creator of the material (or the archivist) at time of ingest or accession and may lack structure or may not be comprehensive at all levels of the archive.

The **Archival Information Package (AIP)** is the SIP combined with Preservation Description Information, such as a unique and persistent identifiers, the history of the object, its relationship to other objects and a fixity or authenticity value.

The **Dissemination Information Package (DIP)** is the result of a user requesting the object from the OAIS and is a combination of the object and its metadata. The metadata at this stage is usually more descriptive than technical.



Standards Supported

The Format Policy Registry (FPR) — a database that allows users to define policies for handling file formats, indicating the actions, tools and settings to apply to files of particular formats (e.g. converting to preservation or access formats). Format policies change as community standards, practices and tools evolve.

Media type	File formats	Preservation format(s)	Access format(s)	Normalization tool
Audio	AC3, AIFF, MP3, WAV, WMA	WAVE (LPCM)	MP3	FFmpeg
Email	PST	MBOX	MBOX	readpst
Email	Maildir**	Original format	MBOX	md2mb.py
Office Open XML	DOCX, PPTX, XLSX	Original format	PDF for PPTX	Tool search in progress
Plain text	TXT	Original format	Original format	None
Portable Document Format	PDF	PDF/A	Original format	Ghostscript
Presentation files	PPT	Original format	PDF	Tool search in progress
Raster images	BMP, GIF, JPG, JP2*, PCT, PNG*, PSD, TIFF, TGA	Uncompressed TIFF	JPEG	ImageMagick
Raw camera files/Digital Negative format**	3FR, ARW, CR2, CRW, DCR, DNG, ERF, KDC, MRW, NEF, ORF, PEF, RAF, RAW, X3F	Original format	JPEG	ImageMagick/UFRaw
Spreadsheets	XLS	Original format	Original format	None
Vector images	AI, EPS, SVG	SVG	PDF	Inkscape
Video	AVI, FLV, MOV, MPEG-1, MPEG-2, MPEG-4, SWF, WMV	FFV1/LPCM in MKV	MP4	FFmpeg
Word processing files	DOC, WPD, RTF	• ODF (WPD and RTF) • Original format (DOC)	PDF	Tool search in progress

The Microservices

Metadata — **BagIt** (packages digital objects and metadata for archival storage), **Zip** (a utility used by Bagit to create AIP packages), and **FITS (File Information Tool Set)** (identifies, validates and extracts technical metadata from files)

Virus Scanning — **Clam AV** (an antivirus scan that detects viruses and other threats)

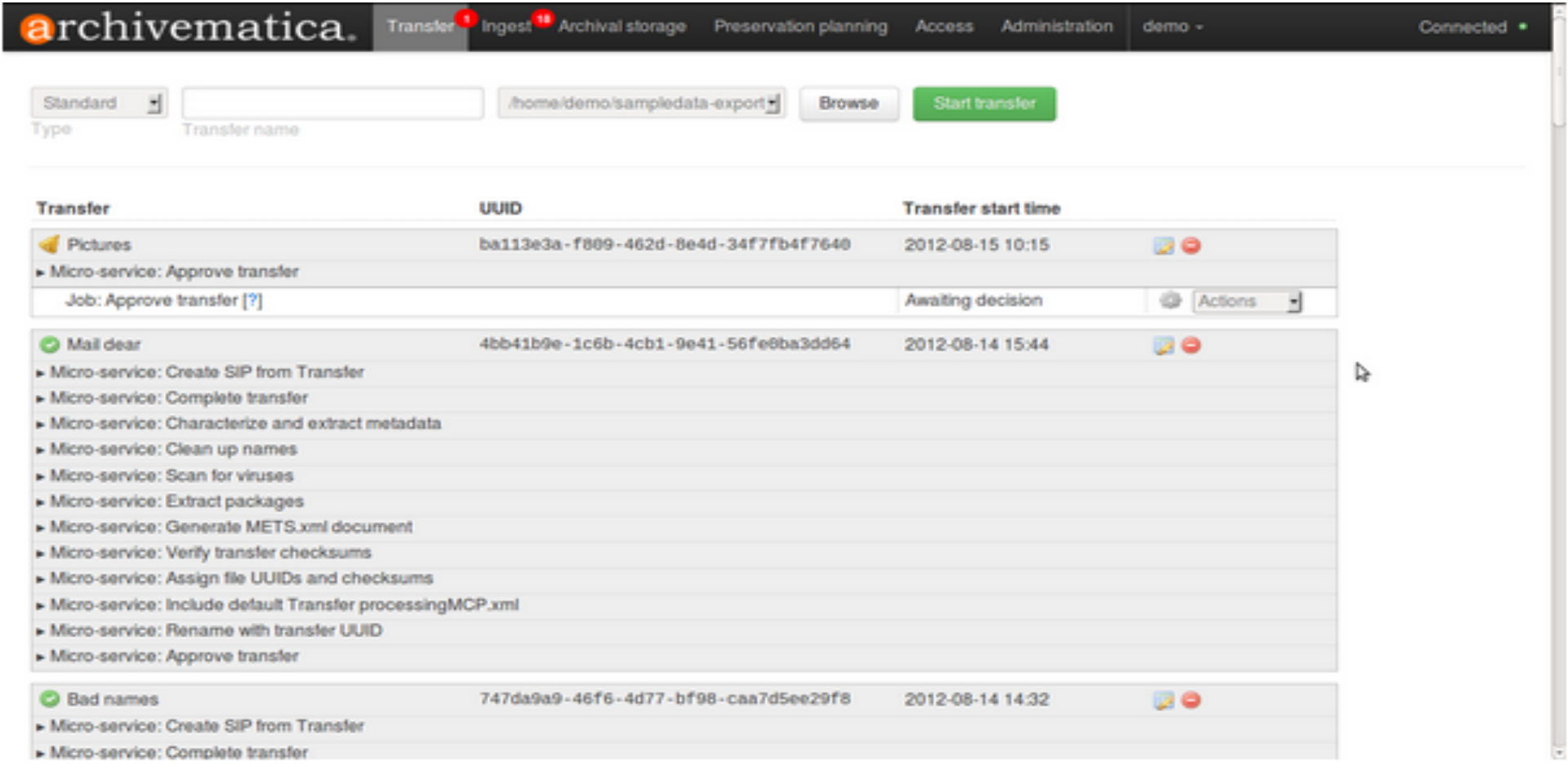
Search — **ElasticSearch** (an indexing, search and analytics tool)

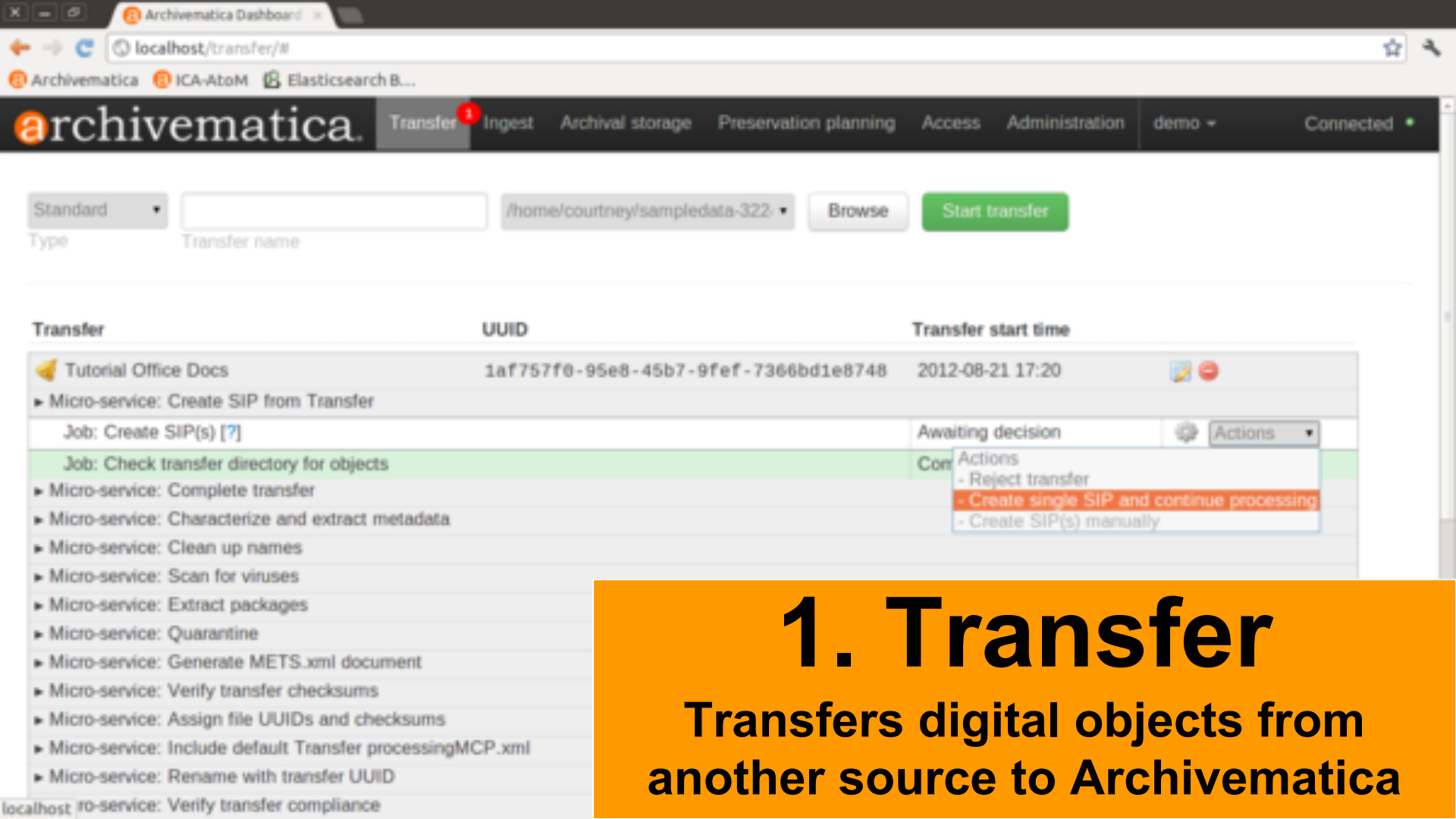
Normalization — **FFmpeg, Imagemagick** and **LibreOffice** (software frameworks that support normalization by converting audio and video, image, and document formats respectively)

Access — **ICA-AtoM** (description and access tool for archives) and **NFS-common** (provides access to files on network storage)

Verification — **MD5** (generates checksums or 128-bit hash values and verifies object integrity), **UUID** (produces unique numbers for objects in order to reliably identify them across a system), and **EXT4 filesystem** (a file system that keeps track of changes that will be made in a log)

The Dashboard





1. Transfer

Transfers digital objects from another source to Archivematica

Transferring digital objects from another source to Archivematica

Files or objects to be transferred can come from —

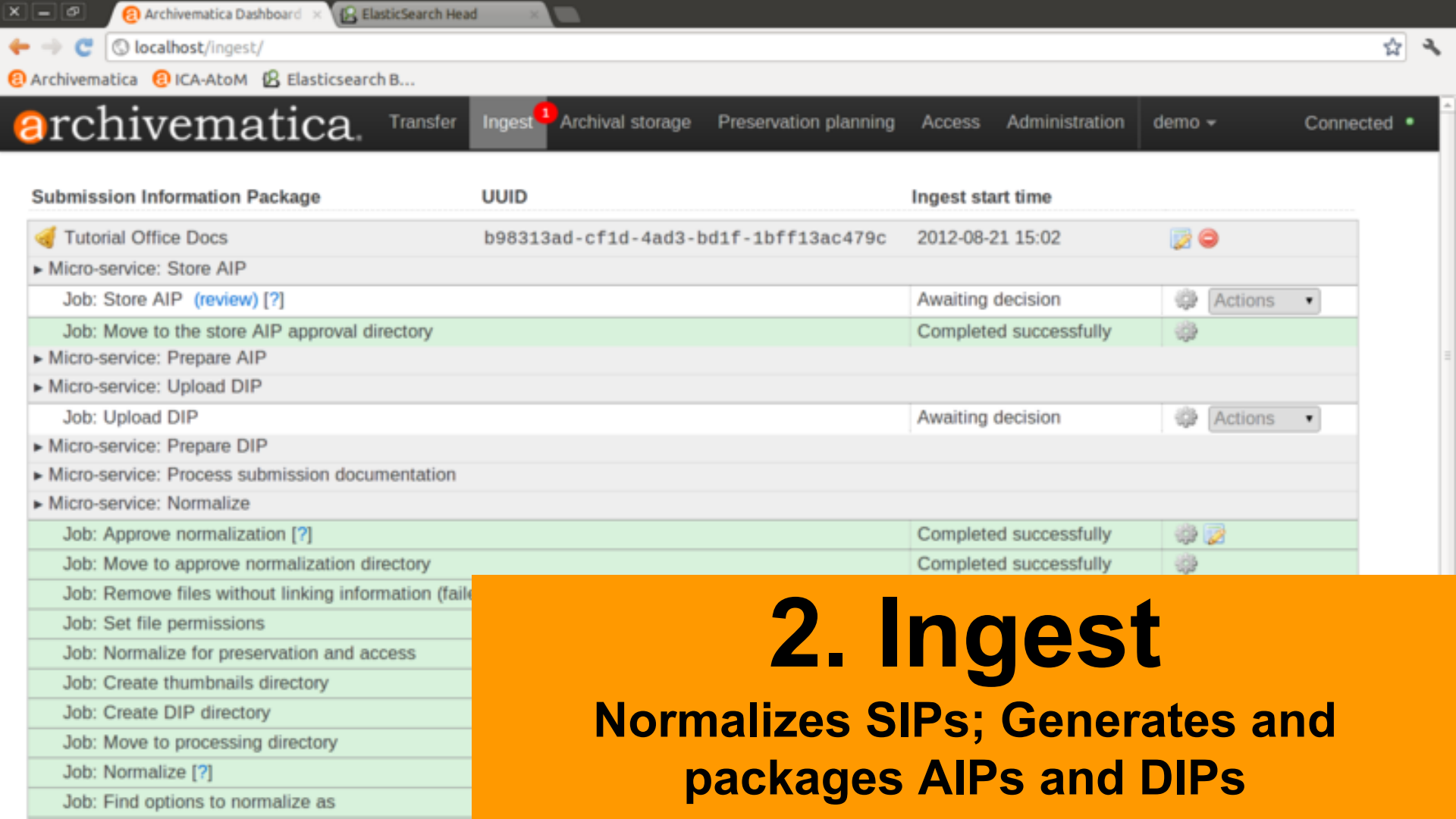
- standard source (desktop folder)
- unzipped or zipped folder
- DSpace or Maildir (a way to store email messages)

but cannot be uploaded from a web source.

The **Transfer** process includes —

- **generating checksums** and **running fixity checks** (verifies that an object has not been changed in a given period by computing checksums and comparing them to a stored value)
- **generating** a **METS.xml** document (captures the original order of the transfer to be automatically added to any generated SIPs, in case users later delete, rename or move files or break the transfer up)
- **renaming** the files with unique identifiers (UIDs)
- **sanitizing** names (removes any special characters in filenames and replaces them with dashes, while preserving the original names in the metadata)
- **scanning** the files for viruses
- **identifying and extracting** object metadata

At the end, a single SIP is created and sent on to Ingest.


Archivematica Dashboard
ElasticSearch Head
localhost/ingest/
Archivematica
ICA-AtoM
Elasticsearch B...
archivematica
Transfer
Ingest
1
Archival storage
Preservation planning
Access
Administration
demo
Connected
Submission Information Package
UUID
Ingest start time
Tutorial Office Docs
b98313ad-cf1d-4ad3-bd1f-1bff13ac479c
2012-08-21 15:02
Micro-service: Store AIP
Job: Store AIP (review) [?]
Awaiting decision
Actions
Job: Move to the store AIP approval directory
Completed successfully
Micro-service: Prepare AIP
Micro-service: Upload DIP
Job: Upload DIP
Awaiting decision
Actions
Micro-service: Prepare DIP
Micro-service: Process submission documentation
Micro-service: Normalize
Job: Approve normalization [?]
Completed successfully
Job: Move to approve normalization directory
Completed successfully
Job: Remove files without linking information (fail
Job: Set file permissions
Job: Normalize for preservation and access
Job: Create thumbnails directory
Job: Create DIP directory
Job: Move to processing directory
Job: Normalize [?]
Job: Find options to normalize as
2. Ingest
Normalizes SIPs; Generates and packages AIPs and DIPs

Normalizing SIPs; Generating and packaging AIPs and DIPs

The **Ingest** process includes —

- **inputting additional metadata** or **adding rights** management, permissions, acts, grants or restrictions
- **normalizing file formats** (ensures digital objects of a particular type (e.g., color images) are converted into a single standard format so the file remains functional and preserved for a long period of time (e.g., all BMP, JPG or GIFs changed to uncompressed TIFFs)
- **preparing the METS.xml file** (transfers all logs created during Transfer to SIP)
- **approving AIP** (with the option to store in a preconfigured storage location)
- **approving DIP** (with option to upload to selected location in the access directory)
- **verifying checksums** (ensures the files and objects have not been corrupted in Transfer or Ingest)
- **indexing** for searchability

At the end, an AIP is created to be approved and stored in archival storage while a DIP is also created to be approved and uploaded to an access system.

archivematica Transfer Ingest 1 Archival storage Preservation planning Access Administration x

Archival storage / Search

Memphis | Any | Keyword | Search archival storage | Show files?

or | jpg | File extension | Keyword

Add New

Found 3 results. Showing 1 to 3.

	File(s)	AIP(s)
	freekitten.jpg 3ae76880-8572-4d0c-be50-6bbf6e44756f	Memphis_pictures 122b7de4-cdbc-4637-8a94-4353f23f8f51 (view raw)
	WorkingWithBug.JPG 11ee753d-994f-4137-a180-41d7432ce055	Memphis_pictures 122b7de4-cdbc-4637-8a94-4353f23f8f51 (view raw)
	Landing_zone.jpg 710cd55e-9393-45dc-8e	

3. Archival Storage

AIP is moved into its storage repository

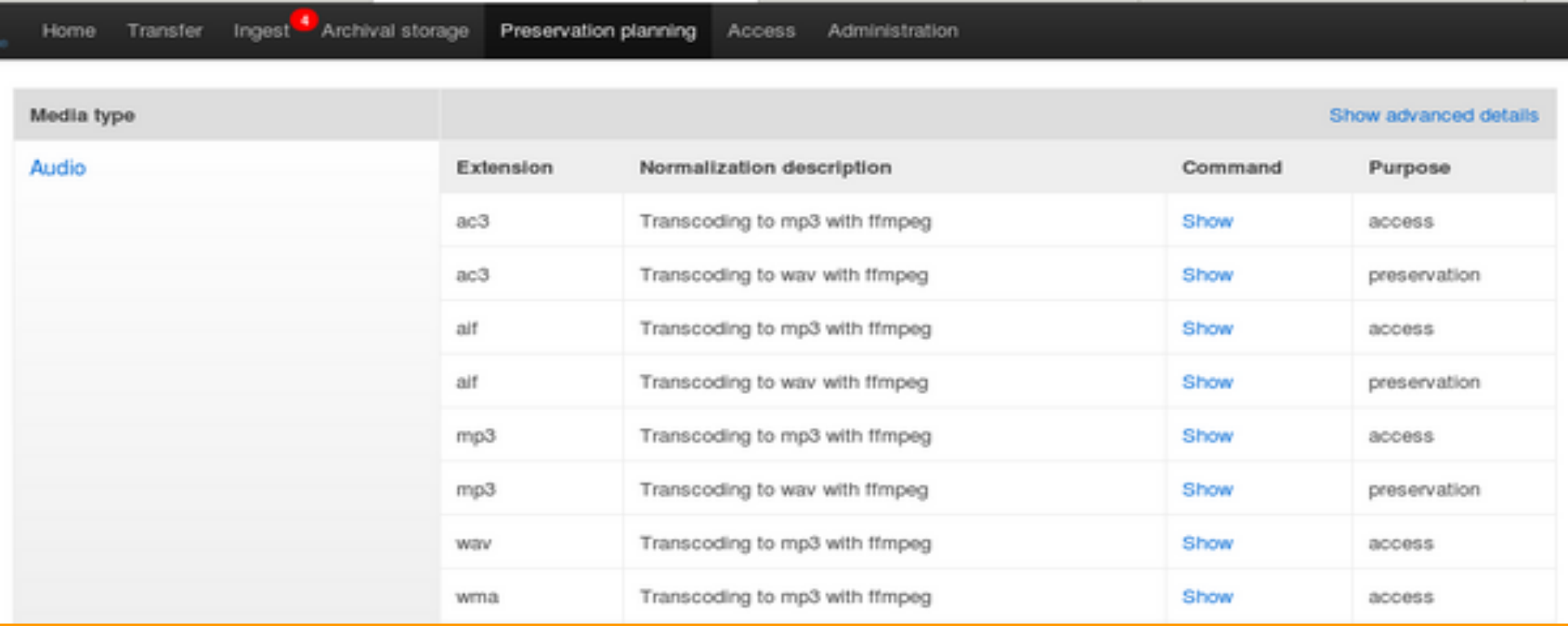
Moving AIP into its storage repository

The **Archival Storage** process (*which was finished in Ingest)* includes —

- **approving and packing AIP** (which is based on Bagit, and consists of a base folder with a tag and a subdirectory. The **tag** is a simple text-file, like a packing slip, that consists of an inventory and checksum. The **data subdirectory** consists of the METS file and three folders — logs, objects, and thumbnails)
- **storing AIP** as a compressed zip file to storage directory (usually a remote network)

In **Archival Storage**, users can —

- **view and search** the contents (all textual content as well as METS metadata) via a table with information about the stored AIPs
- **open, download and unzip** AIPs



4. Preservation Planning

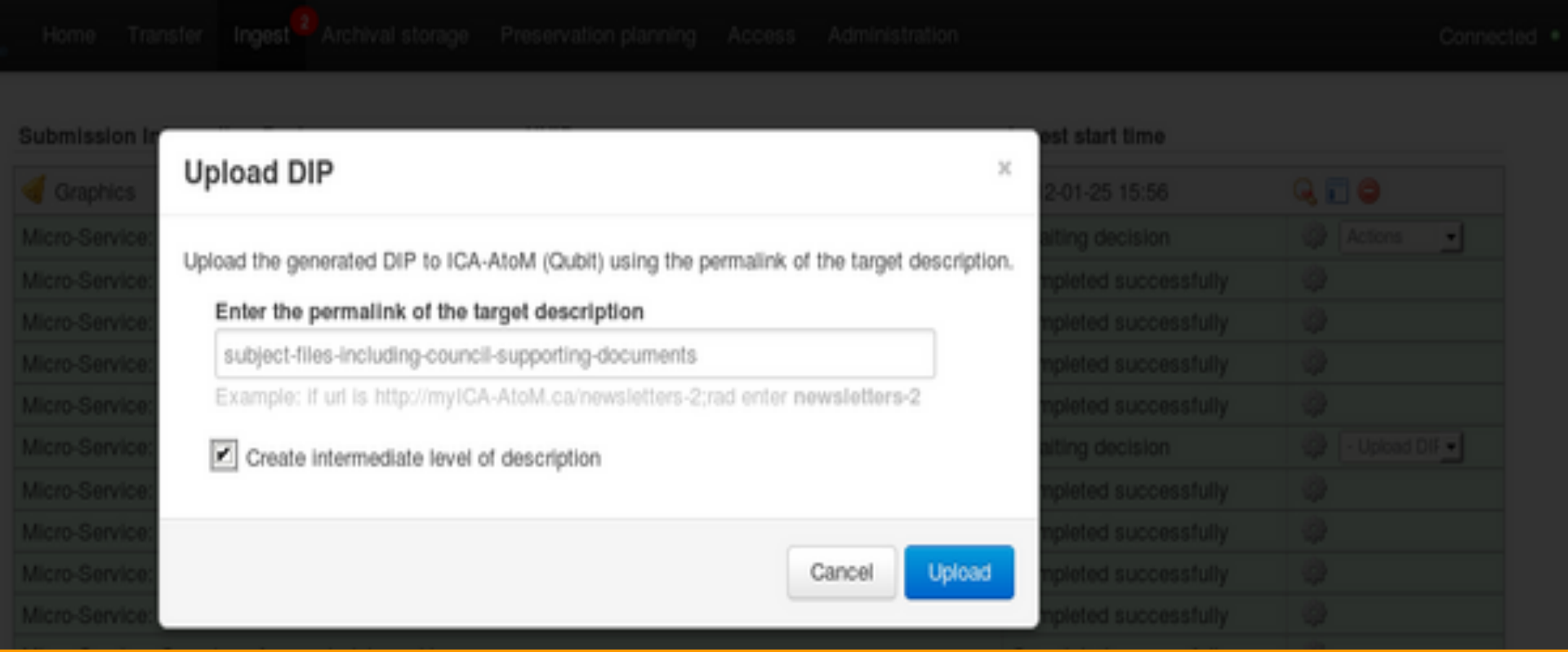
Lists formats by media type and describes their preservation and access normalization paths

Formatting media types and describing their preservation and access normalization paths

In **Preservation Planning**, users can —

- **click on a media type** (such as audio) to see the relevant preservation plan on the Archivematica wiki
- **see the original format** of the object, the normalization action taken by Archivematica, the format the object is now in, and the purpose of the action (either preservation or access)

Archivematica's preservation strategy normalizes files to designated preservation and access formats upon Ingest. The preservation copies are added to the AIP and the access copies are used to generate the DIP for upload to the access system. (Original files are **always** kept, to allow for different actions in the future, such as normalization to different archival formats or emulation.)



5. Access

DIPs generated during Ingest are uploaded to access system

Uploading DIPs to the access system

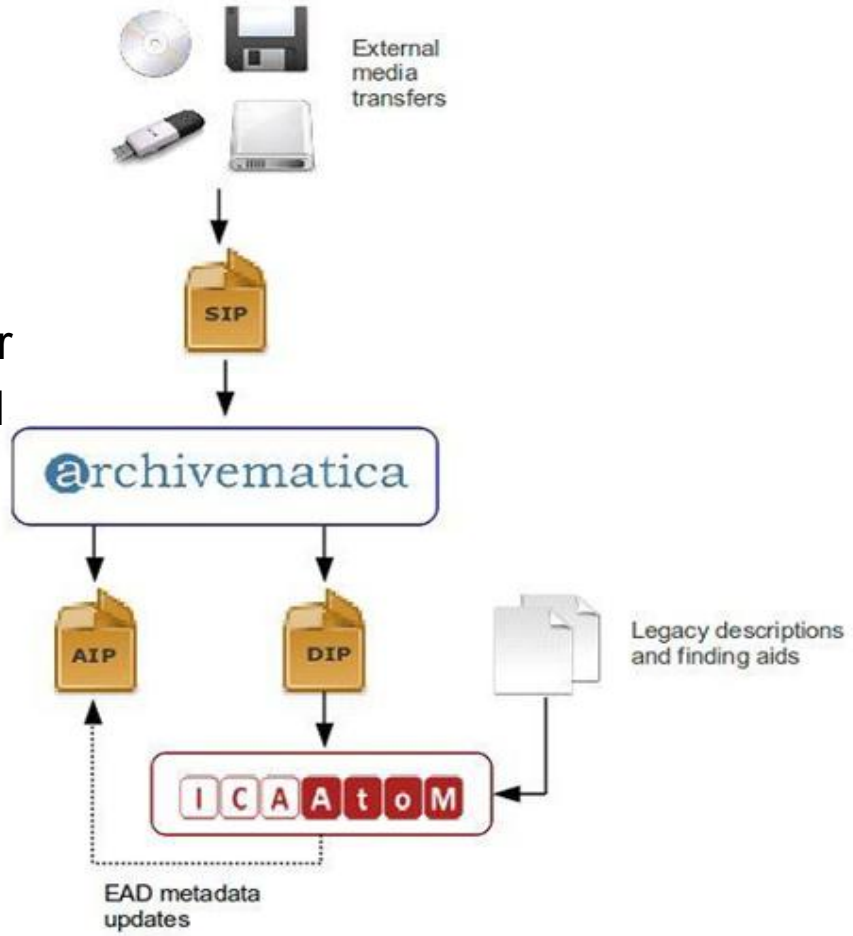
The **Access** process (*which was finished in Ingest)* includes —

- **approving and packing DIP**
- **uploading DIP** to either AtoM, ContentDM or another system *(*AtoM is default and bundled with Archivematica)

The user must have target description ready in AtoM, or the target collection where the DIP will be stored.

In **Access**, users can —

- view a list of AIPs and any DIPs created
- click on DIP URLs to open and view within AtoM or other access system



Administration

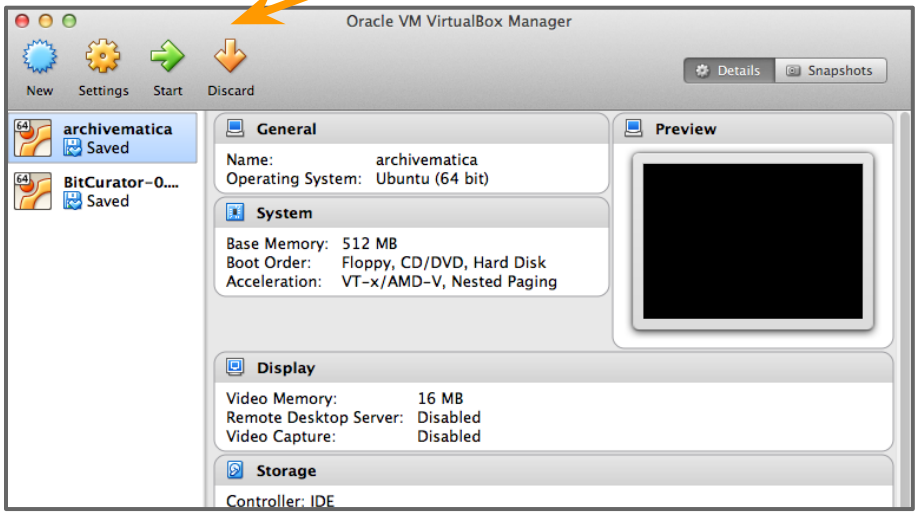
In **Administration**, users can —

- **configure application components** and **manage users** (create, modify access levels and delete)
- **create designated source directories** from which to upload
- **configure server** to allow users to log in to AtoM during the final stages without a password and access the URL where the DIPs will be sent

Using Archivematica: Intro

Open VirtualBox.

Select the Archivematica machine and click **Start**.



Sign in using —

username: demo

password: demo

Using Archivematica: Before Transfer

If you want to include object-level metadata, it should be stored in a **.csv** file in the same folder as your objects. The file should follow Dublin Core, designating DC fields as "**dc.*field***."

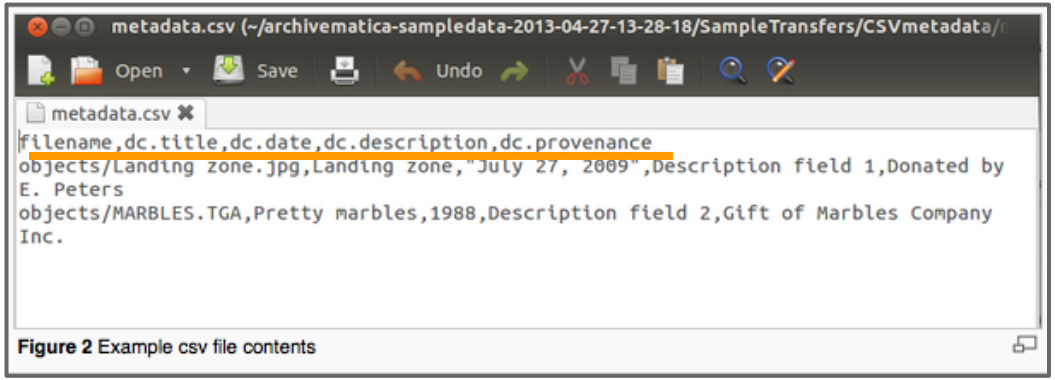


Figure 2 Example csv file contents

https://www.archivematica.org/wiki/UM_Transfer_metadata_import

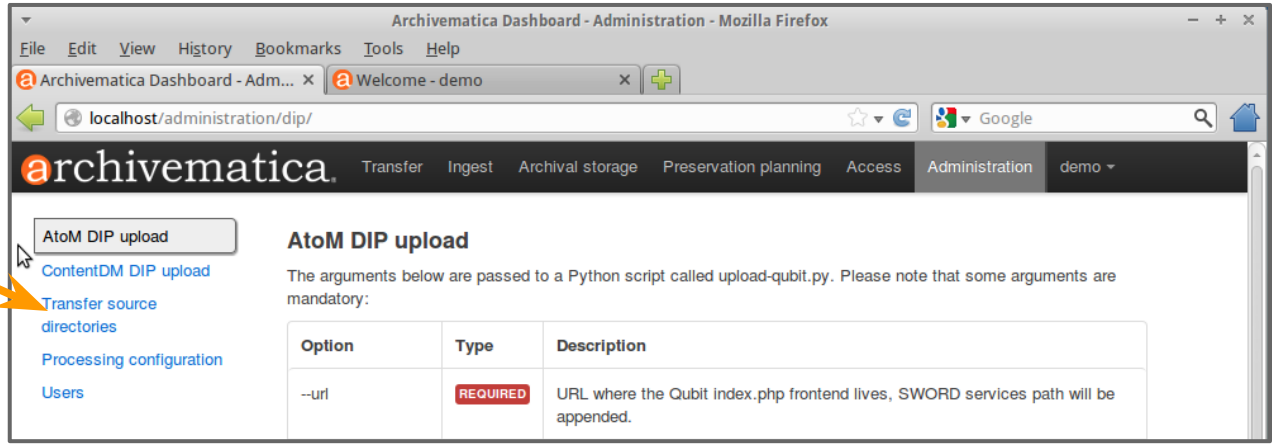
Archivematica only allows for the creation of SIP-level metadata within the program, so object-level metadata has to be treated separately.

Metadata imported as a .csv file will be used in the DIP, will be searchable after storage, and can be edited in an access system like AtoM.

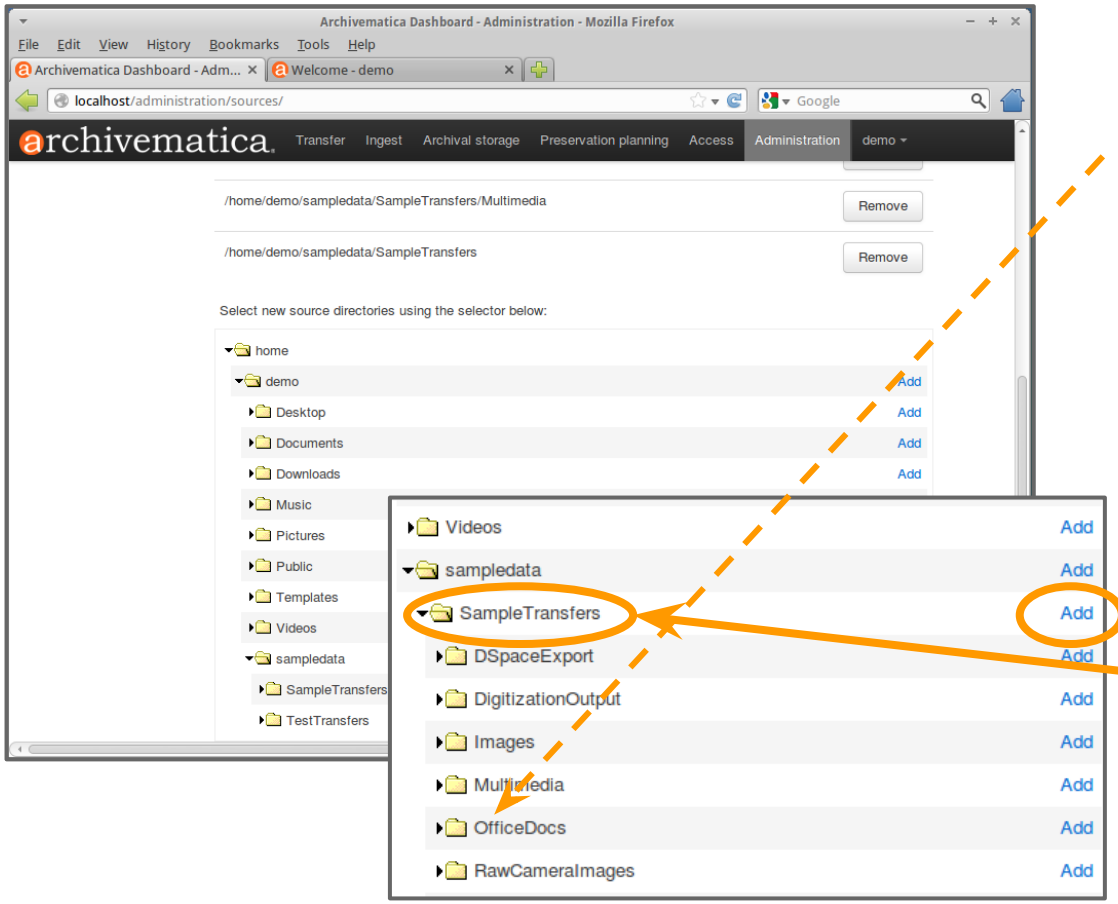
Using Archivematica: Administration Tab

Before you can transfer files, they need to be in a directory accessible to Archivematica.

In the Administration tab, click on **Transfer source directories** to add a source directory.



Using Archivematica: Administration Tab



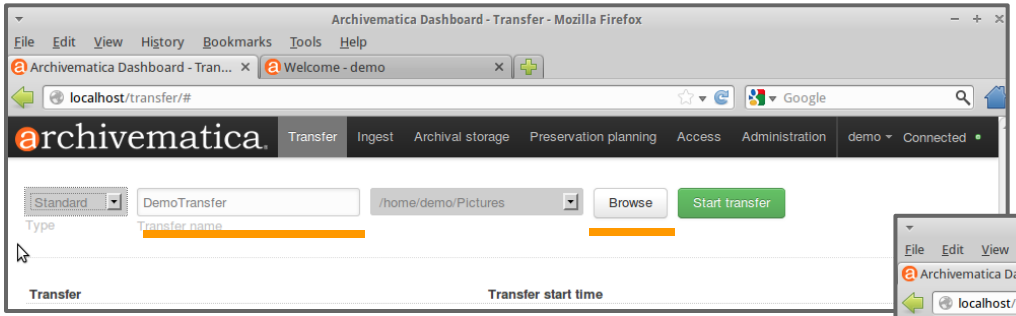
Navigate to the directory you want to add. We are going to use **OfficeDocs** for this transfer:

home > demo > sampledata > SampleTransfers > OfficeDocs

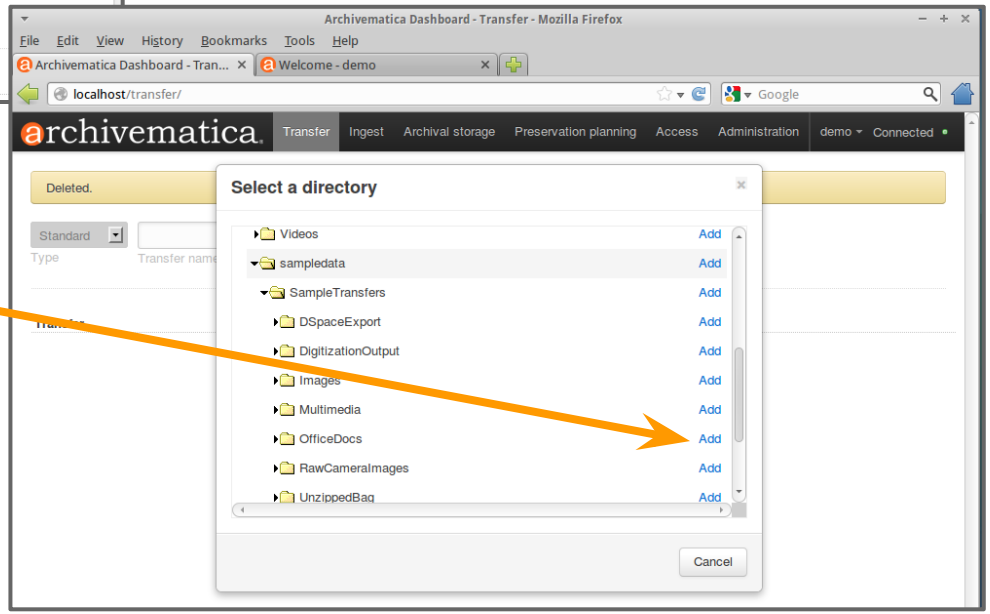
Open OfficeDocs to look at the file structure. Notice that there are object and metadata files.

Click **Add** next to the ***parent folder*** of the directory to be transferred.
In this case, add **SampleTransfers**.

Using Archivematica: Transfer



In the Transfer tab, **name** the transfer and **browse** to the folder you want. (OfficeDocs) Click **Add**.

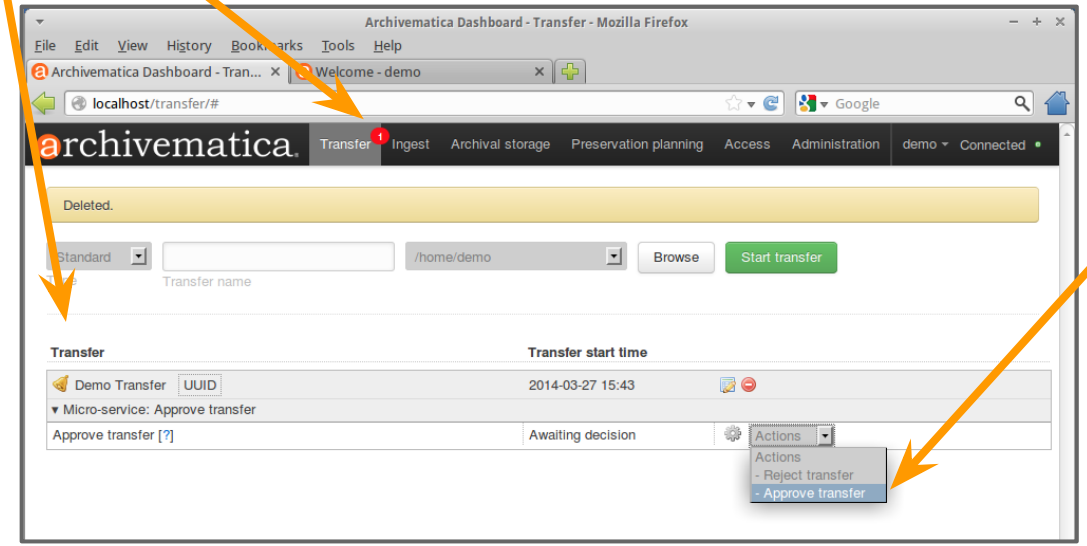


Start the transfer by clicking the green button.

Using Archivematica: Transfer

A **bell** will appear next to the job, prompting you to approve the transfer.

A red indicator also appears next to Transfer in the top bar. It shows the **number of active jobs**. In this case you should see a 



Choose **Approve transfer** in the Actions tab. The transfer will begin to run.

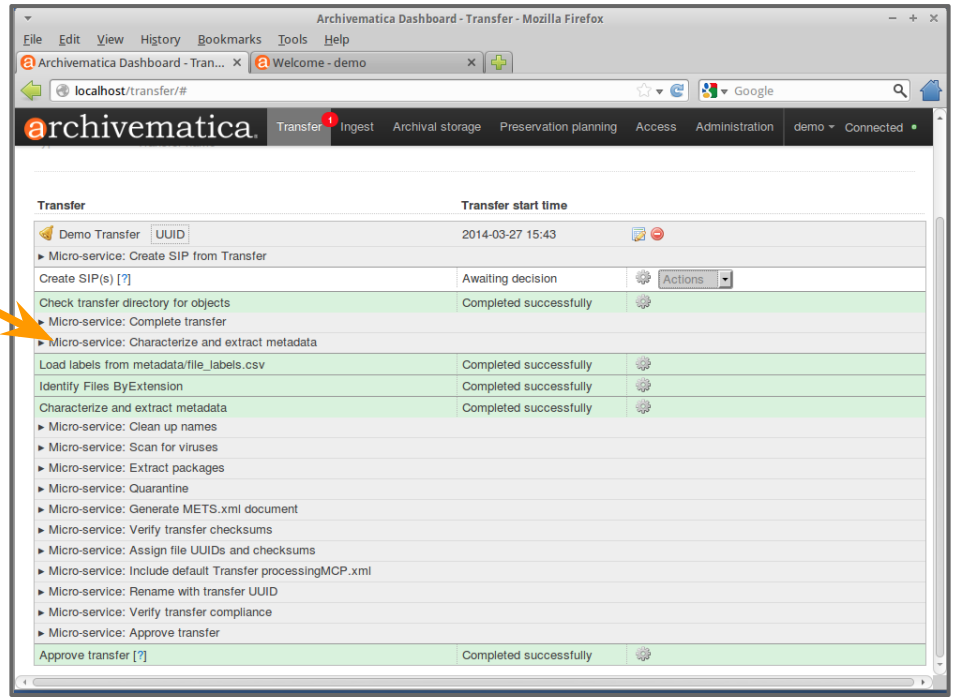
Using Archivematica: Transfer

Microservices appear on the dashboard as they begin.

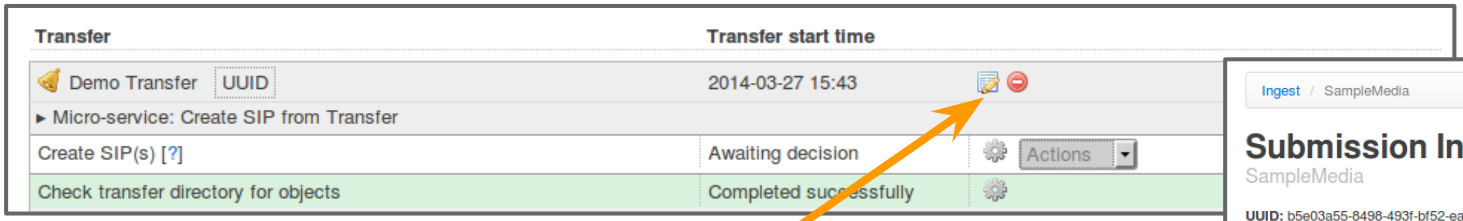
Click on a microservice to see details.

The details of each **microservice** will turn green when they complete, orange when they are in progress, or pink if they fail.

Note: There is no indication given when a complete process is finished. It is a good idea to wait for a few seconds when the service seems to have finished; there may be more microservices running.



Using Archivematica: Transfer



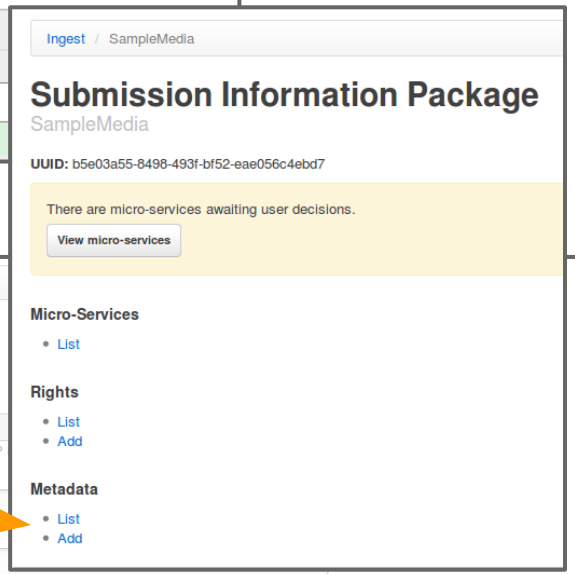
Add SIP-level metadata by clicking the top **Template icon**.

Choose **Add** under Metadata, if it is available.

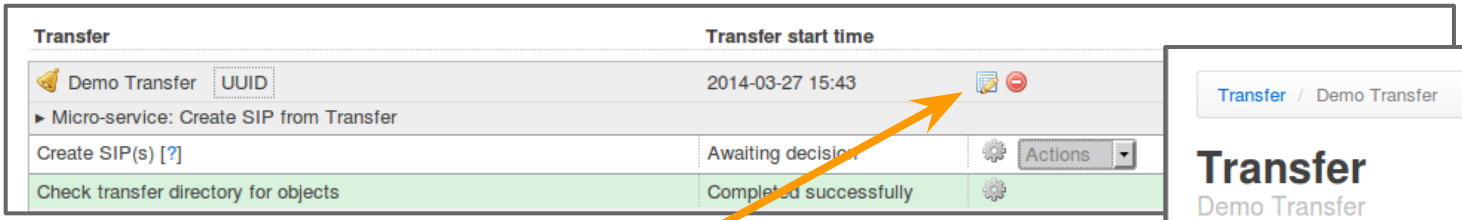
It isn't always, and we haven't determined why. It appears to be a glitch.

Complete the form and click **Create**.

You can view the completed metadata on the **List** page.



Using Archivematica: Transfer

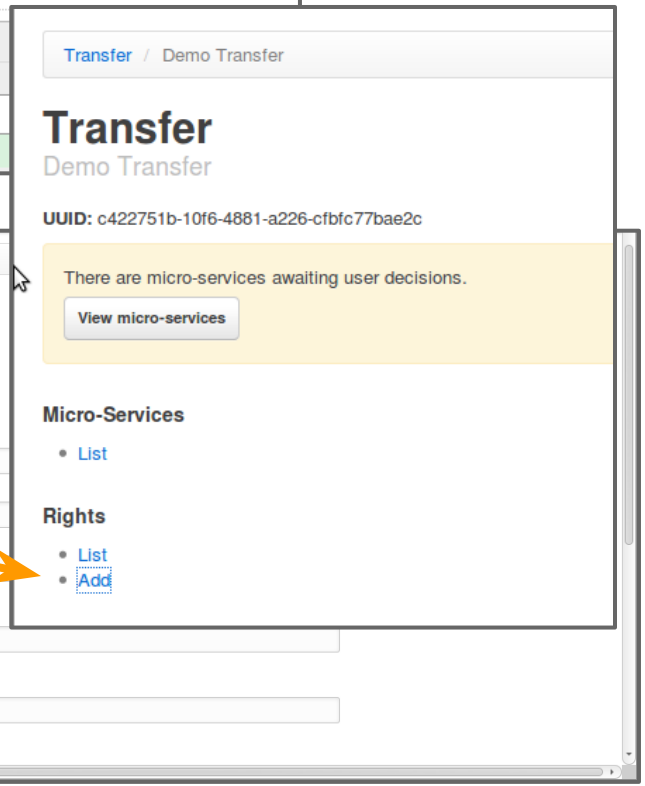


Add SIP-level PREMIS rights by clicking the **template icon** again.

Choose **Add** under Rights.

Complete two pages of rights information and click Save.

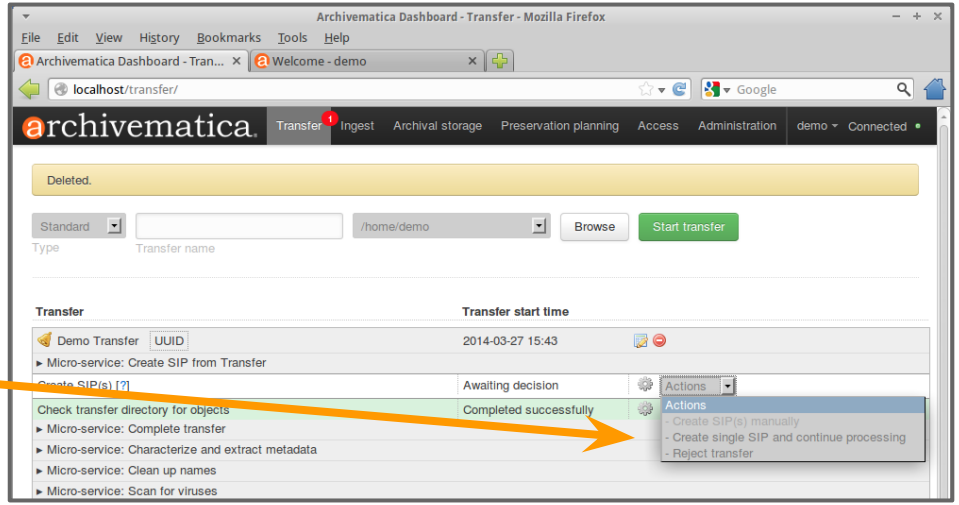
You can view the completed rights information on the **List** page.



Using Archivematica: Transfer

Now that we have created SIP-level metadata, we can package the SIP.

Choose **Create single SIP and continue processing** in the Actions tab.



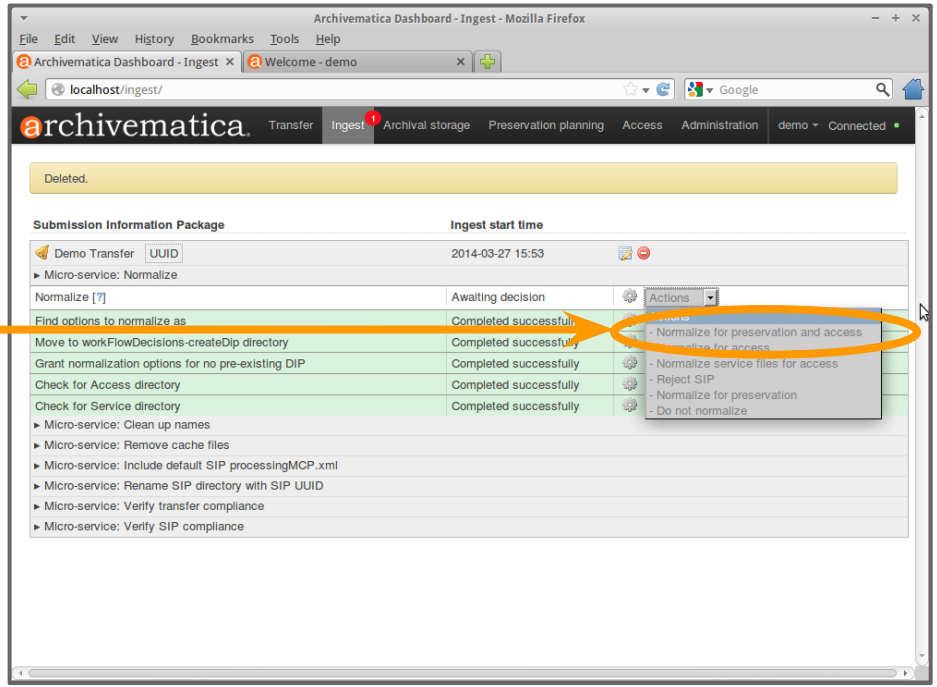
Transfer	Transfer start time	
Demo Transfer UUID	2014-03-27 15:43	
▸ Micro-service: Create SIP from Transfer		
Check transfer directory for objects	Completed successfully	
Move to completedTransfers directory	Completed successfully	
Create SIP from transfer objects	Completed successfully	
Move to processing directory	Completed successfully	
Create SIP(s) [?]	Completed successfully	
Check transfer directory for objects	Completed successfully	

The microservice “Create SIP from Transfer” will run. Each step turns green when complete. When the SIP is finished, the red indicator in the top bar will move to Ingest.

Using Archivematica: Ingest

Now that the SIP is packaged with its associated metadata, we can create a DIP and AIP.

Go to the Ingest page, and select a **normalization** option from the Actions tab —



Normalize for preservation and access: creates preservation copies of the objects plus access copies which will be used to generate the DIP.

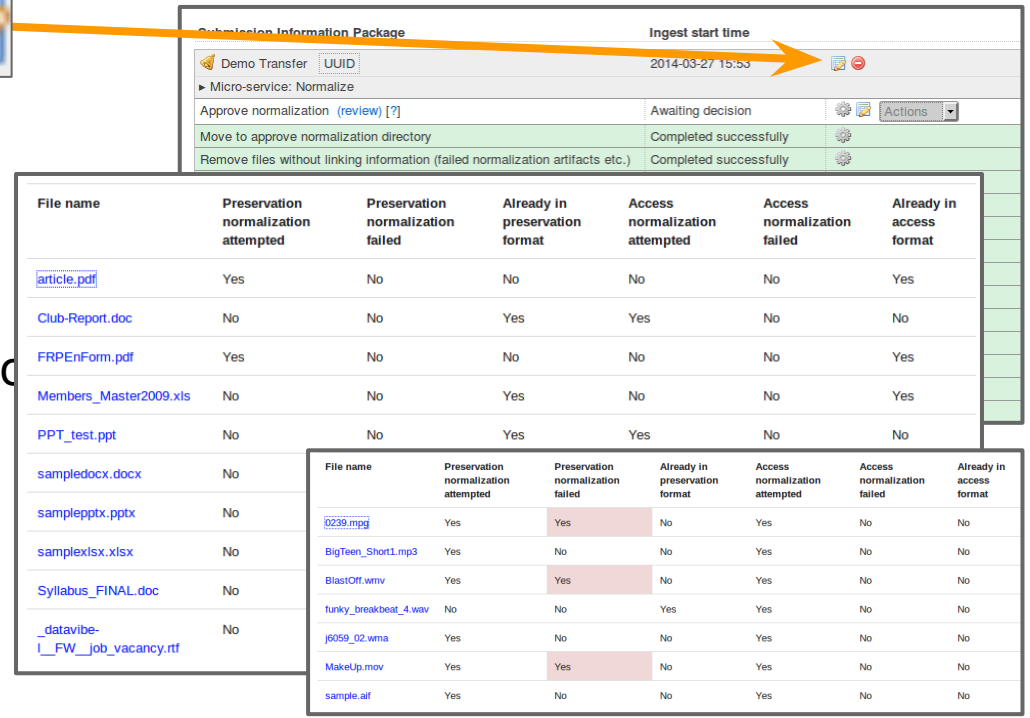
Normalize for access: no preservation copies are created. Creates access copies which will be used to generate the DIP.

Normalize for preservation: creates preservation copies. No access copies are created and no DIP will be generated.

Using Archivematica: Ingest

Click on the **Report icon** to see which objects were normalized for preservation and/or access. The new, normalized files will be used to make the DIP.

You will see red cells in the report if there was a problem with normalizatio You can continue to process the SIP with the non-normalized files or retry normalization using a different tool (Droid vs. Jhove).



Submission Information Package		Ingest start time
Demo Transfer UUID		2014-03-27 15:53
▸ Micro-service: Normalize		
Approve normalization (review) [?]	Awaiting decision	Actions
Move to approve normalization directory	Completed successfully	
Remove files without linking information (failed normalization artifacts etc.)	Completed successfully	

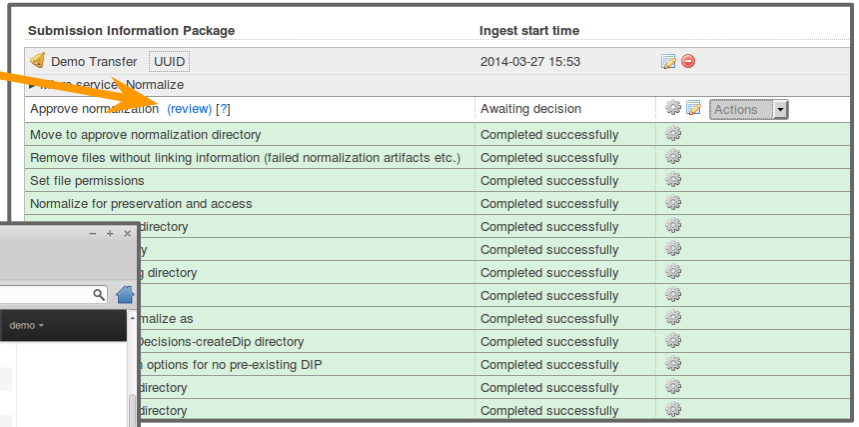
File name	Preservation normalization attempted	Preservation normalization failed	Already in preservation format	Access normalization attempted	Access normalization failed	Already in access format
article.pdf	Yes	No	No	No	No	Yes
Club-Report.doc	No	No	Yes	Yes	No	No
FRPEnForm.pdf	Yes	No	No	No	No	Yes
Members_Master2009.xls	No	No	Yes	No	No	Yes
PPT_test.ppt	No	No	Yes	Yes	No	No
sampledocx.docx	No					
samplepptx.pptx	No					
samplexlsx.xlsx	No					
Syllabus_FINAL.doc	No					
_datavibe-l__FW__job_vacancy.rtf	No					

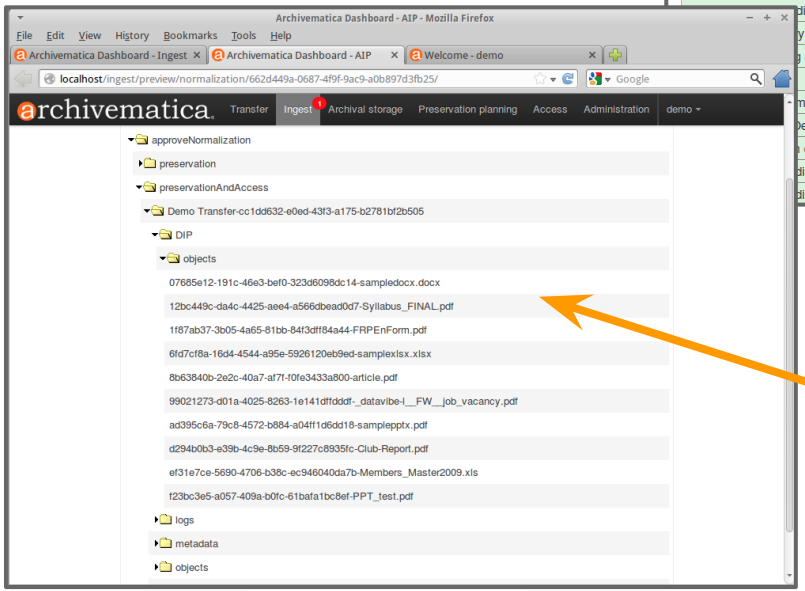
File name	Preservation normalization attempted	Preservation normalization failed	Already in preservation format	Access normalization attempted	Access normalization failed	Already in access format
0239.mpg	Yes	Yes	No	Yes	No	No
BigTeen_Short1.mp3	Yes	No	No	Yes	No	No
BlastOff.wmv	Yes	Yes	No	Yes	No	No
funky_breakbeat_4.wav	No	No	Yes	Yes	No	No
j6059_02.wma	Yes	No	No	Yes	No	No
MakeUp.mov	Yes	Yes	No	Yes	No	No
sample.aif	Yes	No	No	Yes	No	No

Droid and Jhove are File Information Tool Sets that identify and validate files and extract technical metadata (http://code.google.com/p/fits/wiki/tools).

Using Archivematica: Ingest

Click **Review** to access the normalized files in a new tab.

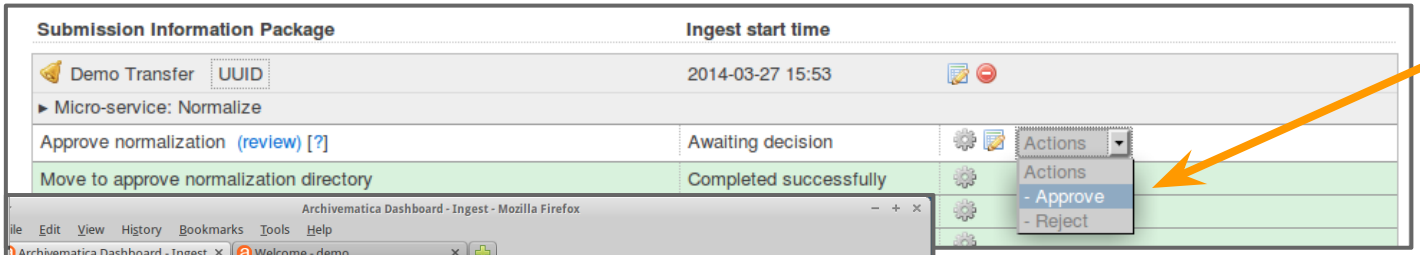




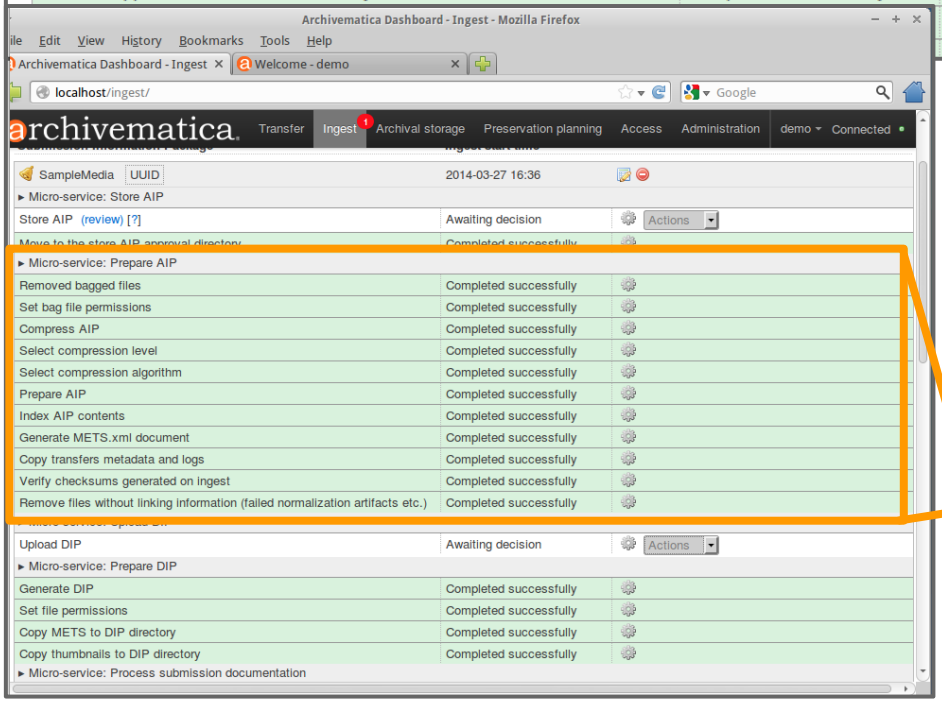
Notice that the file **names** now contain a UUID*, and some files have new **formats**.

*Universally Unique Identifier, generated by MD5 or SHA-1 hash.

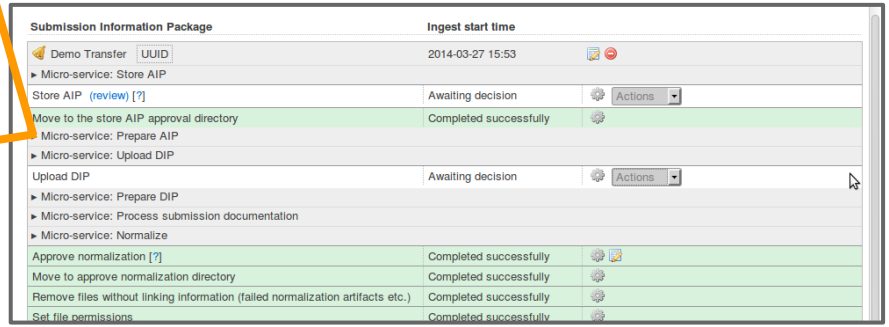
Using Archivematica: Ingest



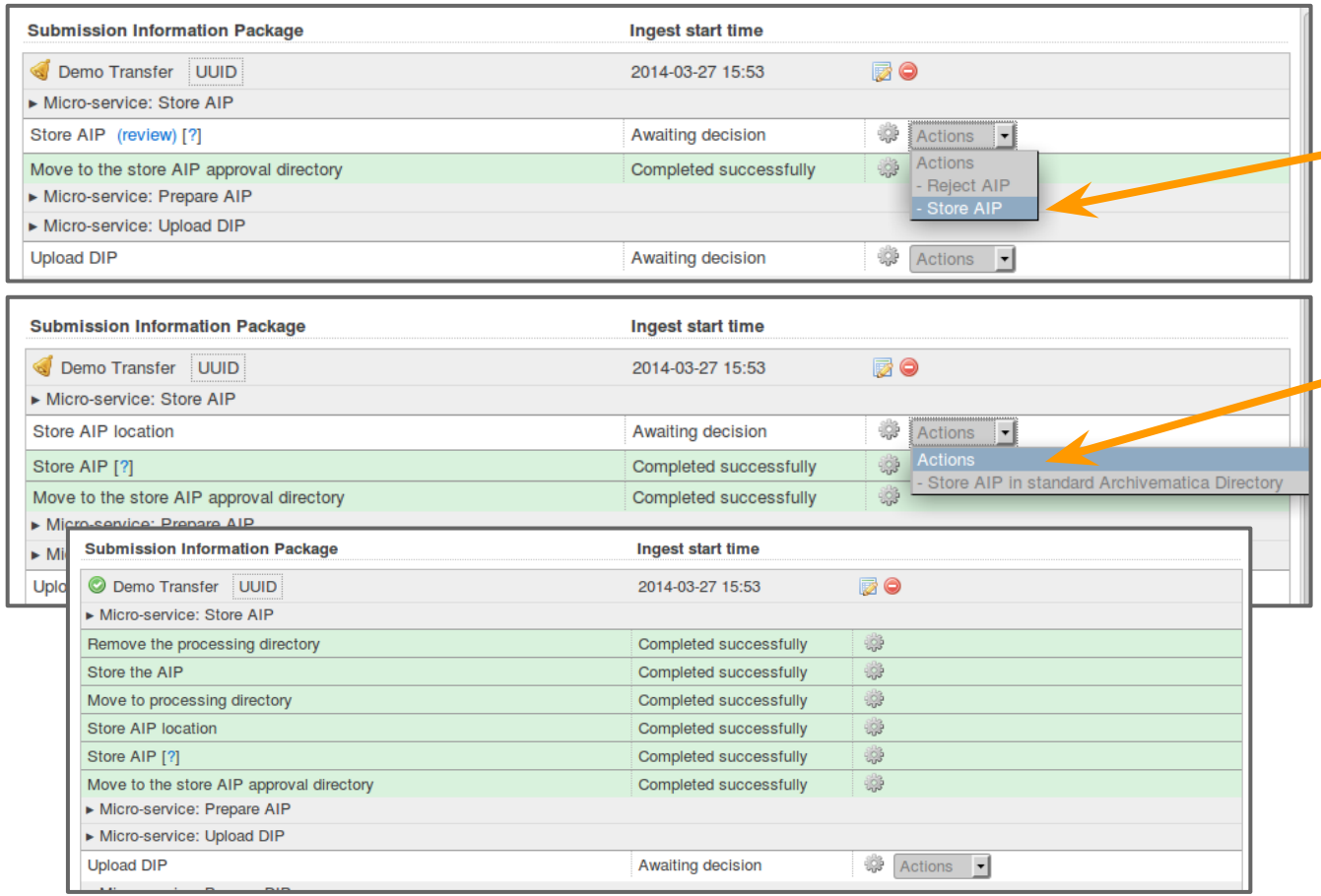
Approve the normalization in the Action tab.



Now the SIP will go through several **microservices**, which create a METS data file, generate a DIP, and package an AIP.



Using Archivematica: Ingest



In the Action tab, select **Store AIP** to send it to Archival Storage.

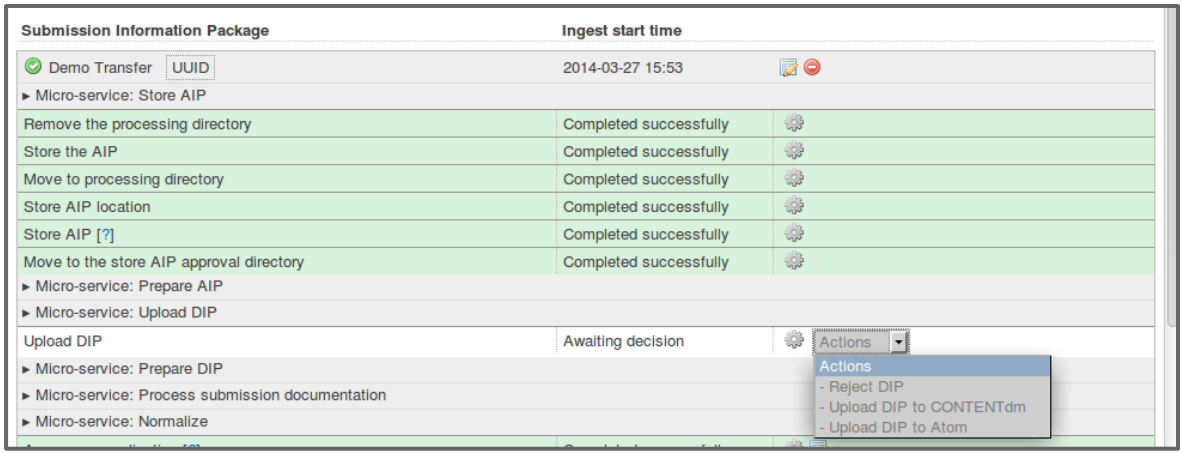
Then select a **location for the AIP** in the Action tab. Our only option in the demo is the **standard Archivematica Directory**, but the directories can be customized.

The microservice will finish running.

Using Archivematica: Ingest to Access

Almost done!

We have created and stored the AIP. We have created the DIP. Now we just need to move the DIP to an access system. You can upload DIPs to AtoM or ContentDM. ContentDM must be configured separately, but AtoM is already integrated.



You must send a DIP to an existing description in AtoM, so for this demonstration we will just pretend.

Use the Actions tab to select **Upload DIP to AtoM**.

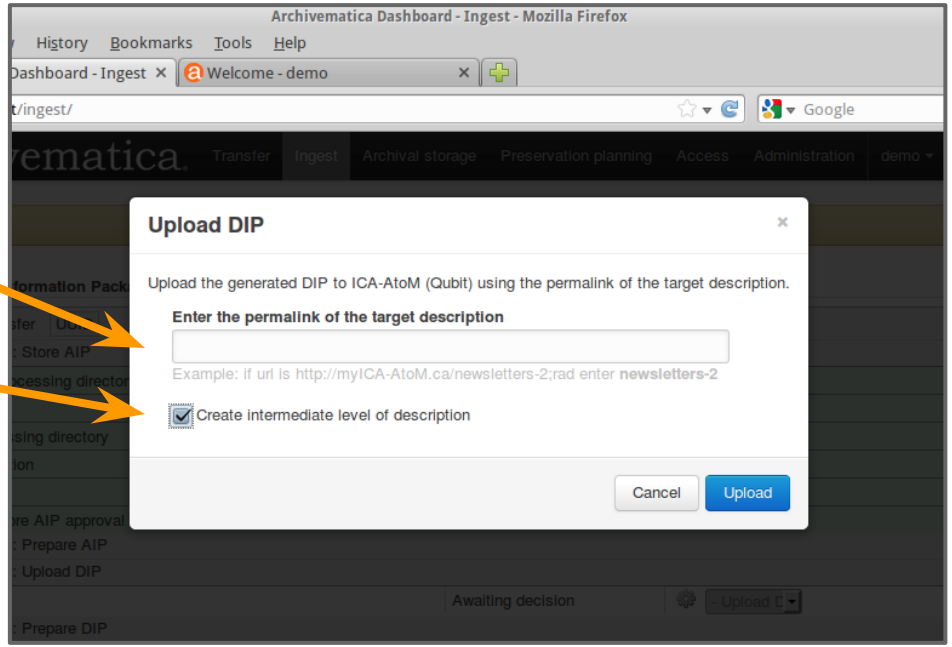
Using Archivematica: Ingest to Access

If you had a destination in AtoM, then you would:

Enter the **permalink** to the AtoM description in the dialog box.

Here you can also choose to create **layered metadata from the Dublin Core description** created on Ingest from your .csv file. Without this step, each object will be stored as a child of the DIP, without any hierarchy.

Click **Upload** to send the DIP and its associated metadata to the AtoM directory.



Using Archivematica: Ingest to Access

Congratulations!

You are done processing!

Once the DIP is uploaded, you can go to its AtoM description via the Access tab.

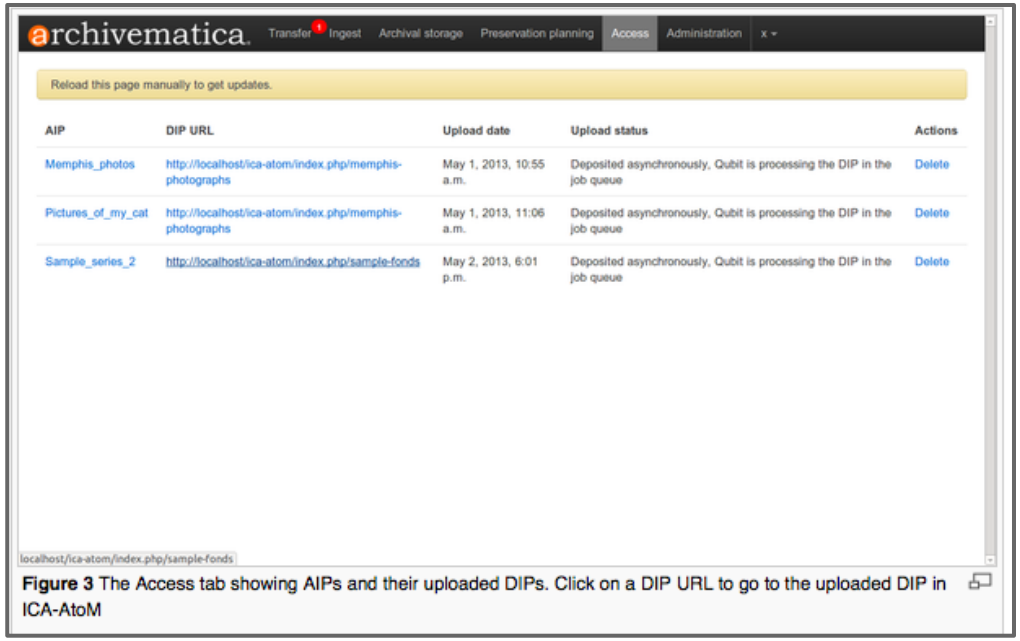


Figure 3 The Access tab showing AIPs and their uploaded DIPs. Click on a DIP URL to go to the uploaded DIP in ICA-AtoM

https://www.archivematica.org/wiki/UM_access

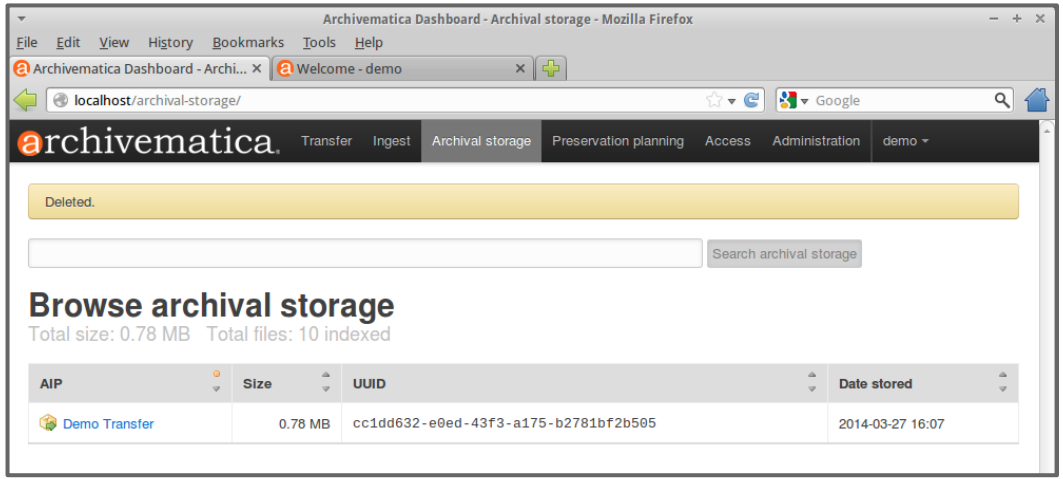
Using Archivematica: Archival Storage

Archival Storage is a table of AIPs, which were stored at the end of Ingest.

Here you can view the **UUID** for each AIP, generated during storage.

The table also shows the size and date of ingest for each AIP.

Clicking on the **name** of an AIP allows you to download or open the zipped file.



Using Archivematica: Archival Storage

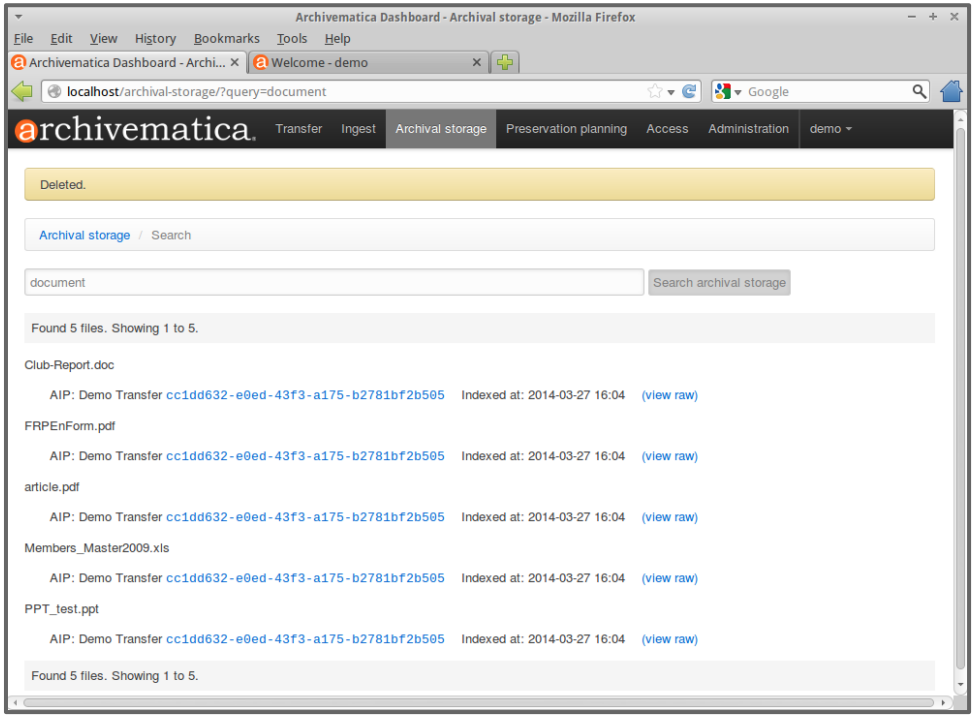
The search bar is (sort of) powerful. It searches the entire index.

The index includes **all of the textual contents** of the AIPs *and* the METS metadata for each AIP.

The results return **AIPs** and/or their **constituent parts** (e.g., a file/object), along with the corresponding UUIDs.

The search bar will not return image results without metadata, or results based on PREMIS metadata.

It also does not indicate where the search term was found.



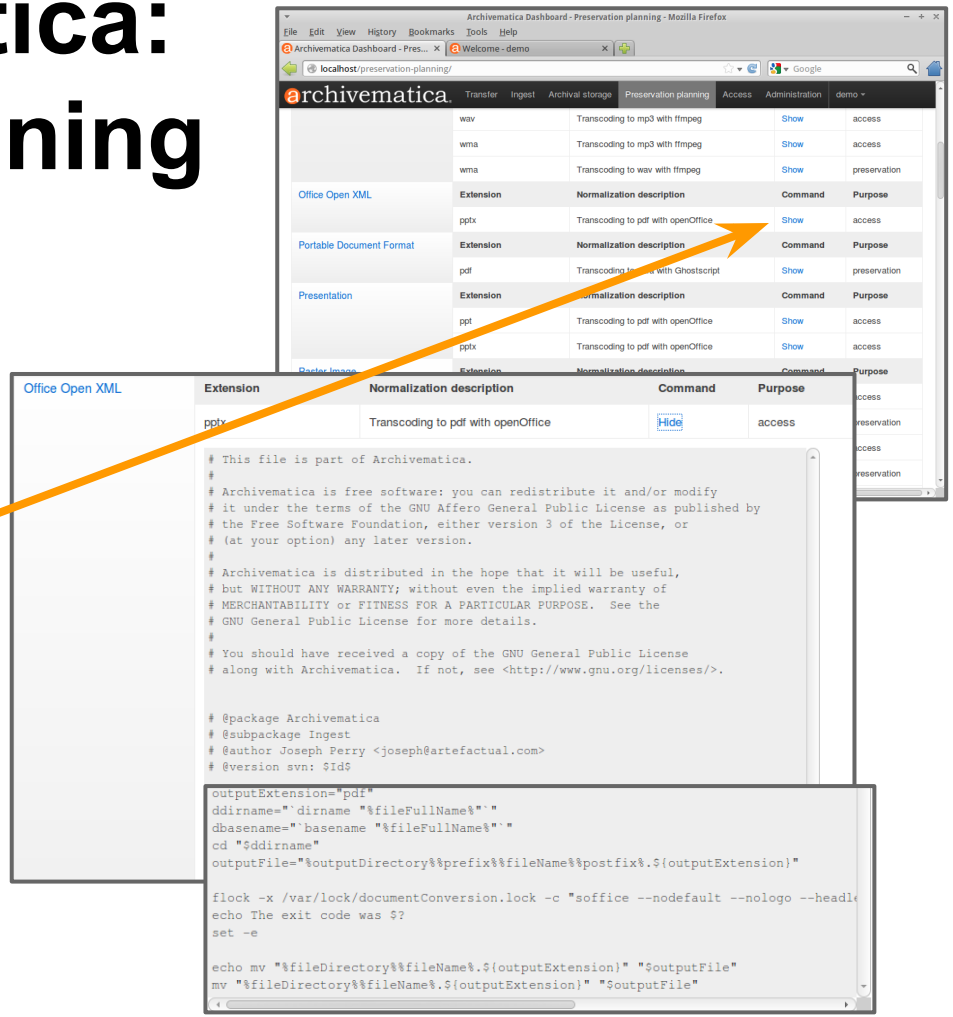
Using Archivematica: Preservation Planning

The Preservation Planning page allows the user to manage the normalization process.

Each format type is a link to the corresponding Archivematica wiki entry.

Click on **Show** to see the normalization command for a file type.

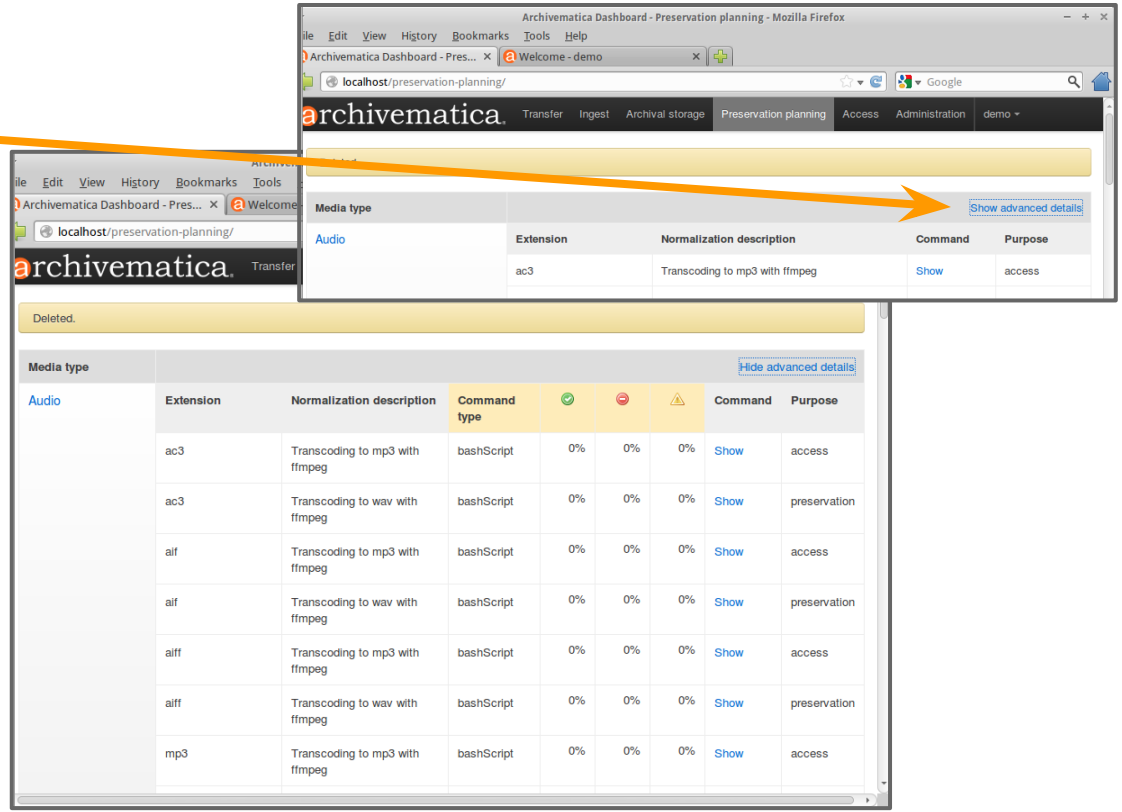
Commands can be edited in mySQL. Future versions aim to have editable commands within the Preservation Planning tab.



Using Archivematica: Preservation Planning

Click on **Show advanced details** to see the success rates for normalization.

Hover over **percentages** to see a success/attempts ratio.



Using Archivematica: Wrap-up

The process is quick when you follow an established workflow.

Customization requires more time and expertise, but it is possible to change almost all the steps.

Responses?

Questions?

An Overview of Our Evaluation

- **Comparing** open-source to commercial software
- **Defining and delineating** archival tools vs. archival management systems: Does Archivematica meet system benchmarks?
- **Strengths and weaknesses** of Archivematica broadly and technically
- **Who is the ideal Archivematica user?** What features should an institution possess in order to successfully implement and adopt Archivematica?

Open Source vs. Commercial Software

The pros of open-source software include —

- less licensing and update related pains
- institutions have more freedom to customize and localize installations
- encourages a collaborative community of users
- change becomes a more democratic and consented upon process

The **cons** of open-source software include —

- supporting documentation can be weak or unreliable
- more onus is placed on users/IT staff to troubleshoot
- accrual of recurring costs versus commercial one-time licensing fees

The divisive reaction to the Oracle and Sun Microsystems merger within the open-source community reveals that this type of software should not be overly idealized.

Archival Tools vs. Archival Management Systems

- **Archival tools** are designed to deal only with a specific task or part of a workflow (e.g., XML editors or normalization).
- **Archival management systems** are fully integrated suites developed to handle all aspects of the archival workflow (e.g., Archivematica or Archivists' Toolkit).

Hence, the evaluation phase of entire management systems is far more critical.

According to Liso Spiro's CLIR Report in 2009, ***ideally*** **archival management systems should —** be integrated, portable, and powerful, but not intimidating; support the inclusion of legacy data; be able to migrate data; have a web-publishing function and collection management capabilities; help in setting up processing priorities; and comply with standards.

Strengths of Archivematica

1. Archivematica is based on OAIS-ISO model and supported by various Trusted Digital Repository and Digital Asset Management Systems.
2. Archivematica's web-based dashboard gives it machine flexibility.
3. Archivematica supports the digital preservation strategies of migration, emulation, and normalization, and archival institutions are assisted in the process of selecting which strategy best meets their needs via an evolving format registry policy.
4. Archivematica supports linking to digital objects and provides an extensive array of microservices.

Archivematica is *very attuned* to the pulse of and rapidity of development within the digital preservation community, and from an agnostic perspective, it is *an excellent piece of software*. Archivematica was envisioned to *grow and change* with digital preservation, itself. Archiving moves at a relatively glacial pace, and one must contemplate if your institution is adequately equipped to meet the challenge of selecting Archivematica.

Weaknesses of Archivematica

1. Archivematica has no built-in automatic upgrade component, meaning that adopting a newer version requires a complete reinstallation of that version every time.
2. Archivematica does not check for duplicate files within its system as it runs through the workflow.
3. Institutions cannot create custom metadata fields beyond the Dublin Core and PREMIS pre-packaged standards.
4. Access rights cannot be customized within a single collection.

Deciding on Archivematica

***Minimum* Requirements for Selecting Archivematica —**

- extensive RAM
- a virtual, and ideally, dedicated server
- working knowledge of Virtualbox
- comfortability with Linux and the command line
- staff willing to learn and engage with the software and consider Archivematica a long-term investment

Deciding on Archivematica

Unfortunately, open source does not always mean free. To get further help with installation, maintenance or training to use the system, Archivematica requires committing to hefty financial packages.

Annual Maintenance Agreement

$24,999

per annum

archivematica	free	basic	premium
Installation	$0	$27,999	$44,999
free software license (AGPL3)	✓	✓	✓
installation documentation	✓	✓	✓
community forum support	✓	✓	✓
installation technician	✗	on-site plus travel expenses	on-site plus travel expenses
dedicated telephone and email support	✗	3 months plus optional annual maintenance	6 months plus optional annual maintenance
secure remote support	✗	3 months plus optional annual maintenance	6 months plus optional annual maintenance
IT department liaison	✗	✓	✓
storage and network integration	✗	✓	✓
backup procedure review	✗	✓	✓
scalability testing & optimization	✗	✓	✓
administrator and end-user training	✗	✓	✓
integration with supported systems	✗	$4,999 per system	✓

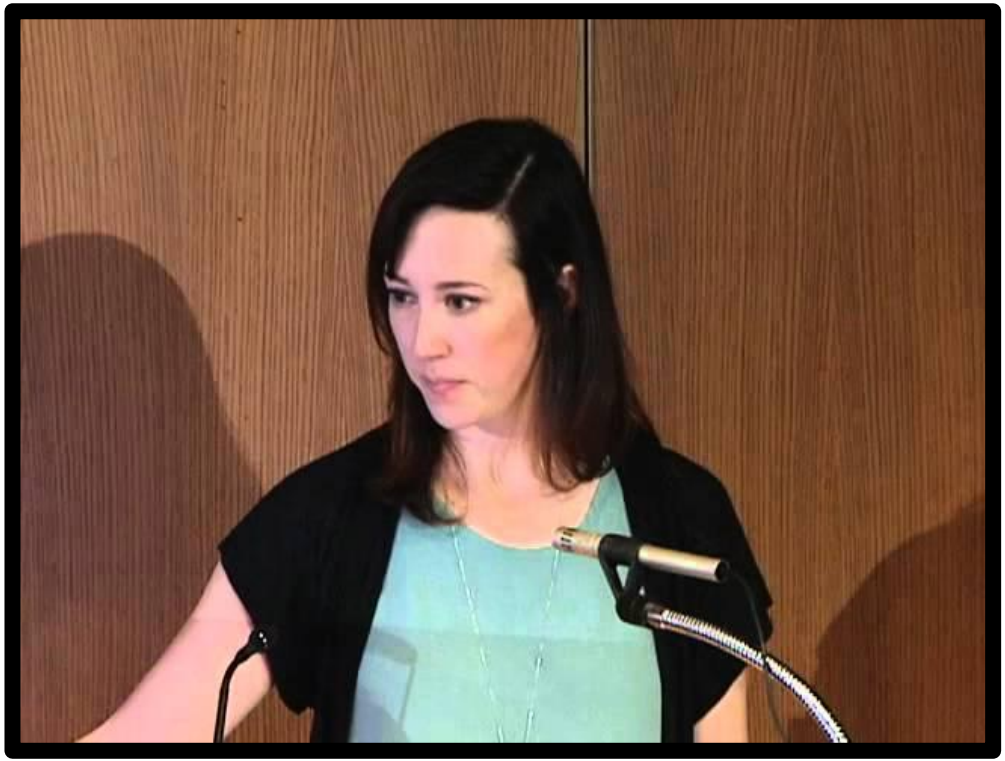
A Commitment to the Archival Community

In 2012, Courtney Mumma, Archivematica's Project Manager, asks users to reconsider the "free" element of open-source software.

Yes, while Archivematica is free, it requires maintenance and upkeep, like a "free" kitten.

"You have to pet it and make it purr."

https://www.youtube.com/watch?v=XfwxWTVsHY0



Who's the Ideal Archivematica User?

Archivematica is **not** merely a software application that facilitates a particular aspect of the archival workflow, but a fully integrated management system, that is equipped to handle all of an institution's preservation and access needs.

Archivematica is **best suited** to serve institutions with **external storage** and **server capabilities** who seek to align Archivematica's preservation policies and technological sophistication with their institution's mission statement and collecting focus.

Sources

Archivematica. (April 24, 2013). *User Manual 0.9.* Retrieved from https://www.archivematica.org/wiki/User_manual_0.9

Archivematica. (April 10, 2014). *Administrator Manual 1.0.* Retrieved from https://www.archivematica.org/wiki/Administrator_manual_1.0#Format_Policy_Registry_.28FPR.29

Archivematica. (July 30, 2013). *Virtual Appliance Instructions.* Retrieved from https://www.archivematica.org/wiki/Virtual_appliance_instructions.

Bradley, K., J. Lei, and C. Blackall (2007). Towards an Open Source Repository and Preservation System: Recommendations on the Implementation of an Open Source Digital Archival and Preservation System and on Related Software Development. Paris: UNESCO. Retrieved from http://freegovinfo.info/node/1415

Caplan, P. (February 1, 2009). *Understanding PREMIS.* Retrieved from http://www.loc.gov/standards/premis/understanding-premis.pdf

Craig, Barbara Lazenby (2000). A Look at a Bigger Picture: The Demographic Profile of Archivists in Canada Based on a National Survey. *Archivaria* 49. Retrieved from http://journals.sfu.ca.ezproxy.lib.utexas.edu/archivar/index.php/archivaria/article/view/12738/13919

Dublin Core Metadata Initiative. (June 14, 2012). *Dublin Core Metadata Element Set, Version 1.1.* Retrieved from http://dublincore.org/documents/dces/

McKenna, Patti. How Virtual Computing Works. Retrieved April 16, 2014, from http://computer.howstuffworks.com/how-virtual-computing-works2.htm

Mumma, C. (2012). Integrating Legacy Repository Systems with Archivematica. Presentation at the NDIIPP Briefing, *Library of Congress*. Retrieved from http://www.youtube.com/watch?v=XfwxWTVsHY0.

Mumma, C., Dingwall, G., & Bigelow, S. (2011). A First Look at the Acquisition and Appraisal of the 2010 Olympic and Paralympic Winter Games Fonds: or, SELECT * FROM VANOC _Records AS Archives WHERE Value="true";. *Archivaria, 72*(72). Retrieved March 29, 2014, from http://journals.sfu.ca/archivar/index.php/archivaria/article/view/13361/14665

Sources

Paradigm. (January 2, 2008). Introduction to OAIS. Retrieved from http://www.paradigm.ac.uk/workbook/introduction/oais-information.html

Spiro, Lisa (2009). Spiro 2009 Report: Archival Management Software A Report for the Council of Library and Information Resources. *Council on Library and Information Resources. Retrieved from* http://www.clir.org/pubs/reports/spiro

The Library of Congress. (September 20, 2011). *METS: An Overview & Tutorial.* Retrieved from http://www.loc.gov/standards/mets/METSOverview.v2.html

The Library of Congress. (March 2008). *PREMIS Data Dictionary for Preservation Metadata.* Retrieved from http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

Van Garderen, P. (2010). Archivematica: Using Micro-Services and Open-Source Software to Deliver a Comprehensive Digital Curation Solution. *iPRES 2010*, 145. Retrieved from https://www.archivematica.org/download/2010-09-20-iPress-Vienna-Archivematica-VanGarderen.pdf

Van Garderen, P. and C. Mumma (2013). Realizing the Archivematica vision: delivering a comprehensive and free OAIS implementation. *10th International Conference on Preservation of Digital Objects.* Lisbon, Portugal. Retrieved from http://purl.pt/24107/1/iPres2013_PDF/Realizing%20the%20Archivematica%20vision%20delivering%20a%20comprehensive%20and%20free%20OAIS%20implementation.pdf

Van Garderen, P, P. Jordan, T. Hooten, C. Mumma and E. McLellan, The Archivematica Project, Meeting Digital Continuity's Technical Challenges. Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Van_Garderen_et_al_26_Workshop1.pdf