



Digital POWRR How-to Tech Tutorial

Archivematica Walk Through

Task 1 - Getting Started

1. For the purposes of the POWRR Institute, we will be utilizing the Sandbox environment provided by Archivematica. Some of Archivematica's features are not available in the Sandbox. If you install Archivematica locally, some of the steps in the walkthrough may change. To login to the Archivematica Sandbox, open a web browser and navigate to <http://sandbox.archivematica.org/>

A screenshot of the Archivematica login interface. The page has the Archivematica logo at the top. Below it is a login form with two input fields: "Username" containing "demo@example.com" and "Password" containing a series of dots. A "Log in" button is located at the bottom right of the form.

Enter demo@example.com as your username, and demodemo as the password.

2. Maximize your browser screen. Many laptops will not display the entire screen by default.

Generally, before you can start using Archivematica, you need to assemble content for Archivematica to ingest and process. Archivematica is able to read data from a number of different storage systems and repositories. Content can be entirely unstructured to start, or it can be pre-arranged in a couple of different ways to facilitate specific workflows. For the purposes of the POWRR Institute, we will be using a set of sample data that has been assembled by Artefactual for use in testing and demonstrating the application.

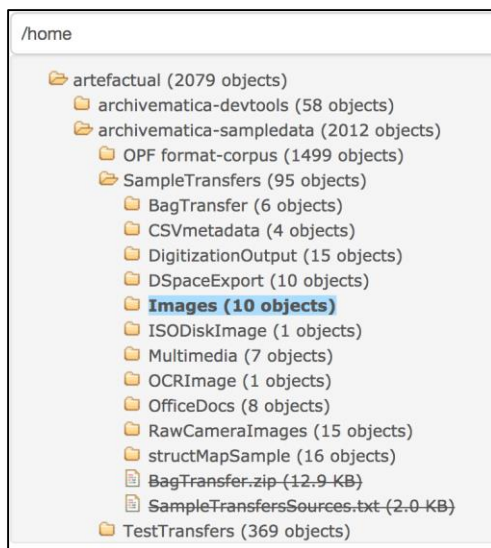
Task 2 - Starting a Standard Transfer

1. Go to the Transfer tab.
2. Make sure 'Standard' is selected from the transfer type box.

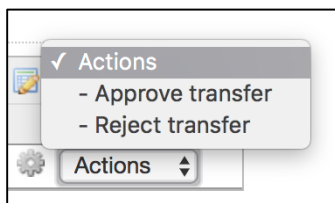
Standard			Browse	Start transfer
Transfer type	Transfer name	Accession no.		

Type in a name for your transfer. Since we are all working in the same sandbox, and can see one another's transfers, it might be best to type in some portion of your name.

3. Click on browse to look through the available content. Make sure to click directly on the small folder icon to expand the directory trees.
4. Find the *images* directory in SampleTransfers and click Add.



5. Click on the green *Start transfer* button to start the transfer.
6. When prompted, make decisions as you wish; however, don't select anything that would stop the transfer (i.e. anything that says "Reject"). For more information about the decision points, check out Archivematica's [transfer tab documentation](#).



7. When you reach the *Identify file format* micro-service, stop and move on to the next section.

Task 3 - Review File Formats

1. You do not need to consult the command for every job, but we do recommend taking a look at the output of the *Identify file format* micro-service. One of Archivematica's most important jobs is to identify file formats and then preserve those files as best as possible.
2. When you are prompted, select Siegfried as your file format identification tool.

▶ Micro-service: Identify file format		
Job: Select file format identification command	Awaiting decision	⚙️ Actions
Job: Move to select file ID tool	Completed successfully	- Identify using Fido - Skip File Identification - Identify using Siegfried - Identify by File Extension
▶ Micro-service: Clean up names		
▶ Micro-service: Generate transfer structure report		
▶ Micro-service: Scan for viruses		

- Once file format identification is complete, click on the gear icon to the right of the job name to open the job page. On the job page, under the heading STDOUT, you will see information similar to the following:

```

STDOUT
IDCommand UUID: 9d2cefc1-2bd2-44e4-8d55-6cf8151eecff
File: (0a137964-58bd-4cf2-a1bb-6e197522683c) /var/archivematica/sharedDirectory/watchedDirectories/workFlowDecisions/selectFormatIDToolTransfer/testtesttest-4012b2f5-fe0a-40b3-8baa-17508cdc1d8d/objects/pictures/MARBLES.TGA
fmt/402

Command output: fmt/402
/var/archivematica/sharedDirectory/watchedDirectories/workFlowDecisions/selectFormatIDToolTransfer/testtesttest-4012b2f5-fe0a-40b3-8baa-17508cdc1d8d/objects/pictures/MARBLES.TGA identified as a Truevision TGA Bitmap

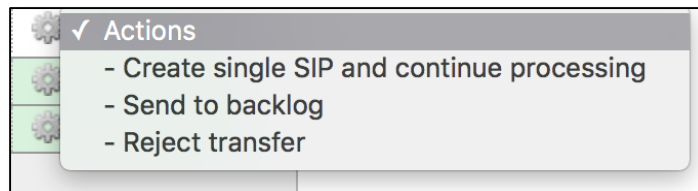
```

- The above tells us that the file MARBLES.TGA was identified as a Truevision TGA Bitmap 2.0. Archivematica uses [PRONOM](#), a registry of technical information maintained by the National Archives of the UK, for file identification and to inform normalization, characterization, and other file-manipulation events. Archivematica identifies a TGA file as [fmt/402](#) (format 402), which is the PRONOM format identifier for a Truvison TGA Bitmap 2.0. There should be a similar STDOUT section for each item in your transfer.

Archivematica will continue processing your transfer in the background.

Task 4 - Create a SIP

- The final micro-service on the Transfer tab is *Create SIP from Transfer*. The final job, *Create SIP(s)*, allows you to either proceed directly to the Ingest tab or to send the transfer to the backlog. For more information on the backlog, check out the [backlog documentation](#).
- When prompted, select *Create single SIP and continue processing*.

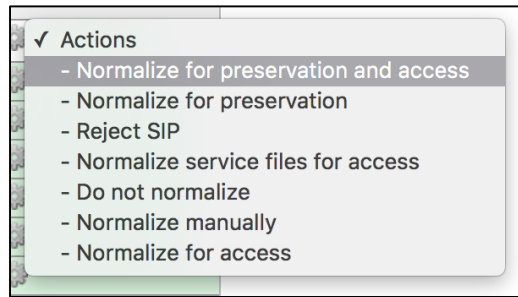


Task 5 - Make AIPs and DIPs

The primary function of Archivematica is to produce Archival Information Packages (AIPs) and Dissemination Information Packages (DIPs) from SIPs. You just created a SIP on the Transfer tab. The Ingest tab is where you run micro-services that create the AIP and the DIP.

Ingest, like Transfer, is also made up of a series of micro-services. The most significant micro-service that takes place during ingest is Normalize. Normalization is the process of converting your digital content into appropriate formats for long-term storage (for an AIP) and access (for a DIP). When you reach the Normalization micro-service, you will be prompted to decide how you would like to normalize your content.

1. Click on the Ingest tab.
2. Select *Normalize for preservation and access* when prompted. By selecting this option, you are telling Archivemata that you would like to create a preservation copy (AIP) and an access copy (DIP) of the contents of your SIP.



3. Once normalization is complete, you will be prompted to approve normalization. Before selecting approve, click on the small page icon next to the drop down menu. A yellow pop up balloon saying "Tasks" will appear when you hover on this icon.



4. The Normalization Report will open in a separate tab. This is what it looks like:

File name	File format	Preservation normalization attempted	Preservation normalization failed	Already in preservation format	Access normalization attempted	Access normalization failed	Already in access format
pictures/MARBLES.TGA	Truevision TGA Bitmap	Yes	No	No	Yes	No	No
BBhelmet.ai	Acrobat PDF	Yes	No	Yes	No	No	No
Vector.NET-Free-Vector-Art-Pack-28-Freedom-Flight.eps	EPS 3	Yes	No	No	Yes	No	No
799px-Euroleague-LE_Roma_vs_Toulouse_JC-27.bmp	Windows Bitmap	Yes	No	No	Yes	No	No
oakland03.jp2	JP2 (JPEG 2000 part 1)	Yes	No	Yes	Yes	No	No
WFPC01.GIF	1989a	Yes	No	No	Yes	No	No
G31DS.TIF	TIFF	Yes	No	No	Yes	No	No
pictures/Landing_zone.jpg	JPEG	Yes	No	No	Yes	No	Yes
Nemastylis_geminiflora_Flower.PNG	PNG	Yes	No	Yes	Yes	No	No
lion.svg	Scalable Vector Graphics	Yes	No	Yes	Yes	No	No

5. In your main tab, click on the Preservation Planning tab at the top of the page. When the Preservation Planning tab is open, search for "SVG" (or whatever file format you would like to review). Click on the name of the file format.
6. You should now have two tabs open - the Normalization Report and the Preservation Planning page. Go back to the Normalization Report and review the next two sections.
7. The Normalization Report details whether or not normalization was attempted on the contents of your SIP. This screenshot shows the report for lion.svg, identified as a Scalable Vector Graphic, with the preservation columns highlighted.

If you return to the Preservation Planning tab where you searched for SVG, you can see that SVG files are considered a preservation format. Therefore, the Normalization Report indicates the following:

- Preservation normalization was attempted.
- Preservation normalization did not fail.
- The image was already in a preservation format.

Essentially, this means that preservation normalization kicked off, but Archivemata realized that the file was already in a preservation format and so no action was taken.

8. Reviewing normalization for access: This screenshot shows the report for lion.svg with the access columns highlighted.

File name	File format	Preservation normalization attempted	Preservation normalization failed	Already in preservation format	Access normalization attempted	Access normalization failed	Already in access format
lion.svg	Scalable Vector Graphics	Yes	No	Yes	Yes	No	No

For access normalization, the report indicates the following:

- Access normalization was attempted.
- Access normalization did not fail.
- The image was not in an access format.

9. To review what this means for lion.svg, we'll dig a little deeper into the Preservation Planning tab.
10. Navigate back to the Preservation Planning tab.
11. Scroll down and find the *Normalization* section in the left-hand sidebar. Click on *Rules*.
12. Search for "Scalable Vector Graphics" (or whatever file format you are analyzing).

The results show the Access and Normalization rules for SVG files. Under the Command column we can see that the preferred access format for an SVG is PDF. Archivemata follows these rules to create access copies, so we can infer from the Normalization Report that a PDF copy of the SVG file has been successfully created for the DIP. You can confirm this by checking the command output for the *Normalize for access* job (similar to how you checked the command output for *Identify file format*, above) or by reviewing the DIP once it has been stored.

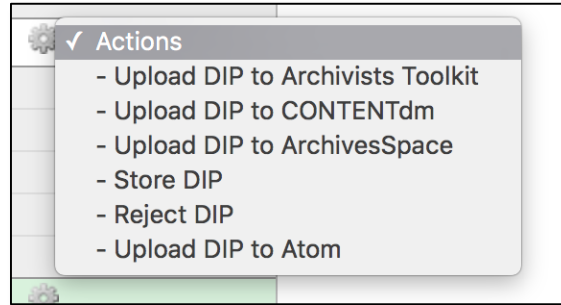
Continue processing your ingest, stopping when you reach the AIP and DIP decision points.

Task 6 - Store AIPs and DIPs

Archivemata is a tool for creating packages. In a production environment, storage occurs externally to Archivemata in a storage system selected by the user or institution, but for the sake of this demo we'll store our AIP and DIP in Archivemata's default internal storage.

AIPs should always be stored first. Because the packages are smaller, storage options for DIPs are usually the first to appear, so it's tempting to store them right away. However, if anything goes wrong with your AIP, you would then have to delete the DIP from the storage and access systems. Dealing with the AIP first allows you to store and provide access to DIPs knowing that that the AIP is secure.

1. Process your ingest until the *Store AIP* and *Upload DIP* micro-services prompt you for a decision point.
2. Select “Store AIP” from the *Store AIP* dropdown.
3. In a moment, another decision point will prompt you to select a storage location for your AIP. There should only be one option - “Store AIP in standard Archivemata directory”. Select this option.
4. Once the AIP is stored successfully, you can move on to dealing with the DIP. Neither a locally-installed Archivemata VM nor the sandbox is hooked up to an access system, so under *Upload DIP* select “Store DIP”.



5. You will be prompted to select a storage location for your DIP. There should only be one option - “Store DIP in standard Archivemata directory”. Select this option.

Your AIP and DIP are now stored in Archivemata’s internal storage. The Archivemata workflow is complete!

Task 7 - Review AIPs

1. Click on the Archival Storage tab. You should see your AIP listed in the search results there, but if not, you can search for it using the name you gave it in Task #1.
2. Click on the filename of your AIP and you’ll see a page with details on your AIP:

testtesttest Archival Information Package	
UUID	4a88a4ce-aa44-486b-b326-ccd6aeebc0b
Size	28.50 MB
Date stored	2017-10-15 21:24
Status	Stored
Location	<input type="button" value="Download"/> [...] / testtesttest-4a88a4ce-aa44-486b-b326-ccd6aeebc0b.7z
Pointer file	<input type="button" value="View"/>

3. Click the “download” button to download your AIP. Once it’s downloaded, open the AIP. You will need to a program capable of opening 7zip files installed on your computer. If required, you can download 7Zip here: <http://www.7-zip.org/download.html>
4. Once you have the AIP extracted, navigate through the folders until you find the objects directory. This directory contains the original images from your transfer as well as the preservation copies. You can compare the file formats in the objects directory to the rules in the Preservation Planning tab.

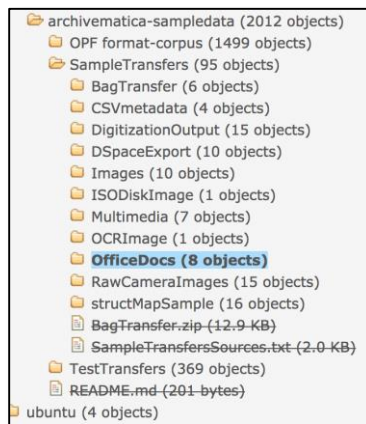
5. Navigate through the folders until you find the METS file and open it in a web browser or text editor. It will be titled something like "METS.7e58760a-e357-4165-9428-26f5bb2ba8ee.xml".
6. Find the <mets:fileSec> tag in the METS. Within the fileSec, you should be able to find information about every item in your original transfer - these are in the section tagged <mets:fileGrp USE="original">. Scrolling down, you can view complementary information for each of the preservation copies - this is in the section tagged <mets:fileGrp USE="preservation">.

The METS.xml file is very long because it contains all of the information about your files as well as information about the processes and tools that acted on those original files. For more information about the contents and structure of the METS file, check out the [Archivematica wiki](#).

The Archivematica sandbox does not allow access to the Storage Service, so we are unable to review DIPs.

Optional Task - Send Transfers to Backlog

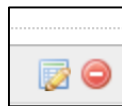
1. Now we are going to create another transfer, but this time, we are not going to complete the process of turning it into an AIP. We will send it to the Transfer Backlog. This is a special storage location, designed to allow partially processed material to be stored for a period of time. This allows the user to make decisions later about what content should be kept and what is not going to be preserved, and to arrange content into AIPs from multiple transfers.
2. Go to the Transfer tab and create a new transfer.
3. This time, select the Office Docs sample transfer.



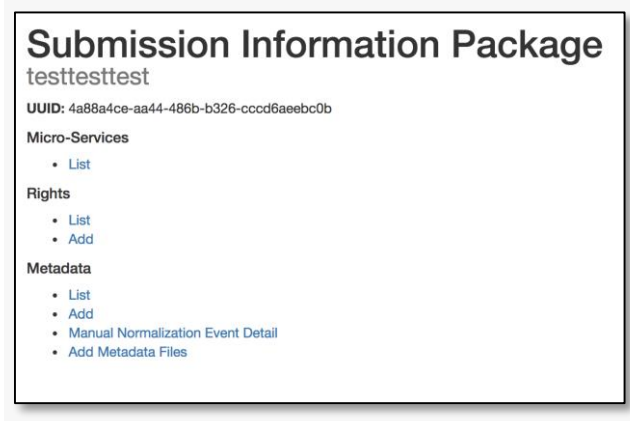
4. When you get to the "Create SIP" question, select "Send to Backlog" option.

Optional Task - Create Metadata for your SIP

1. Click on the Ingest tab.
2. Find your transfer, and hover over the little paper/pencil icon - "Metadata" will pop up. Click on it.



3. You'll be given the choice to edit/provide several kinds of metadata.



4. Click on "Add" under Metadata.
5. Fill out as many fields as you would like, then click "Save."

*adapted from a printed guide for Archivemata Camp, available on Artefactual's [GitHub page](#), and Archivemata's [Quick Start Guide](#).