artefactual
s y s t e m s   i n c .

Archivematica Worksheet

HEAP – AIM25 Network Meeting

Contents

What is Archivematica?

Archivematica is a digital preservation system. It takes your digital files in, along with any metadata you supply, and performs a series of preservation actions, that either generate new metadata about your files, or create derivative files.

All of these files and related metadata are then packaged into an Archival Information Package (AIP) using a standard file layout and standards based metadata. Learn more by looking at the following introductory slide deck:

- http://www.slideshare.net/Archivematica/digital-preservation-with-archivematica

Suite 201 – 301 Sixth Street
New Westminster, BC
Canada V3L 3A7

(tel) 604.527.2056
(fax) 888.607.8022
www.artefactual.com

artefactual
systems inc.

Getting started

By the end of this workshop you should be able to:

- Create a standard transfer in Archivematica
- Automate an Archivematica pipeline
- Identify processing errors
- Arrange multiple transfers into a single sip

Documentation

Throughout the training, we will be including links to the documentation with each training task. You can find all of Archivematica's documentation here:

- https://www.archivematica.org/docs/latest/

Logging in to the application

All the following tasks, except for Task #0, expect that you will be logged in. The credentials for the Administrator account we will be using are:

- User: admin
- Pass: aim25

Task #0 - Assemble some testing data

Description:
Before you can start using Archivematica, you need to assemble some content for Archivematica to 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 this workshop, we will be using a set of sample data that has been assembled by Artefactual for use in testing and demonstrating the application. This sample dataset includes a copy of the Open Preservation Foundation (OPF) format corpus. We have already downloaded the sample data to the demo servers, but you can also look at the data online.

Task:
Go to the Artefactual organization on github, using a web browser (hint: https://github.com/artefactual)

Suite 201 – 301 Sixth Street
New Westminster, BC
Canada V3L 3A7

(tel) 604.527.2056
(fax) 888.607.8022
www.artefactual.com

Find the archivematica-sampledata repository.
Browse through the repository, to get a bit of familiarity with the kind of content that is available.
Find the Images folder of the sample transfers.

Question:
How many files will be included in an Archivematica Transfer if you use this folder?

Task #1 – Start a Standard Transfer

Description:
In order to get your files into Archivematica for processing, you need to start by creating a Transfer. A Transfer is any set of one or more files that you decide to process as a group. They can come from any storage system that Archivematica has been hooked up to.

Archivematica has a few different transfer types - these are used to describe files that are organized in a particular way in advance. The transfer types supported are:

- Zipped or Unzipped bag - content in the Library of Congress Bag format
- DSpace - DSpace AIP's, exported from DSpace in zip format
- Disk Image - e.g., ISO, AFF and other disk image formats
- Standard Transfer - any other content

We will start by creating a Standard Transfer using the Images folder you found in Task #0.

Relevant documentation:
- Create a Transfer

Task:
- Go to the Transfer tab.
- Make sure 'Standard' is selected from the transfer type box.
- Type in a name for your transfer (anything you like).
- Click browse, to look through the available content.
- Click on folder for 'sample transfers'. Make sure to click on the actual folder icon to expand the tree.
- Find the images directory and click on the name 'images' to select the entire directory, then click 'Add'. The selected objects will be displayed above the selection box.

Question:
How many digital objects will get included in your Transfer?

- Click 'Start Transfer'. A new Transfer should be created, and it will stop at the first Micro-service, waiting on user input. In this first step, you are being asked to approve the transfer you just started.
- Choose 'Approve transfer' from the drop-down list.

Suite 201 – 301 Sixth Street
New Westminster, BC
Canada V3L 3A7

(tel) 604.527.2056
(fax) 888.607.8022
www.artefactual.com

Task #2 – Make an AIP

Description:
Now that you have a Transfer started, you can continue processing until you have an AIP. After having accepted the Transfer in task #1, you will see a series of microservices appear in the Transfer tab. The processing will stop occasionally, waiting for user input. You can answer anything you want, except for 3 of the questions as noted.

Relevant Documentation:
• Process the Transfer

Task:

• When you get to the 'Create SIP' question, choose 'create single SIP and continue processing'. Your Transfer has now been turned into a SIP, and processing will continue in a new Tab.
• Switch to the 'Ingest' Tab.
• When you get to the 'Normalize' question, choose 'do not normalize'.
• Finish by storing the aip.

Question:
How many options were presented in 'Normalize' question?

Task #3 – Review your First AIP

Description:
The primary function of Archivematica is to produce Archival Information Packages. In this task, we will explore the structure of an Archivematica AIP.

Task:
• Once your AIP from Task #2 has been stored, go to the Archival Storage Tab. You should see your AIP listed in the search results there.
• Click on the name of your AIP to open the AIP Details page.
• Click on the 'download' button.
• Open the AIP up. You will need to have 7zip installed on your computer or another program capable of opening 7zip files (if you can't open a 7zip file, don't worry move on to Task #4).
• Once you have the AIP extracted, drill down until you find the METS file.
• Open the METS file in a web browser or text editor.

Question:
Can you determine from the METS file, how many digital objects there are in this AIP? (hint look for the <mets:fileSec>).

Suite 201 – 301 Sixth Street
New Westminster, BC
Canada V3L 3A7

(tel) 604.527.2056
(fax) 888.607.8022
www.artefactual.com

Task #4 – Automate Workflow through Configuration

Description:
Once you have gone through the Archivematica workflow, there are questions that you may want to preconfigure the answers for, so you don't have to answer the same thing each time. This can be done in the Processing Configuration screen.

Task:
- Go to the Administration tab. You should see a single processing configuration listed, called 'default'. In Archivematica 1.6.0 you can create multiple configurations and give them each a different name. For this task, edit the default configuration.
- Review the different options, and select as many defaults as you would like.
- Do not select anything for the 'Create SIP' question.

Task #5 – Send Transfers to Backlog

Description:
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 AIP's from multiple Transfers.

Task:
- Go to the Transfer Tab and create another Transfer.
- This time, select the Sample Transfers -> Multimedia folder.
- When you get to the 'Create SIP' question, choose the 'send to backlog' option.

Question:
Did you notice any errors or problems during the processing of this Transfer?

Task #6 – Examine a microservice failure

Description:
Sometimes, in digital preservation, things go wrong. Archivematica attempts to protect the user from as much of the guts of the digital preservation process as possible, while still providing a way to peer under the hood when required.

In this task we will look at a specific error during a microservice.

Task:

- While processing your last transfer, there should have been an error in the File Identification microservice.
- Click on that microservice to expand the list of jobs that happened during that microservice.
- Find the one that says 'Failed'.
- Click on the gear icon beside the word 'Failed'.

A new browser tab should open up, displaying details about each file that was processed in this job. Find the one that failed.

Question:
Which file failed file identification and why?

Task #7 – Use the Backlog Tab

Description:
Even though there was an error during file identification, Archivematica did not consider this to be a fatal error, and processing continued. Your Multimedia transfer should be in backlog now.

Task:
Go to the Backlog tab. Try searching to find the file that failed in the previous task.
Delete this transfer from the backlog.

Question:
Do you think Deleting the Transfer is the best approach? How would you prefer to handle this situation?

Task #8 – Hide a Transfer (or SIP)

Description:
Once you have completed processing, you may not need to keep the full detail of each Transfer showing in the Transfer tab. You can hide Transfers, if you like. This doesn't delete anything from the database, it just removes the package from the user interface.

Task:
Go to the Transfer Tab. Find the Multimedia transfer you made earlier. Click the red icon at the far right of the transfer name. Remove the transfer from the dashboard.

Task #9 – Arrange a SIP

Description:
Let's try again with the same transfer (the Sample Transfers / Multimedia example) and instead of deleting it from the backlog, keep it in the backlog.

There is another feature of Archivematica, available in the Ingest Tab, called SIP Arrange. This is a way to select individual files from one or more Transfers, and arrange them into one or more SIP's. Instead of a one to one relationship between a Transfer and an SIP (which ultimately becomes an AIP) it is possible to create many to many relationships.

Task:
Create a Transfer as in Task #5. Send it to backlog.
Go to the ingest tab and click the search button. You should see the Transfer you sent to Backlog appear on the left in the Originals Pane. On the right hand side, there is the Arrange Pane. Click on the top level of that pane (the word 'arrange') and then click the Add Directory button. This will add a new element under arrange in that Pane.   You can continue to add as many directories and subdirectories as you want. You can then drag files from the Originals Pane to directories in the Arrange Pane.

Once you have decided on the arrangement of the files you want, click on directory in the Arrange Pane and then click 'create sip'. This will move the content out of the Transfers it came from and add a new SIP ready for processing in the Ingest Tab.

Conclusion

Thank you for participating in the AIM25 HEAP Network Meeting AtoM training event. There is plenty more to explore in Archivematica!

Take a look at http://wiki.archivematica.org and https://www.archivematica.org for more.

Feel free to get in touch with us at Artefactual as well!

www.artefactual.com
info@artefactual.com
@archivematica