archivematica®

Archivematica 0.8 tutorial

Home	Transfer	Ingest ¹	Archival storage	Preservation planning	Access	Administration		Connected ●

| Submission Information Package	UUID	Ingest start time		
✦ OfficeDocs	179869b9-bd81-401f-84a4-d5e8c3ddd7c2	2012-01-10 17:08	🔍 📄 ⛔	
Micro-Service: Normalize [?]		Awaiting decision	⚙	Actions ▾
Micro-Service: Find options to normalize as		Completed successfully	⚙	Actions
Micro-Service: Move to workFlowDecisions-createDip directory		Completed successfully	⚙	- Preservation and access
Micro-Service: Grant normalization options for no pre-existing DIP		Completed successfully	⚙	- Reject SIP
Micro-Service: Check for Access directory		Completed successfully	⚙	- Access only
Micro-Service: Check for Service directory		Completed successfully	⚙	- No normalization
Micro-Service: Sanitize SIP name		Completed successfully	⚙	- Preservation only
Micro-Service: Sanitize object's file and directory names		Completed successfully	⚙	
Micro-Service: Set file permissions		Completed successfully	⚙	
Micro-Service: Remove thumbs.db files		Completed successfully	⚙	
Micro-Service: Include default SIP processingMCP.xml		Completed successfully	⚙	
Micro-Service: Rename SIP directory with SIP UUID		Completed successfully	⚙	
Micro-Service: Verify SIP compliance		Completed successfully	⚙	
Micro-Service: Move to processing directory		Completed successfully	⚙	
Micro-Service: Set file permissions		Completed successfully	⚙	
Micro-Service: Approve SIP Creation [?]		Completed successfully	⚙	

Exercise 1: OfficeDocs transfer

In this exercise you will be working with a sample transfer called OfficeDocs. You will place the transfer into Archivematica, prepare a Submission Information Package from it, run it through a number of ingest processes, package it as an Archival Information Package, generate a Dissemination Information Package, and upload the Dissemination Information Package into the web-based access system (Qubit/ICA-AtoM).

If your web browser is open, minimize it to view the Desktop. Open the file browser by double-clicking the Home icon on your Archivematica Desktop:
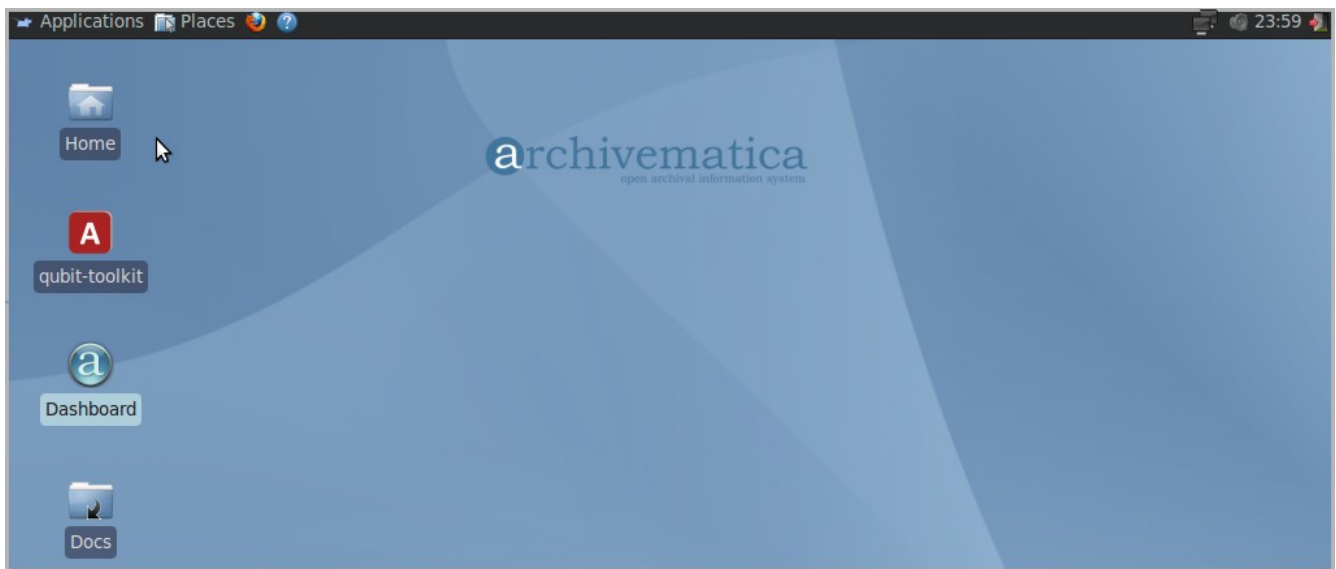
In the file browser, navigate to sampleData/SampleTransfers/OfficeDocs using the shortcuts in the panel on the left. This is the sample transfer you will be running through Archivematica during this exercise.
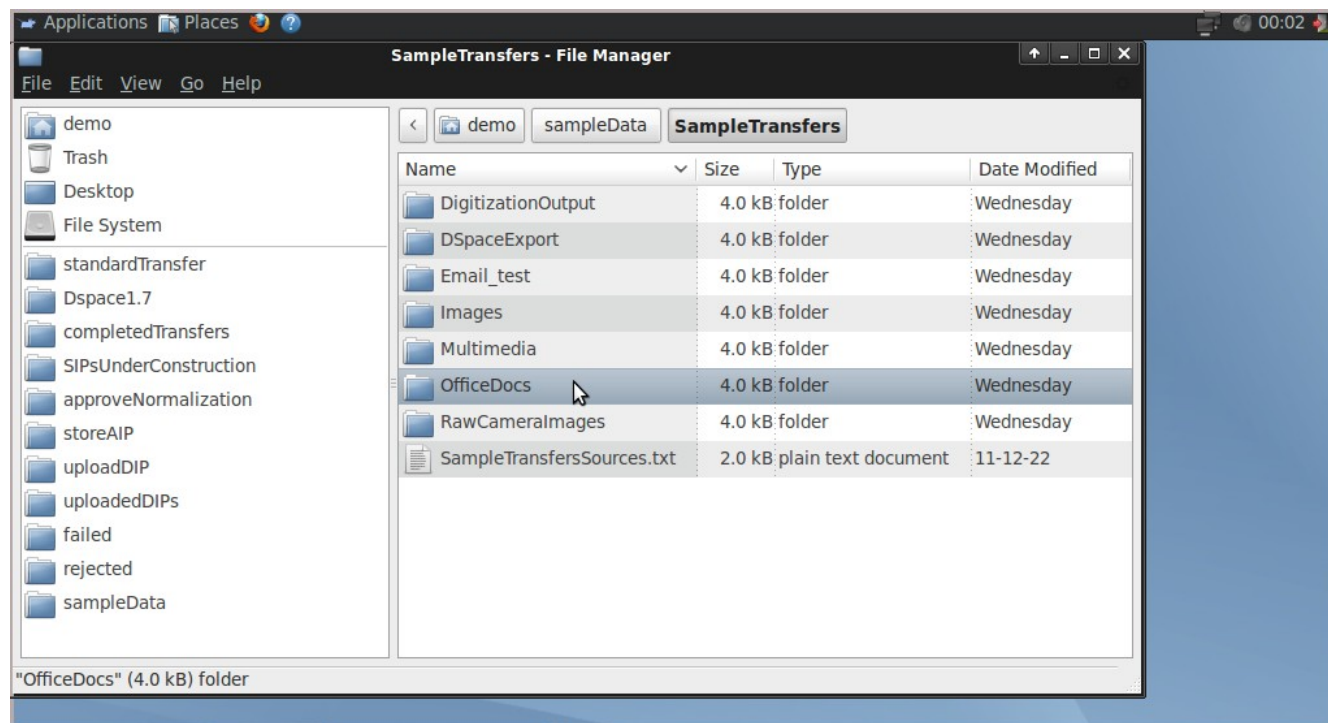
Open OfficeDocs. You will see the sub-folders shown in the screenshot below: feel free to open them and view the contents. The objects directory contains the digital objects that are to be preserved. The metadata directory contains a submissionDocumentation subfolder, which can be used for transfer forms, donation agreements or any other documents that relate to the acquisition of the records (you'll notice a sample in there - Records_transfer.rtf). The logs folder will eventually contain logs generated when processing the transfer in Archivematica.
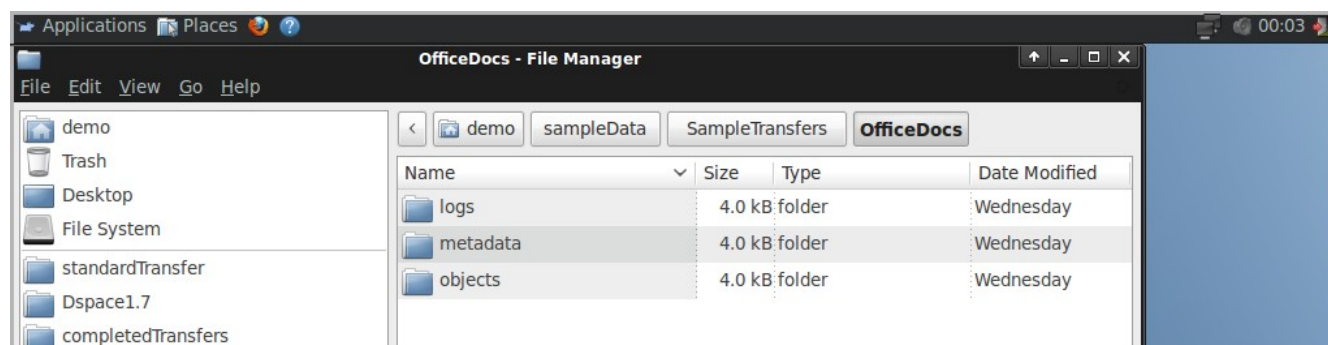
Copy the OfficeDocs transfer and paste it into standardTransfer.
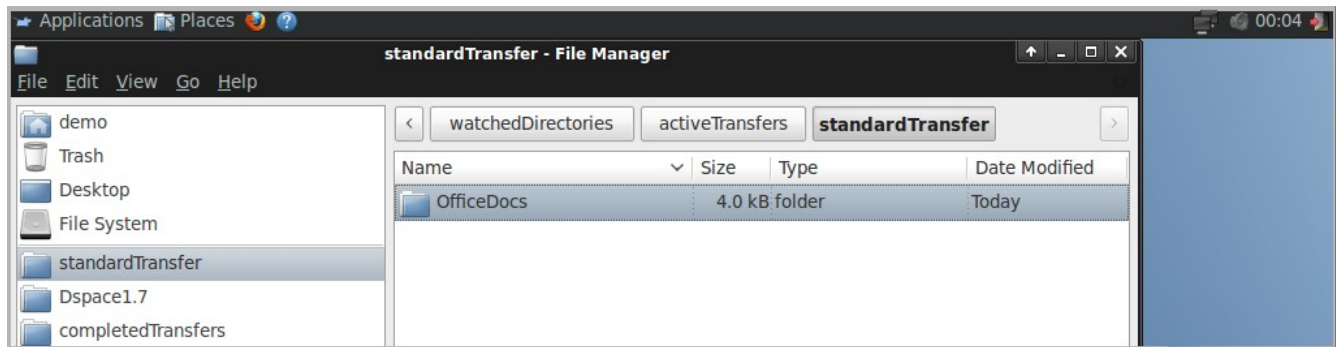
Close or minimize the file browser. Open the Archivematica dashboard by clicking on the Dashboard icon on the Desktop. Click on the Transfer tab. You will see the transfer on the page with a bell icon next to it. This means that Archivematica is asking you to make a decision about the transfer.

Click on the spyglass on the right-hand side to open the list of micro-services:
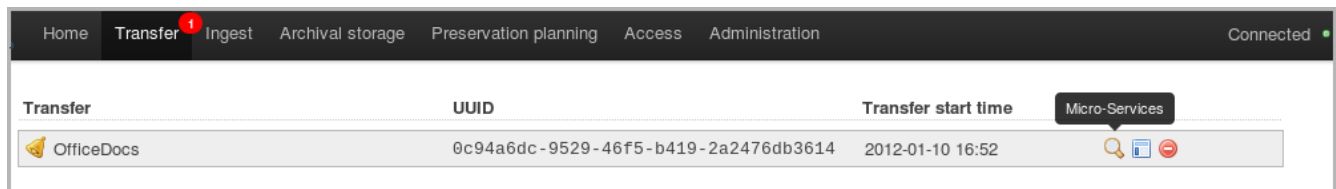
Archivematica is awaiting to your decision to approve the transfer. Select "Transfer complete" in the Actions drop-down menu:
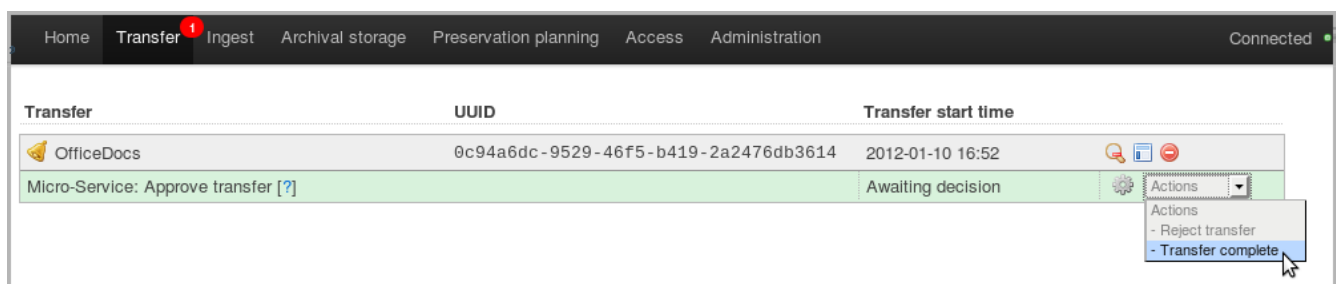
Once the transfer is approved, Archivematica moves it through a series of micro-services. These include:

➢ Assigning universal unique identifiers (UUIDs)
➢ Assigning checksums

- ➢ Generating a METS file (to capture the original order of the transfer)
- ➢ Extracting packages (unzipping any zipped or otherwise packaged files)
- ➢ Scanning for viruses
- ➢ Characterizing files (identifying them and validating them against known specifications)
- ➢ Extracting technical metadata

Once these micro-services are completed, The transfer will show a bell icon next to its name, indicating that you have reached another approval step. Click on the Actions drop-down menu and select "Create single SIP":


Click on the Ingest tab. You will see the SIP with a bell icon next to it. You are now ready to add descriptive metadata to the SIP. Click anywhere in the grey bar highlighting the SIP:

This opens a detail panel for the SIP. To add metadata, click Add under Metadata on the left-hand side of your screen:
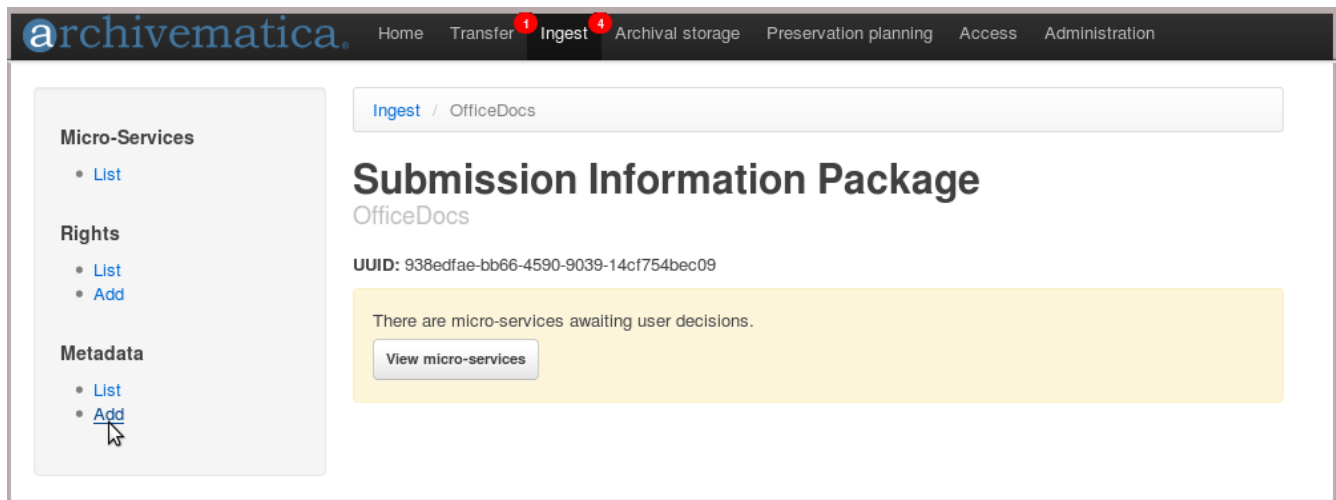

Enter metadata as desired. At a minimum, add a title to the SIP. When you're done, click Save at the bottom of your screen.
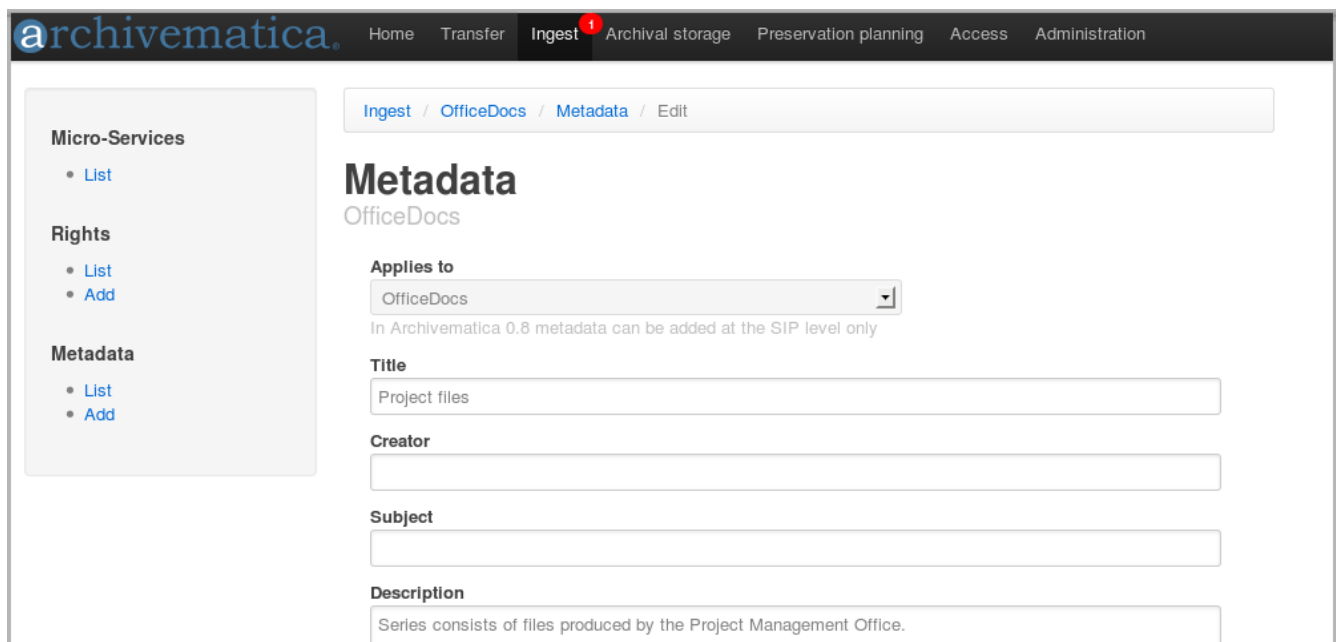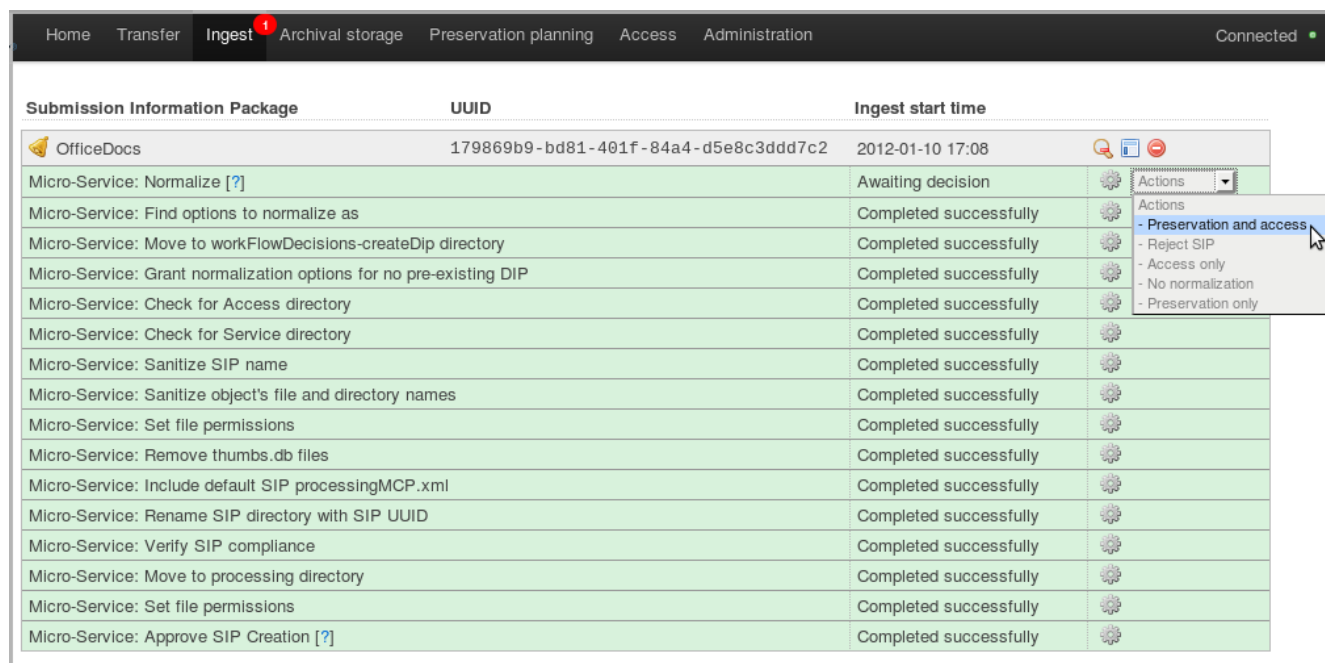
Your screen will show that metadata has been added to the SIP. If you would like to edit your metadata, click Edit on the right-hand side and enter metadata as before.
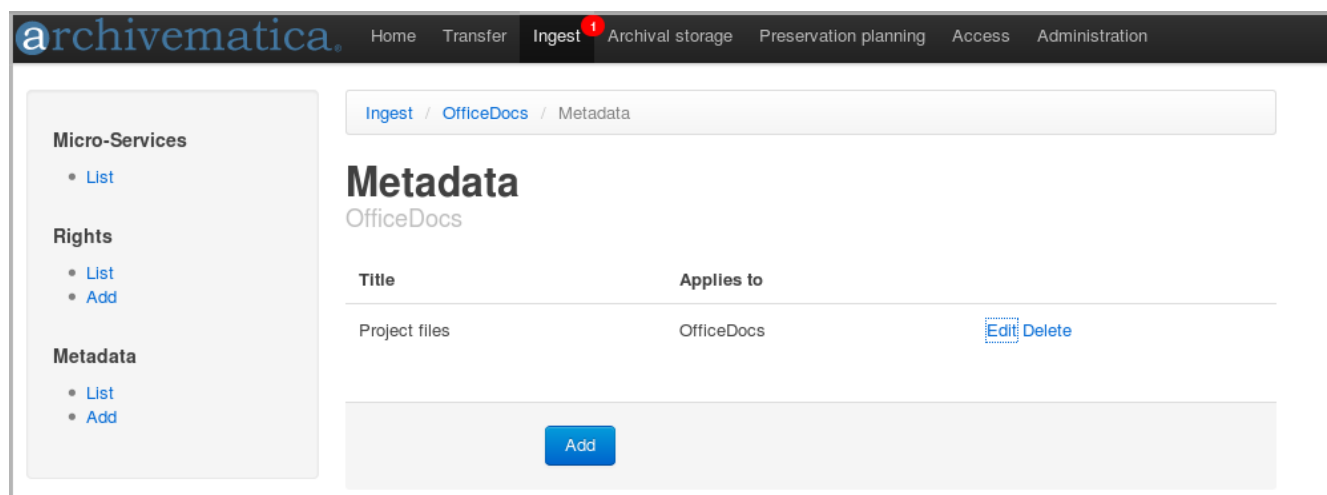
When you are done, click the Ingest tab. You are now ready to move your SIP through ingest. Click on the spyglass icon to open the micro-services grid and select "SIP Creation complete":
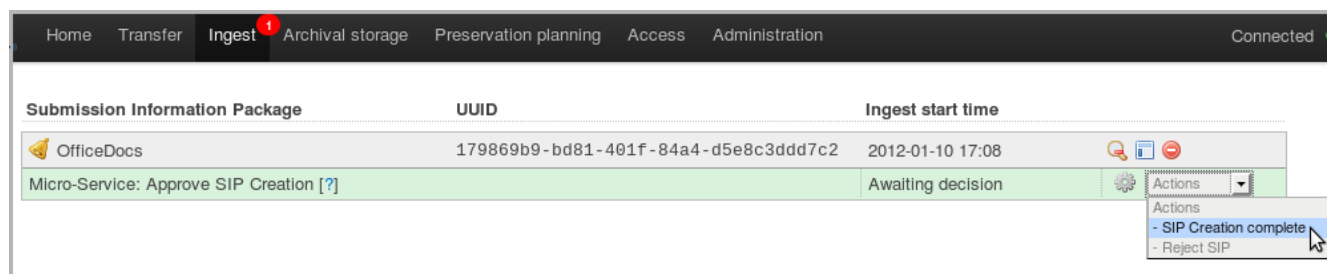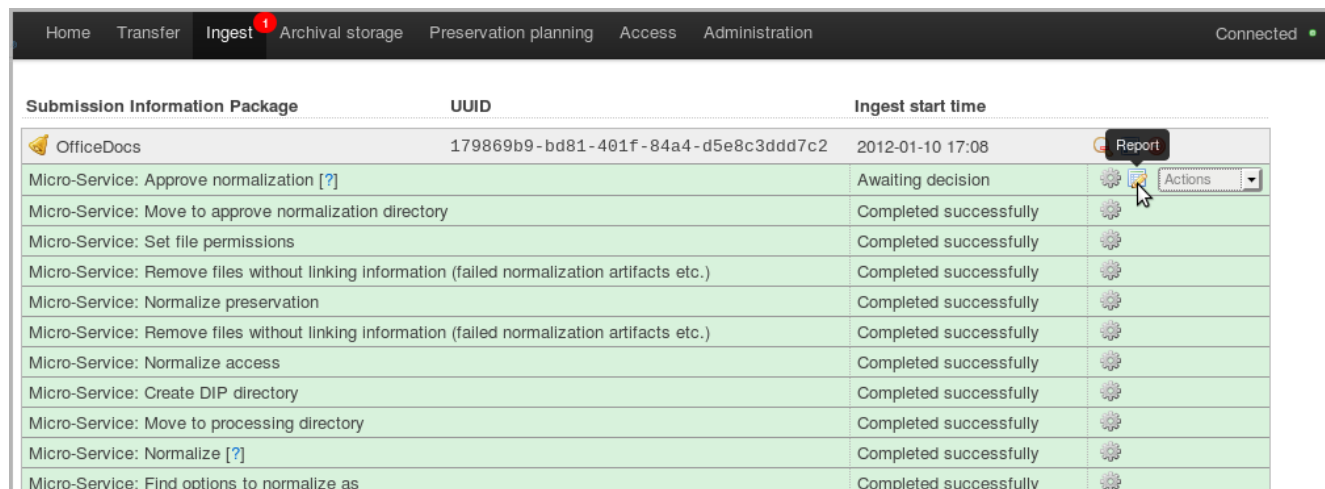
The next approval step is normalization. At this point, the user can select from a number of options. Select "Preservation and access" to create preservation masters of the ingested objects and generate access copies for the DIP (Dissemination Information Package).

Once normalization is completed you can review the results of normalization by clicking on the report icon next to the Actions drop-down menu:
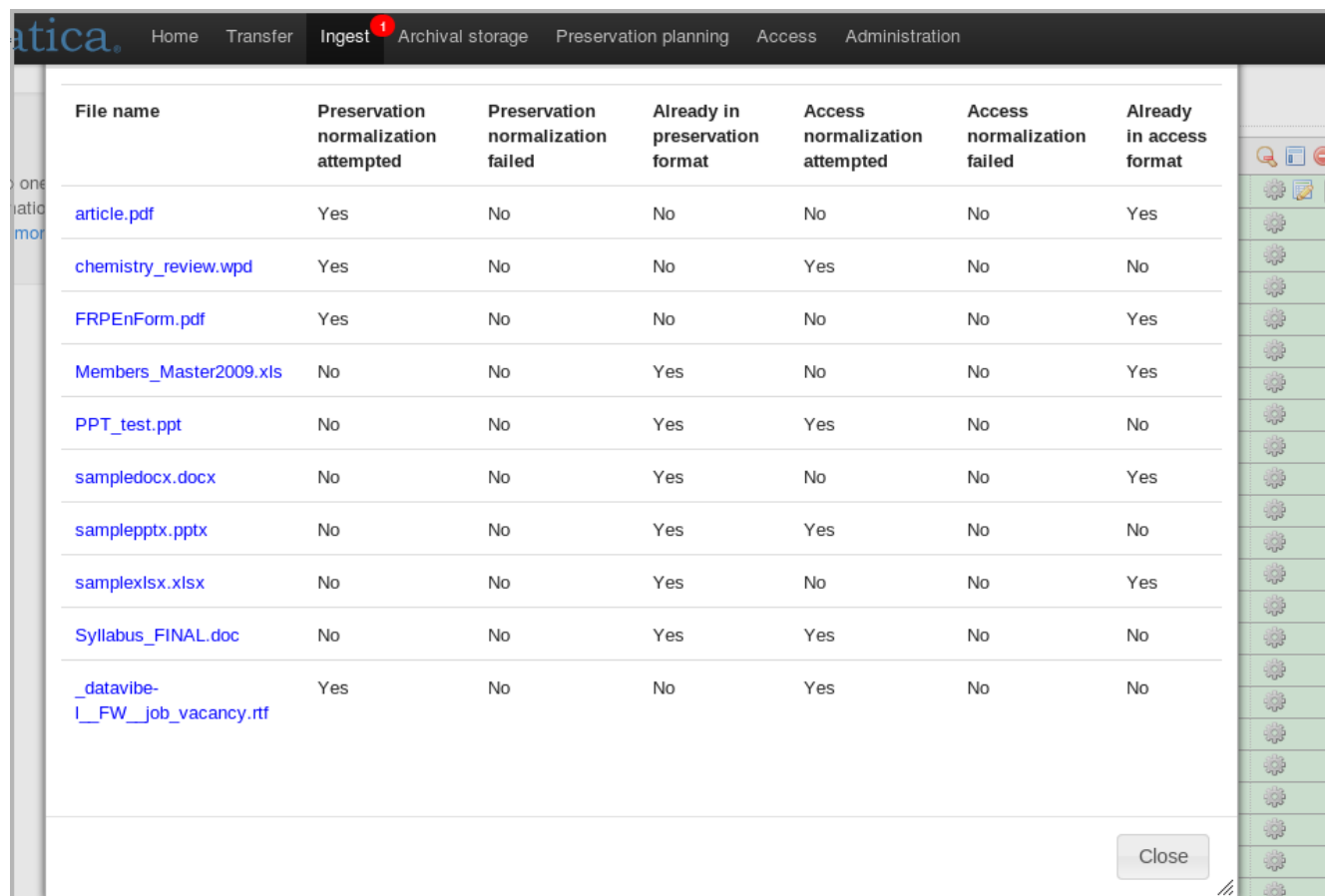

This will open a normalization report. The report lists all the files in the SIP and indicates whether normalization has been attempted and whether or not it was successful. For example, the report shows that the first file, article.pdf, was normalized to a preservation

format not to an access format, since pdf is already an acceptable access format.

If there had been any normalization failures the report would have shown the errors in red.

Close the normalization report by clicking on the Close button at the bottom right-hand corner of the screen. Once the report is closed, select "Approve normalization" from the Actions drop-down menu (not shown).

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
chemistry_review.wpd	Yes	No	No	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	No	Yes	No	No	Yes
samplepptx.pptx	No	No	Yes	Yes	No	No
samplexlsx.xlsx	No	No	Yes	No	No	Yes
Syllabus_FINAL.doc	No	No	Yes	Yes	No	No
_datavibe-l__FW__job_vacancy.rtf	Yes	No	No	Yes	No	No

After normalization is approved, the SIP moves through a number of micro-services. These include:

➢ Processing submission documentation

➢ Preparing the METS file

➢ Generating the DIP

➢ Verifying checksums generated on ingest

➢ Packaging the AIP in accordance with the Library of Congress Bagit specification

When these micro-services are completed, the user is presented with two final approval steps, Store AIP and Upload DIP:

In the Actions drop-down menu in store AIP, select "Store AIP". This will move the AIP into archival storage.

Click on the Archival storage tab. This tab shows a list of AIPs currently in storage which can be sorted by clicking on the arrows in the column headers. At the bottom of the Size column the total storage of all AIPs appears. To view the contents of an AIP, click on a blue link.

The last stage of processing is to upload the DIP. Return to the Ingest tab and in the Upload DIP Actions drop-down menu, select "Upload DIP".

A dialogue box will appear which will allow you to select an archival description to which the DIP will be uploaded. Enter *vancouver-city-clerk-fonds*, just as shown in the screenshot, and click on the box next to "Create intermediate level of description":
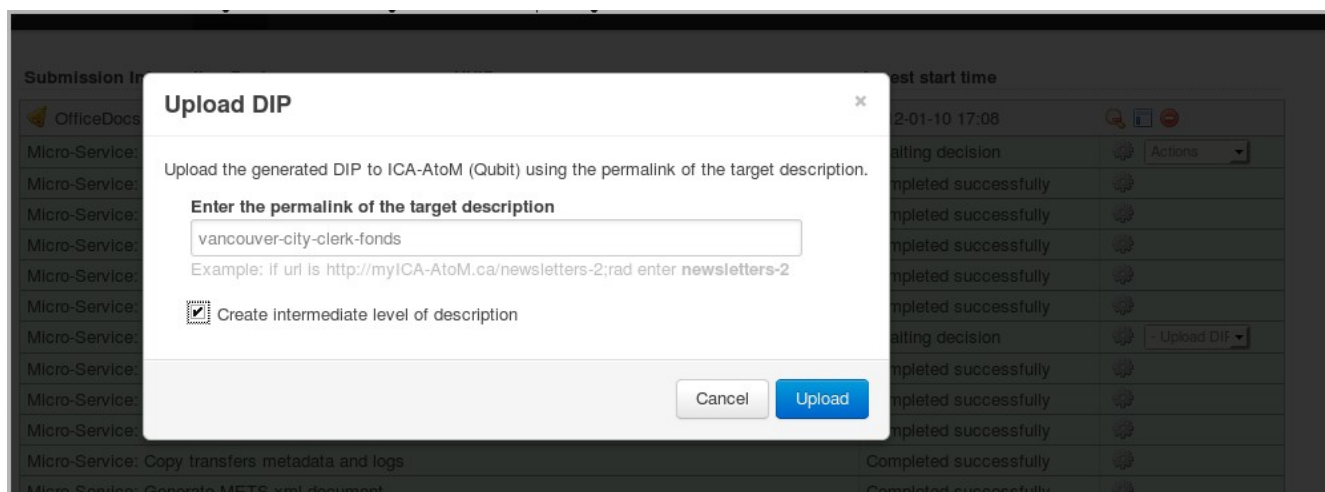
This will upload the digital objects to the Vancouver City Clerk fonds in ICA-AtoM, the default web-based access system that is bundled with Archivematica. ICA-AtoM has been pre-populated with some sample data, including the Vancouver City Clerk fonds. Note that clicking on "Create intermediate level of description" creates a parent archival description populated with the metadata you added when you were creating the SIP. If you had not clicked on "Create intermediate level of description" the objects would have been uploaded directly to the fonds as immediate child-levels of the fonds.

Click on the Access tab. This tab shows a list of AIPs and any DIPs that have been created from them. Right-click on the DIP URL and open the link in a new tab to view the uploaded objects in Qubit.
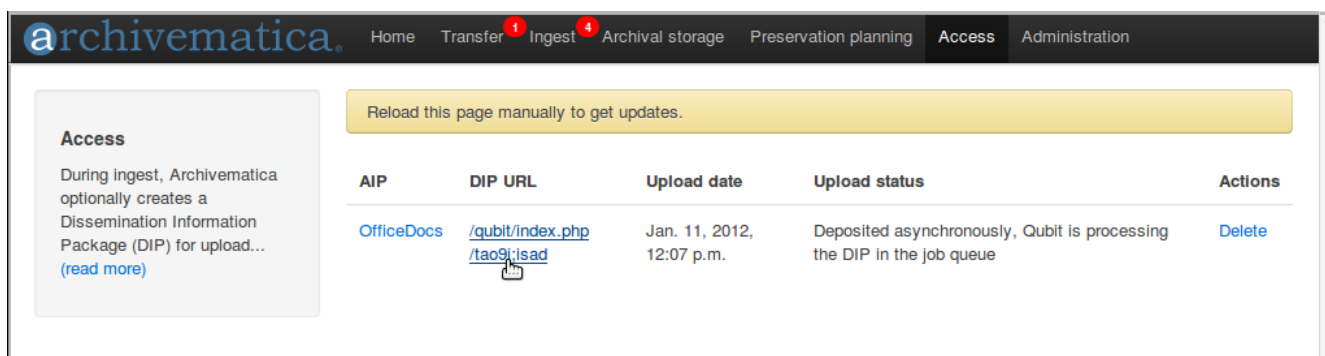
Since you are not currently logged into Qubit, you will see a message that you do not have access permission. Click on the Log in link in the upper right-hand corner:
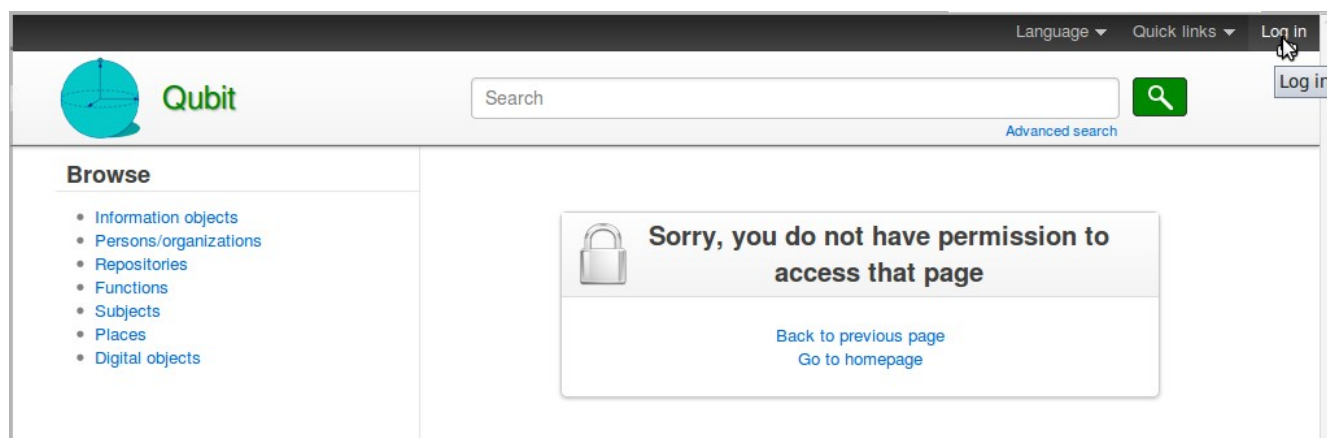
Enter the email demo@example.com and the password demo, then click the blue Log in button underneath.
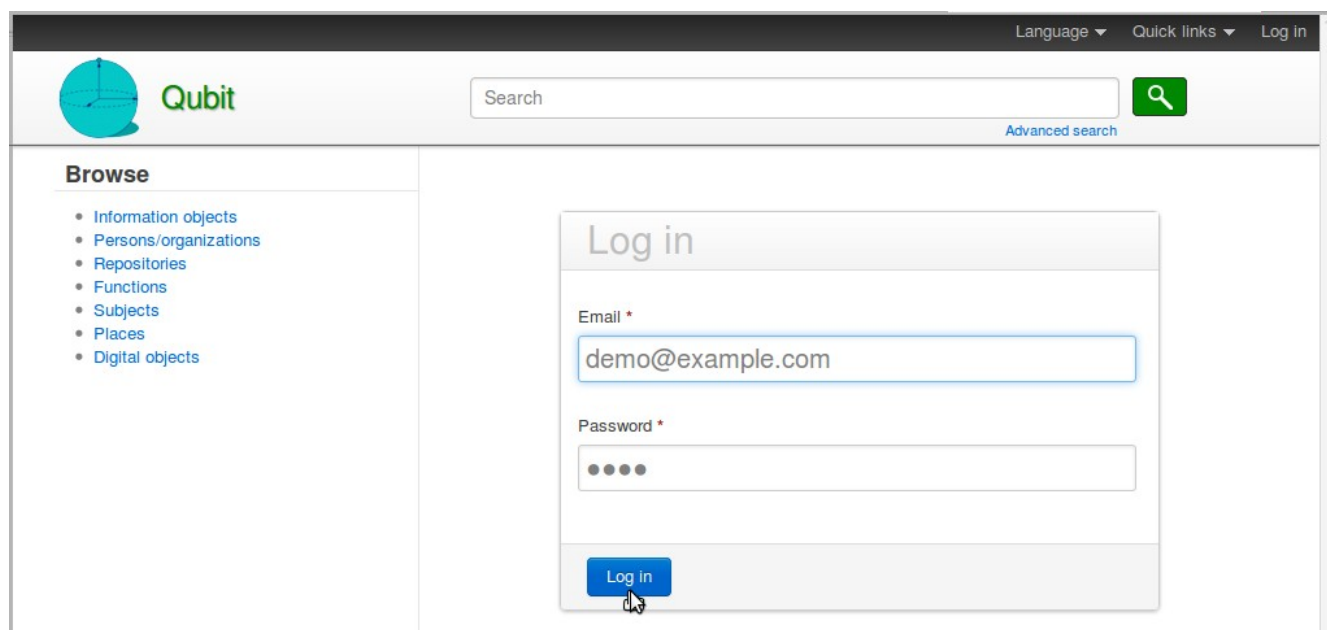
You should now see the uploaded DIP: an archival description containing the metadata you entered when you prepared the SIP, plus the uploaded objects. On the left-had side is a cover-flow viewer showing thumbnail images. You can scroll through the objects by using right and left arrows on your keyboard.

Click on a thumbnail in the cover-flow viewer to go to that object.

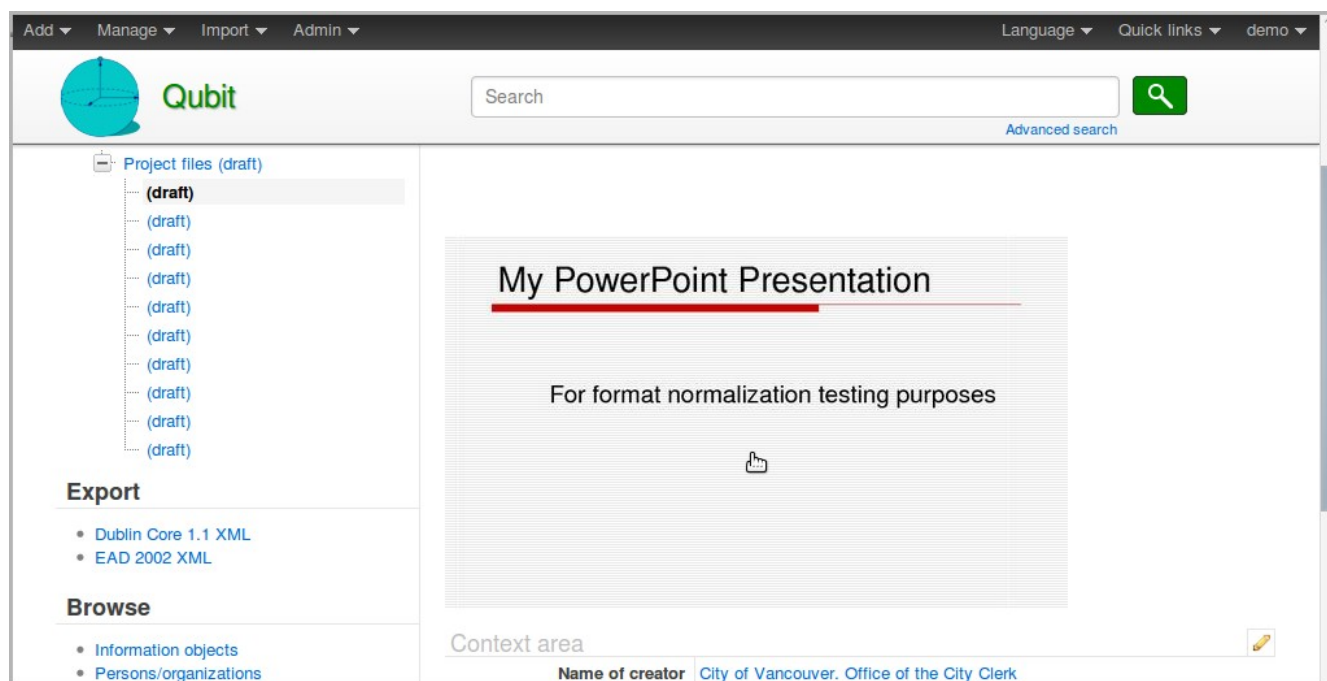This is an example of a page showing an uploaded digital object. The user can download the object by clicking on the image.

Now that the AIP has been stored and the DIP uploaded, transfer and ingest of OfficeDocs is complete.

Exercise 2: Images transfer
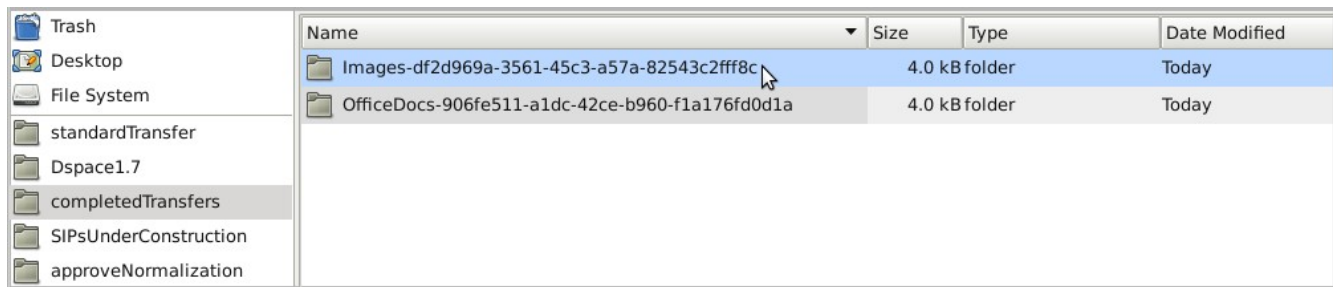
Now that you have run OfficeDocs through, you are ready to try another transfer. In this case, select sampleData/SampleTransfers/Images. Copy and paste Images to the standardTransfer folder and approve the Transfer in the Transfer tab of the dashboard, just as you did with OfficeDocs. However, once the transfer is finished processing and has reached the Create SIP(s) approval step, select "Manual SIP creation" instead of "Create single SIP". This allows the user to create more than one SIP from a single transfer (all log files, submission documentation and the METS file from the transfer will be copied to all the SIPs created from the transfer).
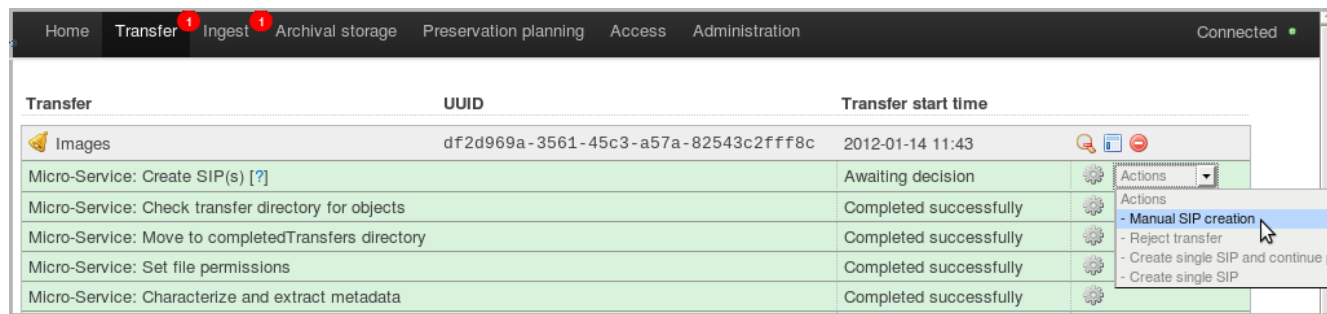
To create the SIPs manually, navigate to the SIPSUnderConstruction directory, and right-click. You will see a list of scripts to choose from. Select "Create a structured directory". Do this twice, creating two SIPs with the names SIP1 and SIP2 (or other names if you prefer).
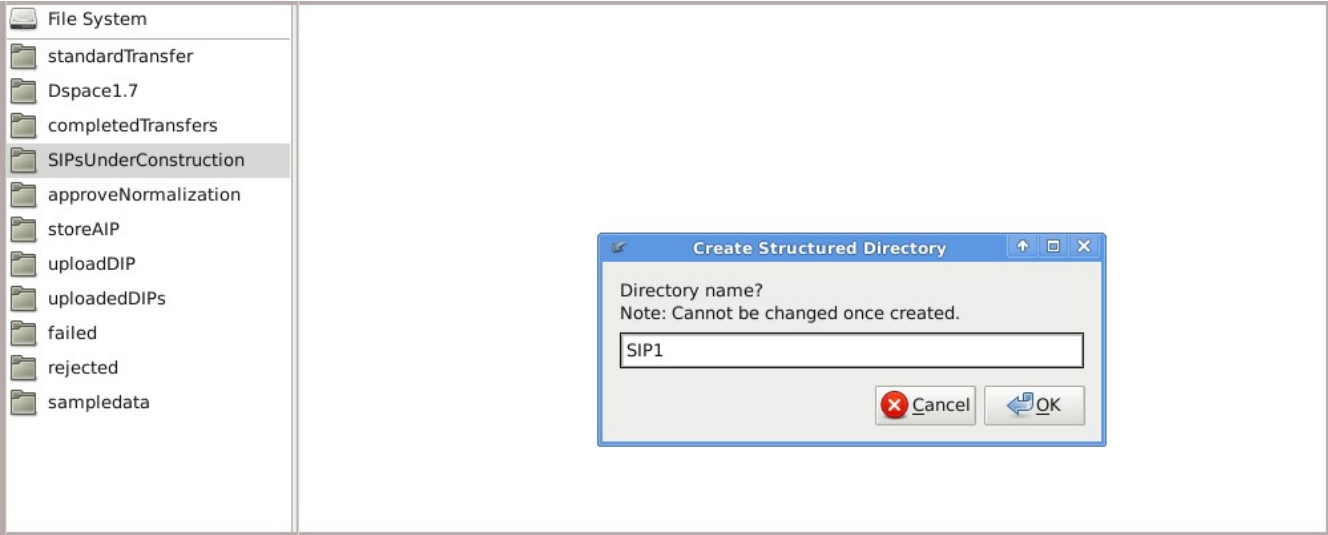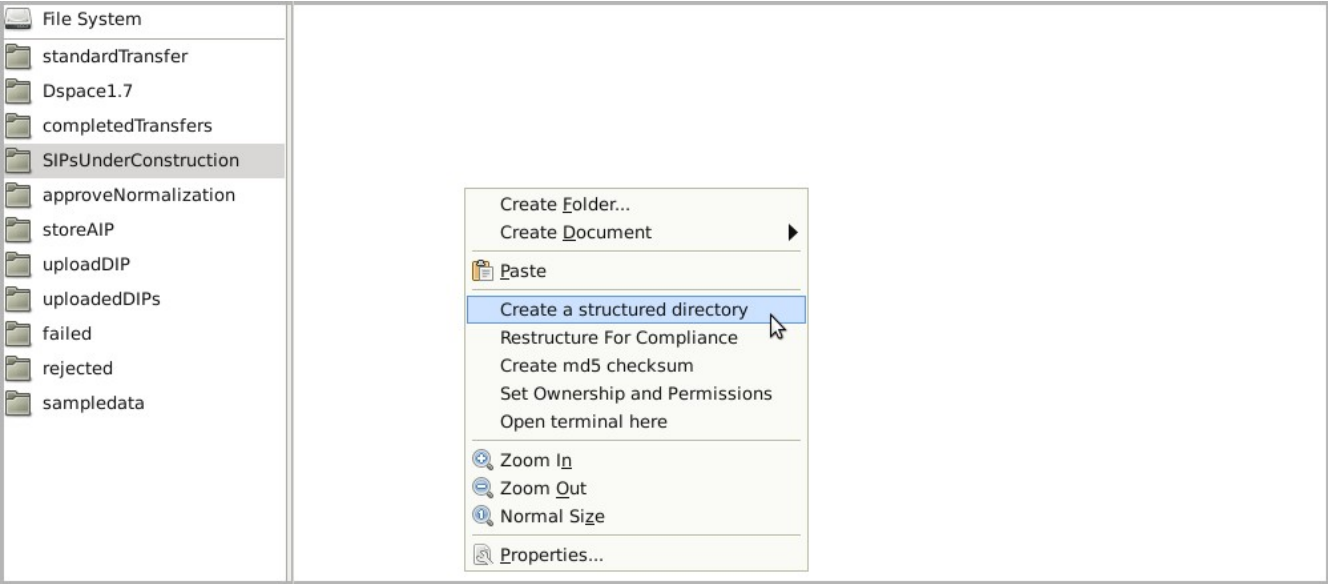
Now go to completedTransfers. You will see the Images transfer there (with a unique universal identifier added to the filename):

Open the transfer, navigate to the objects directory, and cut and paste some files to SIP1 and the rest to SIP2. Note that you must cut and paste, not copy and paste: this will prevent you from accidentally copying the same file to more than one SIP.

Once you have finished adding objects to the SIPs, open the dashboard, click on the Ingest tab and process these SIPs the same way you did OfficeDocs.

Exercise 3: Other transfers

Now that you are familiar with the transfer and ingest process, you may wish to try processing other transfers in sampleData/SampleTransfers, noting the following:

➢ If you process the DigitizationOutput transfer, at the normalization step you should select "Preservation only" or "No normalization". Archivematica will automatically create a DIP from the access directory. The purpose of this is to allow institutions to generate master, service and/or access copies during digitization and to have Archivematica recognize the derivative relationships between them and generate a DIP from the access copies only.

➢ Archivematica is designed to act as a back-end preservation system for DSpace. If you process the the DSpaceExport transfer, copy the transfer to the DSpace1.7 directory in the left panel, not the standardTransfer directory. At the normalization step, select "Preservation only" or "No normalization".

➢ If you process Email_test, at the normalization step select "Preservation only" or "No normalization". Archivematica extracts the contents of PST files and converts them to MBOX, creating one large MBOX file for access and individual MBOX messages plus linked attachments for preservation. The access version of the MBOX file is designed to be imported into open-source email tools such as Evolution, where the messages are displayed and can be navigated and read in the same way as in their original email program.