Automated Ingestion Agent

Table of contents

Introduction

ONIXEDIT Automated Ingestion Agent (ONIXEDIT AIA) is an option that you purchase with your ONIXEDIT Server Edition (hosted or "on premises" model). ONIXEDIT AIA is a special service that can be configured to check for "hot folders" (FTP folders or Amazon S3 "buckets") where metadata and asset files are uploaded. Each hot folder is associated with a specific ONIXEDIT Server Edition catalog. Consequently, each time a metadata file is available in a hot folder, it is imported to the corresponding catalog while respecting the import rules defined for that catalog. As well, the Automated Ingestion Agent will search for cover image files (jpg, tiff, png, gif, bmp) and ebook files (pdf or ePub) and based on their name (ISBN or EAN in the file name), those files will be automatically associated with the corresponding title and uploaded to your Media Space or simply archived. The Agent is able to process ONIX files as well as flat files (Excel, or CSV) with automatic conversion to ONIX, providing that an ONIXEDIT import script is defined in the corresponding import script folder. After each ingestion the metadata and asset files are archived and a processing report is sent to you by email. This report is also saved at the server in order to keep track of every file ingested.
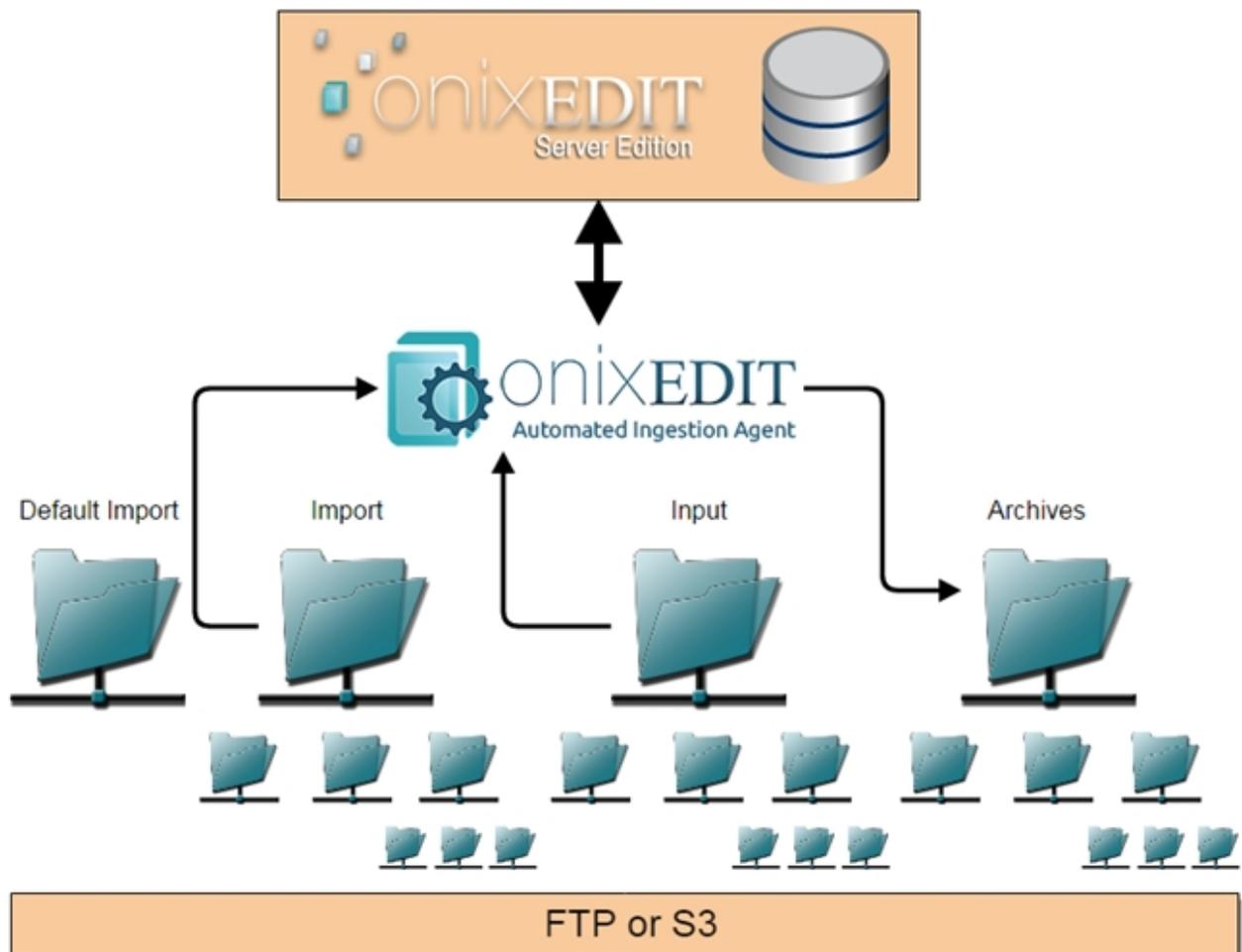
Components of ONIXEDIT AIA

ONIXEDIT AIA is composed of 3 components:

- [FTP/S3 Repository](#)

- [The Windows Service](#)
- [The AIA Dashboard](#)

FTP/S3 repository

ONIXEDIT AIA needs to have access to a file repository which can be provided by an FTP server or [Amazon Simple Storage Service](#) (S3). When we install ONIXEDIT AIA, we configure it following your file repository preferences.

No matter if you are using FTP or S3, the folder/bucket structure will be same. Indeed, at the root level of the file repository, you will find 4 folders/buckets:

- **Input**: The input folder will contain one "hot folder" for each catalog defined in your ONIXEDIT Server account. Those hot folders will be created automatically by the Windows Service of AIA. To be considered by ONIXEDIT AIA, a catalog must have [import rules](#) defined and the admin user (the owner of the ONIXEDIT Server account) must have Read, Write, Update and Delete access rights. Consult [ONIXEDIT Server user manual](#) for more information.
- **Default import**: In this folder, you can upload two ONIXEDIT import scripts that will be used by default by ONIXEDIT AIA when a non ONIX metadata file will be uploaded to an Input folder having no specific import script in its corresponding Import folder. You can set one default import script for ONIX 2.1 catalogs and another for ONIX 3.0 catalogs. For instance, if an Excel file is uploaded to an input folder, ONIXEDIT AIA will check if there is a specific ONIXEDIT import script in the corresponding Import folder. If there is no ONIXEDIT import script, ONIXEDIT AIA will check in the default import folder and will then use the import script corresponding to the ONIX version of the catalog. Consult the [ONIXEDIT User Manual](#) to learn more about import scripts.
- **Import**: Like for Input hot folders, ONIXEDIT AIA will automatically create an import folder for each of your catalogs. Each of these folders can be used to define a specific ONIXEDIT Import script to process non ONIX metadata files like CSV, tabulated, fixed length or Excel. If no specific import script is defined and a flat file is uploaded to a hot folder, ONIXEDIT AIA will use the import script defined in the Default Import folder.
- **Archive**: This is where all metadata and asset files uploaded to your hot folders will be moved after having been processed.

The Windows Service

ONIXEDIT AIA Windows Service is the black box running in the background on the ONIXEDIT Server that you are using. Its role is to process metadata and asset files uploaded to your hot folders. If ONIXEDIT Server is installed on your own server ("on premises" model), you will see a service named "AIA_Service" that is set to start automatically if your server is restarted:
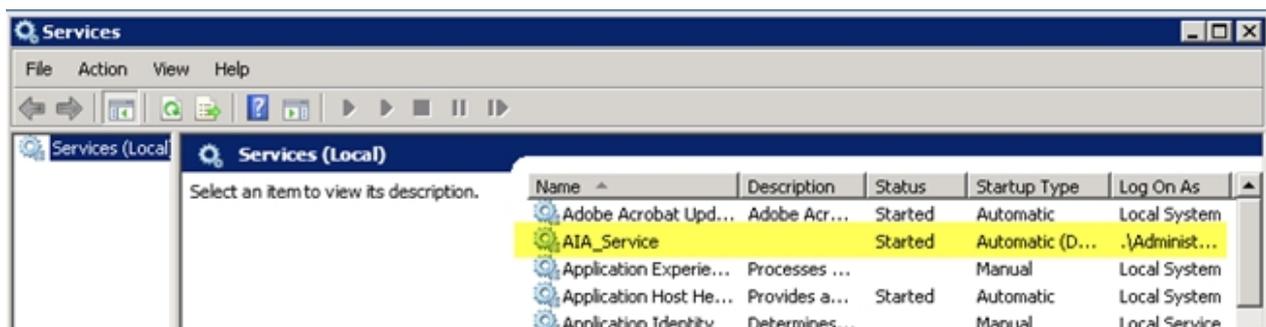
AIA Dashboard

AIA Dashboard was created to help ONIXEDIT Server administrators to configure ONIXEDIT AIA. You can access it with your preferred browser at the URL that was provided to you by ONIXEDIT after your ONIXEDIT Server system implementation. You will need your ONIXEDIT Server admin user and password to login:


After having successfully entered your user name and password, you will be automatically routed to the Dashboard Summary.

Dashboard Summary

The Dashboard Summary, is the first thing that you see after login to AIA Dashboard:
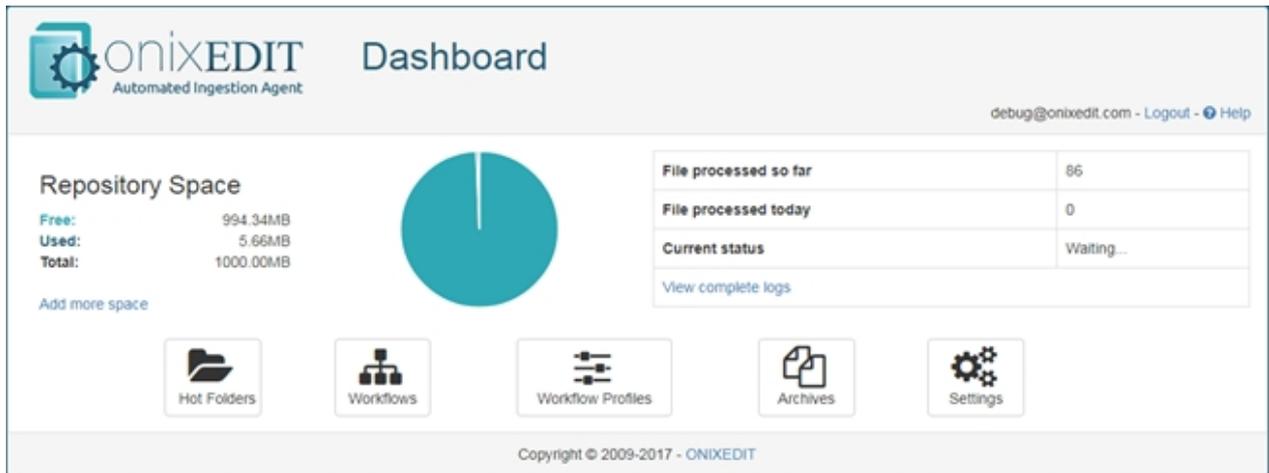
The Dashboard Summary allow you to quickly see:

- The remaining space in your files repository.
- The number of files processed.
- The current status of ONIXEDIT AIA.
- And the log of your file processed.
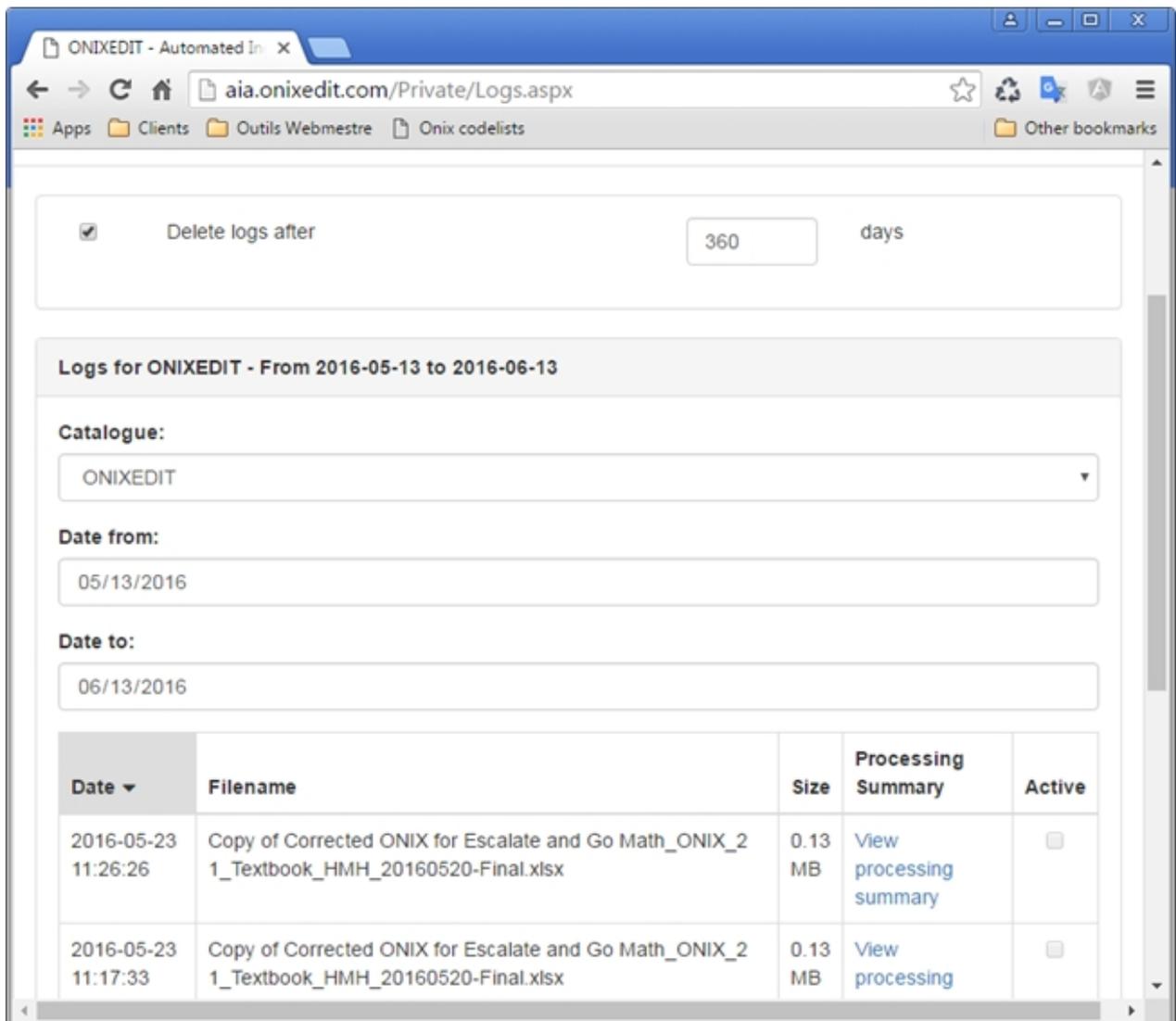
From the Dashboard Summary, you can get access to your Hot Folders, their work-flows, the Profiles, Archived files and Folder Settings:

Hot Folders

For each of your ONIXEDIT Server catalogues:

You will find a corresponding "Hot Folder" having the same name providing that your ONIXEDIT Server catalogues have "Import Rules" defined and are authorized for the Admin user. See Troubleshooting section for details.

The number shown in the folder icon of each Hot Folder indicates if the corresponding catalog is compliant to ONIX 2.1 or ONIX 3.0.

In the upper section of that page, you will find buttons that will allow you set default ONIXEDIT import scripts and corresponding Excel templates (optional):

Default import scripts will be used automatically by AIA when a non-ONIX metadata file will be uploaded to an hot folder for which there is no ONIXEDIT import script defined. Using that import script, AIA will be able to convert the flat file to an ONIX file. Consult the ONIXEDIT User Manual to learn more about import scripts.

You can also set specific ONIXEDIT import script for each hot-folder. You just have to select the hot-folder for which you want to set an import script and click on the button "Set ONIXEDIT import script for ...":

Only valid ONIXEDIT import scripts (files with extension ".import" or ".import3") are authorized.

Once an import script is set for an hot folder, you will have to click on the button "Reset to default script" prior to change the script:

If you have different types of flat files to import to a specific ONIXEDIT Server catalogue, you can create Sub-Folders and assign different import scripts for each. To add a sub-folder to an hot-folder, follow these steps:

1- You select the hot-folder for which you want to add a sub-folder.
2- You click on the button "+ Add a sub-folder":

You can define as many sub-folders as you want:

Each sub-folder can be set with a distinct and specific ONIXEDIT Import script. And like regular hot folders, each sub-folders can receive metadata (ONIX or flat files) and asset files (cover images, eBooks, etc).

Working with "Drop Zone"

When you select an hot folder or a sub folder, a drop zone is automatically activated:

You can drag & drop metadata and asset files from your desktop to that zone to upload those files to the selected folder. ONIXEDIT Automated Ingestion Agent will then process the uploaded files at its next run. By default, there is a limit of 20MB for each file but this can be increased upon request for ONIXEDIT Server "on premises" implementation.

You can also provide a dedicated drop zone to your partners without having to let them know your ONIXEDIT credentials. To do so, simply copy the drop zone URL dedicated to each single folder:

You can then send that link by email to your partner to let him upload metadata and asset files directly to a specific hot folder or sub folder:
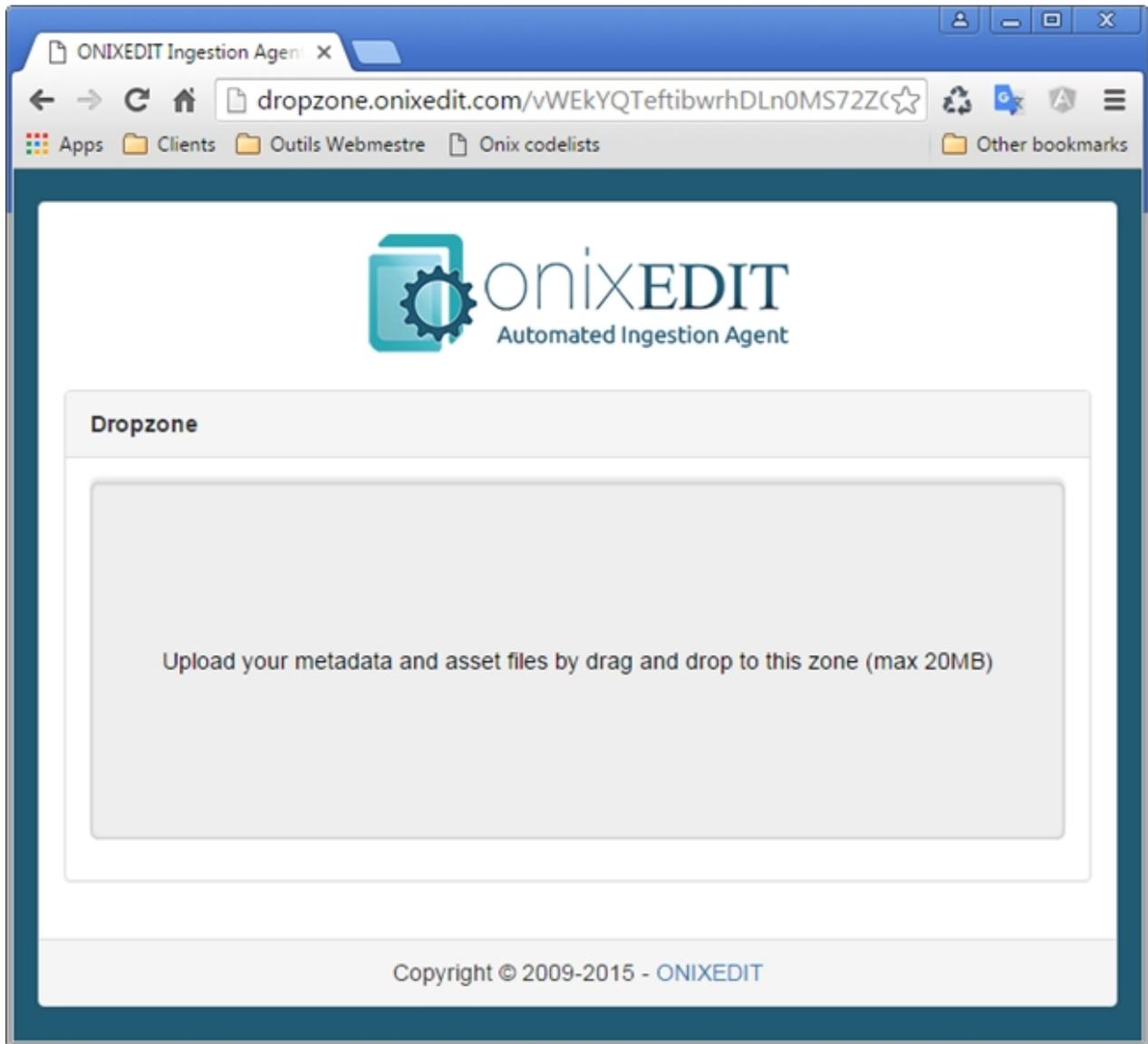
If ever you want to cancel a drop zone URL, you just have to renew it:
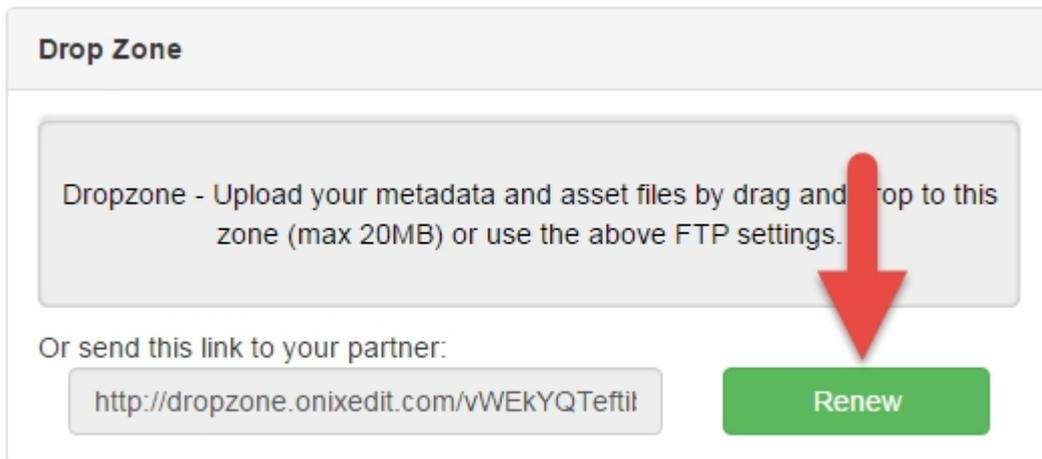
After having been renewed, the former URL of the drop zone won't work anymore:
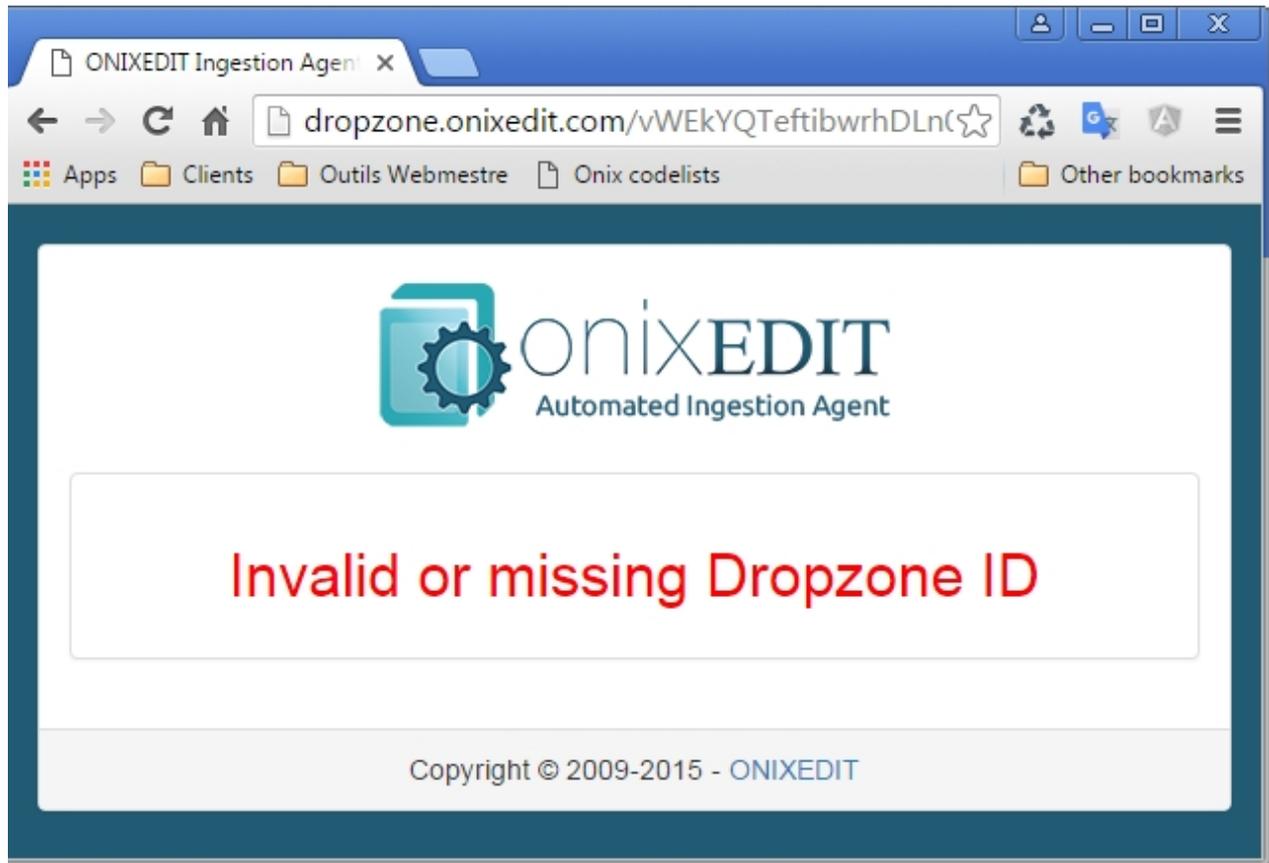
Workflow

You can activate a series of actions for each hot folders and sub-folders from the Workflows page with the help of Workflow Profiles:

By default, each Hot Folder is associated to the "Default" Workflow Profile. But you can change it easily by selecting the Hot Folder you want to change and then, by choosing a different profile:

Once selected, the Workflow Profile will determine the Metadata Files Workflow:

Metadata Files Workflow from Profile Default

	Action	Settings
☑	Send email upon Success	support@onixedit.com
☑	Send email upon Failure	support@onixedit.com
☐	Save HotFolder path to Custom Field	
☑	Keep Ingested File in Archive	
☐	Copy Ingested File to FTP	
☐	Enhance metadata with Amazon API (only for inserted titles)	
☐	Send Summary File to FTP upon batch completion	
☐	Send Summary Email upon batch completion	
☐	Send ONIX Acknowledgment file to FTP	
☐	Post ONIX Acknowledgment to URL	

And the Asset Files Workflow:

Asset Files Workflow from Profile Default

	Action	Settings
☐	Retrieve external covers (from metadata) to local media space *(Runs every 12 hours)*	
☑	Make thumbnails from cover image	
☐	Process orphan media files *(Media files older than 12 hours and no metadata file present)*	
☐	Avoid duplicate media files *(Applied to file with the same name and for same type)*	
☐	Extract full text from eBook (PDF only)	☐ Page by page
☐	Archive eBooks with extension	
☐	Copy cover images to FTP	
☐	Copy eBooks to FTP	

The only section you can change from the Workflow page is the External Fetch Workflow:

External Fetch Workflow

	Action	Settings
☐	Fetch files from external FTP *Current server time: 18 juillet 2017 09:11*	
☑	Move processed files to Archives *(Archives folder will be created in FTP directory for this, please be sure that FTP user has write access)*	
	Scheduled day for fetch	Everyday ▼
	Scheduled time for fetch	13 ▼ : 10 ▼

The Metadata Files Workflow and Asset Files Workflow are Read Only in the Workflow page. To change it you must go to the [Workflow Profiles](#) page. But the External Fetch Workflow is specific to each Hot Folder. It allow AIA to automatically fetch files from an external FTP source an drop it directly to the hot folder following a particular schedule. To not process the same files again and again, you can optionally move processed files to an "Archives" folder:

External Fetch Workflow

	Action	Settings
☑	Fetch files from external FTP *Current server time: 18 juillet 2017 09:24*	ftp.acme.com
☑	Move processed files to Archives *(Archives folder will be created in FTP directory for this, please be sure that FTP user has write access)*	
	Scheduled day for fetch	Everyday ▼
	Scheduled time for fetch	13 ▼ : 10 ▼

You can choose the day of the week when the External Fetch function will execute or you can also set it daily:

Scheduled day for fetch	Everyday ▼
	Sunday
Scheduled time for fetch	Monday
	Tuesday
	Wednesday
	Thursday
	Friday
	Saturday
	Everyday

You also need to set an hour when the fetch function will start. The time used is the one of the server and current server time is shown here:

External Fetch Workflow

	Action	Settings
☑	Fetch files from external FTP *Current server time: 18 juillet 2017 09:24*	ftp.acme.com

Note that the External Fetch mechanism is not recursive. It will process files at the first level of the path provided in the FTP settings:

If no Path is provided, only the root level of your FTP account will be processed.

Workflow Profiles

Workflow profiles allows you to easily manage different group of workflow actions to be assigned to your hot folders. You can define as many Workflow Profiles as you want. Every AIA installation have at least a default workflow profile having the most common group of actions already set. To create a new workflow profile you simply have to click on "Create a new Profile from this one" button:

Then you should give a name to the new profile and click on Save:

You can also rename or delete an existing profile:

If you delete a profile, all the Hot Folders associated to it will be switched to the Default profile automatically.

You can't rename or delete the Default profile.

Once created, you will be able to set the workflow actions that you need for the current (selected) profile.

There is 3 types of workflow actions:

- **Metadata files workflow actions** are those applied to metadata files after their ingestion by AIA:

Metadata Files Workflow for My new Profile

	Action	Settings
☑	Send email upon Success	support@onixedit.com
☑	Send email upon Failure	support@onixedit.com
☐	Save HotFolder path to Custom Field	
☑	Keep Ingested File in Archive	
☐	Copy Ingested File to FTP	
☐	Enhance metadata with Amazon API (only for inserted titles)	
☐	Send Summary File to FTP upon batch completion	
☐	Send Summary Email upon batch completion	
☐	Send ONIX Acknowledgment file to FTP	
☐	Post ONIX Acknowledgment to URL	

- o Send email upon Success: Activate transmission of a processing report send by email to one or more email addresses. An email report is sent after ingestion of each metadata file uploaded to the corresponding folder. To set more than one email, use the comma as a separator. For instance: jesse@acme.com,simon@acme.com.
- o Send email upon Failure: Activate transmission of an error report send by email to one or more email addresses. That error report is sent only if a problem occurs during processing of a metadata file. To set more than one email, use the comma as a separator.
- o Save HotFolder path to Custom Field: When activated for a folder, each title ingested by AIA will get a special Custom Field named **IA_HotFolderPath** that will be automatically set to the folder path where the metadata file was uploaded. For instance, if we set that action for our sub folder "CSV files", all titles contained in a metadata file uploaded to the CSV files folder will get "ONIXEDIT/CSV files" as a value for the custom field IA_HotFolderPath. This may be useful later to find titles having been ingested from a specific hot folder.

You will be able to search/see/edit that custom field using ONIXEDIT as well as ONIXEDIT Scheduler allowing transmission of metadata based on their origins.

- o Keep Ingested File in Archive: When checked, the files ingested will be saved in the Archive. Otherwise, those files will be deleted after having been processed.
- o Copy Ingested File to FTP: This allow you to define a specific FTP site to which each file will be uploaded after ingestion. When you check this option, you will have to click on the "Edit" link that will appear to set the ftp address and your credentials. If you set a destination path, you must insure that it exist prior to save the configuration. Use the "Test FTP connection" button to check if everything is ok prior to save.

- o Enhance metadata with Amazon API: If you are using Amazon already to sell

your books, you can obtain the metadata of your books from Amazon using their API. You will need some IDs form Amazon to use their API:

☑ Enhance metadata with Amazon API (only for inserted titles)

Associate Tag:

Access Key ID:

Secret Access Key:

Marketplace:

US ▾

Save

Once set, you will be able to import very basic ONIX or flat files and the API will complete your metadata for any new titles added to an ONIXEDIT Server catalog. For instance, you can ingest an Excel file having only a series of ISBN and you will get as a result, complete ONIX titles enhanced with the metadata coming from Amazon.

o Send Summary File to FTP upon batch completion: This option will allow you to send a text file at a specific FTP location when every files uploaded to the selected folder and its sub-folders will be ingested. The file will include the name of each file ingested with the date and time of the processing. Like the previous option, you will have to set the address and credentials of the FTP site where you want to send the summary file. This may be useful to send an acknowledge signal to external systems about completion of files ingestion.

o Send Summary Email upon batch completion: same as the previous option except that the batch summary report is sent by email at a specific email address. To set more than one email, use the comma as a separator. For instance: jesse@acme.com,simon@acme.com.

o Send ONIX Acknowledge File to FTP: ONIXEDIT Server is compliant with the ONIX for Books Acknowledgement format established by EDItEUR. Here is the EDITEUR description for this standard:

The ONIX for Books Acknowledgement Format Specification, released January 2015, is a separate and optional message intended to be sent to a data supplier in response to receipt of a standard ONIX for Books message. It allows an ONIX for Books recipient to:

- confirm receipt of the original ONIX message
- report to the original sender a summary of Product records processed and updated into the recipient's database
- report details of any errors encountered, or any queries about the product data supplied, or
- return to the sender details of proprietary identifiers assigned by the recipient

o Post ONIX Acknowledge to URL: Same as the above but instead of sending the acknowledge file to a FTP site, this action post it to a specific URL address.

- **Asset files workflow actions** are those applied to asset file during and after their

ingestion by AIA:

Asset Files Workflow for My new Profile	
Action	**Settings**
☐ Retrieve external covers (from metadata) to local media space *(Runs every 12 hours)*	
☑ Make thumbnails from cover image	
☐ Process orphan media files *(Media files older than 12 hours and no metadata file present)*	
☐ Avoid duplicate media files *(Applied to file with the same name and for same type)*	
☐ Extract full text from eBook (PDF only)	☐ Page by page
☐ Archive eBooks with extension	
☐ Copy cover images to FTP	
☐ Copy eBooks to FTP	

o Retrieve external covers (from metadata) to local media space: This option when checked, will activate an automated mechanism that will verify each Media File defined in your ONIX metadata. Each media file having an external url (not in your ONIXEDIT Media Space) will be fetched and saved to your own ONIXEDIT Media Space. As a result, you will find in your ONIX title, a new media file having the usual url address of your media space. For instance, "http://www.amazing.com/book_photos/cover-image-249.jpg" will become something like this "http://www.onixedit.com/MediaSpace/87CAF7E4-D61B-4E8E-9999-1273214FEACC/5228ab27-a5cd-42de-b646-e32a773796cd/9781553792260.jpg". When you check this option, you will have the choice to keep the original reference to the media file or remove it leaving just the new fetched one in place:

☑ Retrieve external covers (from metadata) to local media space *(Runs every 12 hours)*	Keep external covers in metadata ▼ Keep external covers in metadata Remove external covers from metadata

o Make thumbnails from cover image: When this option is checked, AIA will automatically create a thumbnail from the cover images ingested and it will add it to your Media Space as well as in your ONIX metadata.

o Process orphan media files: When this option is activated, AIA will automatically

import PDF and ePUB files that will be uploaded to the hot folder. Each file will have to be named using ISBN or RecordReference. If the corresponding title exist, a new Media File Link leading to the final location of the eBook in your Media Space will be added to that title. It is strongly recommended that you secure your eBooks. To do so, set eBook file types to be placed in your Private Media Space and not in the Public section from your ONIXEDIT Server preferences. Consult the [ONIXEDIT Server manual](#) for more details.

o Avoid duplicate media files: When activated, this action will check each incoming Asset file and will compare it to files already added to your Media Space for the same title. Identical files based on their filename and content, won't be ingested.

o Extract full text from eBook (PDF only): With this option, you can extract the text from PDF (uploaded with your metadata) to store in ONIXEDIT Server database. This provide an additional searching capability through the complete book text instead of ONIX metadata only. The "Page by page" checkbox is an option that will extract and save text from your PDF on a page per page basis instead of saving a very large text of all the pages (for better performance when dealing with very large eBooks).

o Archive eBooks with extension: This option will keep all eBooks having the extensions defined to the AIA Archive.

☑	Keep eBooks with extension	.pdf, .epub ▼
		Save

o Copy cover images to FTP: With this option checked, you will be able to transfer automatically all cover images ingested by AIA to a specific FTP site:

☑	Copy cover images to FTP	**Host:**
		Username:
		Password:
		Path:
		Save
		Test FTP connection

o Copy eBooks to FTP: Same as for the option "Copy cover images to FTP", but for eBooks ingested by AIA.

- **Custom Asset Extensions Routing**: The Custom Asset Extension Routing is a new optional concept of AIA. By default, there is no Asset Extension routing defined and the

system behave as before. Thus, it will ingest Cover images (jpeg, png, tif, etc) as well as eBooks (PDF and EPUB). But you can change how AIA will process the Asset files by defining the extension you want to support and to which ONIX metadata element it should be added.

For instance, the following Asset Extensions Routing configuration will take care of eBooks and cover images:

Custom Asset Extensions Routing

Every file uploaded to the Hot Folder with this Workflow Profile having ISBN/EAN/RecordReference and the Filename string marker in its name will be kept in your MediaSpace according to the following routing criteria:

Add new asset routing criteria

Extension	Filename string marker	Keep as	
pdf	cover	Media File - Image: front cover	Delete
pdf		Other Text - Full text	Delete
jpeg,jpg,png,tif,tiff		Media File - Image: front cover	Delete
epub		Other Text - Full text	Delete
azw,azw3,kf8		Other Text - Full text	Delete

Cover images files will goes to the ONIXEDIT Server Media Space and each file will produce an entry in the ONIX metadata Media File section of type "Image: front cover". The extensions that will allow AIA to recognize media files are "jpeg,jpg,png,tif,tiff" and "pdf". But for the latest, the file name must also have the string "cover" in its name:

Extension	Filename string marker	Keep as	
pdf	cover	Media File - Image: front cover	Delete
pdf		Other Text - Full text	Delete

For instance, "9780006391555_cover.pdf" will be recognized as a cover image, but "9780006391555.pdf" will be recognized as an eBook (Full text).

About ebooks, the above configuration will instruct AIA to process PDF that have no "cover" word in their filenames as well as files with extension "epub" or those with extensions "azw,azw3,kf8" (Kindle).

Archives

Each file ingested by AIA is automatically saved to the Archive folder of AIA FTP site. The archive page is where you will find pruning settings for those files as well as the archived files classified by hot folders. Thus, by clicking on a hot folder, you will see at the right all the files archived for that folder:

You can then click on a specific file to download it to your PC.

Settings

The Settings page allow you to choose Amazon Simple Storage Service (S3) as an alternative to the default FTP folders used by AIA for hot folders, Media space and archive.
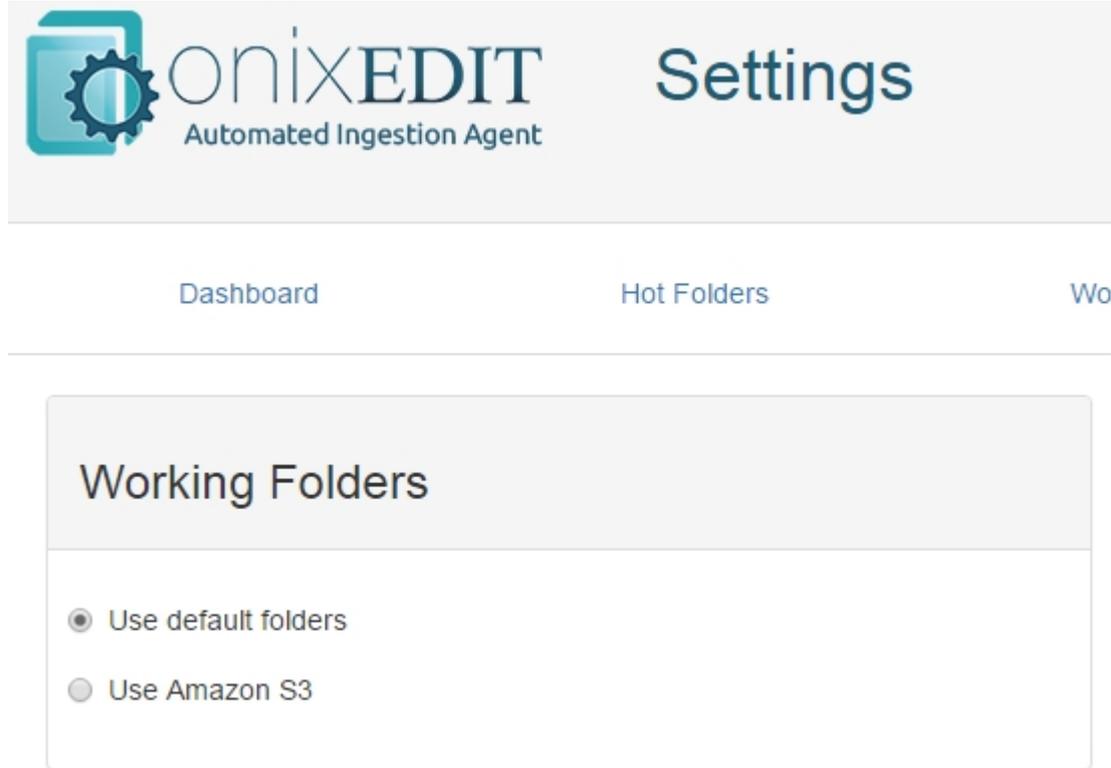
onixEDIT
Automated Ingestion Agent

Settings

28 / 31

Dashboard   Hot Folders   Wo

Working Folders

⦿ Use default folders

◯ Use Amazon S3

Warning: If you switch to S3, all your default FTP folders will be removed and S3 buckets will be used instead. But the names of the hot folders will stay the same:
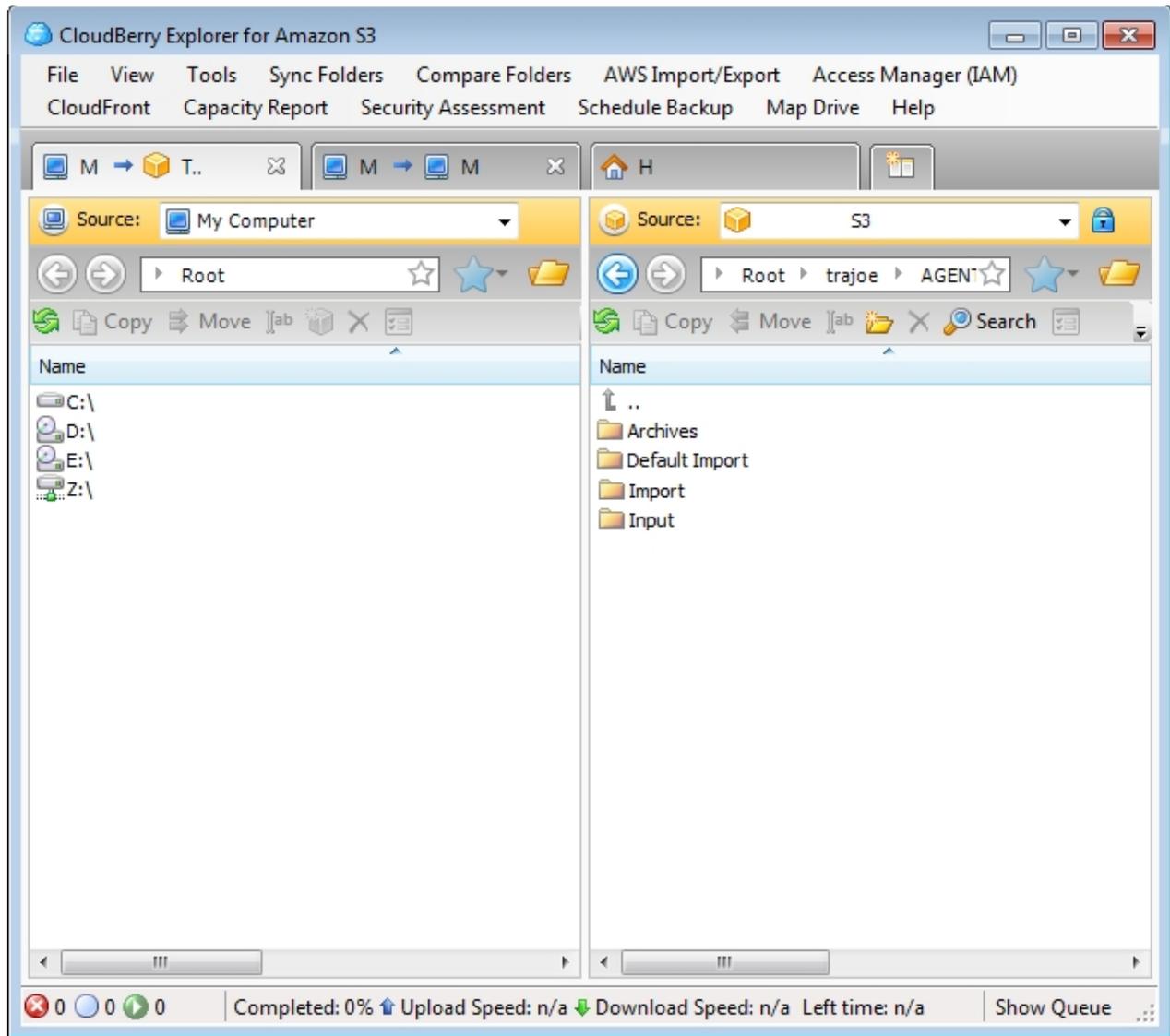
Custom Asset Extensions Routing

Custom Asset Extension Routing are defined at the Workflow Profiles page.

Troubleshooting

I don't see the hot folder for my new catalogue!

Check that your catalogue have "Import Rules" defined in your ONIXEDIT preferences/Server Edition/Catalogues:
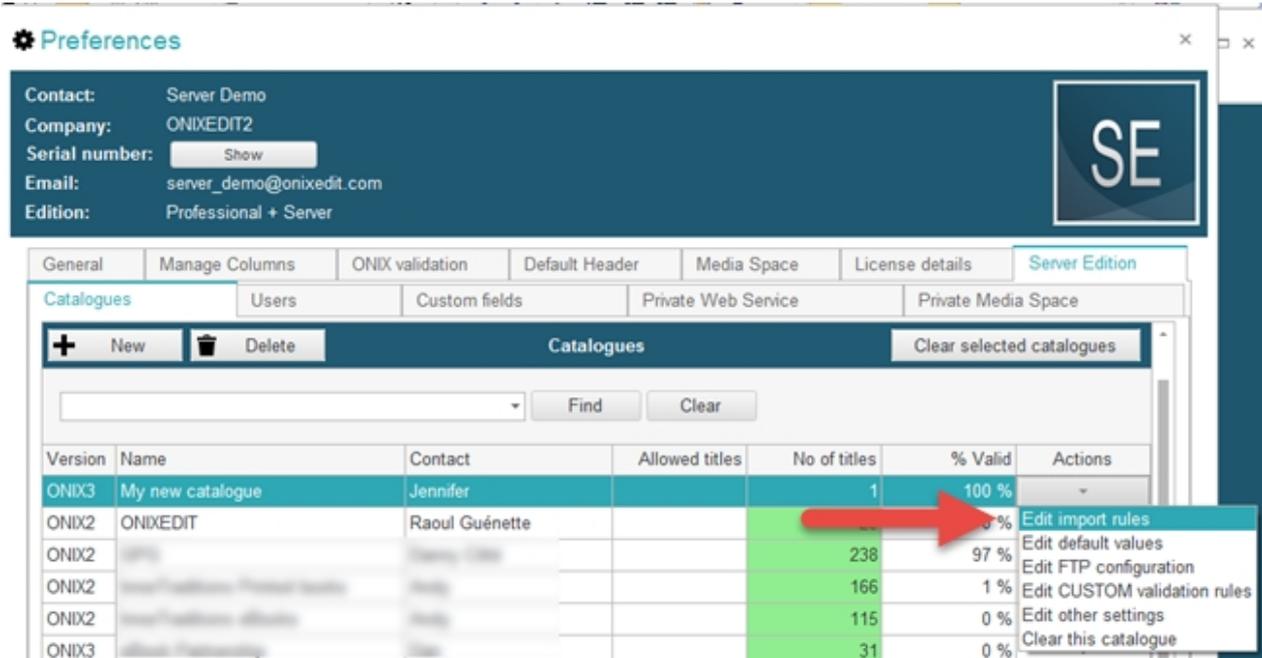
If the catalogue do not have import rules, ONIXEDIT AIA won't create the corresponding hot folder.

You must also insure that the ONIXEDIT Server owner (the admin user) has the user rights to Read, Add, Modify and Delete titles in your catalogue.
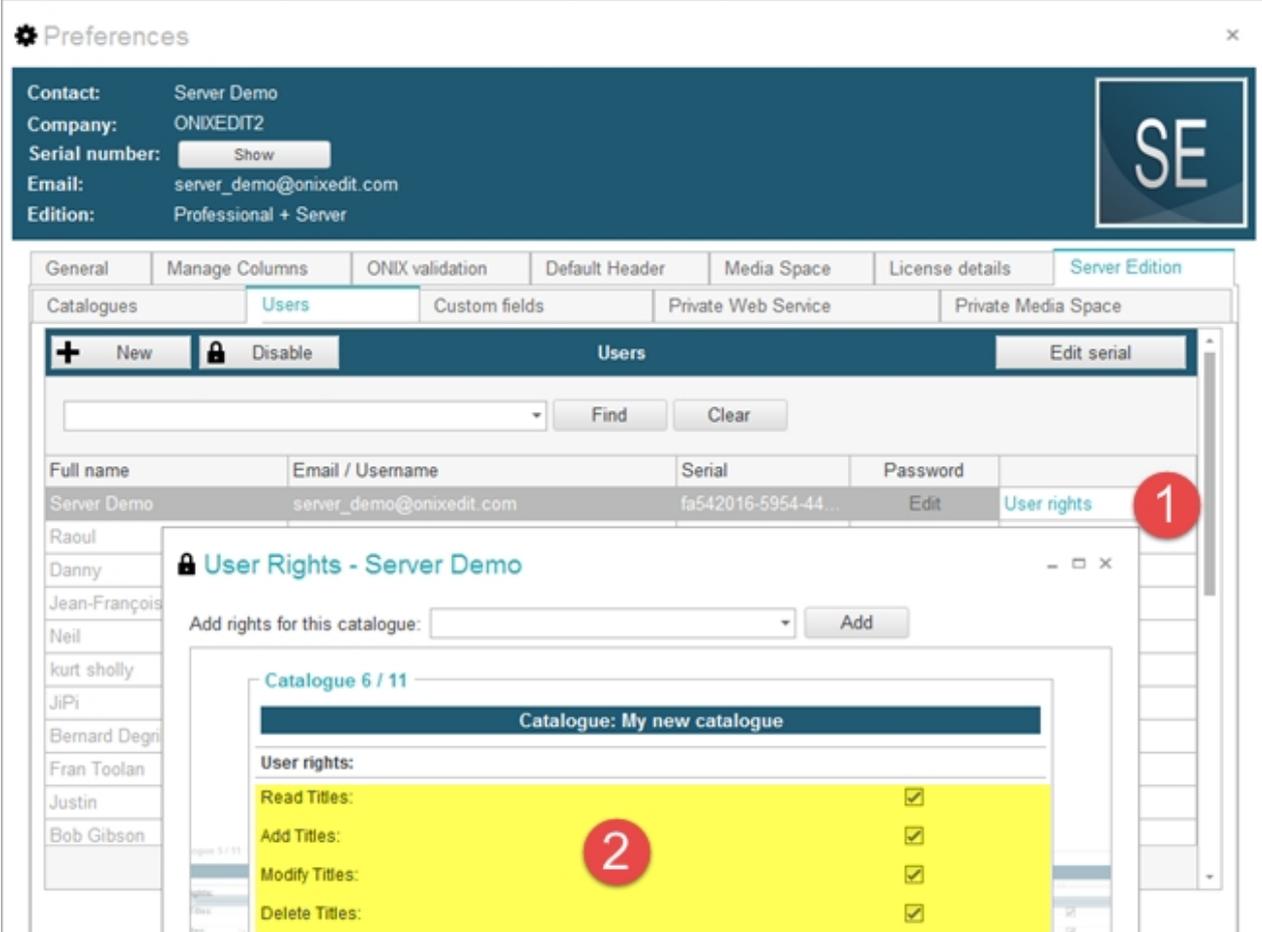
The metadata file that I uploaded to my hot folder is ingested but not the cover images!

metadata file is the trigger ONIXEDIT AIA is looking for to begin the ingestion of the hot folder. If you upload your asset files (cover images or eBooks) at the same moment as your metadata file it is possible that the metadata file upload completes prior the asset files. ONIXEDIT AIA will then begin the title ingestion before the cover images or eBooks upload is completed. As a result, your assets files will be ignored and will stay in the hot folder until another corresponding metadata file is uploaded. However, a possible option for this situation may be to activate the workflow action "Process orphan media files".