



Rosetta 8.1 Highlights

February '24



Contents

Data Management	3
Automate generation of preservation copies with replicated structMap.....	3
Harvesting Multiple OAI Sets in a Single Job	4
Preservation.....	5
Search by format validation automatic rule	5
Display Rule ID in Rule List and Rule Details.....	6
Index Format Library Version Field.....	6
Validate source metadata on ingest	7
Basic Preservation Plan – Enhancements	7
Documentation	8
New Online Help	8
Infrastructure.....	9
Change S3 configuration without impacting METS	9
Deposit from an S3 source - Manual Support.....	9
CDATA Management.....	10
3 rd Party Upgrades	10

Data Management

Automate generation of preservation copies with replicated structMap

This major enhancement provides multiple new and useful workflows, designed at automating generation of preservation copies, based on the original master preservation copy. These include:

1. “Create Copy Representation” process will allow generating an additional copy of any type, based on the original copy. The new copy can either be migrated from the original, using the format migration tools in the system, or simply copy the original. Any migrated file will include an event with details of the originating file, similarly to the existing Rosetta migration.
2. Users can define whether to copy the original structMap over to the new copy, adapting the pointers to the new files.
3. Users can define whether the new representation is a preservation or derivative copy. In case of a preservation copy, validation stack will run on the new files which will be sent to the technical analysis workbench if required.
4. Job can either run on ingest or as an ongoing update process.

Create Copy Representation - Create Copy Representation

* Preservation Type (Input) Preservation Master * Representation Entity Type (Input) None

File Extension Filter

1 Items selected Remove all Add all

tif	+
tiff	+
8td	+
000	+
123	+
3dd	+
3dm	+
3dmf	+
3ds	+

Copy other extensions

Copy logical structmap

Plug-in Type

Transformation Profile TIFF to JP2

Stream Handler Jpg

* Preservation Type (Output) Modified Master

* Representation Code None

* Representation Entity Type (Output) None

* Validation Stack Validation Stack - Full

Access Rights Policy None

* File Entity Type (Output) None

Figure 1 Create Copy Representation



Harvesting Multiple OAI Sets in a Single Job



This feature will allow users to harvest multiple sets in a single OAI Harvesting Job as long as the base URL, Material Flow, and Update Metadata Job are common for all sets. This will avoid the need to create a dedicated job per set.



The job page will allow users to run a test harvest on any selected set.

The screenshot displays the 'Job Parameters' and 'Test Area' sections of the OAI Harvesting interface. The 'Job Parameters' section includes a 'Base URL' field with the value 'http://na02.almalibrary.com/view/oai/EXLDEV1_INST/request', a 'Connect and Edit' button, and a 'Sets' list. The 'Sets' list shows 7 items selected, including 'gould_outputs', 'lappel_outputs', 'word_keywords_feminismo', 'kerupd', 'weber_outputs', 'ford_outputs', and 'Rapid_Books_Electronic_LendableInternational'. The 'Test Area' section is highlighted with a red box and contains a 'Test' button and a dropdown menu for selecting a set. A red arrow points from the 'Sets' list to the 'Test Area' dropdown menu.

Figure 2 Harvest multiple OAI sets

Preservation

Search by format validation automatic rule

Rosetta includes a wide variety of rules which allow automating decision making for technical issues. These are the existing rule types:

1. **Format Identification Auto Correction Criteria** (*event 198*)
2. **File Extension Mismatch Ignore Criteria** (*event 404*)
3. **Metadata Extraction Error Ignore Criteria** (*event 406*)
4. **Virus Check Error Ignore Criteria** (*event 422*)

Following the splitting of the Format Validation and Metadata Extraction, a new search option has been added for the format validation error rules: **Format Validation Error Ignore Criteria** (*event 204*)

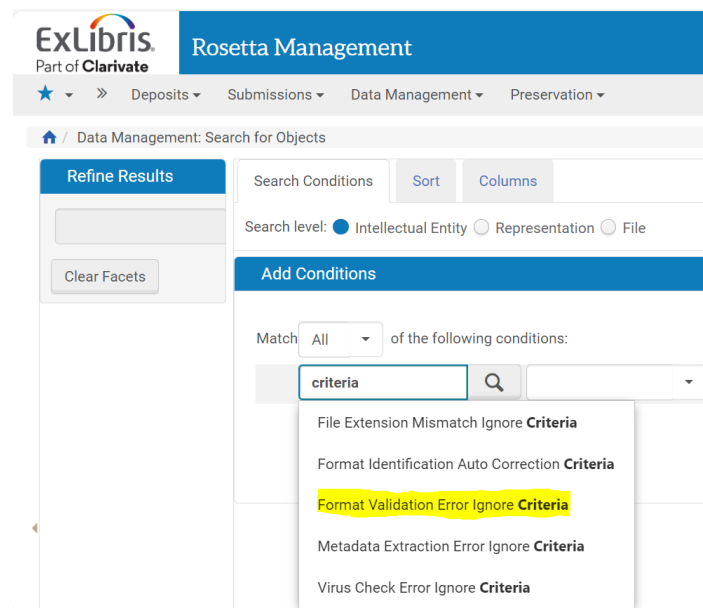
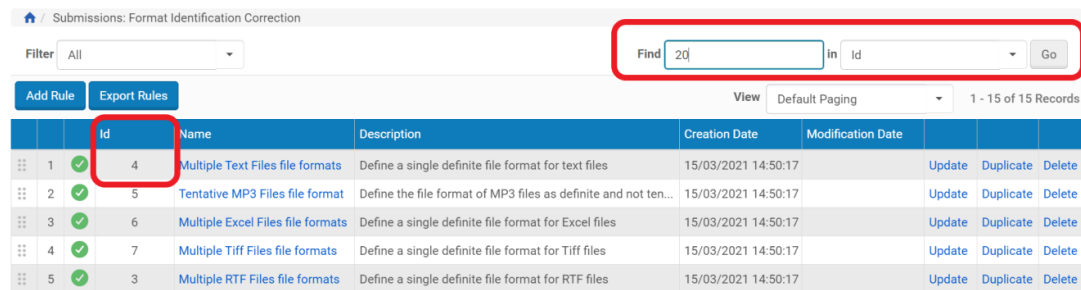


Figure 1 Search by format validation rule

Display Rule ID in Rule List and Rule Details

Multiple improvements in managing Rosetta technical rules, including:

1. Display rule ID of Format Identification Correction, Metadata Extraction Error and Virus Check Error rules in the respective rule lists.
2. Enable search for rule ID in the "Find" field right above the rule list.
3. Add rule ID column to the csv file that is generated by the "Export Rules" function.
4. Display rule ID in the rule details view



The screenshot shows the 'Submissions: Format Identification Correction' page. At the top, there is a search bar with 'Find' and 'Go' buttons, and a dropdown menu set to 'id'. Below the search bar are 'Add Rule' and 'Export Rules' buttons. The table below has columns for 'Id', 'Name', 'Description', 'Creation Date', 'Modification Date', and actions. The 'Id' column is highlighted with a red box, and the value '4' in the first row is also highlighted with a red box.

	Id	Name	Description	Creation Date	Modification Date			
1	4	Multiple Text Files file formats	Define a single definite file format for text files	15/03/2021 14:50:17		Update	Duplicate	Delete
2	5	Tentative MP3 Files file format	Define the file format of MP3 files as definite and not ten...	15/03/2021 14:50:17		Update	Duplicate	Delete
3	6	Multiple Excel Files file formats	Define a single definite file format for Excel files	15/03/2021 14:50:17		Update	Duplicate	Delete
4	7	Multiple Tiff Files file formats	Define a single definite file format for Tiff files	15/03/2021 14:50:17		Update	Duplicate	Delete
5	3	Multiple RTF Files file formats	Define a single definite file format for RTF files	15/03/2021 14:50:17		Update	Duplicate	Delete

Figure 2 Rule ID Display

Index Format Library Version Field

As our community phrased it - The version of the File Format Library is vital information to understand how ground truth about a file format was gathered. Previously visible in the UI and captured in the METS techMD section, the format library version is now also indexed and searchable in the UI.



Validate source metadata on ingest

Rosetta will validate all sourceMD sections against their respective customizable XSD schemas during METS ingests, in order to allow submitting valid metadata only into the archive.

Basic Preservation Plan – Enhancements

Enhanced functionality for the basic preservation plan, including:

1. Users can choose to create Logical or Itemized Preservation Set for the basic preservation planning workflow
2. Improved search capabilities in *refine execution set* by adding “or” operator

Documentation



New Online Help

We are thrilled to announce a significant milestone in our ongoing commitment to enhancing customer experiences—the [Rosetta Online Help](#) is now directly accessible on the Knowledge Center! As part of a large-scale project, we have switched from the traditional PDF format to a more dynamic and user-friendly online platform.

A link to the new landing page was added to the Rosetta Help menu

This change reflects our commitment to providing smooth and effective support to our customers and simplifying internal processes for our staff. The transition to the Knowledge Center marks a significant step forward in our efforts to improve accessibility, searchability, and overall user experience.

Key Benefits:

1. **Accessibility:** The Rosetta online help is now just a click away on the Knowledge Center, making it easily accessible to our customers. This shift eliminates the need to download and navigate through PDFs, saving time and effort.
2. **User-Friendly Interface:** The Knowledge Center offers an intuitive and user-friendly interface, enabling users to quickly find the information they need. The improved layout and navigation will create a more positive and efficient user experience.
3. **Real-time Updates:** The online platform allows for real-time updates, ensuring the information provided is always current and reflects the latest developments. This dynamic approach enhances the accuracy and relevance of the content available.
4. **Feedback:** Users can provide feedback directly on the platform. This fosters a sense of community and enables continuous improvement based on user input.

We encourage everyone to explore the enhanced Rosetta online help on the Knowledge Center and take advantage of its new features.

NOTE Due to the new technology, the in-app documentation will display previous documentation from 8.0. For documentation on 8.1 features, please refer to the online help. This gap will be adjusted in 8.2

Infrastructure



Change S3 configuration without impacting METS

The S3 storage endpoint can now be updated without creating new IE versions.

A new checkbox “Use Storage Prefix” was added to the S3 storage configuration UI. Once this is enabled, the METS XML will store a persistent storage prefix instead of the explicit endpoint (in the FLocat).

The screenshot shows a 'Plugin Parameters' configuration window. The 'Use Storage Prefix' checkbox is checked. Other fields include Endpoint, Storage Prefix, Bucket name, Secret Access Key, Access Key ID, Region, and Maximum waiting time (Milliseconds) set to 1800000.

Plugin Parameters	
* Endpoint	<input type="text"/>
Use Storage Prefix	<input checked="" type="checkbox"/>
Storage Prefix	<input type="text"/>
* Bucket name	<input type="text"/>
* Secret Access Key	<input type="text"/>
* Access Key ID	<input type="text"/>
Region	<input type="text"/>
* Maximum waiting time (Milliseconds)	1800000

Figure 3 S3 Prefix configuration

Deposit from an S3 source - Manual Support

Support for manual deposits from S3 sources, allowing for dynamic selection of a deposit folder.

A new configuration “Allow Navigation” defines whether the content provider can navigate between the “folders” to select the desired SIP location.



CDATA Management

A major improvement has been applied to CDATA handling in Rosetta, providing consistent management across submission methods and CDATA structures.

A new general parameter `LOAD_SAVE_CDATA_BOOKMARKS` under the "general" module.

will cause XmlBeans to utilize XmlOptions and parse elements and CDATA elements in a consistent manner.

This Parameter is false by default to allow controlled application by the customers.

3rd Party Upgrades

As always, Rosetta is aligned with the latest 3rd party components in order to provide the most secure environment as possible. 8.0 includes, amongst others:

- Apache 2.4.58
- Tomcat 9.0.84
- OpenSSL 3
- Java 17.0.8





ExLibris
Rosetta