



# Rosetta 7.4 Highlights

January '23



## Contents

Preservation .....	3
Bulk risk assignment to file formats.....	3
Multiple file format identification .....	5
Basic preservation planning – Improved Management.....	5
Data Management .....	6
Migrate collections to structural IEs .....	6
Enhanced date field search.....	7
Integrations.....	8
Deposit from an S3 source.....	8
REST APIs - Data Management .....	8
Infrastructure .....	9
3 <sup>rd</sup> Party Upgrades .....	9
Additional.....	9

# Preservation

## Bulk risk assignment to file formats

Single risks can now be assigned to multiple formats in bulk, by assigning a risk to multiple (or all) PUIDs at once in the Format Library. For this purpose risks have been decoupled from formats, saving significant time in cases such as when a single risk is applicable to an entire classification of formats, e.g video files.

In addition, it is now possible also to create, edit and delete a risk identifier from the format library risk identifiers page.

The following enhancements are included:

- New 'Edit' column in the Risk Identifiers page under
- The edit page includes a “Related Formats” tab with an “Add Format” action and deactivate action for one or more formats
- “Add Format” displays a searchable list of all formats with multi-select option
- From the risk identifiers page, it is also possible to add a new risk identifier using the button ‘Add Risk Identifier’

	Name	Risk Identifier	Risk Type	Risk Source	Format			
1	LateModificationD...	-	PROPERTY_DRIVEN	Local	ExL-Fmt-21	View	Edit	Delete
2	Large ISO block	-	PROPERTY_DRIVEN	Local	fmt/468	View	Edit	Delete
3	NLB FT	-	PROPERTY_DRIVEN	Local	fmt/44	View	Edit	Delete
4	Encrypted risk	-	PROPERTY_DRIVEN	Local	fmt/494	View	Edit	Delete
5	CCITT Risk	-	PROPERTY_DRIVEN	Local	fmt/4	View	Edit	Delete

Figure 1 Add risk identifier from Risk Identifiers page

🏠 / Preservation: Risk Identifiers / Details

<b>Name</b>	LZW compression in image	<b>Created on</b>	21/12/2022
<b>Description</b>	-	<b>Created by</b>	Ex Libris
<b>Classification</b>	-		

General Details Risk Parameters Related Formats History

Add Format Deactivate All

	Activated	Name	Description	Version
1	✓	fmt/353	Tagged Image File Format	
2	✓	fmt/7	Tagged Image File Format	3
3	✓	fmt/8	Tagged Image File Format	4
4	✓	fmt/9	Tagged Image File Format	5

Figure 2 Add risk page, with 'Add Format' action

🏠 / Preservation: Risk Identifiers / Details

### List of Formats

Find:

View 20 Records ▾

	<input type="checkbox"/>	Name ▲	Description	Classification
1	<input type="checkbox"/>	ExL-Fmt-21	Finale Notation Format	Generic
2	<input type="checkbox"/>	ExL-Fmt-22	CDX File Format	Text (Unstructured)
3	<input type="checkbox"/>	ExL-Fmt-23	ENIGMA Transportable File	Generic
4	<input type="checkbox"/>	ExL-Fmt-41	Open Publication Structure (OPS) 2.0	eBook
5	<input type="checkbox"/>	ExL-Fmt-61	MPEG-4 Media File	Video
6	<input type="checkbox"/>	ExL-Fmt-62	Microsoft Office Open XML	Generic
7	<input type="checkbox"/>	ExL-Fmt-161	WebM	Video
8	<input type="checkbox"/>	ExL-Fmt-241	Genealogy GEDCOM File	Text (Structured)
9	<input type="checkbox"/>	ExL-Fmt-261	ARTIST 2 Interchange Format	eBook

Figure 3 Link one-to-many formats to the risk

## Multiple file format identification

It is now possible to setup Rosetta to use multiple file format identification tools.

This allows users to include additional tools for performing file format identification, potentially increasing the accuracy of the file format identification and allowing to search and analyze according to the additional data.

Since the additional tools results are regarded as complementary information at this point, these results and formats do not have operational impact on Rosetta, i.e. they do not trigger an error in the TAW, they are not used for other tasks such as technical MD extraction, format validation, risks analysis, etc.

The results of any additional tools added are regarded as additional information recorded in the METS and indexed for searches.

The format identification task has been enhanced to allow to choose more than one plug-in, where Siegfried has been added as an additional OTB identification tool:

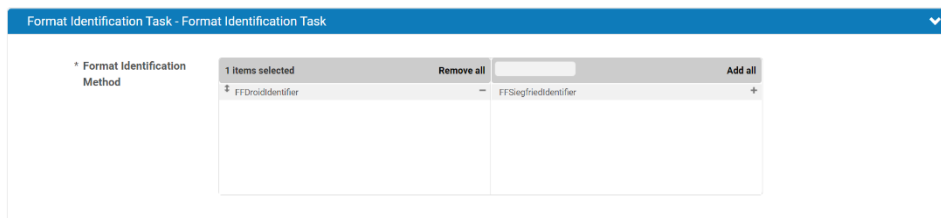


Figure 4 Format identification task with multiple tools

## Basic preservation planning – Improved Management

Basic preservation plans which have not yet been executed can now be deleted by the preservation analyst. This will allow cleaning up of redundant plans and / or reverting back to the original preservation set and recreating the plan.

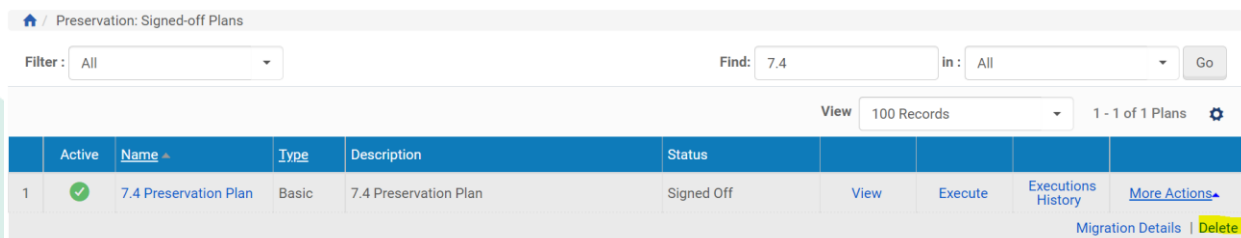


Figure 5 Delete action on signed off basic plan

# Data Management

## Migrate collections to structural IEs

Rosetta supports migrating collections to structural IEs, allowing for a quick and easy conversion with minor manual intervention. The process includes the following steps:

- Collections management page includes a service to export collections as structural IEs. Users can select which top-level collections require export / migration and to which location
- Optional manual manipulation of the export XSL via collection2StructuralIE.xsl
- Submission of the resulting export into Rosetta as a structural IE hierarchy

One pre-requisite is that a single IE can be assigned to one parent IE only.

The new service can also be useful for exports of structural IEs to external sources.

The screenshot displays the 'Data Management: Collection Management' interface. On the left is a tree view of collections, with 'Denliquin Gaol (17614329)' selected. The main area shows details for this collection, including a thumbnail image of a building, a 'View Collection' button, and tabs for 'Contents', 'Metadata', and 'History'. Below the tabs is a table of records with columns for 'Title', 'PID', 'View', and 'Editor'. The table contains 8 records. At the bottom right, a blue button labeled 'Export Collections As Structural IEs' is highlighted with a blue arrow.

Title	PID	View	Editor
1 Name John O'BEIRNE Born 1865 Birth Place Ireland Photo Number 1 Gaol Denliquin	IE56278	View	Editor
2 Name Charles JOHNSON Born 1862 Birth Place Norway Photo Number 2 Gaol Denliquin	IE56281	View	Editor
3 Name Michael WHELAN Born 1864 Birth Place Victoria Photo Number 3 Gaol Denliquin	IE56284	View	Editor
4 Name John Frederick NICHOL Born 1856 Birth Place Victoria Photo Number 4 Gaol Denliquin	IE56287	View	Editor
5 Name Charles Henry BROWN Born 1872 Birth Place Victoria Photo Number 5 Gaol Denliquin	IE56290	View	Editor
6 Name Michael MCEVOY Born 1846 Birth Place Ireland Photo Number 6 Gaol Denliquin	IE56293	View	Editor
7 Name Frederick HOWARD Born 1861 Birth Place America Photo Number 7 Gaol Denliquin	IE56296	View	Editor
8 Name William PHELAN Born 1868 Birth Place Victoria Photo Number 8 Gaol Denliquin	IE56299	View	Editor

Figure 6 Export collections as structural IEs



Home / Data Management: Collection Management / Details

\* Name

---

**Job Parameters**

Top Level Collections

0 items selected	Remove all		Add all
		Singapore Collection	+ ▲
		Litvak Special Collection 2	+ ▲
		Israel Photos	+ ▲
		Australia (9587890) (9587890)	+ ▲
		Test	+ ▲
		Boston Collection	+ ▲
		Philology.Theses	+ ▼

\* NFS Path

---

**Email Notification**

Send email?



Figure 7 Export job page

## Enhanced date field search

New search options added in the UI and SRU for DC date fields:

- Search only by month and year
- Search only by year

The new search options now align with the dmdSec in DC which accepts all dates conforming to the formats DD-MM-YYYY, MM-YYYY and YYYY.

# Integrations

## Deposit from an S3 source

A new submission format of type S3 supports direct deposits from an S3 source, either cloud or local object storage, to all content structures. S3 folder structure can then be navigated to select a specific folder for deposit.

The screenshot shows the 'Submission Format Details' page for an S3 submission format. The page is divided into several sections:

- Name:** A text input field.
- Description:** A large text area.
- Handling Method:** Radio buttons for 'Move' (selected) and 'Copy'.
- Endpoint:** A text input field.
- Region:** A text input field.
- \* Access Key ID:** A text input field.
- S3 Path (Prefix):** A text input field.
- Max. Number of Files:** A text input field with the value '1'.
- \* Bucket name:** A text input field.
- \* Maximum waiting time (Milliseconds):** A text input field with the value '1800000'.
- \* Secret Access Key:** A text input field.
- Max File size (MB):** A text input field with the value '1'.
- Min. Number of Files:** A text input field with the value '1'.
- File Extensions:** A table with columns for 'Remove all' and 'Add all'. The table lists file extensions: .9ml, .000, .123, .355, .3dm, .3dmf, and .3ds. The 'Remove all' column is currently empty, and the 'Add all' column has a plus sign for each extension.

Figure 8 S3 submission format page

Note: Remote files in deposit are not supported

## REST APIs - Data Management

A new set of RESTful APIs provides the following services:

- Get a list of processes  
`rest/v0/conf/processes`
- Get a process' details  
`rest/v0/conf/processes/{process_id}`
- Run a process  
`rest/v0/conf/processes/{process_id}?op=run`
- Get a list of process instances  
`rest/v0/conf/processes/{process_id}/instances`

# Infrastructure



## 3<sup>rd</sup> Party Upgrades

As always, Rosetta is aligned with the latest 3<sup>rd</sup> party components in order to provide the most secure environment as possible. 7.4 includes, amongst others:

- Tomcat 9.0.68



## Additional

- New DNX elements in Metadata form, expanding possible input fields from content providers: 'dnx:generalIECharacteristics.objectIdentifierType' and 'dnx:generalIECharacteristics.objectIdentifierValue'.