Rosetta Product Update and Roadmap
Rosetta Advisory Group 2017, University of Sheffield

Dvir Hoffman | Corporate VP, Resource Management Solutions
Adi Alter | Director of Product Management, Digital Solutions
2017-2018 Main Focus Areas
Main Focus Areas Since the Last RAG Meeting

1. Enhance Core Functionalities
2. Increase Synergy with other Ex Libris Products
3. Streamline the Implementation Process
Main Focus Areas for 2017-2018

1. Enhance Interoperability and OOB Integrations and Tools
2. Improve User Experience and System Performance
3. Enrich Core Functionalities Through Streamlined Collaboration with the RUG
Continue Investing in an Open Architecture

OAI-PMH/SRU/SRW
Publisher Plugins
Primo
Converter Registry Plugins

Metadata Extraction Plugins
Virus Check Plugins
Custom Fixity Plugins

ILS Synchronization

Voyager  Alma  Aleph
Other ILS

PUBLISH
DELIVER
MANAGE
DEPOSIT
PRESERVE

Collection Management Web Services
BIRT reports
Data Management Web Services
Startups Check Plugins
CRUD Access Rights Web Services
Access Rights Plugins
IE Update Web Services
Persistent Identifier Plugins

Producer Management Web Services
Field Validator Plugins
Deposit Web Services
Decomposer Plugins

Storage Abstraction Layer
Repository Task Plugins
Comparison Plugins
Risk Extractor Plugins
Migration Plugins

ExLibris Rosetta

© 2017 Ex Libris | Confidential & Proprietary
While Enhancing Rosetta’s Out-of-the-Box Offering
Improve User Experience and System Performance

- Continue the New UI Design Uplift
- Simplify Workflows
- Support User Preferences

PROVIDING A SEAMLESS USER EXPERIENCE

- New Dedicated Infrastructure Team
- Enhanced Performance Lab
- New System Monitoring Tools

ROBUST, EFFICIENT & SCALABLE DIGITAL PRESERVATION PLATFORM
Enrich Core Functionalities Through Streamlined Collaboration with the RUG

- Delivery & Integrations
- Digital Preservation
- Format Library
- System Operations
From New York to Sheffield: A Year in Review
Victoria University of Wellington Selects Rosetta
New Rosetta Dedicated Infrastructure Team
Rosetta SUSU Team – Speed Up, Scale Up

• New Performance Lab with Millions of Records
  • Based on the production DB and METS files of one of our largest customers

• New DB Monitoring Tool
  • Installed at production environments at some large customers

• Systematic benchmark and review of core components and main workflows
  • Database
  • Solr
  • Ingest
  • Publishing
  • Process automation
  • Reports
  • And more

• Security
  • Monthly penetration scans and security hardening
Rosetta Version 5.1 Released

Quick Stats

<table>
<thead>
<tr>
<th>Status</th>
<th>SIPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Draft</td>
<td>12</td>
</tr>
<tr>
<td>Rejected</td>
<td>25</td>
</tr>
<tr>
<td>Declined</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
</tr>
</tbody>
</table>

Quick Stats

SIPs In Process

<table>
<thead>
<tr>
<th>Status</th>
<th>SIPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>In process</td>
<td>1</td>
</tr>
</tbody>
</table>

Technical Analyst

Stage

<table>
<thead>
<tr>
<th>Status</th>
<th>SIPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation</td>
<td>62</td>
</tr>
<tr>
<td>Loading</td>
<td>21</td>
</tr>
<tr>
<td>Enrichment</td>
<td>3</td>
</tr>
<tr>
<td>Bytestream</td>
<td>16</td>
</tr>
<tr>
<td>Deposit</td>
<td>80</td>
</tr>
<tr>
<td>ToPermanent</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>196</td>
</tr>
</tbody>
</table>

Manual Assessment

Stage

<table>
<thead>
<tr>
<th>Status</th>
<th>SIPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arranger</td>
<td>23</td>
</tr>
<tr>
<td>Assessor</td>
<td>260</td>
</tr>
<tr>
<td>Approver</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>321</td>
</tr>
</tbody>
</table>

Top Formats at Risk

Files by Classifications
Rosetta Version 5.1

**Digital Asset Management**
- SIP Deposit Prioritization
- Improved Deposit Management
- Enhanced Search – ‘Not’ Operator Support and Smart Search
- Collection Ordering and History
- Improved Collection Management
- New Graphical XML Editor
- Derivative Copies Publishing and Creation
- New HTML5 Viewers
- Enhanced Viewers Customization

**Preservation**
- Create a Preservation Plan for more than One Source Format
- Share Technical Analyst Rules Between Institutions
- File Comparison Plug-in
- Consistency Checks
- Validation Stack Enhancements - SHA-256 Fixity, Virus Check Exit Codes and Automatic Rules

**Infrastructure**
- Update Software Stack - Tomcat, PDS, Aurigma Uploader
- New User Interface
- Performance Improvements
- Enhanced Indexing
- New APIs
Alma and Rosetta Work Together at the National Library of New Zealand
Workflow I: Two-Way Metadata Synchronization Using OAI-PMH

1. Deposit

2. Store

3. Publish (Rosetta) / Harvest (Alma) (OAI-PMH)

4. Create Digital Representation / Bib Records

5. Manage Metadata

6. Publish (Alma) / Harvest (Rosetta) (OAI-PMH)

7. Update Metadata

8. Publish (OAI-PMH)

9. View

Librarian
The Metadata in Alma and Rosetta

MARC Record in Alma

DC Record in Rosetta

LDR 02593 am 22001693u 4500
001 99130008400121
005 20160621144854.0
024 8  _a oai:d4i1-pubam:IE46890
024 8  _a alma_99117009500191
024 8  _a http://rosetta.exlibrisgroup.com:1801/delivery/DeliveryManagerServlet?dps_pid=IE46890
035  _a oai:rosetta.exlibrisgroup.com:IE46890
042  _a dc
245 00  _a The Potato Eaters
260  _c 1885

MID 42906  Metadata Type descriptive:dc  Description The Potato Eaters (Dutch: De Aardappelen) Is a painting

Created by admin1  Creation Date 11/11/2013 09:20:50
Updated by Victoria Holmes  Update Date 11/12/2016 22:54:50

<table>
<thead>
<tr>
<th>Tag</th>
<th>Qualifier</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td></td>
<td>The Potato Eaters</td>
</tr>
<tr>
<td>Creator</td>
<td></td>
<td>Vincent Willem van Gogh</td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td>text</td>
</tr>
</tbody>
</table>
Workflow II: Digitization Requests’ Processing

1. Create Digitization Request

2. Deposit

3. Store

4. Create Remote REP (API)

5. Close Digitization Request (API)

Librarian

or

Automatic Process

ExLibris Rosetta

ExLibris Alma
Workflow III: Digitization – Alma Submits to Rosetta (Roadmap)

1. Create Digitization Request
2. Deposit
3. Create Digital Representation / Bib Records
4. Deposit (API)
5. Store
6. Close Digitization Request (API)
Workflow IV: Rosetta Preserves Alma Digital (Roadmap)

1. Deposit
2. Create Digital Representation / Bib Records
3. Store
4. Manage Metadata
5. Publish (Alma) / Harvest (Rosetta) (OAI-PMH)
6. Store
7. Publish (OAI-PMH)
8. View
More than 40 hbz Member Institutions
Join the Rosetta Community
The Most Popular Idea on Idea Exchange

The Idea Exchange idea with the highest number of supporters (13) from 8 institutions casting 28 votes in total

Enhance UpdateMD web service to support DNX update

Currently, the updateMD web service is designed only for updating DC metadata. Updating DNX metadata is not supported. But tests have shown that DNX metadata can perfectly be updated using the updateMD web service – with the drawback, however, of losing some other DNX metadata such as Producer, Producer Agent, and Access Rights Policy. See attached file with a comparison of the IE METS before and after the updateMD process.

The getIE web service works fine in this regard, allowing you to choose between DC, DNX and source MD and giving you the complete metadata output.

We strongly argue that... more

Compare_METS_pre_and...
294 KB
1 comment • APIs & Plug-ins

COMPLETED • Adi Alter (Rosetta Product Manager, Ex Libris, Ex Libris) responded

I'm happy to let you know updateMD was enhanced to support editing of any MD that can be edited via the UI. This will be available in the upcoming version of Rosetta (version 5.3).
Ten Rosetta Implementations in Parallel
Streamlined the Implementation Process

- Dedicated cross-regions Rosetta implementation team
- Revised all implementation processes and deliverables
- New recorded training materials
- Developments supporting implementation needs
State Library of Victoria Selects Rosetta

https://www.flickr.com/photos/fairerdingo/8464811126
Rosetta Version 5.2 Released

March 2017
Rosetta Version 5.2

**Digital Asset Management**
- Alma Digital Inventory’ Repository Tasks
- Alma Digitization Submission Application
- Access Rights Check for Remote Repositories
- OAI-PMH Harvester Test UI
- Improved TA Workbench and TA Error Visibility
- Submission Job Reports
- Search and SRU Improvements
- Delete Derivative Copy Representation Task
- Full Text Viewer Pre-Processor (VPP)

**Preservation**
- Deposit Content using BagIt ‘Bags’
- Storage Migration
- Indexing Support for Long Fields
- Access Rights Policies Audit Trail
- Retaining Timestamps for Digital Content Uploaded to Rosetta

**Infrastructure**
- SAML / Local Authentication
- Improved User Experience
- Performance Improvements
- Enhanced Clean-Up Jobs
Rosetta Joins the Open Preservation Foundation (OPF) and Participates in the Jhove Hackathon
59 Support Cases Closed by the Support Team

Number of Open Support Cases

- Jan-17: 142
- Feb-17: 126
- Mar-17: 109
- Apr-17: 77
- May-17: 81
Celebrating 10 Years of Rosetta
2017-2018 Roadmap Overview
Rosetta Roadmap Schedule

Version 5.3
August 2017

Version 5.4
December 2017

Version 6.0
End of H1 2018

Version 6.1
End of H2 2018
SIP Metadata Search

• SIP metadata will be indexed early in the ingest process enabling searching through ingested content in different stages (TA workbench, 3As)

• This metadata will be searchable during the ingest process via
  • A new search screen
  • SRU

<table>
<thead>
<tr>
<th>SIP ID</th>
<th>Deposit ID</th>
<th>Producer Id</th>
<th>Title (DC)</th>
<th>Date Submitted</th>
<th>Date Updated</th>
<th>Location</th>
<th>Search Location</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 973</td>
<td>1002</td>
<td>10000</td>
<td>Books 1</td>
<td>31/05/2017</td>
<td>31/05/2017</td>
<td>Permanent</td>
<td>Search Permanent</td>
<td>Report</td>
</tr>
<tr>
<td>2 974</td>
<td>1003</td>
<td>10000</td>
<td>Books 2</td>
<td>31/05/2017</td>
<td>31/05/2017</td>
<td>Assessor</td>
<td>Show in Workbench</td>
<td>Details Report</td>
</tr>
<tr>
<td>3 975</td>
<td>1004</td>
<td>10000</td>
<td>Books 3</td>
<td>31/05/2017</td>
<td>31/05/2017</td>
<td>Permanent</td>
<td>Search Permanent</td>
<td>Report</td>
</tr>
<tr>
<td>4 983</td>
<td>1013</td>
<td>10000</td>
<td>Books 4</td>
<td>31/05/2017</td>
<td>31/05/2017</td>
<td>Permanent</td>
<td>Search Permanent</td>
<td>Report</td>
</tr>
</tbody>
</table>
1. SIP Metadata Search

Poll locked. Responses not accepted.

Very helpful, was always struggling to find my ingested content: 38%

Will probably be helpful in some cases: 50%

No need, I am able to manage my ingested content using the existing tools: 12%

Total Results: 26
TA Workbench Improvements

- Enable selecting multiple files within a SIP and applying a single action on them
  - e.g. assign file extension, assign format, ignore error, etc.

- Enables ad-hoc rules for the current SIP

- Manual format and extension assignment

```
Ingest

The format ID of the file: adm_ditarc-0064-unesco_actions_haiti-eng-short.mp4 will be updated. Please choose the new format ID and add user note.

<table>
<thead>
<tr>
<th>Format ID</th>
<th>Format Mime Type</th>
<th>Format Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>fmt/199</td>
<td>video/mp4</td>
<td>ExL-Fmt-61</td>
</tr>
<tr>
<td>fmt/596</td>
<td></td>
<td>fmt/596</td>
</tr>
<tr>
<td>unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reason: Problem not relevant for preservation
Optional Note
```
2. TA Workbench Improvements

Poll locked. Responses not accepted.

This is extremely important. I find the current workbench hard to work with. 81%

Nice improvements that will probably help me sometimes with my regular work. 8%

I’m hardly using the TA Workbench but rather trying to handle technical issues pre-ingest. 12%

Total Results: 26
Duplicates Report

• Identify duplicate objects by:
  • Checksum value
  • Metadata identifier

• Both on the consortium and institution levels
3. Duplicates Report

- **Poll locked.** Responses not accepted.

- **This is very useful**
  - 61%

- **Might come out handy in some cases**
  - 39%

- **Duplicates should be managed before they are deposited into Rosetta**

Total Results: 23
Event History Report

• Enables users to generate a report by any event:
  • Takes an event ID and date range
  • Returns a list of PID, event date, and event key values

• Examples:
  • Indexing events (to track failed indexing)
  • CRUD reporting for set and process management
figshare Integration

- Enhance the OAI-PMH harvester to match records of external origin based on a dc:identifier or an IE ObjectIdentifier

- Repository Task plug-in to notify figshare upon AIP creation
4. figshare Integration

- **Poll locked.** Responses not accepted.

  - We have figshare and will probably be using this: 26%
  - Would like to see such an integration also with other repositories: 70%
  - We are managing everything in Rosetta, so such integrations are not needed: 4%

Total Results: 27
5. Which systems have you integrated Rosetta with (multiple answers are allowed)?

Poll locked. Responses not accepted.

Total Results: 39
6. Which other systems would you like to see Rosetta integrated with (multiple answers are allowed)?

Poll locked. Responses not accepted.

Total Results: 27
Out-of-the-Box Preservation Tools

• Enhance Rosetta’s out-of-the-box preservation tools with tools such as:

  • Metadata Extractors:
    • ExifTool – for image, audio, video, and PDF content
    • MediaInfo – for media content
    • Apache Tika – for over a thousand different file types

• Migration Tools
  • FFMPEG - for multimedia content
7. Which preservation tools would you like to see integrated with Rosetta (multiple answers are allowed)?

Poll locked. Responses not accepted.

Total Results: 42
Multiple Preservation Tools Support

• Support Multiple MD Extractors
  • Enable receiving properties from multiple extractors for the same format

• Support Multiple Format Identification Tools
  • Enable crosschecking the identified format
8. Multiple Preservation Tools Support

🔒 Poll locked. Responses not accepted.

- This is crucial for a preservation system: 88%
- This is nice to have in a preservation system: 8%
- Currently, in most cases one tool satisfies most of my needs: 4%
- Having multiple tools doing the same thing is a waste of resources! Tools of the world, unite!: 4%

Total Results: 25
Support Custom Events

• Enable providing custom (non-Rosetta events) as part of the SIP

• Focus on PREMIS events
9. Support Custom Events

- **Poll locked.** Responses not accepted.

- I have multiple events waiting in line for this feature (33%)
- I can think of cases in which I’ll be using this feature (67%)
- Rosetta is my only event generator; Thou shalt have no other event generator before me.

Total Results: 3
Responsive, mobile-friendly, configurable, similar look and feel as the recently added REP viewer
11. What OOB viewer / format support would you like to see in Rosetta?

Poll is full and no longer accepting responses
IIIF Support – What is IIIF?

The International Image Interoperability Framework (IIIF) provides a standardized method of describing and delivering images over the web by defining the following APIs:

- **Image API** – provides the image with characteristics such as region, size, rotation, quality and format
- **Presentation API** - describes the structure and layout of a digitized object
- **Search API** - allows for "searching annotation content within a single IIIF resource"

These APIs can be used to develop image servers and web clients that provide a world-class user experience in viewing, comparing, manipulating and annotating images.
IIIF Support - What Will Rosetta Do For You?

• **IIIF Manifest**
  • Describes the file structure and contains references to Image API endpoints.

• **Out-of-the-box IIIF Server - Cantaloupe**
  • An open-source image server with operations such as cropping, scaling, and rotation, it can support deep-zooming image viewers, as well as on-the-fly thumbnail generation.

• **Out-of-the-box IIIF Viewer – Universal Viewer**
  • An open source viewer with wide range of formats supported providing advanced features such as zooming, navigation and more.

• **Out-of-the-box File Viewer – OpenSeadragon**
  • An open-source, web-based viewer for high-resolution zoomable images.

It will be possible for institutions to use all components provided by Ex Libris or to take some of them and use their own server or client.

Support for **IIIF Open Annotation Extensions** - 2018
10. IIIF Support

Poll locked. Responses not accepted.

- This is exciting! Will use it asap! 13%
- Will be probably using it in the future 33%
- We have already developed our own IIIF implementation 4%
- No need, we are using Rosetta as a dark archive 42%
- Sorry, can you please explain once more what IIIF is? 8%

Total Results: 24
11. What new out-of-the-box viewers / format support would you like to see in Rosetta (multiple answers are allowed)?

Poll locked. Responses not accepted.
Extend SRU Capabilities

• Add Authentication to SRU/SRW Requests

• Add additional query operators such as ‘<’, ‘>’, ‘!=’, etc.

• Add query operators between fields – such as ‘or’ and ‘not’

• Add ability to search by partial date (e.g. year only)

• Enable configuration of the SRU response with an xsl file
12. Extend SRU Capabilities

Poll locked. Responses not accepted.

- We’re using SRU extensively, so any improvement here will be great (43%)
- We’re using SRU sometimes, so some of these improvements may be helpful one day (14%)
- We’re not using SRU at all (10%)
- I really want to participate in this poll, but unfortunately I have no idea what SRU is (33%)

Total Results: 21
Improved UI and UX

• Continue the new UI improvements
• Simplify workflows
• Drag-and-drop rows in tables
• New BIRT UI
• And more...
13. Improved User Interface

Poll locked. Responses not accepted.

- I love the new UI. I have a Rosetta screenshot hanging in my living room! 8%
- The new UI looks good and additional improvements are important 33%
- Nice... 50%
- I want the old UI back!!!
- It’s a back office system. Who cares about the UI? 8%

Total Results: 24
• Provide customized functionalities per user
  • List page size
  • Facets display
  • Search result columns
  • Quick links
  • Quick search
  • Recent search
  • Favorite menu items
  • Dashboard
  • ...and more
# Version 5.3 Highlights – August 2017

## Digital Asset Management
- IIIF Support
- SIP Metadata Search
- Collection Management Improvements
- Duplicates Report
- AIP Update Enhancements
- OAI-PMH Harvesting

## Preservation
- Store TA Rule ID as Part of the TA Rule Event
- Store Identification Method in File Format Section in DNX
- Keep Original File Format Identification Event in DNX

## Infrastructure
- Copy Configuration
- Improved User Experience
- Drag-and-Drop Rows in Tables
- Performance Improvements

## Performance Improvements
Version 5.4 Highlights – December 2017 (Planned)

Digital Asset Management
- Extend SRU Capabilities
- Index Derivative Copies Metadata
- Enable Sending Emails for Jobs Statuses
- Enhance the UI with Process Automation Logs

Preservation
- TA Workbench – Bulk Operations
- TA Workbench – Manual Format and Extension Assignment
- Out-of-the-Box Tech MD Extractors
- Out-of-the-Box Migration Tools

Infrastructure
- Update Software Stack - Solr 6.0 – Solr as a server, BIRT 4.7, Tomcat
- New BIRT Reports UI
- Performance Improvements
Versions 6.X Highlights – 2018 (Initial List)

**Digital Asset Management**
- Export to FTP
- Delete OAI sets
- IIIF Annotations
- Improved Collection Management
- New IE / METS Viewer

**Preservation**
- Support Multiple MD Extractors
- Support Multiple Format Identification Tools
- Support Custom Events
- Add Provenance Information about Original Checksums
- Improved Custom Fixity Support
- BagIt Support – Export
- Out-of-the-Box Tech MD Extractors
- Out-of-the-Box Migration Tools
- Improved Preservation Flow

**Infrastructure**
- Support for Oracle 12.2
- Support for User Preferences
- Favorite Menu Items
- Performance Improvements
15. Wishlist (multiple answers are allowed)

Poll locked. Responses not accepted.

- “better support” 5 days ago
- “much, much better support” 5 days ago
- “better support” 5 days ago
- “User guides to help plugin dev” 5 days ago
- “fresh install options” 5 days ago
- “More exposure to logs from UI” 5 days ago
- “Moving configuration server to server” 5 days ago
- “better support” 5 days ago
- “Collection management” 5 days ago
- “better support” 5 days ago
- “Moving configuration server-to-server” 5 days ago
- “better support” 5 days ago
- “https without load balancer” 5 days ago
- “Multicopy aware” 5 days ago
- “better support” 5 days ago
- “Easier plugin development” 5 days ago
- “Multicopy awareness” 5 days ago
- “API’s” 5 days ago
- “Stringing multiple API actions (for example updateRepresentation) together without multiple versions” 5 days ago
- “REST API” 5 days ago
- “browser back button support” 5 days ago
- “Web deposit using JSON” 5 days ago
- “better support” 5 days ago

"Ability to have Open MoRe than One tab" 5 days ago

"Easier plugin development" 5 days ago

"Stringing multiple API actions (for example updateRepresentation) together without multiple versions" 5 days ago

"REST API" 5 days ago

"browser back button support" 5 days ago
Thank You!

Adi.Alter@exlibrisgroup.com
Dvir.Hoffman@exlibrisgroup.com