Rosetta’s Ingest Methods – What’s the Right Path to Take?

Introductions

• Zvi Fass - Rosetta Development Team Leader
  • In Ex Libris since 2006
  • Part of the Rosetta Development team since day one
  • Development Team Leader for the past two years
Objectives and Target Audience

- Brief Session Description:
  - We will review different Rosetta ingest methods – such as Web deposit, CSV, METS, OAI-PMH, BagIt and Deposit APIs - along with the advantages & limitations of each method.

- Session Objective(s)
  - **What’s the right path to take?** By the end of the session you will be able to decide what’s the best path to take with your specific data and needs.

- Session Target Audience:
  - All Rosetta users.

Agenda

1. Introduction
2. Manual Ingest
3. Automated Ingest
4. Bulk Ingest
5. Next Steps, Support Resources and Survey
Introduction

Depositing into Rosetta
What’s the right path to take?

Files
- DC
- Bagit
- OAI-PMH
- Web Services

Metadata
- Mets
- CSV
- Submission Application

Rosetta
Introduction

Basic terminology

Material Flow Components

- Material Flow
- Metadata form
- Submission format
- Content structure
- Access rights
- Retention Periods

Agenda

1. Introduction
2. Manual Ingest
3. Automated Ingest
4. Bulk Ingest
5. Next Steps, Support Resources and Survey
Manual Ingest

Manual Material Flow

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• User friendly – wizard</td>
<td>• Non bulk</td>
</tr>
<tr>
<td>• Customizable Metadata Form</td>
<td>• Only supports one Rep per IE</td>
</tr>
<tr>
<td>• Customizable Submission Format</td>
<td>• Content Structure - IE per file or single IE</td>
</tr>
<tr>
<td>• Bulk uploader</td>
<td>• Only PC upload</td>
</tr>
<tr>
<td>• Http</td>
<td>• No DNX Metadata</td>
</tr>
<tr>
<td>• Detailed</td>
<td></td>
</tr>
</tbody>
</table>
Agenda

1. Introduction
2. Manual Ingest
3. Automated Ingest
4. Bulk Ingest
5. Next Steps, Support Resources and Survey

Automated Ingest
Automated Ingest

Automated Material Flow

• Non web ingest (can also be done via web)

• Supports multiple Representations in multiple IEs.

• Customizable Submission Format
  • NFS
  • FTP
  • SFTP

• Customizable Content Structures
  • Dublin core Converter
  • Mets Converter
  • CSV Loader Converter
  • Bagit Converter

Advantages

Limitations

• Automated – Bulk Ingest
• Simple - Need only DC or no metadata
• Support of multiple IEs
• Relative/Absolute/Http Paths
• Supports only DC metadata
• No DNX metadata
• No Source metadata
• No support of multiple representations
## Automated Ingest

### Mets Converter

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Automated – Bulk Ingest</td>
<td>• Complex</td>
</tr>
<tr>
<td>• Multiple IEs support</td>
<td>• Need a Submission Application to prepare the Mets for Ingest</td>
</tr>
<tr>
<td>• Multiple Representations support</td>
<td></td>
</tr>
<tr>
<td>• Relative/Absolute/Http Paths</td>
<td></td>
</tr>
<tr>
<td>• Metadata update – full flexibility</td>
<td></td>
</tr>
<tr>
<td>• Collection support (assignment only)</td>
<td></td>
</tr>
</tbody>
</table>

## Automated Ingest

### CSV loader Converter

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Automated – Bulk Ingest</td>
<td>• Need a tool to create the CSV Files</td>
</tr>
<tr>
<td>• Multiple IEs support</td>
<td>• No support for Logical Struct Map</td>
</tr>
<tr>
<td>• Multiple Representations support</td>
<td></td>
</tr>
<tr>
<td>• Relative/Absolute/Http Paths</td>
<td></td>
</tr>
<tr>
<td>• Metadata update – full flexibility</td>
<td></td>
</tr>
<tr>
<td>• Full Collection support</td>
<td></td>
</tr>
<tr>
<td>• Simple to create</td>
<td></td>
</tr>
</tbody>
</table>
### Automated Ingest

#### Bagit Converter

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Automated – Bulk Ingest</td>
<td>• No metadata support for files</td>
</tr>
<tr>
<td>• BagIt format support</td>
<td>• No Logical Struct Map support</td>
</tr>
<tr>
<td>• IE DC and DNX metadata</td>
<td></td>
</tr>
<tr>
<td>• Rep DC and DNX metadata</td>
<td></td>
</tr>
</tbody>
</table>

---

### Agenda

1. **Introduction**

2. **Manual Ingest**

3. **Automated Ingest**

4. **Bulk Ingest**

5. **Next Steps, Support Resources and Survey**
Bulk Ingest

Bulk Ingest – Submission Job

- Pre-prepared load packages (according to the Material Flow)
- Automatic scheduling
- Locking mechanism
The end-to-end process of harvesting records is comprised of four stages:

1. Harvesting
2. Matching
3. Transformation
4. Based on the matching results generates a:
   - Submission Job folder for new records
   - Metadata Update folder for existing records.

Scheduled Submission and/or Metadata Update Jobs run independently of the Harvesting Job.
Bulk Ingest – OAI-PMH Harvester

Test Area

Bulk Ingest - Deposit Web Services

Deposit Web Services (wsdl)

Allow institutions to develop applications over Rosetta

- Customers would like to use their own deposit frontend, which includes creation of IEs, depositing and retrieving deposit Information (for example Indigo)
- Bulk Deposits – creation of automated tools to deposit massive numbers of IEs
- Migrate existing metadata into Rosetta
Bulk Ingest - Deposit Web Services

Submission Application

- SDK – development tools
- IE Creation – Mets deposit (or any other structure)
- Deposit – using the Deposit web services
- Monitoring - using the SIP web services

Six Step Submission Application

1. Create a SIP directory
2. Create a Rosetta Compatible Mets
3. Synchronize Deposit Directory with Submission Format
4. Authentication with PDS/Header authentication
5. Submit Deposit
6. Retrieve the SIP status

The source code of the example can be found under:
dps-sdk<version>/src/com/exlibris/dps\sdk\examples\FullFlowExample.java
Recap

<table>
<thead>
<tr>
<th></th>
<th>Web</th>
<th>Bulk</th>
<th>DC Update</th>
<th>DNX IE/REP</th>
<th>DNX File</th>
<th>Source MD</th>
<th>Collections</th>
<th>Struct Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Dublin Core Converter</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Mets Converter</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓ (only assign)</td>
<td>✓</td>
</tr>
<tr>
<td>CSV loader Converter</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Bagit Converter</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>

Recap

Depositing into Rosetta
What’s the right path to take?

What's Your Business?

Migration
Ingest Flows
End user
Next Steps and Support Resources

• Include documentation links relative to the topic:
  • SDK
  • Deposit Web Services
  • OAI-PMH Harvesting

• Additional support resources within the ExLibris Ecosystem:
  • Idea Exchange
  • Developer Network

• Technical Seminar Presentations
  (located in the Cross-Product section of the CKC)
Any Questions?

Session Survey Evaluation

Please use the following link https://www.surveymonkey.com/r/techsem2017 to provide feedback on your sessions.
THANK YOU