Ex Libris Rosetta - Multidimensional Scalability

Adi Alter | Digital Resources Product Manager

13th International PASIG Meeting

September 18th 2014
CONFIDENTIAL INFORMATION

The information herein is the property of Ex Libris Ltd. or its affiliates and any misuse or abuse will result in economic loss. DO NOT COPY UNLESS YOU HAVE BEEN GIVEN SPECIFIC WRITTEN AUTHORIZATION FROM EX LIBRIS LTD.

This document is provided for limited and restricted purposes in accordance with a binding contract with Ex Libris Ltd. or an affiliate. The information herein includes trade secrets and is confidential.

DISCLAIMER

The information in this document will be subject to periodic change and updating. Please confirm that you have the most current documentation. There are no warranties of any kind, express or implied, provided in this documentation, other than those expressly agreed upon in the applicable Ex Libris contract. This information is provided AS IS. Unless otherwise agreed, Ex Libris shall not be liable for any damages for use of this document, including, without limitation, consequential, punitive, indirect or direct damages.

Any references in this document to third-party material (including third-party Web sites) are provided for convenience only and do not in any manner serve as an endorsement of that third-party material or those Web sites. The third-party materials are not part of the materials for this Ex Libris product and Ex Libris has no liability for such materials.

TRADEMARKS

"Ex Libris," the Ex Libris Bridge to Knowledge, Primo, Aleph, Voyager, SFX, MetaLib, Verde, DigiTool, Rosetta, bX, URM, Alma, and other marks are trademarks or registered trademarks of Ex Libris Ltd. or its affiliates.

The absence of a name or logo in this list does not constitute a waiver of any and all intellectual property rights that Ex Libris Ltd. or its affiliates have established in any of its products, features, or service names or logos.

Trademarks of various third-party products, which may include the following, are referenced in this documentation. Ex Libris does not claim any rights in these trademarks. Use of these marks does not imply endorsement by Ex Libris of these third-party products, or endorsement by these third parties of Ex Libris products.

Oracle is a registered trademark of Oracle Corporation.

UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company Ltd.

Microsoft, the Microsoft logo, MS, MS-DOS, Microsoft PowerPoint, Visual Basic, Visual C++, Win32, Microsoft Windows, the Windows logo, Microsoft Notepad, Microsoft Windows Explorer, Microsoft Internet Explorer, and Windows NT are registered trademarks and ActiveX is a trademark of the Microsoft Corporation in the United States and/or other countries.

Unicode and the Unicode logo are registered trademarks of Unicode, Inc.

Google is a registered trademark of Google, Inc.

Web address:  http://www.exlibrisgroup.com
Rosetta is a complete digital asset management and preservation solution that addresses the ever-growing need to collect, archive and preserve the digitally-born and digitized materials stored at academic institutions, research organizations, and government institutions, ensuring data integrity and access over time.
GROWING NUMBER OF ROSETTA CUSTOMERS
ARCHITECTURAL SCALABILITY
The Challenges

• Support entire range of institutions – from small universities to national libraries
• Ingest and process millions of files and TBs per day
• Manage PBs of data
• Flexible configuration to allow for growth
• Ability to dedicate servers to particular roles
• Minimize manual intervention
• Serve multiple end users accessing stored content simultaneously
ARCHITECTURAL SCALABILITY
The Solution: Scalable Distributed Environment

End Users

All-in-One: PDS
Deposit
Staging
Permanent
Delivery
Database
Storage

Depositors

Staff Workstation
ARCHITECTURAL SCALABILITY
The Solution: Horizontal Growth

End Users

Depositors

Load Balancer

Staff Workstation

Storage

All-in-One  All-in-One  All-in-One  All-in-One

Database

Rosetta

ExLibris
ARCHITECTURAL SCALABILITY

The Solution: Vertical Growth

Staff Workstation

Deposit+PDS

Staging

Permanent

Storage

Database

End Users

Depositors

Delivery

Rosetta
OPERATIONAL SCALABILITY
The Challenges

- Multiple formats
- Multiple metadata standards
- Multiple metadata extractors
- Multiple format identifiers
- Multiple virus scanners
- Multiple publisher types
- Multiple viewers
- Multiple storage providers
- Multiple access rights policies
- Multiple report types
- Multiple collection types
- Multiple data types
- Multiple institution types
- Multiple ILSs
- Multiple discovery systems
- Multiple migration rules
- Multiple browsers
- Multiple users
- Multiple user types
- Multiple processes
- Multiple representations
- Multiple material flows
- Multiple delivery rules
- Multiple migration plans
- Multiple risks
OPERATIONAL SCALABILITY
The Solution: An Open, Integrative and Extendable System

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

DEPOSIT
- Migration Plugins
- Repository Task Plugins
- Risk Extractor Plugins

Primo

Gabriel

Ex Libris
Rosetta

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

PUBLISH
- Delivery Web Services
- Viewer Pre-Processor Plugins
- Viewer Plugins

Deliver

INGEST
- ILS Synchronization
- Other ILS
- Virus Check Plugins
- Custom Fixity Plugins
- Format Identification Plugins
- Metadata Extraction Plugins

Preserve

Ingest

Alma
Aleph
Voyager

NFS server

Storage Abstraction Layer

Startups Check Plugins

Collection Management Web Services
Data Management Web Services
IE Update Web Services
Access Rights Plugins
Rosetta connects your past, present and future. Learn how to connect to Rosetta using our APIs.

Learn
In depth documentation of our APIs:
- Collection Management
- Data Management
- Delivery
- Deposit
- IE Update
- Producer Management

Integrations
Rosetta integrates with Aleph, Primo, and Voyager via SRU, OAI-PMH, and specifically developed APIs. Learn more...
- Viewers
- Delivery (Deep Linking)
- Search (SRU)

SDK Projects

Share
Share your experience and read other success stories at the Rosetta Developers Blog

Latest entries:
07 Jul 2014
Using the IA Book Reader to Deliver PDF Content
30 Jun 2014
Creating Metadata Tool Exit Point
INFORMATIONAL SCALABILITY
The Challenges

- Rendering applications
- Format identification tools
- Metadata extraction tools
- Format migration options
- Format types
- Preservation strategies
- Preservation policies
- Significant properties
- Risk identifiers
- Research Data
### INFORMATIONAL SCALABILITY

The Solution: Close Collaboration with the Academic Community

<table>
<thead>
<tr>
<th>Based on International Standards</th>
<th>Rosetta Advisory Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research and Collaboration</td>
<td>Community-driven Knowledge Base</td>
</tr>
</tbody>
</table>

**Rosetta**
The Solution: Community-driven Knowledge Base

- Rosetta’s **Format Library Workgroup** is a central hub for all institutions using Rosetta to share between them knowledge that can help preservation activities.
- Library of formats based on the PRONOM global library with numerous data on formats, risks, applications and more added by the Rosetta community.
ORGANIZATIONAL SCALABILITY
The Challenges

For smaller institutions:
Funding regulations and need for increased access to data mandate a data management and preservation solution

But...
TCO entry barrier is perceived to be high
ORGANIZATIONAL SCALABILITY
The Solution: Shared Services due to Economies of Scale

Institutions

Departments

Consortium

TIB, Hannover

ZB Med, (Köln/Bonn)

ZBW, Leibniz

Rosetta

ExLibris
CASE STUDY – THE NEED
National Library of Israel

First 11 months:
• Total # of IEs: 2,200,000
• Total # of Files: 5,900,000
• Max IEs/ Day: 60,000
• Max Ingest size/Day: 3 TBs
• Total size of files: 103 TB

Forecast for the next 12-18 months:
• Ingest ~2,000,000 IEs
• Ingest ~10,000,000 Files
• Ingest ~250TBs of Data

Step I: Prepare
• Digital Projects
• Digitization Processes

Step II: Deposit
• Manual Deposit - Web UI
• Partial-Automatic Deposit - CSV/ Excel
• Automatic Deposit - Digitization Centre Automatic Daemons

Step III: Ingest
SIP Generation

Rosetta
ExLibris
CASE STUDY – ARCHITECTURAL SCALABILITY
National Library of Israel

Satellite Servers
- IIP Image Server
- File Processing Server

Remote Site
- Access Files - NetApp
- Original Preservation Files - IBM

Local Site
- Oracle Database - NetApp

Rosetta
- Delivery
- Load Balancer
- Delivery

DEP, REP, PER, IDX
DEP, REP, PER, IDX
DEP, REP, PER, IDX, BO
CASE STUDY – OPERATIONAL SCALABILITY
National Library of Israel

- Virus Check Plugins
- Custom Fixity Plugins
- Format Identification Plugins
- Metadata Extraction Plugins
- ILS Synchronization

Primo

INGEST

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

ExLibris Rosetta

PUBLISH

DEPOSIT

- Migration Plugins
- Repository Task Plugins
- Risk Extractor Plugins

Preserve

- Storage Abstraction Layer

Manage

- Viewer Pre-Processor Plugins
- Viewer Plugins
- Delivery Web Services

- BIRT Reports
- Collection Management Web Services
- Data Management Web Services
- IE Update Web Services
- Startups Check Plugins
- Access Rights Plugins
INFORMATIONAL SCALABILITY
Case Study - National Library of Israel

Need: Delivery method for high-detailed digital objects

Discovery: Search for tools, consult Rosetta’s community

Decision: IIPImage - Image Server for Pyramidal TIFF

Review: Data shared with Rosetta’s Format Library Workgroup

FLWG Discussion: Format, PRONOM, JHOVE, DROID support, etc.

Agreement: Add PTIFF to Rosetta’s Format Library

Deploy: PTIFF and IIPImage used in Rosetta

Release: Change was made in the next Format Library release
THANK YOU!

Adi.Alter@exlibrisgroup.com

http://www.exlibrisgroup.com/category/Rosetta_resource_pack