The Digital Preservation System and the Open-Platform.

Carl Grant, President, Ex Libris North America
Agenda

• Preservation Overview
• System Architecture
• Open Platform
• Charter Program
• Going Forward
• Wrap-Up
Preservation Overview

Our Digital Preservation System

- Handles digital collections
- OAIS Compliant
- Total solution
- Enterprise-level solution
- Open Platform
- Vision/Roadmap/Support
Preservation Overview

System built by working with partners and industry leaders:

• Development partner:
  The National Library of New Zealand

• A Peer Review Group
• Research and collaboration
• Sun Microsystems
Preservation Overview

Peer Review Group

- Industry leaders in digital preservation
- Different size/types of organizations
- The goal: objective review of Preservation
Preservation Overview

Preservation vs. Repositories:

- Objects stored in a permanent repository
- Provenance
- Risk analysis and preservation actions
- Characterization
- Workflows
- Supports very large collections
- Preservation does not offer discovery
Agenda

- Preservation Overview
- System Architecture
- Open Platform
- Charter Program
- Going Forward
- Wrap-Up
Characteristics of Permanent Repository

- Preservation
- Management
- Search tools

Manual process → Automatic process

Deposit → Working Area → Permanent Repository → Publishing → Delivery

Ingest

SIP → AIP → DIP

Manual process

© Ex Libris Group--Confidential
System Architecture

- Scalable
- Flexible and extensible
- Secure
- Based on standards
System Architecture

Scalable

• Millions of objects, petabytes of storage
• Ingest/access to large # of objects/day
• Scalable - vertically and horizontally
• Distributed deployment
• No single point of failure
System Architecture

Flexible and extensible

• Embedding of third-party tools:
  • Validations tools
  • Enrichment tools
  • Delivery conversion tools
• Embedding of pre-transforms
• SDK to allow external Submission Applications
System Architecture

Secure

• Manages audit trails
• Routine checks to ensure system integrity
System Architecture

Based on standards

• Preservation standards/reference models:
  • OAIS
  • TDR certification checklist

• Library standards:
  • Metadata standards
  • Exchange standards

• Format identification and registry
• SOAP
Agenda

• Preservation Overview
• System Architecture
• Open Platform
• Charter Program
• Going Forward
• Wrap-Up
Open Platform

- Designed to be open
- Public API
- Easy to use, Easy Integrations
- Based on standards
- Documentation & Examples
- Backward compatibility
Spectrum of Interfaces/Integration Points

Matching functionality to interface type
Integration Points

Customer systems uses

- Access API
- Tech MD Extraction
- CMS Integrations
- V.S. Plug-In
- Enrichment Plugin
- Access API
- Viewers
- SRU/SRW
- Publishing OAI
- Migration Tools

Digital Preservation

- PDS
- Deposit
- Staging
- Operational
- Permanent

© Ex Libris Group--Confidential
Authentication/Authorization API

• Using PDS – SSO System

• From Java - Simple to use

PdsClient pds = PdsClient.getInstance();
pds.init("http://pds-server/pds");
String pdsHandle = pds.login("INS01", "ADMIN01", "ADMIN01");
PdsUserInfo user = pds.getPdsUserByPdsHandle(pdsHandle);

• **API Documentation**

• Can be used from other development platform – URL based API
IEParser / DNXDocument

- Utilities to create/parsing xml representation of an intellectual entity.
- Ensuring valid DPS Mets profile
  - IEParser API Documentation
  - DNX Utilities API
Deposit API

- Simple API
  - Submit Deposit Activity
  - Get Deposit Status
- Used by Preservation Deposit UI
- Staging workflow can be customized for deposits that comes via Deposit API
- NLNZ customized Web Curator (Tool that harvest web sites), to automatically deposit arc files to DPS
- Indigo - NLNZ develop a Desktop based deposit application for their staff user.
Indigo – Deposit Application, Uses Authentication API, IEParser, Deposit API
Access API

• Abilities
  • Access to Metadata records
  • Access to file streams

• Used By
  • Harvesting Process
  • Customized Viewers

• API Examples
  • IEParse getIE(String iePid, String profile)
  • IEParse getIEByDVS(String dvs)
  • InputStream getFileStream(String filePid)
  • String exportFileStream(String filePid, String namedFolder)
  • DnxDocument getFileInfo(String filePid)
## Viewers Integration

**Ex Libris Preservation**

<table>
<thead>
<tr>
<th>Title</th>
<th>Picasso Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors/Creators</td>
<td>Diana</td>
</tr>
<tr>
<td>Department</td>
<td>Arts</td>
</tr>
<tr>
<td>Thesis Type</td>
<td>Masters</td>
</tr>
<tr>
<td>Institution</td>
<td>Ex Libris University</td>
</tr>
<tr>
<td>Date of issue</td>
<td>07/10/2007</td>
</tr>
</tbody>
</table>

**Logical Structure (LOGICAL)**

1. **IE (1001, View)**
   - Representation (1002, View)
   - Representation (1003, View)
   - Representation (1004, View)
Process Automation Framework

- Process Automation
  - Framework for IE related tasks processing.
  - Bulk processing based
  - Scaleable – more threads, more machines as needed.
  - Manageable & Monitorable
  - Insuring Data integrity

- Set – Logical or Itemized

- Task Chain – A List of Tasks

- Task – Pluggable Unit
Task Plug-Ins – Different Types

• Validation related task
  • Checksum
  • Format validation
  • Virus
  • Other...

• Enrichment task
  • Thumbnail
  • Full text Extraction
  • Representation Creation
  • Other...

• Migration task
  • Format conversions
Agenda

• Preservation Overview
• System Architecture
• Open Platform
• Charter Program
• Going Forward
• Wrap-Up
Going Forward

Charter Customer Program

- Currently seeking interested organizations.
- Purpose.
- Areas of collaboration interest:
  - Workflows
  - Scaling
  - Reference Architecture
  - Large Data Sets
  - Rich Media
  - Email/Wiki/Blog Archives
  - Migration
  - Preservation Consortiums
  - Others?
Going Forward

Charter Customer Program

• Benefits to Members include:
  – Participation in industry leadership group.
  – Early access.
  – Work with developers.
  – Help define best practices.
  – Preferred business terms.
Agenda

• Preservation Overview
• System Architecture
• Open Platform
• Charter Program
• Going Forward
• Wrap-Up
Development time line

• Phase I
  • April 2007: Project Started
  • October 2008: system goes live at the National Library of New Zealand
  • November 2008: begin Charter Customer Program
  • December 2008: general availability

Going Forward
Going Forward

Development time line

• Phase II
  • May 2009: Beta delivery
  • October 2009: new functionality goes live at the National Library of New Zealand
  • December 2009: general availability
Agenda

• Preservation Overview
• System Architecture
• Functional Characteristics
• Open Platform
• Going Forward
• Wrap-Up
How can DPS help you?

• **Manage** and **preserve** digital collections
• Ensures data is **safely stored** and is **available**
• Provides **workflows** of all types.
• **Open platform** - add other OSS software components
• **Standards based** - OAIS, TDR
• **Total Solution** - From SIP to DIP
• **Commercial solution** – support, vision, roadmap, regular enhancements
• **Enterprise-level solution** – scalable, distributed architecture, flexible, secure
THANK YOU!

To become a charter member or buy DPS, contact: carl.grant@exlibrisgroup.com
mike.thuman@exlibrisgroup.com