Towards Interoperable Preservation Repositories

iPRES 6
2009-10-05

Priscilla Caplan (FCLA), Bill Kehoe (CUL), Joseph Pawletko (NYU)
Presentation Overview

- Motivation & Approach
- TIPR Partners: Heterogeneous Repositories, AIPs
- Repository Exchange Package (RXP)
- Transfer Tests
- Progress & Issues to Date
- Next Steps
- Questions
Motivation & Approach

- **Motivation:**
  - Repositories need to exchange AIPs for:
    - redundancy
    - software migration
    - succession planning

- **TIPR Approach:**
  - define a common exchange package format: the Repository Exchange Package (RXP)
Florida Center for Library Automation (FCLA)
- runs the DAITSS Preservation Repository
- **Team:** Priscilla Caplan (PI), Franco Lazzarino, Marly Wilson, Randy Fischer
TIPR Repositories

- Cornell University Libraries (CUL)
  - runs an aDORe-based repository
  - migrating to Fedora

- Team: Oya Rieger, Bill Kehoe, Rick Silterra, Adam Smith
New York University Libraries (NYU) runs a DSpace-based repository.

Team: Dr. Michael Stoller, Joseph Pawletko, Rasan Rasch
• One AIP per Intellectual Entity

• Retain First and Latest Representations

• Discard Intermediate Representations

TIPR
AIP Structures
FCLA

Intellectual Entity

Representation $R_0$

Representation $R_1$

Representation $R_n$
One AIP per Representation

Retain All Representation AIPs (including “intermediates”)
One AIP per Intellectual Entity

Retain all Representations (including “intermediates”)

Intellectual Entity

- Representation $R_0$
- Representation $R_1$
- Representation $R_2$
- Representation $R_n$

NYU Archive
Repository / AIP Heterogeneity:

<table>
<thead>
<tr>
<th>Repository Implementation</th>
<th>FCLA</th>
<th>CUL</th>
<th>NYU</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAITSS</td>
<td>1</td>
<td>n</td>
<td>1</td>
</tr>
<tr>
<td>aDORe -&gt; Fedora</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>DSpace-based</td>
<td>yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- # of AIPs per n representations:
  - FCLA: 1
  - CUL: n
  - NYU: 1

- Retain intermediate representations?
  - FCLA: no
  - CUL: yes
  - NYU: yes
Repository Exchange Package (RXP)

- RXP design goals:
  - use standards:
    - METS and PREMIS
  - be able to accommodate different AIP structures
  - contain sufficient data for receiving repository at:
    - RXP level
    - Representation level
  - identify data that receiving repository must understand
RXP: Minimal Structure

./rxp.xml

./rxp-digiprov.xml

./rxp-rep-1.xml

./rxp-rep-1-digiprov.xml

./files/
RXP: Minimal Structure

- ./rxp.xml
- ./rxp-digiprov.xml
- ./rxp-rep-1.xml
- ./rxp-rep-1-digiprov.xml
- ./files/

RXP-level information
RXP: Minimal Structure

- ./rxp.xml
- ./rxp-digiprov.xml
- ./rxp-rep-1.xml
- ./rxp-rep-1-digiprov.xml
- ./files/

Representation-level information
RXP: Minimal Structure

- `./rxp.xml`
- `./rxp-digiprov.xml`
- `./rxp-rep-1.xml`
- `./rxp-rep-1-digiprov.xml`
- `./files/`

METS document containing Source Repository info. References RXP provenance, (optional) rights, and representations
RXP: Minimal Structure

- ./rxp.xml
- ./rxp-digiprov.xml
- ./rxp-rep-1.xml
- ./rxp-rep-1-digiprov.xml
- ./files/

PREMIS document containing RXP-level digital provenance
RXP: Minimal Structure

./rxp.xml

./rxp-digiprov.xml

./rxp-rep-1.xml

./rxp-rep-1-digiprov.xml

./files/

METS document describing representation 1
RXP: Minimal Structure

- ./rxp.xml
- ./rxp-digiprov.xml
- ./rxp-rep-1.xml
- ./rxp-rep-1-digiprov.xml
- ./files/

PREMIS document containing representation 1 digital provenance
RXP: Minimal Structure

- ./rxp.xml
- ./rxp-digiprov.xml
- ./rxp-rep-1.xml
- ./rxp-rep-1-digiprov.xml
- ./files/

directory containing the representation files
RXP: Optional Files

- ./rxp-rights.xml
  PREMIS document containing RXP-level rights information

- ./rxp.xml.sig
  a stand-alone digital signature in OpenPGP format generated using sender’s private key, and rxp.xml
RXP: Optional Files

Each `rxp-rep-n.xml` must be accompanied by a corresponding `rxp-rep-n-digiprov.xml`.

Files:
- `./rxp-rep-2.xml`
- `./rxp-rep-2-digiprov.xml`
- `./rxp-rep-n.xml`
- `./rxp-rep-n-digiprov.xml`
Transfer Tests

- Two different transfer scenarios
  - Broadcast Transfer
  - Ring Transfer

- Analyzing test results against expectations
  - using results to improve RXP structure
FCLA sends $\text{RXP}_{\text{FCLA}}$ to CUL and NYU. CUL and NYU Ingest $\text{RXP}_{\text{FCLA}}$. 
CUL sends $\text{RXP}_{\text{cul}}$ to FCLA and NYU. FCLA and NYU Ingest $\text{RXP}_{\text{cul}}$. 

TIPR
Testing: Broadcast Transfer
TIPR Testing: Broadcast Transfer

NYU sends $\text{RXP}_{\text{NYU}}$ to FCLA and CUL. FCLA and CUL Ingest $\text{RXP}_{\text{NYU}}$. 
FCLA sends $\text{RXP}_{\text{FCLA}}$ to CUL.
CUL Ingests $\text{RXP}_{\text{FCLA}}$. 
CUL sends $\text{RXP}_{\text{FCLA,CUL}}$ to NYU.
NYU Ingests $\text{RXP}_{\text{FCLA,CUL}}$. 
NYU sends $\text{RXP}_{\text{FCLA, CUL, NYU}}$ to FCLA. FCLA Ingests $\text{RXP}_{\text{FCLA, CUL, NYU}}$. 
Progress to Date

- Wrote and revised initial RXP Specification
  - Require **info**: URIs for PREMIS identifiers
- Identified and resolved issue with RXP history tracking
  - now tracking RXP history in `rxp-digiprov.xml`
- Modified Partner Repositories to export and import RXPs
- Wrote & used Schematron validation scripts to test RXPs
- Completed Broadcast Transfer
- Wrote requirements for Inter-repository agreements
Issues:

- **PREMIS:**
  - Repositories need rights and digiprov at RXP-level
  - Highest unit of description in PREMIS is `representation`
  - RXP can contain multiple representations
  - RXP more like an Intellectual Entity
  - Asked PREMIS Editorial Committee to consider allowing PREMIS elements to describe Intellectual Entities when applicable
  - TIPR project still using PREMIS for RXPs
Issues:

- Identified need for Inter-repository Service Agreements (SA):
  - **SA documents** inter-repository relationship:
    - RXP composition vis-à-vis optional files
    - RXP transfer logistics
    - target repository actions upon RXP receipt
    - inter-repository rights and permissions agreements
    - archiving & preservation treatment at target repository
    - financial arrangements between source and target repositories
    - legal aspects of source and target relationship
Next Steps:

- Run Ring Transfer test
- Exchange sufficiently-varied source-AIPs
- Analyze results
- Refine transfer format
- Communicate findings
Acknowledgements

- TIPR is funded by a generous grant from the Institute of Museum and Library Services (IMLS)
Questions?

- Thank you for your time...
TIPR:
Interchange you can believe in!

http://wiki.fcla.edu:8000/tipr/