Towards Interoperable Preservation Repositories

US Digital Preservation Interoperability Framework Workshop

2010-03-31

Joseph Pawletko (NYU), Priscilla Caplan (FCLA)
Presentation Overview

- Motivation & Target Audience
- Other Approaches
- TIPR Approach
- TIPR Partners: Heterogeneous Repositories & AIPs
- Repository Exchange Package (RXP)
- Testing, Issues, & Lessons Learned to Date
- Future Plans
- Questions
Motivation & Target Audience

**Motivation:**
- Facilitate AIP exchange between heterogeneous preservation repositories
- AIP exchange important for:
  - redundancy
  - software migration
  - succession planning

**Target Audience:**
- those interested in exchanging AIPs between heterogeneous preservation repositories
Other Approaches

- Other Approaches to AIP exchange between heterogeneous repositories
  - NDIIPP ECHO DEPository Hub and Spoke (HandS)
    - R&D at University of Illinois at Urbana-Champaign
    - Hub adds preservation metadata during exchange
      - http://www.ndiipp.illinois.edu/
  - Open Archives Initiative’s Object Reuse and Exchange (OAI-ORE)
    - Fedora/EPrints demonstration project by Tarrant, et. al.
    - Not formally about AIP exchange, but may be applicable in preservation context
      - http://journal.code4lib.org/articles/1062
TIPR Approach

- Define a common exchange package format: the Repository Exchange Package (RXP)
- No reliance on transport protocol
- RXP accommodates heterogeneous AIP structures and heterogeneous repository implementations
- Exchanging Repositories can Ingest and Disseminate RXP
- Interchanging repositories agree on exchange parameters
- Peer-to-peer repository exchanges
TIPR
Repositories

- FCLA Archive
  - producer
  - producer

- RXP Format

- Cornell Archive
  - producer
  - producer

- NYU Archive
  - producer
  - producer

Repositories:
Florida Center for Library Automation (FCLA)

- runs the DAITSS Preservation Repository

Team: Priscilla Caplan (PI), Franco Lazzarino, Marly Wilson, Randy Fischer
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
TIPR
AIP Structures
FCLA

- One AIP per Intellectual Entity
- Retain First and Latest Representations
- Discard Intermediate Representations
TIPR
AIP Structures
CUL

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

Intellectual Entity

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

Cornell Archive

producer

producer

producer
• **One AIP per Intellectual Entity**
• **Retain all Representations (including “intermediates”)**
<table>
<thead>
<tr>
<th>Repository Implementation</th>
<th>FCLLA</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></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSpace-based</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of AIPs per n representations</th>
<th>FCLLA</th>
<th>CUL</th>
<th>NYU</th>
</tr>
</thead>
<tbody>
<tr>
<td>retain intermediate representations?</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
Repository Exchange Package (RXP)

- RXP design goals:
  - use standards familiar to Preservation community:
    - 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

Representation-level information
RXP: Minimal Structure

- 
  ./rxp.xml
  METS document containing Source Repository info.
  References RXP provenance, (optional) rights, and representations

- ./rxp-digiprov.xml

- ./rxp-rep-1.xml

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

- ./files/
RXP: Minimal Structure

-PREMIS document containing RXP-level digital provenance

- ./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/

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

`./rxp-rep-2.xml`  
`./rxp-rep-2-digiprov.xml`

`./rxp-rep-n.xml`  
`./rxp-rep-n-digiprov.xml`

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

Information for additional Representations
Transfer Tests

- Two different transfer scenarios
  - Broadcast Transfer
  - Ring Transfer

- Analyzing test results against expectations
  - using results to improve RXP structure
TIPR
Testing:
Broadcast Transfer
TIPR
Testing:
Broadcast Transfer
TIPR
Testing: Broadcast Transfer

- FCLA Archive
- NYU Archive
- Cornell Archive
Testing: Ring Transfer
TIPR
Testing:
Ring
Transfer
TIPR
Testing: Ring Transfer

- Cornell Archive
- FCLA Archive
- NYU Archive
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
Lessons Learned to Date

- Effort required to generate RXPs is reasonable
- Ingesting foreign RXPs more difficult, but not prohibitive
- Maintaining cross-repository provenance is tricky
- Transfer format is only one part of solution
- Inter-repository agreements are important
Future Plans

- Finish grant-related work
- Continue to talk and write about TIPR
- Respond to feedback from the community on RXP spec
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/