Integrating High-Latency Storage with Samvera

Don Brower
University of Notre Dame

Samvera Virtual Connect, July 2017
● Institutional Repository formally launched Feb. 2015
● Built using the *curate* gem
Our institution has a tape library system.

- We have access to 250 TB of storage
- But content on tape can take minutes to retrieve
Goals

● Use tape for preservation storage
● Integrate tape system with our Samvera application
● Allow patrons to get items from tape on demand without human intervention
● Change as little as possible of the existing web application
Strategy

- Make a "preservation layer" to manage the tape system, versioning, and fixity checking.
- Use our batch ingest system to put content onto tape
- Make targeted changes to Samvera app where needed
Fedora

- Maintain current use as object registry. Used by access and discovery layers
- Holds copy of metadata (descriptive, access, technical)
- Uses R datastreams to point to URL for file content
Tape Manager

- Very simple semantics: only sees bitstreams
- Assigns a unique URL to each version of each file
- Does fixity checking
- Keeps own disk cache of content stored on tape
- Store a serialization of each Fedora object on tape
Architecture

Red = Things we Created
Changes to Samvera Application Itself

- Display cache status on item show pages
- New “Download Waiting” page
- Send item updates and edits back to tape store (via cron job)
- Turn off some default background jobs, such as fixity checking
Teotihuacán: Architectural sculpture in the shape of an abstracted form of the feathered serpent.

Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Teotihuacán</td>
</tr>
</tbody>
</table>

Files

Image: 3113.xml

Assets:
- Mexico_Chaman_Aze_Cross Quintanaracul
- Mexico_Chaman_Aze_Cross Quintanaracul.png

Collections

Contact Information:

National Museum of Anthropology, Mexico City
Item Page

File Download Section
• Download button to get content

• These three files are cached on disk and can be downloaded without delay
● **“Please Note”**

● The "Begin Download" button indicates a file needs to be retrieved from tape.

● Caching is at the file level (not item level)
Retrieving File

Retrieving the file may take several minutes. The download will start automatically when the file is ready.

The file is 254 MB.

Loading...
Retrieving File

The file is 254 MB.

Success!

Manual Download

CurateND is a service of the Hesburgh Libraries of Notre Dame.
Questions? Call (574) 631-6258 or email curate@nd.edu.
Future

- Will port this to Hyrax as part of our Hyrax migration
- Use Fedora 4 instead of Fedora 3
- Abstract the tape manager API as a general “file store”? Useful to others? Maybe S3 or other cloud storage interface?
Thank You

Don Brower
dbrower@nd.edu

Github: https://github.com/ndlib/curate Nd
https://github.com/ndlib/bendo