Recollection:
Sharing and Visualization of Digital Collections

Leslie Johnston
Digital Library Federation 2010
November 3, 2010
NDIIPP: A Network of People, An Architecture for Preservation

**Vision**
Ensures the long-term access to a rich body of digital content through the establishment of a national network of committed partners.

**Preservation Network**
Two key components of infrastructure:
- **Preservation Network**: Partners collaborating to preserve and provide long-term access to digital content.
- **Preservation Architecture**: Technical components that enable digital preservation.

**Preservation Infrastructure**
- Support the needs of multiple communities over long periods of time.
- Respond to rapidly changing technologies and innovative behaviors.
- Be transparent and trustworthy.
Content Types

Images & Text
Audio Visual
Geospatial
Web Sites
Content Topics

NDIIPP Content by Topic

- Government, politics & law
- Arts & culture
- Science, mathematics & technology
- World history and culture
- Social sciences
- Religion & philosophy
- Maps & geography
- News, media & journalism
What is the Problem?

- There are thousands of memory institutions with online collections.
- There are many disparate data standards to work with.
- How can users compare data, make connections, and federate across collections and institutions?
- How do users make use of what they’ve found and share it?
In order to explore more useful tools and processes for sharing diverse content across partners’ collections, the Library began a pilot project in 2009 with Zepheira to develop an environment that can be used to collect and explore information about digital collections. The result is a software-as-a-service platform that we are calling Recollection.

http://zepheira.com/
What is Recollection Meant to Do?

• Rapidly combine private and public information, style it, and share it.
• Make collections easier to find, access and share.
• Enable new insights into patterns and relationships inherent to data.
• Build a network of trust and participation around a common goal.
What Data Types Can Recollection Work With?

- For the moment, three types of files can be imported into Recollection: rectangular spreadsheets, RSS feeds, and XML MODS. XML MODS files must be valid XML and follow the Metadata Object Description Schema (MODS).

- Best results will be obtained from data with less than 1,000 records and smaller than 2 Mb. Larger amounts of data will result in significant time delays.
What are the Recollection Standards?

• Data is translated to and exposed as RDF for you during upload.
• Recollection stores a converted copy of uploaded data in the JSON file format, augmented with some additional information used to make Collections.
• Recollection does not store the original data files. Do not rely on Recollection to store them for you.
• Users can download your data in that format, in RDF/XML, semantic wikitext, tab-separated values, or Exhibit JSON.
And Linked Data?

- Recollection exposes data as RDF for machine use, for downloading, and for individual re-use elsewhere.
- RDF provides structure and relationships that make it easier to compare data and both understand and create semantic linkages.
- Zepheira and Recollection are a part of the Linked Open Data (LOD) community initiative for moving the Web from the idea of separated documents to a wide information space of data. The key principles of LOD are that it connects public and open data sets together in a way that is simple and readily adaptable by Web developers, complementing many other popular Web trends.
- As an example, Recollection uses GeoNames, a database, Web service, and source for geographical information for users and developers. It has a rich, RESTful API and offers Semantic Web features using LOD conventions. Recollection uses GeoNames to augment geographical preserved materials with geographical context.
- In the process of developing Recollection, the team has contributed data-processing tools and best practices to the community, especially through the enhancement of open-source tools used in Recollection.
Core Recollection Technologies

• Freemix is a social networking platform for digital content, which supports data transformation and the creation of web pages with facets.

http://freemix.it/
Core Recollection Technologies

- The MIT Simile project is focused on designing tools to facilitate interoperability among digital asset distributed across individual, community, and institutional stores. Freemix builds on Simile's Exhibit project to provide interfaces to stitch together data and display the data visualizations.

http://www.simile-widgets.org/exhibit/
Core Recollection Technologies

- Freemix leverages Sesame, a framework for inferencing and querying RDF data along with Elmo, a role based interface and component oriented design infrastructure. This platform provides a flexible means for managing data and developing modular domain driven designs.

http://www.openrdf.org/
Core Recollection Technologies

- Akara is a Web framework specialized for RESTful data services designed to work with XML and other semi-structured formats. You express data models and create transforms based on these data models, plugging required inputs and outputs (XML, JSON, CSV, Atom, etc.) together into pipelines. The data is exposed in RESTful wrappers to Freemix.

http://xml3k.org/Akara/
Core Recollection Technologies

• Mulgara is a transactional XA Triplestore – a scalable, transaction-safe, purpose-built database meant for the storage, retrieval and analysis of RDF data. Mulgara is not based on a relational database. Mulgara is optimized for metadata management, as its models hold metadata in the form of short subject-predicate-object statements (triples) rather than using the large numbers of table joins usually required by relational systems when dealing with metadata.

http://mulgara.org/
Create, share, and use digital resources available from the Library of Congress and our NDIIPP partners.

Find new sources, new audiences, new collaborators.

Explore the Gallery
- NDIIPP Partners
- Featured Views
- Browse NDIIPP Collections

Try it for Yourself
- Sign up to publish data
- Upload data to get started
- Find a Data Set to create a View

Learn More
- Welcome message from the Recollection team
- User’s Guide
- View Recollection Overview Video
Recollection

lesliej

Data Sets

Conferences
Loaded by: lesliej
Loaded on: July 20, 2010
Last Modified: July 20, 2010
Basis for: 1 Data View

Folklife and Landscape in Southern West Virginia
Loaded by: lesliej
Loaded on: July 2, 2010
Last Modified: July 2, 2010
Basis for: 1 Data View

Voices From the Days of Slavery
Loaded by: lesliej
Loaded on: July 2, 2010
Last Modified: July 2, 2010
Basis for: 1 Data View

See all 4 sets

Data Views

Notices

[Other Connection] abgr and bfor are now connected. 2010-08-27 13:39:36

[Other Connection] carla_miller and abgr are now connected. 2010-08-11 16:10:09

[Acceptance Sent] You are now connected with carla_miller. 2010-08-11 15:09:11

[Invitation Received] carla_miller has requested to connect with you (see invitations) 2010-08-11 14:42:21

[Other Connection] Recollection Guide and carla_miller are now connected. 2010-08-11

Edit
Your Connections' Data Views

Your connections have 51 data views.

Sort by: Name | Creation | Modified

- **Work Leisure**
  - Created by: emad
  - Created on: July 20, 2010
  - Last Modified: July 20, 2010
  - Original data set

- **Winter Olympics 2**
  - Created by: tracy_seneca
  - Created on: July 20, 2010
  - Last Modified: July 20, 2010
  - Original data set

- **The Alaska-Yukon-Pacific Exposition Digital Collection**
  - Created by: vickimueller
  - Created on: May 2, 2010
  - Last Modified: May 2, 2010
  - Original data set

- **test DPOE**
  - Created by: abpo
  - Created on: June 28, 2010
  - Last Modified: June 28, 2010
  - Original data set
**Library of Congress Tool Directory**

<table>
<thead>
<tr>
<th>Enabled</th>
<th>Field Name</th>
<th>Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Category</strong></td>
<td>text</td>
<td>Data Integrity</td>
</tr>
<tr>
<td></td>
<td><strong>Written in</strong></td>
<td>text</td>
<td>Java</td>
</tr>
<tr>
<td></td>
<td><strong>Description</strong></td>
<td>text</td>
<td>ACE is a prototype tool that validates the integrity of digital files through mathematical techniques. Its purpose is to ensure the authenticity of digital objects in long term archives. ACE consists of a third-party Integrity Management Service (IMS) which issues integrity tokens for digital objects, and a local archive Audit Manager (AM) that periodically validates the repository. Consistency in ACE is guaranteed through the use of the archive independent IMS to validate integrity tokens and with the publication of witness values to prove the correctness of the system.</td>
</tr>
<tr>
<td></td>
<td><strong>License</strong></td>
<td>text</td>
<td>To be decided</td>
</tr>
<tr>
<td></td>
<td><strong>OS and run-time environment</strong></td>
<td>text</td>
<td>Web-based and platform independent. Requires Java 1.6 or later; Tomcat 6.0 or later; MySQL 4.0 or later; and MySQL connector 5.0.7 or better</td>
</tr>
<tr>
<td></td>
<td><strong>Tool</strong></td>
<td>text</td>
<td>ACE (Audit Control Environment)</td>
</tr>
<tr>
<td></td>
<td><strong>Application</strong></td>
<td>url</td>
<td><a href="https://scm.umiacs.umd.edu/redmine/adapt/projects/list_files/ace">https://scm.umiacs.umd.edu/redmine/adapt/projects/list_files/ace</a></td>
</tr>
<tr>
<td></td>
<td><strong>Last Update</strong></td>
<td>text</td>
<td>2010/06/25</td>
</tr>
<tr>
<td></td>
<td><strong>NDIIAPP Project</strong></td>
<td>text</td>
<td>No Value</td>
</tr>
<tr>
<td></td>
<td><strong>id</strong></td>
<td>text</td>
<td>No Value</td>
</tr>
<tr>
<td></td>
<td><strong>Developer</strong></td>
<td>text</td>
<td>University of Maryland</td>
</tr>
</tbody>
</table>
ACE is a prototype tool that validates the integrity of digital files through mathematical techniques. Its purpose is to ensure the authenticity of digital objects in long term archives. ACE consists of a third-party Integrity Management Service (IMS) which issues integrity values to prove the authenticity of objects. The IMS in ACE is independent IMS to validate objects integrity values to prove the authenticity of objects.

Data Augmentation

New Field Name*: DateTime

View for Field*:

- Timeline
- Map
- List

Based on*:

1 items selected

<table>
<thead>
<tr>
<th>Remove all</th>
<th>Add all</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last Update</td>
<td>+</td>
</tr>
<tr>
<td>Category</td>
<td>+</td>
</tr>
<tr>
<td>Written in</td>
<td>+</td>
</tr>
<tr>
<td>Description</td>
<td>+</td>
</tr>
<tr>
<td>License</td>
<td>+</td>
</tr>
<tr>
<td>OS and run-time environment</td>
<td>+</td>
</tr>
<tr>
<td>Tool</td>
<td>+</td>
</tr>
<tr>
<td>Documentation</td>
<td>+</td>
</tr>
<tr>
<td>Application</td>
<td>+</td>
</tr>
<tr>
<td>NDIIPP Project</td>
<td>+</td>
</tr>
<tr>
<td>Developer</td>
<td>+</td>
</tr>
</tbody>
</table>

Base the new field on the existing ones above.

Create
<table>
<thead>
<tr>
<th>Enabled</th>
<th>Field Name</th>
<th>Type</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conference</td>
<td>text</td>
<td>ALA Winter Meeting</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>text</td>
<td>Denver, CO</td>
</tr>
<tr>
<td></td>
<td>Date</td>
<td>text</td>
<td>2009/01/23</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
<td>text</td>
<td>2 Days</td>
</tr>
<tr>
<td></td>
<td>Type</td>
<td>text</td>
<td>library</td>
</tr>
<tr>
<td></td>
<td>URL</td>
<td>url</td>
<td><a href="http://www.ala.org/ala/conferencesevents/upcoming/midwinter/home.cfm">http://www.ala.org/ala/conferencesevents/upcoming/midwinter/home.cfm</a></td>
</tr>
<tr>
<td></td>
<td>Label</td>
<td>text</td>
<td>_0</td>
</tr>
<tr>
<td></td>
<td>ID</td>
<td>text</td>
<td>No Value</td>
</tr>
</tbody>
</table>

**ConferenceLocation**
Derived from these fields:
- F1

**ConferenceDate**
Derived from these fields:
- F2
California Wildfires - 2007

EO Natural Hazards: Fires in Southern California (link)
Title: EO Natural Hazards: Fires in Southern California
Date Captured (start): 2007-10-24T23:00:24
Date Captured (end): 2007-10-30T23:02:22

Domain:
- 34 government
- 1 map
- 30 news
- 4 non profit

Type:
- 22 city
- 3 county
- 4 federal
- 5 state

Created by abpo based on California Wildfires - 2007.
United States Elections, 2004-2008
Campaign Web Sites

Search

View by:
- Party
- State
- Election
- Race

Election
- 2004
- 2006
- 2008

State
- Alabama
- Alaska
- American Samoa
- Arizona
- Arkansas
- California
- Colorado
- Connecticut
- Delaware
- District of Columbia
- Florida
- Georgia
- Guam
- Hawaii
- Idaho
- Illinois
- Indiana
- Iowa
- Kansas
- Kentucky
- Louisiana
- Maine
- Maryland
- Massachusetts
- Michigan
- Minnesota
- Mississippi
- Missouri
- Montana
- Nebraska
- Nevada
- New Hampshire
- New Jersey
- New Mexico
- New York
- North Carolina
- North Dakota
- Northern Mariana Islands
- Ohio
- Oklahoma
- Oregon
- Pennsylvania
- Puerto Rico

Items grouped by Party

Independent
184
5.1%
Voices From the Days of Slavery
Former Slaves Tell Their Stories


Search

State
Georgia  South Carolina

County
Beaufort County  Glynn

City
St. Helena Island

Form
12" disc, 78 rpm

Genre
Ethnography and Interviews

Topic
Slaves--South Carolina--Social conditions--History--19th century. Sea Islands Creole dialect--South Carolina--St. Helena Island, Slave narratives--South Carolina, Corn--South Carolina--St. Helena--History, and Gullahs--South Carolina--St. Helena Island--Social conditions

Note
Recorded by Dr. L. Turner, St. Helena Island, South Carolina (Gullah), June 27, 1932., Sound Recording, Non-Music., and On jacket and incised on disc: "Recorded by Dr. L. Turner". Per Waterman manuscript: "Disks (sic.) #1277-1292 were made in the Gullah areas of Georgia and South Carolina. Not read for DARE." See also Publication of the American Dialect Society 3.13-28. Skip on Side A about 1 inch from outer rim.

URL
http://hdl.loc.gov/loc.afc/afr9999001.25656b
What Does Recollection Provide?

Resilience

- The Linked Data approach allows for flexibility in adapting solutions to the community as they're surfaced. While we may have a solid foundation of the community requirements, there are always things that get missed, and there are always great ideas that get surfaced as the platform is used.
- Currently, we're providing preference in the interface to highlight people, data, and views. What if later the community wanted to highlight organizations, mappings between data elements, or augmentation or curation of data? Treating these entities all as first class objects in the contextual, linked data platform allows for quick re-adjustment and delivery of new functionality. The platform becomes an reflection of the changing community needs.
What Does Recollection Provide?

**Increased Data Portability & Usability**

- The linked data approach in Recollection allows for more effective means of enabling the data to be surfaced and used beyond the application that originally created it.
- Not only is this enabled from the translation services -- translating a spread-sheet to a more web-friendly architecture, giving a spread-sheet an identifier so it can be contextually related to other things -- but it is also enabled by the users of the system. Recollection supports the ability to create human readable labels, normalize and augment data, and create comprehensible data visualizations.
- The Recollection platform allows for the ability of one person to do the "heavy lifting" of data augmentation / curation and others to build off of this effort.
What Does Recollection Provide?

Trust

- When anyone can upload data, map it, enhance / augment it and curate it, connecting it to various controlled vocabularies and reconciliation services, who you trust to do this becomes increasingly important. The Linked Data architecture behind Recollection allows for a means to endorse or trust various tasks and then others to subscribe to those believes. One might for example choose to accept all of the enhancement/augmentations that someone from the Library of Congress has put in place because of their authoritativeness.
Can I Use Recollection?

Yes, But.

• Recollection is still in development. It is being rolled out to the NDIIPP partner institutions and DLF attendees as a soft launch in November 2010 for testing and refinement of its development use cases.

• We expect to make it more openly available in early 2011.