Recollection:
Sharing and Visualization of Digital Collections
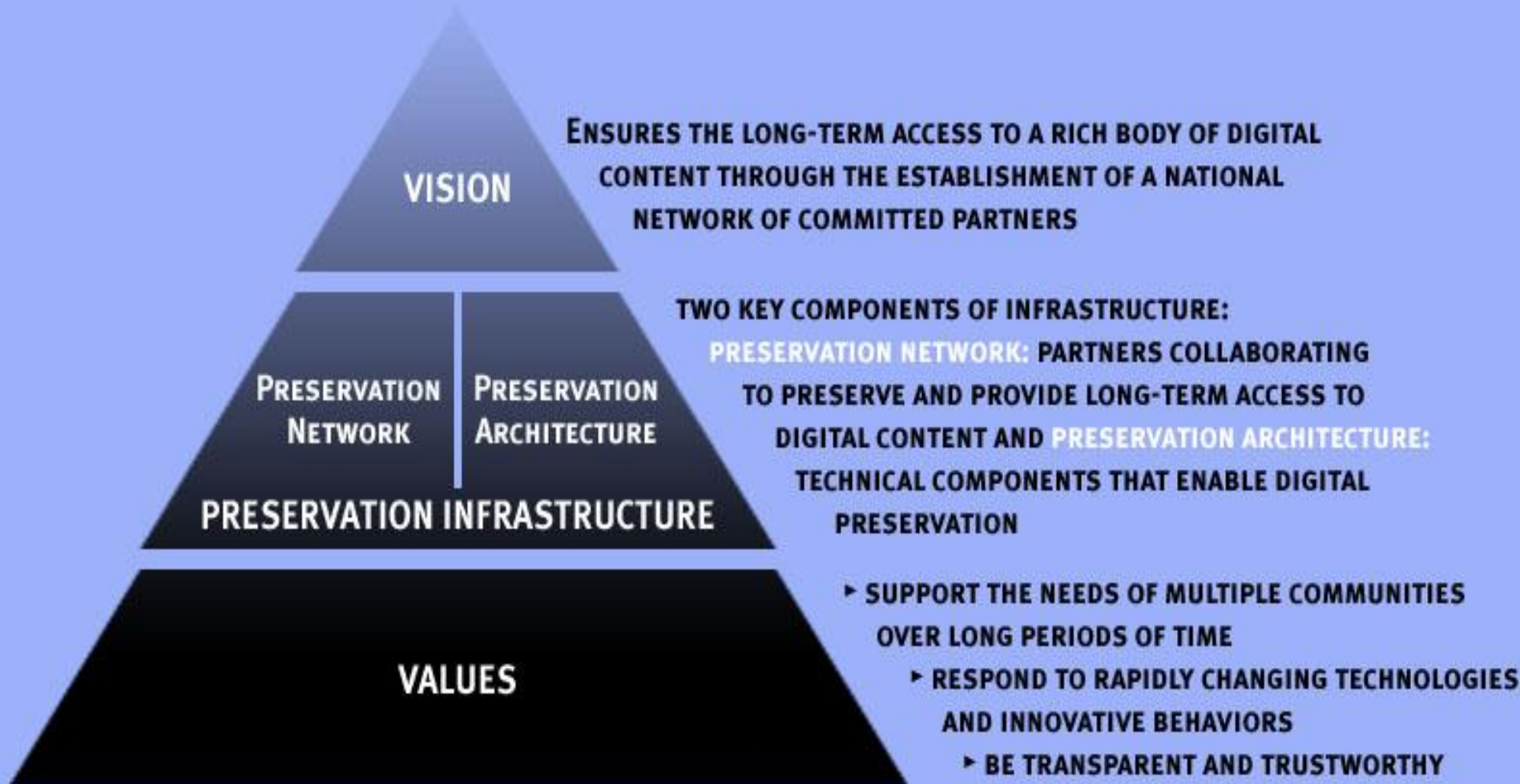
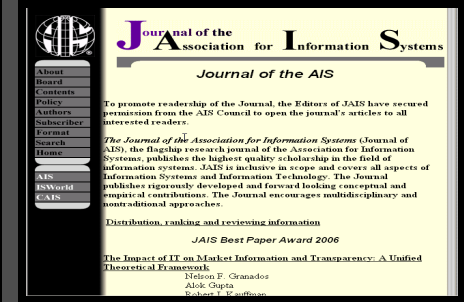**LIBRARY OF CONGRESS**

Office of Strategic Initiatives

Leslie Johnston
Digital Library Federation 2010
November 3, 2010

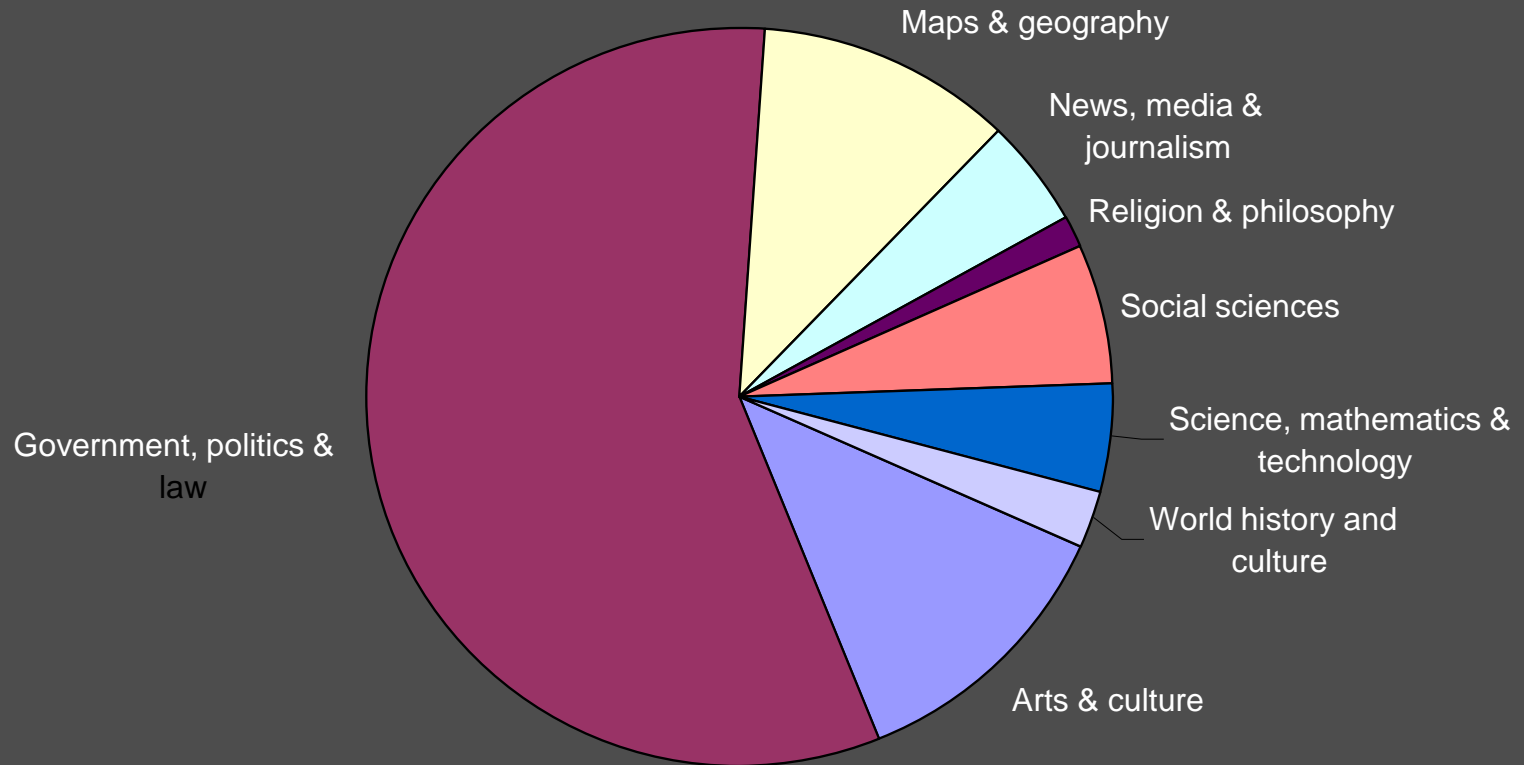# NDIIPP: A Network of People, An Architecture for Preservation

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

PRESERVATION ARCHITECTURE

PRESERVATION INFRASTRUCTURE

TWO KEY COMPONENTS OF INFRASTRUCTURE:
PRESERVATION NETWORK: PARTNERS COLLABORATING TO PRESERVE AND PROVIDE LONG-TERM ACCESS TO DIGITAL CONTENT AND PRESERVATION ARCHITECTURE: TECHNICAL COMPONENTS THAT ENABLE DIGITAL PRESERVATION

VALUES

► 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

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

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?

Recollection

- 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/

Simile

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/

Recollection

Create, share, and use digital resources available from the Library of Congress and our NDIIPP partners.

Find new sources, new audiences, new collaborators.

Battling California's Wildfires

California Wildfires

Jean Thomas

Wellstone!

Elections

Explore the Gallery	**Try it for Yourself**	**Learn More**
NDIIPP Partners	Sign up to publish data **Sign Up**	Welcome message from the Recollection team
Featured Views	Upload data to get started **Upload**	User's Guide
Browse NDIIPP Collections	Find a Data Set to create a View	View Recollection Overview Video

Recollection

Users | Data | Views | Groups | Messages (2)

Your Details | Your Connections | All Users

lesliej

Recollection Home | User's Guide | FAQ | Ask the Guide | **Upload Data**

lesliej

Edit

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

Data Sets

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

Build | Export | Edit | Delete

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

Build | Export | Edit | Delete

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

Build | Export | Edit | Delete

See all 4 sets

Data Views

Welcome, lesliej

Users | Data | **Views** | Groups | Messages (2)

Your Views | **Your Connections' Views**

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

Describe Your Data

✏ Library of Congress Tool Directory

Enabled	Field Name	Type	Value
☑	✏ Category	text ▾	Data Integrity
☑	✏ **Written in**	text ▾	Java
☑	✏ [Description] **Rename** **Cancel**	text ▾	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.
☑	✏ License	text ▾	To be decided
☑	✏ **OS and run-time environment**	text ▾	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
☑	✏ Tool	text ▾	ACE (Audit Control Environment)
☑	✏ Documentation	url ▾ text url image datetime location	https://wiki.umiacs.umd.edu/adapt/index.php/Ace:Audit_Manager_User_Guide
☐	✏ label		_0
☑	✏ Application		https://scm.umiacs.umd.edu/redmine/adapt/projects/list_files/ace
☑	✏ **Last Update**	text ▾	2010/06/25
☑	✏ **NDIIPP Project**	text ▾	*No Value*
☐	✏ id	text ▾	*No Value*
☑	✏ Developer	text ▾	University of Maryland

Enabled	Field Name		Type	Value
☑	🖉	Category	text ▾	Data Integrity
☑	🖉	**Written in**	text ▾	Java
☑	🖉	Description	text ▾	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

Data Augmentation ✕

New Field Name*:

DateTime

View for Field*:

 Timeline

 Map

 List

1984-04-26 1986-09-19
1993-07-24 2003-12-20
2004-07-26

Select which type of view this field will be used to create.

Based on*:

1 items selected	Remove all		Add all
↕ Last Update	−	Category	+
		Written in	+
		Description	+
		License	+
		OS and run-time environment	+
		Tool	+
		Documentation	+
		Application	+
		NDIIPP Project	+
		Developer	+

Base the new field on the existing ones above.

Create

☑	🖉	Licen		va 1.6 or later;
☑	🖉	OS a		SQL connector 5.0.7
☑	🖉	Tool		
☑	🖉	Docu		
☐	🖉	label		
☑	🖉	Applic		cts/list_files/ace
☑	🖉	Last		
☑	🖉	NDIIP		
☐	🖉	id		
☑	🖉	Devel		

+ Augment Data

of 39 Next Record →

Enabled	Field Name	Type	Value
☑	✏ **Conference**	text ▾	ALA Winter Meeting
☑	✏ **Location**	text ▾	Denver, CO
☑	✏ **Date**	text ▾	2009/01/23
☑	✏ **Duration**	text ▾	2 Days
☑	✏ **Type**	text ▾	library
☑	✏ **URL**	url ▾	http://www.ala.org/ala/conferencesevents/upcoming/midwinter/home.cfm
☐	✏ **Label**	text ▾	_0
☐	✏ **ID**	text ▾	*No Value*
☑	✏ ConferenceLocation	location	39.739,-104.985

Derived from these fields:
- F1

Delete

| ☑ | ✏ ConferenceDate | datetime | 2009-01-23T00:00:00+0000 |

Derived from these fields:
- F2

Delete

+ Augment Data   ↻ Refresh Augmented Data   🗑 Delete This Record   ← Previous Record   **1 of 36**   Next Record →
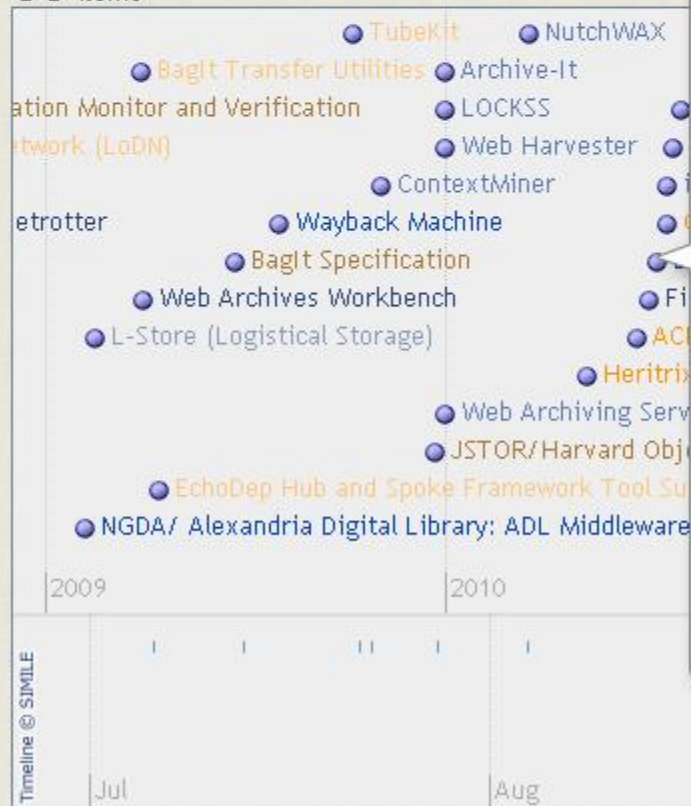
About Recollection | Help

Recollection

| Users | Data | Views | Groups | Messages |

1. Describe   2. Build   **3. Preview**   4. Publish

Theme: Pepper Grinder ▾

Library of Congress Tool Dire

LIST · T

39 Items

- TubeKit
- NutchWAX
- BagIt Transfer Utilities
- Archive-It
- ation Monitor and Verification
- LOCKSS
- twork (LoDN)
- Web Harvester
- ContextMiner
- etrotter
- Wayback Machine
- BagIt Specification
- Web Archives Workbench
- Fi
- L-Store (Logistical Storage)
- AC
- Heritrix
- Web Archiving Serv
- JSTOR/Harvard Obj
- EchoDep Hub and Spoke Framework Tool Su
- NGDA/ Alexandria Digital Library: ADL Middleware

BWF MetaEdit

Tool	BWF MetaEdit
Developer	Federal Agencies Digitization Guidelines Initiative (FADGI)
Category	Metadata
Description	BWF MetaEdit permits embedding, validating, and exporting of metadata in Broadcast WAVE Format (BWF) files.
Written in	C++
OS and run-time environment	Windows, Mac, Linux
License	Public domain
Documentation	http://sourceforge.net/projects/bwfmetaedit/
Application	http://sourceforge.net/projects/bwfmetaedit/
Last Update	2010/07/13
NDIIPP Project	Federal Agencies Digitization Guidelines Initiative (FADGI)

igital PRESERVATION

ss Data Exchange
tegrity **Data**
er **Format**
tion **Ingest**
lata
servation
ge Repository
Archiving

2009   2010

Timeline © SIMILE

Jul   Aug   Sep   Oct

Recollection

Users | Data | Views | Groups | Messages

California Wildfires - 2007

MAP · **TIMELINE** · LIST

34 Items filtered from 69 originally (Reset All Filters)

Ventura County Fire Department Homepage

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
URL - archive	http://wayback.archive-it.org/877/*/http://earthobservatory.nasa.gov/Naturalhazards/natural_hazards_v2.php3?img_id=14587

Official Web S

EO Natural Hazards: Fires in Southern California

23   Oct 24   Oct 25   Oct 26

Welcome to the City of Carlsbad Official Website

Timeline © SIMILE

Oct 14   Oct 21   Oct 28   Nov 4

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.
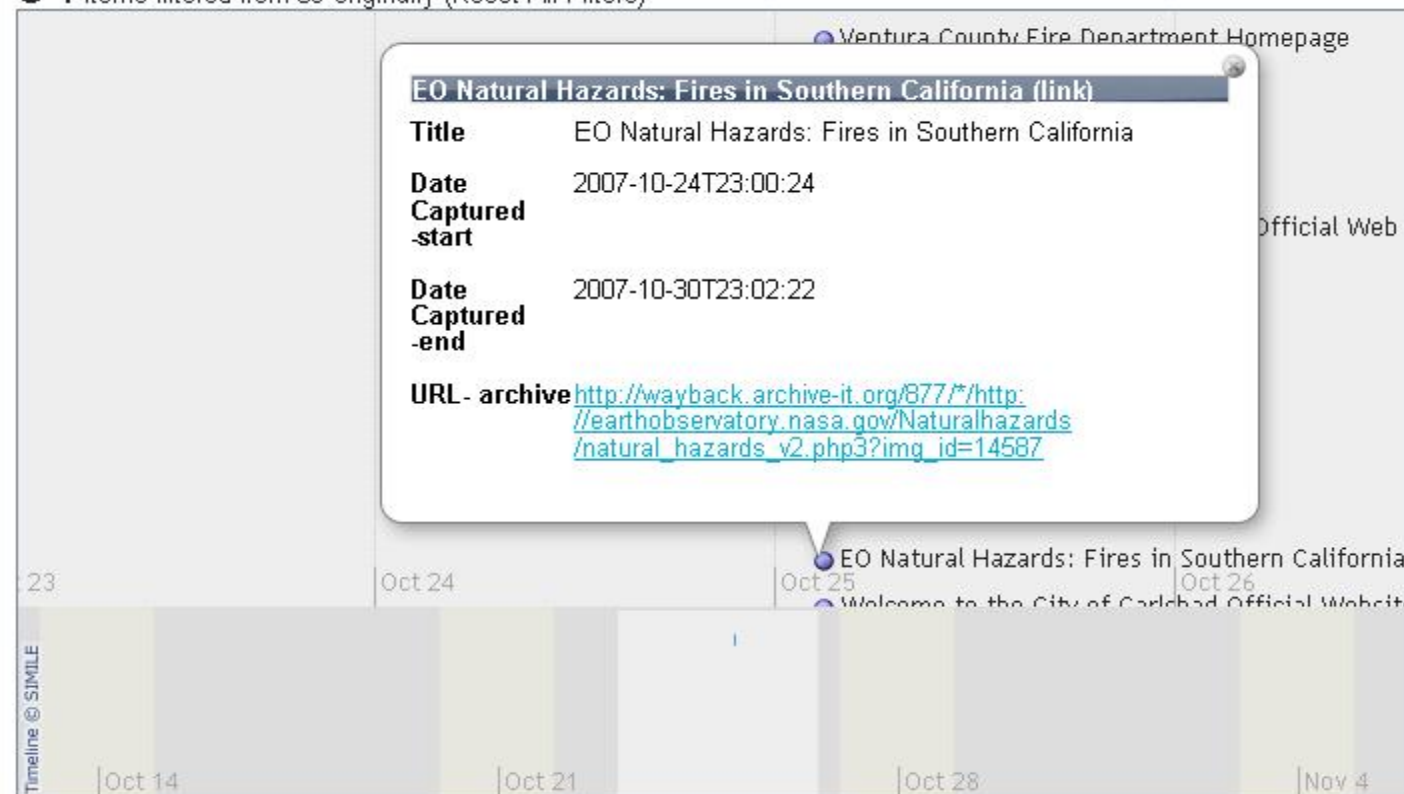
United States Elections, 2004-2008
Campaign Web Sites

Search

Election

2004 **2006 2008**

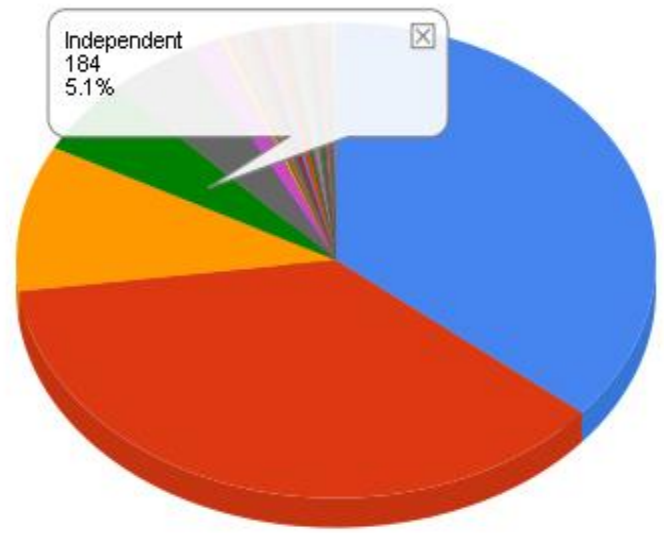**State**

Alabama Alaska American Samoa Arizona Arkansas **California** Colorado Connecticut Delaware District of Columbia **Florida** Georgia Guam Hawaii Idaho Illinios **Illinois** Illnois 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

View by :
Party

- State
- Party
- Election
- Race

Items grouped by Party

Independent
184
5.1%

Embed

Voices From the Days of Slavery
Former Slaves Tell Their Stories

Contraband of War: Slaves of the Rebel General, Thomas F. Dryton, Hilton Head, S.C., May, 1862.

Search

gullah

State

Georgia   South Carolina

County

Beaufort County

Charleston County   Glynn County

26 Items filtered from 1

State	South Carolina
County	Beaufort County
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/afc9999001.25656b

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 an 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 to 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.

http://www.digitalpreservation.gov

http://recollection.zepheira.com/

Leslie Johnston
lesliej@loc.gov