Recollection: Integrating Data through Access

Laura E. Campbell

The Library of Congress
Office of Strategic Initiatives
101 Independence Ave SE
Washington, DC 20540-1310
lcam@loc.gov

Abstract. The National Digital Information Infrastructure and Preservation Program will demonstrate a pilot tools platform called Recollection that supports access to distributed NDIIPP collections. The Recollection platform uses semantic technologies to make NDIIPP collections easier to find, access, share, and integrate with other digital information sources.

Keywords: User Interfaces, Interoperability & Data Integration, Digital Archiving and Preservation, Collection Development & Management and Policies, Knowledge Organization Systems, Semantic Web Issues in Digital Libraries

1 Description

The National Digital Information Infrastructure and Preservation Program at the Library of Congress is an initiative to develop a national strategy to collect, archive and preserve the burgeoning amounts of digital content for current and future generations. It is based on an understanding that digital stewardship on a national scale depends on active cooperation between communities. The NDIIPP network of partners have collected a diverse array of digital content, including social science datasets; geospatial information; Web sites and blogs; e-journals; audiovisual materials; and digital government records [1].

These diverse collections are held in the dispersed repositories and archival systems of over 130 partner institutions where each organization collects, manages, and stores at-risk digital content according to what is most suitable for the industry or domain that it serves. This practice is necessary in a federated network of heterogeneous infrastructures but creates challenges in providing meaningful access across collections. However, it is clear that digital content grows in value exponentially as it is integrated and interconnected. As the Library of Congress and its partners develop a framework for a national digital collection, they have recognized a requirement to share and integrate partner collections in the interest of a coherent strategy. Early in 2009, a pilot project was initiated. The Recollection

project is developing a platform to support access to the dispersed NDIIPP collections.

NDIIPP partners understand through experience that aggregating and sharing diverse collections is very challenging. Transfer of data and accompanying metadata from one institution to another for integration into a different system is resource-intensive. The aggregation and consolidation of information about NDIIPP collections for web resources have, to date, been a manual process. This is not a scalable strategy and information becomes outdated as the Partners evolve.

Project Recollection seeks to provide the platform, tools and environment that enables the community of NDIIPP Partners to share their collections and data on an ongoing basis. In addition, NDIIPP collections can be showcased from a central point through the activities of the Partners and not the manual labor of the Library. This allows NDIIPP to maintain the benefits of a distributed network of partners and also take advantage of the collections speaking to one another. Specific goals for the Recollection project are:

• Focus on providing value through content rather than highly-specialized metadata.

• Look for structure patterns of what is working well and leverage these principles to document best practices for digital preservation of content.

• Build trust among NDIIPP Partners. Give the Partners incentive, tools and methods to work together on Recollection effectively.

The Recollection platform will use semantic technologies to enhance discoverable access for NDIIPP collections, making them easier to find, access, and share, and especially to integrate with other digital information sources. The data that Recollection uses will come from a variety of sources. Excel spreadsheets, databases, XML, bibliographic records, and HTML are just of few examples. The majority of the data that will be included in Recollection already exists from local workflows and will not require massive customization; some of the data will be harvested right from web sites. Data guidelines focusing on key access points will be shared and an open interface will enable third parties to plug services and applications into the Recollection framework, encouraging community participation. Exposing the hooks between collections and enabling reuse of the collections can support education, research, policy analysis and other unforeseen uses.

This platform will also provide administrative tools for the NDIIPP network. Because the platform will provide exportable data from dispersed collections in a common format, it will be possible to know, for example, which institutions are collecting which web sites and can help the network plan and execute its collection activities. See the example below from the Recollection platform that identifies NDIIPP partners, plotted on a map, that are active in collecting and preserving web archives.
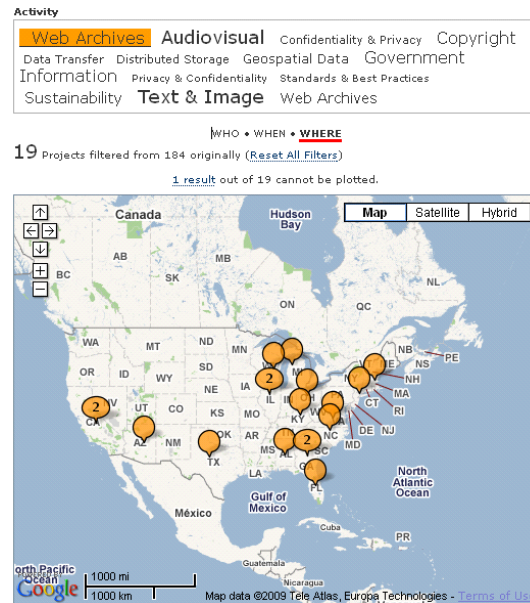
Fig. 1. This shows a screen shot from the demonstration Web site.

The demonstration of the Recollection platform will show how NDIIPP collections have been stitched together to provide broad access to distributed, heterogeneous digital objects from a single platform. The NDIIPP Program at the Library of Congress is working closely with Zepheira, an information management solutions company providing services applying semantic technology and web architecture to support collaboration and social computing [2].

2   Demonstration Outline

The demonstration will include a live web site demonstration of the Recollection tool platform. Technical requirements are: a computer connected to a projector, PowerPoint software, and an Internet connection. Browser requirement is the latest version of Mozilla Firefox.

3   Target Audience

This demonstration of Recollection, an interactive tool platform to enhance access to preserved digital materials is targeted toward digital collection curators, metadata managers, web service designers, digital archivists, and anyone interested in collaborative approaches to information services.

4   Recollection: Integrating Data through Access

References

1. National Digital Information Infrastructure and Preservation Program, http://www.digitalpreservation.gov
2. Zepheria, http://zepheira.com/