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. This demonstration of the Recollection project of the National Digital Information Infrastructure and Preservation Program at the Library of Congress will showcase a prototype platform, tools and environment for sharing and access to diverse born-digital collections. As the Program has addressed the development of distributed preservation through a national community of partner institutions, the challenge of access and interoperability has become more urgent. The network needs to be able to strategically bring collections under stewardship and keep an inventory without excessive burden on the collecting organizations. The data under stewardship is very diverse and follows standards acceptable within each content domain. These circumstances require an infrastructure that enables the community of NDIIPP Partners to share their collections and data on an ongoing basis. This allows NDIIPP to maintain the benefits of a distributed network of partners and also take advantage of the collections speaking to one another.

Keywords: User Interfaces, Interoperability and Data Integration, Digital Archiving and Preservation, Collection Development and 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 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; websites 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.

The Recollection platform uses 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 demonstration data comes from a variety of sources and formats including Excel spreadsheets, databases, and XML datasets. Data guidelines focusing on key access points are shared and an open interface enables third parties to plug services and applications into the Recollection framework, encouraging community participation [2].

2 Demonstration Outline

The demonstration will include a live website demonstration of the Recollection tool platform. The demonstration will show how data is loaded and managed in the system, how data managers can select data attributes and create tag clouds, timelines, maps and graphs to provide access to complex digital content such as geospatial data, social science datasets, and web archives. It will demonstrate how data can be augmented using web services.   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.

References

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