

Enhance Library Catalog Searching with Geospatial Technology, Phase I

Final Report

David Siegel

Geospatial Data and Information Software Engineer

Office for Information Systems

dave_siegel@harvard.edu

Marc McGee

Geospatial Resources Catalog

Harvard College Library Technical Services

mmcgee@fas.harvard.edu

Project summary

The Library Lab project, Enhance Catalog Searching with Geospatial Technology, was proposed with the intention of exploring the feasibility and utility of taking library catalog data and making it geospatially searchable. We took a test set of 1,700 HOLLIS records, sent them to MetaCarta Labs, who geocoded the data and returned the records on a custom map search interface. Using this proof-of-concept system we were able to demonstrate how catalog data could be geocoded, placed in an interactive map and presented in a way that it could be searched spatially while staying true to its roots as library catalog data. The proof-of-concept system we created provided a means to demonstrate spatial searching in an intuitive way and we found that those that did not understand what benefits GIS or geocoding catalog data could provide were immediately informed.

We also demonstrated the value added benefit of this in a number of ways. Through the map interface, information about the catalog data **and relationships between** the data were exposed in ways that traditional library catalogs don't provide. The system demonstrated how data holdings which are geocoded become discoverable in a new and exciting paradigm.

Based on the user feedback we received through one-on-one conversations, through email we received and through the Office of Scholarly Communications (OSC) demonstration web site (<http://osc.hul.harvard.edu/geo/>) it was clear that the concept of enhancing catalog searching by spatially enabling the catalog data was embraced by both those familiar and unfamiliar with geospatial concepts. In fact, we believe that the results had an added benefit of allowing those that did not understand what spatially enabling catalog data meant to become familiar with the concept. Some groups that reached out to us who already work in the geospatial world were excited about the opportunities that this technology could bring to their work as well.

Since the Phase I was a success we reached out further to professionals both at Harvard and outside the Harvard community to explore what directions developing this technology should take to become more mainstream. The Phase II proposal hopes to build on the findings and feedback from Phase I and produce a production-level spatial search engine for various special collection at Harvard, starting with smaller special collections but with an eye toward spatially enabling entire repositories of library data.

Accomplishments

field-dev.metacarta.com/~bmajor/UI_Scratch/index2.html

Apple Yahoo! Google Maps YouTube Wikipedia News Popular

Query: chamberlain
Search Records

Map
Book
Recording
Journal / Serial
Reset This Form

Result Page Page: 1

Current Page 1

Relevance: 0.576043
General Joshua L. Chamberlain electronic resource / Greystone Communications- Inc. produced by Melissa Jo Peltier director- Craig Haffner. Lincoln and Gettysburg / produced by Arthur Drooker director- Donna E. Lusitana.
Movie undefined
New York A&E Home Video-c1994.

Title	Format	Author	Publisher	Geo Refs
General Joshua L. Chamberlain electronic resource / Greystone Communications- Inc. produced by Melissa Jo Peltier director- Craig Haffner. Lincoln and Gettysburg / produced by Arthur Drooker director- Donna E. Lusitana.	Movie	undefined	New York A&E Home Video-c1994.	Gettysburg Appomattox Bowdoin College Harrisburg New York New York Calkins
In the hands of Providence Joshua L. Chamberlain and the American Civil War / by Alice Rains Trulock.	Book	Trulock- Alice Rains.	Chapel Hill University of North Carolina Press-c1992.	Chapel Hill ill. Maine Providence
Ceremonies in commemoration of the one hundredth anniversary of the birth of Abraham Lincoln- Philadelphia- February 12- 1909. Military order of the loyal legion of the United States- Commandery of the state of Pennsylvania.	Book	Military Order of the Loyal Legion of the United States. Pennsylvania Commandery.	Philadelphia Press of John T. Palmer company 1909.	Philadelphia Nicholson
Through blood & fire selected Civil War papers of Major General Joshua Chamberlain / edited- with commentary by Mark Nesbitt.	Book	Chamberlain- Joshua Lawrence- 1828-1914.	Mechanicsburg- Pa. Stackpole Books-1996	Mechanicsburg, P ill. Maine
The passing of the armies an account of the final campaign of the Army of the Potomac- based upon personal reminiscences of the Fifth army corps- by Joshua Lawrence Chamberlain ... with portraits and maps.	Book	Chamberlain- Joshua Lawrence- 1828-1914	New York and London- G.P. Putnam's sons- 1915.	Library of Congress Potomac
Joshua Chamberlain a hero's life and legacy / John J. Pullen.	Book	Pullen- John J.	Mechanicsburg- PA Stackpole Books-c1999.	Gettysburg, Pa Mechanicsburg, P Maine
Conceived in liberty Joshua Chamberlain- William Oates- and the American Civil War / Mark Perry.	Book	Perry- Mark-1950-	New York- N.Y.- U.S.A. Viking-1997.	Gettysburg, Pa ill.
Ceremonies in commemoration of the one hundredth anniversary of the birth of Abraham Lincoln. American				

A screen shot from the proof-of-concept system, showing some of the geocoded HOLLIS records (<http://osc.hul.harvard.edu/geo/>)

In Phase I we demonstrated that it was possible to geocode existing HOLLIS catalog data without making changes to the native MARC records, thus enabling spatial search. By envisioning several use case scenarios we pulled data from HOLLIS and provided a screen cast showing how spatially enabling the catalog data enhances the library catalog search experience. The feedback was nearly all positive (collected by OSC) and groups from around the Harvard community and beyond could see the benefits of expanding this work.

As part of our work for the proof-of-concept phase we used a subscription service from MetaCarta Labs to geocode selected MARC records. We also investigated the availability of open source tools for doing the geocoding. We learned that the open source community is working hard on the issue, combining the

use of gazetteers with natural language processing to enable fast and accurate geocoding. The problem is that an open source solution is still a couple of years away from being at production-level capacity for libraries.

During Phase I, the map interface was demonstrated to a wide audience, including other institutions that we are already collaborating with to share metadata and coding resources to create a common geoportal. The other institutions have expressed strong interest in this technology, and more importantly a willingness to participate in building it. Most of this Library Lab's proposers are also currently working with other institutions on a platform called OpenGeoportal. We believe that the cross institutional support behind the OpenGeoportal could be leveraged to make this project successful beyond the walls of Harvard University.

Use case scenarios were identified, including:

- Geographic browsing/searching of HOLLIS and other catalog data
- Ability to refine searches using facets, better exposing relationships between places and subjects, authors, etc.
- Desire to have separate publication, provenance and acquisitions location searching
- Ability to geospatially track history of publishing/knowledge using catalog data
- Potential as a collection development tool, exposing absences in collection through data density analysis
- Creating a GeorSS feed of library data that can be filtered by clients
- Displaying results from searches on a map base where information from other Harvard catalogs are shown. Also includes integrating the results of a catalog search on a map with other media such as Picassa.

User feedback showed us that spatially enabling catalog searching is useful in expanding access to library resources. It also showed us the importance of creating a user interface that is more specifically geared to the library community. The proof-of-concept interface was acceptable for discovery but did not integrate well with the way people think when searching catalogs. It lacked such features as faceted browsing and well thought out presentation of spatial relationships between resources. It also had no advanced indexes that allowed users to refine searches to explore other catalogs at Harvard and elsewhere. To that end we have proposed a Phase II project, which includes the necessary work to meet those requirements.

Facets from search results; use to further limit results and to discover topical relationships.

Panning and zooming the map would re-populate the facets based on records overlapping the map extent.

Title	Format	Author	Publisher	Geo Refs
Gettysburg heroes perfect soldiers: hallowed ground / Glenn W. LaFantasie.	Book	LaFantasie- Glenn W.	Bloomington Indiana University Press- c2008.	Gettysburg, Pa Raymarket
The grand old man of Maine selected letters of Joshua Lawrence Chamberlain: 1863-1914 / edited by Jeremiah E. Gouka foreword by James M. McPherson.	Book	Chamberlain- Joshua Lawrence- 1828-1914.	Chapel Hill University of North Carolina Press- c2004.	Maine Chapel Hill
The Gettysburg nobody knows / edited by Gabor S. Boritt essays by Harry Pfanz ... et al..	Book	undefined	New York Oxford University Press-1997.	Gettysburg Aldie Pennsylvania Shepherdstown Shepherdstown
His proper work: a biography of Gen. Joshua Lawrence Chamberlain / by Sis Deans.	Book	Deans- Sis Boulos.	Kearny- N.J. Belle Grove Publishing Co.-1996.	Kearny, N.J. Maine Ill.
Courage on Little Round Top a novel / Thomas M. Eshen.	Book	Eshen- Thomas M.- 1955-	Dallas- Tex. Skyward Pub.- c2005.	Gettysburg, Pa
Bayonet! Forward my Civil War reminiscences / Joshua Lawrence Chamberlain.	Book	Chamberlain- Joshua Lawrence- 1828-1914.	Gettysburg- Penn. S. Clark Military Books-1994.	Gettysburg, Pa Penn.
To Gettysburg and beyond the parallel lives of Joshua Lawrence Chamberlain and Edward Porter Alexander / Michael Golay.	Book	Golay- Michael- 1951-	New York Crown Publishers- c1994.	Gettysburg, Pa New York
Chamberlain at Gettysburg electronic resource / video produced by Robert Senkel.	Movie	undefined	New York- NY A&E Television Networks-2005.	Gettysburg, Pa New York, NY

Mock-up of a spatial search tool integrating facets.

Challenges

Relying on MetaCarta's generic user interface

- Slowness/reliability of MetaCarta server
- Lack of control of design and functionality
- Infrastructure to index georeferenced data was not specifically geared towards libraries

Geocoding issues

- Lack of control over which MARC metadata fields are geocoded and how they're indexed (e.g. separate indexing for place of publication, geographic subject headings, acquisition information)
- Problems with geocoding: coordinate fields not identified; false hits: e.g. "ill." for illustrated material mapping to Illinois, "supt." for superintendent mapping to "Supt, France".

Pulling sample MARC records from HOLLIS

- De-duping records from HOLLIS required writing custom macros
- Too much human intervention needed to capture data from the catalog API

Cooperation among disparate Harvard entities

- Since most of the time was spent creating the proof of concept, getting time to meet with others at Harvard was challenging.

Next steps

Next steps for this Library Lab project are thoroughly outlined in the Phase II follow-up proposal submitted to the Office for Scholarly Communications (OSC) on October 1, 2011. The main next steps as outlined in the Phase II proposal are to:

- Create an open source infrastructure to index and search records that were not originally encoded for geospatial applications.
- Geocode a limited set of special collection records (e.g. Thomas Hollis collection, Map Collection, others)
- Build a production-level user interface and back-end to enable geospatial search and discovery. We envision creating an API that allows users to integrate geospatial catalog searching in their own applications. While there could be a “common” user interface, users should not be limited to using it.
- Create a scalable back end that can streamline the data ingest process
- Create new indexing methods and APIs that allow integration with other systems such as Library Cloud
- Include metadata facets search results to allow limiting for: format type, dates, subjects, authors, languages, etc.
- Allow for custom defined geographic searches of: place of publication, geographic subject headings, or acquisition data
- Leverage the latest open source technology and share and collaborate with peer institution using the OpenGeoportal platform (e.g. <http://geodata.tufts.edu/>)

Budget spent

The proof of concept phase cost approximately \$28,000. This covered staff time and a small amount of customization work contracted to MetaCarta.

Publicity

A web page providing access to the proof-of-concept system, along with supporting documentation and a screencast demonstrating potential use case scenarios, was sent out to the Harvard Library community via the hcomms listserv for feedback on Aug. 23, 2011. <http://osc.hul.harvard.edu/geo/>

In addition to sending out project information to the library listserv, we contacted members of the Harvard geospatial community individually, and in some cases met one-on-one with faculty and researchers to solicit specific feedback. Also, we did outreach via e-mail to OpenGeoportal collaborators, including GIS and library professionals from Tufts, Berkeley, Yale, and more.

Presentations

The project and proof-of-concept systems were presented in person at the OSC sponsored events:

Library Lab Showcase – lightning talks (Aug. 4, 2011)

Library Lab Showcase – project demos (Oct. 27, 2011)

ABCD-Technology in Education – project demos (Oct. 3, 2011)

The project was presented to a group of library technical services staff members at the:

HCL Technical Services staff 10-minute Tech Talk (Nov. 10, 2011)

The project was presented to OIS staff; librarians at Princeton, Tufts, Mass. State Government and the Boston Public Library.