

LibraryCloud final report

Project summary

The aim of the project was to develop LibraryCloud as a multi-library metadata server able to make library data of various sorts openly available. LibraryCloud is intended to be middleware that enables broader and innovative use of library metadata.

Accomplishments

- Extended the LibraryCloud list of partners. There are currently about 35 institutions represented on the LibraryCloud mailing list.
- Mounted LibraryCloud on new hardware suitable for Harvard use.
- Developed data models for catalog data and event data. (We are continuing to work on and extend these models.)
- Imported catalog and circulation data from four libraries in addition to Harvard's:
 - Darien Public Library
 - Northeastern University
 - San Jose Public Library
 - San Francisco Public Library
- Imported metadata for Web content, including:
 -
- LibraryCloud now has almost 15 million catalog records, and 17 million circulation event records.
- We have received but not yet integrated about one million social event data points from LibraryThing
- Created an API Query Builder (hls17.law.harvard.edu/dev/matt/librarycloud/librarycloud.v.3/web/query_builder.htm) as a tool to educate developers in our API. You can construct live API queries via pulldowns. The query itself is shown to you, as are the results.
- Engaged in a continuing project to find pages in Wikipedia that are about books. We parsed Wikipedia categories via DBpedia, winnowed the list, and used Amazon Mechanical Turk to come up with a list of 57,000 pages that we are confident are about books. This quadruples the number we were able to find simply by querying Wikipedia via DBpedia.
- Developed an advisory circle.
- Engaged linked open data pioneer Dan Brickley in a continuing project to help discover how to slot Web content into existing book categories. Dan has reported on that here:
- Proposed LibraryCloud as a component of the Digital Public Library of America

- Began discussions on the circumstances under which we can begin to allow limited alpha access to LibraryCloud.
- Submitted a draft privacy policy for Library Board approval
- Worked with Berkman to develop scripts for scraping and harvesting metadata
- Created sample apps (not including ShelfLife) running on top of LibraryCloud, including a “long tail browser,” a dynamic map showing circulation trends in multiple libraries, and a way of seeing top circulating items for any keyword in all participating libraries.
- Enabled LibraryCloud for Linked Open Data access.

Challenges and next steps

- Develop automated processes for keeping LibraryCloud data up to date
- Further develop and document the API, especially in response to a soft alpha launch
- Work on strategies for integrating catalog entries from across multiple libraries
- Get a privacy policy in place
- Continue research into how to automatically associate book categories with non-book content
- Research into de-duplication. (We brought in Corey Harper from NYU for a few days to discuss his work on this.)
- Develop a strategy for maintaining a LibraryCloud server as an integral part of the Harvard Library system.

Budget spent

Total: \$69,810

- Equipment
- Consulting on clustering
- Amazon Turk work on Wikipedia book pages

A list of any publicity you did, e.g., articles, blog posts, podcasts, etc.

I have not kept a list of the times I have written about, demo’ed, or spoken about LibraryCloud. It’s been a pretty constant stream.

Here is a list of posts on the LiL blog categorized as “LibraryCloud”: <http://librarylab.law.harvard.edu/blog/category/projects/librarycloud/>

I have given presentations in Rio, Bonn, Berlin, Helsinki, Dublin, Madrid, Barcelona, Ottawa, and elsewhere where I have talked about LibraryCloud.

I have spoken with numerous media folks about it.

I have demo’ed it at the US State Department and to the head of design at Nokia.

I have given invited lectures at U. of Michigan, a Berlin media studies grad school, University of Illinois communications school, and a talk Forum d'Avignon in which I have discussed LibraryCloud, either in the talk or more informally.

Kim Dulin and I gave a Berkman Tuesday talk about ShelfLife and LibraryCloud

I have discussed LibraryCloud at the DPLA Technical Workstream and on the DPLA mailing list.