Nominee:  Harvard Law Library
Project:   LibraryCloud: An Open Metadata Server

Stanford Prize for Innovation in Research Libraries (SPIRL) Entry

Nominator:  Jonathan Zittrain, Vice Dean for Library and Information Resources, Harvard Law School; Professor of Law at Harvard Law School and the Harvard Kennedy School of Government; Professor of Computer Science at the Harvard School of Engineering and Applied Sciences; Co-founder of the Berkman Center for Internet and Society.

Nominator’s Statement

I am pleased to nominate LibraryCloud (http://librarycloud.harvard.edu) for the Stanford Prize for Innovation in Research Libraries.

Libraries know far more than they’ve been able to make usable so far. For example, they can now see how the ideas in their materials are being discussed and appropriated, for many of those discussions are now done in the networked public. They also have rich usage metadata, much of which can be anonymized. They know which works have been put on reserve for courses and could know which ones are mentioned in syllabi in their own institution and institutions worldwide. Libraries thus can be far more involved in the development of the ideas and knowledge that their resources and services support.

LibraryCloud is an open metadata server that enables libraries to make available to developers more of what libraries and their communities know. The aim is to enable and encourage libraries and unaffiliated developers to create applications that are highly useful to specific groups of users, and to make it easier for other sites to integrate library-based knowledge. LibraryCloud is designed to support a “virtuous circle” of metadata out to and back from a community.

In short, LibraryCloud aims to turn libraries into platforms on the grounds that libraries cannot by themselves create their future. Rather, we need a planet full of smart 14 year olds (and their older siblings) to figure out how to put what libraries know to full work.

While LibraryCloud is currently a Harvard project, it has been conceived from its beginning as a multi-library infrastructure. We have ingested catalog and usage data from a number of other libraries already, and in its next iteration plan on providing channels by which individual instances of LibraryCloud can automatically share data. (LibraryCloud is an Open Source project.)

LibraryCloud supports Stacklife as one of its applications. I am also nominating Stacklife as a SPIRL entry. I would like to suggest that the two be considered together.
The innovation: What LibraryCloud is

LibraryCloud is a platform by which a library can make its metadata available to developers via open APIs.

That metadata can currently include information about the library’s holdings, the usage of its items, the structure of its subject classifications, and links out to related materials.

The open APIs enable developers to create applications that use that metadata and mash it up with other data and existing sites.

The current version of LibraryCloud was designed and implemented by the Harvard Library Innovation Lab in the Harvard Law Library. Another version is being prototyped for the Harvard Library.

Contents

The current version has access to the following metadata:

- Harvard’s complete catalog of almost 13M items spread across 73 libraries and an off-campus repository
- Usage data including:
  - Check-outs faceted by library and by faculty, grad student, or undergrad
  - Times an item has been put on reserve
  - How many libraries have acquired an item
  - How many times an item has been a required text book in a course
- Additional data from:
  - Hathi Trust public domain e-book catalog
  - Open Library’s public domain e-book catalog
  - DPLA’s Biodiversity Heritage Library public domain e-text catalog
  - Harvard Law Library’s Nuremberg Trials Project Case records
  - Harvard Business School faculty’s books and book chapters
- Links to related information from:
  - Google Books
  - National Public Radio
  - Canadian Broadcasting Corp.
  - Wikipedia
- Identifiers from:
  - Harvard’s HOLLIS catalog
Process

- Metadata is ingested from a variety of sources and in a variety of formats and serializations, including MARC21, JSON, XML, and CSV.
- The metadata is normalized to a schema based on Dublin Core, with extensions. (The original source metadata is always maintained and is retrievable through the API.)
- The metadata is enhanced, including by:
  - Clustering all editions of the same work
  - Linking out to Wikipedia book articles
  - Linking out to National Public Radio and Canadian Broadcasting Corporation audio broadcasts
  - Doing a rough calculation of their overall usage (“StackScore”)
  - Algorithmically supplying Library of Congress call numbers when missing, when possible
- The metadata is placed into MongoDB
- A Solr index is created
- The metadata is queryable through LibraryCloud’s RESTful API.

Services

- Item API:
  - Query the API for any indexed data points.
  - The original source metadata in its entirety is by default included in the JSON returns.
  - Supports filtering, facets, boolean queries, paging and sorting.
- Event API (in use internally but not yet publicly available)
  - Query for information about usage events in the life of an item.
- Subject API
  - Query for parents, children, and siblings of any Library of Congress subject class.
  - Supports filtering, facets, boolean queries, paging and sorting.
Mission statement
LibraryCloud was developed by the Harvard Library Innovation Lab. Our mission is threefold:

• We think in public.
• We build software that demonstrates how libraries can bring yet more value to scholars and researchers.
• We amplify our effect by eagerly partnering with other groups with similar passions.

Or, as our founder John Palfrey put it, “Hack libraries!”

Project history
LibraryCloud began in 2010 as a multi-library collaboration with the aim of aggregating interesting metadata and making it available through open APIs. Many libraries expressed interest, and about a half dozen contributed metadata — typically catalog and usage data — to the project. The Library Innovation Lab created the first instance of LibraryCloud. Stacklife became the first substantial application to depend upon its API.

After the founding of the Digital Public Library of America, our Lab was given the task of creating the initial prototype of its platform. We demonstrated this platform at the second national meeting of the DPLA, featuring applications written on top of the API during a hackathon. The current DPLA platform reflects some of the design decisions we made for LibraryCloud.

Since then we have continued developing LibraryCloud at our Lab. We have brought in more sorts of data, have added APIs, and have created new apps.

At this time, Harvard Library has funded a year of development for a new, more scalable version of LibraryCloud. Harvard Library Cloud is designed as a series of workflows that ingest, normalize, and enrich metadata, and then make it openly accessible; the Library at this time is considering the future development of the Harvard Library version. (David Weinberger is leading the development effort for the Harvard Library version, as well as serving as the PI for the Lab’s version.)

The Library Innovation Lab continues to develop its version of LibraryCloud. If and when the Harvard Library version becomes part of the Library infrastructure, the Lab’s version will serve as a sandbox for research and development, some of which we expect would be absorbed by the Harvard Library version.

The version this application concerns is the Library Innovation Lab’s.

Intended clientele
Because LibraryCloud is a platform and not an application, there are four related classes of intended clientele:
• Developers can use LibraryCloud to create applications that make use of the metadata it aggregates.
• End users benefit from the applications developers create.
• Librarians will use applications written especially for (or by) them to help them manage their libraries better.
• Web sites can use LibraryCloud to integrate library knowledge into their own services.

Principal players’ biographies
Here are brief biographies of the people in the Harvard Library Innovation Lab who contributed to this project:

• Annie Cain, developer. A one-time reference librarian who gave it all up to code, Annie Cain hacks away at pushing ideas out to the Harvard Library community and beyond. She has a B.S. in Informatics from the University of Washington and an M.L.S from Simmons College.

• Ingrid Cheung, developer. Ingrid is the newest member of the Harvard Library Innovation Lab. She is an applications developer with a particular interest in developing tools for other developers. She works on many of our back end tools and does integration work between Stacklife and LibraryCloud. She is interested in augmented reality, home brew electronics, and programming outreach. She teaches circuitry and crafts to middle school and high school students, and spends her free time in the air.

• Paul Deschner, developer. Paul has been an applications developer at the Harvard Library Innovation Lab since 2009. He has been working mostly on the Lab’s LibraryCloud data platform and other data projects, developing ingestion, object representation and metadata enhancement implementations for a wide range of library-related item and usage metadata from Harvard and other institutions and web services. Prior to his arrival at the Lab, Paul worked as a web developer at Harvard and software engineer at an internationalization services startup. He has traveled widely, studying and teaching in Europe for over a decade, and is fluent in multiple languages. He has an M.A. in English from Tufts University.

• Kim Dulin, co-director. Kim also serves as Associate Director for Collection Development and Digital Initiatives at the Harvard Law Library. In addition to her experience as an academic law librarian, Kim has served as practicing attorney and an adjunct professor of law. Kim has a JD from the University of Iowa College of Law, an MS from the University of Illinois Graduate School of Library and Information Science, and a BA from the University of Iowa.

• Jeff Goldenson, designer. Jeff works at the intersection of libraries, technology and fun. He is the designer in the Library Innovation Lab where new library projects are imagined and built. He’s Co-Teacher, Harvard Graduate School of Design Seminar 09125, The Library Test Kitchen, a workshop where - with the financial support of the Harvard Library
- students design and build their own library projects. Jeff earned a Masters of Science from the MIT Media Lab.

- Matthew Phillips, developer. Matt works on a number of different projects in the Lab. Recently his focus has been on the Awesome Box – an alternate returns box for awesome checkouts, and Perma.cc — an archiving tool that aims to stop link rot in academic publications and beyond. He holds a BS in Computer Science and earned a Masters in Computer Science from Virginia Tech. He loves libraries, the Internet, and bicycles.

- David Weinberger, co-director. David writes about the effect of technology on our ideas. He has been a philosophy professor, high tech marketer and entrepreneur, and freelance journalist. He is currently a senior researcher at Harvard's Berkman Center for Internet & Society and is leading Harvard Library's Interoperability Initiative. He has a Ph.D. in philosophy from the University of Toronto.

Links

- Harvard Stacklife: http://stacklife.law.harvard.edu
- DPLA-Hathi-Internet Archive mashup: http://stacklife-dpla.law.harvard.edu
- Integrated DPLA version: http://dp.la [click on “bookshelf” at the top]
- LibraryCloud home: http://librarycloud.harvard.edu
- About the StackView widget: http://librarylab.law.harvard.edu/blog/stack-view/
- StackView technical documentation: https://github.com/harvard-lil/stackview
- Stacklife Open Source code: https://github.com/harvard-lil/stacklife

Media