Harvard LibraryCloud

The Harvard LibraryCloud is a beta service that provides a rich set of Harvard Library data, including:

- Complete bibliographic information for the almost thirteen million items in the Harvard Library collection
- Metadata for approximately two million items in Harvard's image collection
- Metadata about millions of items in the Harvard Archives, including links to several hundred thousand that have been scanned and are available online.
- Stackscore, a metric that gauges the "community relevance" of items based on usage and other factors.

Topics

- About
- Inside LibraryCloud
- More Information
- Applications

About

LibraryCloud operates as a metadata hub that makes the included data openly available for use by applications and websites at Harvard and beyond, and enables easier management of the data flowing among the Library's various systems. With LibraryCloud, anyone at Harvard and beyond can use Harvard Library's data to fuel an app or integrate with a website.

The metadata that Harvard LibraryCloud makes available is offered under open licenses, as is the source code (http://library.harvard.edu/https://github.com/harvard-library) for LibraryCloud itself.

The LibraryCloud was developed with the generous support of the Harvard Library Lab (http://lab.library.harvard.edu/), the Harvard Library Office for Scholarly Communication (http://osc.hul.harvard.edu/), the Berkman Center for Internet & Society (http://cyber.law.harvard.edu/), and the Arcadia Fund (http://www.arcadiafund.org.uk/). The initial version of LibraryCloud was created by the Harvard Library Innovation Lab as an experiment in combining metadata from multiple libraries. The current version was built by Harvard Library Technology Services (http://hul.harvard.edu/ois/) and Harvard Library Innovation Lab

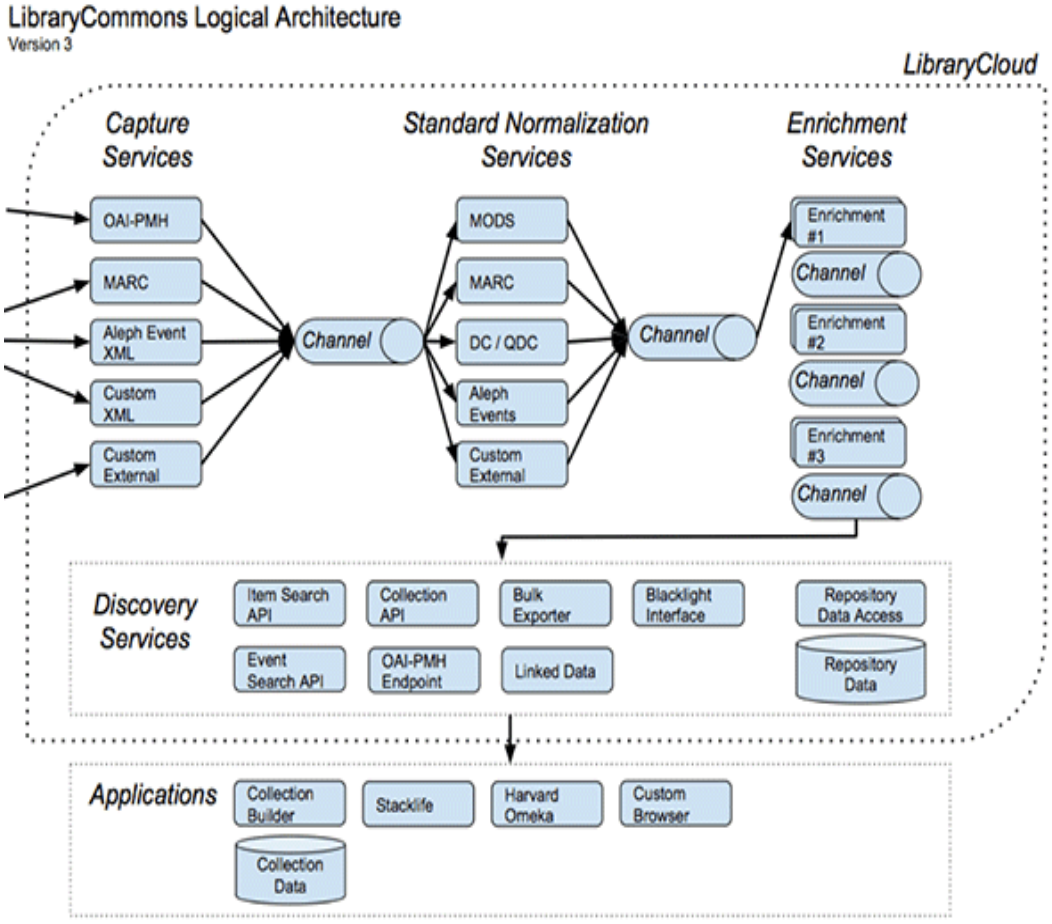
(http://librarylab.law.harvard.edu/).

Back To Top

Inside LibraryCloud

LibraryCloud consists of four sets of services connected by automated workflows.

1. *Ingest*: Connect to data sources and pull them in. This can be triggered by a particular event (e.g., a source updates) or on a schedule
2. *Normalize*: Once the data is in, it is converted into one of the standard formats LibraryCloud uses. For example, internally bibliographic data is represented using the MODS standard (with a few tweaks).
3. *Enrich*: LibraryCloud enhances its data. For example, dates can be converted into standard formats, links to external sources can be added, etc.. Over time, more enhancement services will be added.
4. *Discover*: The output of LibraryCloud can be through an API, direct download, or other interfaces. It is currently available through APIs in XML or JSON format, using the MODS or Dublin Core schema.



Back To Top

More Information

- FAQ
- API documentation
- API tutorial for beginners
- Contact: Randy Stern, Library Technology Services, randy_stern@harvard.edu

Back To Top

Applications

Here are some initial prototype applications, done by a very early adopter and a hobbyist programmer on staff. More are on the way!

- Marc Duby's native Android 4.2 that lets you search by title, scroll through results, and display details. This is a work in progress, expected to be ready in very early 2015. Available at GitHub.
- David Weinberger is a hobbyist programmer. He promises you that his code is embarrassing. Nevertheless: the BoogyWoogy Browser is a colorful browser of the Harvard collection, intended to encourage serendipity.
- David's similarly hobbyist wordWalker displays the text from an item's record, lets the user select up to ten words, and then searches the Library for the works that have those words the most often in their records.
- mediaCount (also from David) draws pie charts showing the types of media found in up to ten subject searches.

Back To Top