



LibraryCloud server

Library Lab Grant Proposal

Note to the Library Lab committee

We understand that this proposal is out of scope for what the Library Lab has in mind. But we believe there is a case to be made that LibraryCloud is enabling technology for library innovation across the university (and broader), and thus ought to be considered for extraordinary funding.

If the Library Lab disagrees, we hope you will still consider the far more modest proposal we are submitting for funding the developing of the Library Cloud APIs. The development of those APIs is covered by the present proposal, but deserves independent consideration (we believe) if you decide against this present proposal.

Overview

Library Cloud is a proposed data service that will provide open, programmatic access to the metadata available within the Harvard Library system, and, eventually, across multiple libraries. We believe LibraryCloud is essential to the work of the Library Innovation Lab, should be a core service offered by the Harvard Library Lab to enable innovation, and will be a major part of the public, open library ecosystem.

Why LibraryCloud?

We believe a publicly available metadata server could have a tremendously positive impact on library innovation at Harvard. These benefits include:

1. A powerful, open platform for developers creating innovative library-related tools and apps
2. A wealth of information open and available to researchers investigating patterns of library use

The first version of the server will include metadata from the Harvard's entire bibliographic and holdings databases as well as related information, such as checkout, recall and course reserve information. In a subsequent release, we also intend to include collections information in the metadata server (in the form of collection objects and metadata), which would support more effective discovery and analytic access to Harvard's special collections archives.

The server will use an extensible database schema to merge and manage different types of metadata. In addition to providing a metadata repository, LibraryCloud will also incorporate a faceted search engine, based on Lucene/Solr, that will support microsecond response times to search queries across a wide range of metadata. All data services will be available through a documented and fully supported web-services API.

Developing LibraryCloud

LibraryCloud further develops work we are doing for the ShelfLife application. We are currently building an initial version of the metadata server to support the data needs of ShelfLife.

The data schema and API of the server we are building are based on the data requirements in ShelfLife. To build and deploy an open metadata server to support other library applications at Harvard, we would need to do the following:

1. Migrate to a more extensible data schema based on a publicly supported standard (e.g. Dublin Core, MODS)
2. Design and implement an appropriate set of APIs to support queries and data requests
3. Document and publish the data architecture and the API protocols
4. Set up a keyed access system to control data access and to track usage by application
5. Purchase, set up, and maintain a server for public access
6. Investigate the legal liabilities and develop appropriate policies for data access

Prior Art

Early examples of web-service API's for library data are becoming available on sites such as OpenLibrary, where one can access catalog records in JSON and RDF formats. LibraryCloud aims to make a broader set of data available (e.g. bibliographic, scrubbed circulation, acquisitions, collections, etc.), and make it available via protocols and in formats that can be easily accessed by other software applications (using REST, XML, JSON, etc.)

Metadata standards have emerged over the years from efforts such as Dublin Core, METS, MarcXML, and MODS. The Hydra Project, a collaborative effort between the Universities of Hull, Stanford, and Virginia, is attempting to build a multi-institution framework for creating library apps that includes a metadata repository and related components. Serial Solutions, the company behind the AcquaBrowser, has introduced a web discovery tool for libraries, called Summon, that includes a proprietary metadata repository with substantial data.

Match with Library Lab grant values

This will be open source enabling technology that will both lower the bar to and stimulate library innovation across Harvard. It can be accomplished well within a year, but it would be, however, a continuing part of the library innovation infrastructure.

Effect on daily operations

Paul Deschner will be the "principle investigator." He will oversee the contract developers as part of his work at the Library Innovation Lab.

Proposed funding

Contract developer:

240 hours at \$85/hour = \$20,400

One year of operational support:

60hours/month for 12 months @ \$85/hour = \$61,200

Server (32gb RAM):

\$5,0000

TOTAL: \$86,600

