LibraryCloud

Harvard LibraryCloud is a metadata hub that provides granular, open access to a large aggregation of Harvard library bibliographic metadata. The public LibraryCloud Item API supports searching LibraryCloud and obtaining results in a normalized MODS or Dublin Core format. LibraryCloud contains records from Harvard's Alma ILS (over 16M bib records), JSTOR Forum (6.7M image records), and ArchivesSpace finding aids (2.6M finding aid components). Records for materials in Harvard's Digital Repository Service (DRS) are enriched with a subset of administrative metadata from that repository. Some records are also marked for inclusion in curated sets accessible via OAI-PMH.

LibraryCloud also contains an alpha release of a Collections API, that is planned for use as a digital collection definition and export service. The Collection API allows a group of LibraryCloud records to be labeled as part of a named collection. The collection may then be harvested through OAI-PMH in order to import metadata into on-line digital exhibit platforms, such as Omeka or DPLA. The full build out of the collection API and a collection builder web application is still a work in progress.

LibraryCloud documentation

- LibraryCloud APIs
- LibraryCloud Metadata
- LibraryCloud Support

Technical Information

LibraryCloud's backend is built on a scalable metadata processing pipeline. The technology stack includes:

- Amazon Web Services (including the Simple Queue Service, Simple Notification Service, and Apache camel), deploying up to 10 EC2 servers to quickly update the data from source catalogs 3 times/week.
- A Java RESTful API Service that uses solr/lucene as the search index supporting Item API search functions
- A Java RESTful API Service that uses a simple Amazon RDS database supporting the Collections API
- An OAI-PMH data provider

LibraryCloud is open source software that can be downloaded from Github at https://github.com/harvard-library

An overview of the metadata processing pipeline is shown below.