Project Rationale

Archived

LD4L 2014, which was the Linked Data for Libraries original grant running from 2014-2016, has been completed. This page is part of the archive for that grant.

The Cornell University Library, the Harvard Library Innovation Lab, and the Stanford University Libraries have all been exploring new approaches to dramatically improve the discovery experience for users seeking scholarly information resources, such as traditional monograph and journal publications, archival materials, research datasets, images, recordings, cultural artifacts, newspapers and magazines, web archives, and much more. All three institutions have been looking at ways to gather context and relationships about these resources that go far beyond traditional metadata approaches. Drawing from a wide variety of sources, including traditional catalogs, usage information, research guides, author information, semantically interoperable subject headings, geospatial references, collection metadata at multiple granularities, and expert annotation and selection, we plan to gather contextual information about resources such as books, articles, serials, datasets, and multimedia into a semantic-web-based Scholarly Resource Semantic Information Store (SRSIS). Based on our initial explorations and related work (see Background above), we believe that the SRSIS approach will enable significant improvements for both focused and serendipitous discovery by scholars and students in all disciplines and at every level.

The Cornell University Library will build on its major success in creating VIVO, a semantic-web-based system for creating interlinked and interoperable profiles of faculty, scholars, and researchers in the context of their institutions and activities, with the goal of enhancing discovery of researchers and research. The VIVO system already supports the creation and use of extensive researcher-related context around publications, and it is being extended to do the same for research datasets. The VIVO system has been widely adopted by universities and research organizations, including Penn, Brown, Duke, and the USDA, and the VIVO ontology has been recommended by vote of the 60 members of the NIH-funded Clinical and Translational Science Awards (CTSA) collaboration as the vehicle for networking research information across platforms and institutions.

The Harvard Library Innovation Lab will build on its work in creating the LibraryCloud system[1], which makes library metadata from many institutions available through open APIs and as Linked Open Data; and StackLife[2], which improves information resource discovery by enabling libraries to leverage the behavior of their users, the relationships among their assets and among their users, and the socially-generated knowledge that emerges from them.

The Stanford University Libraries will leverage its ongoing efforts to model and publish scholarly assets as Linked Data. In addition to expressing MARC records and journal article metadata as RDF, and seeking to cross-link them to provide cross-domain discovery, Stanford is designing a local authority file for Stanford authors. Drawing on information in its enterprise identity management system, faculty profiles application, and library management system, the Libraries seek to build a Linked Data-based author file that will draw from, and enrich, international efforts such as VIAF and ORCID (see Appendix B for a description of ORCID), while populating local University systems with knowledge about researchers' publication activities. Stanford will also expand on its leading work in the Fedora[3] and Hydra[4] open source communities (see Appendix B for descriptions of these projects), by modeling digital objects and their metadata with Linked Data rather than more traditional XML-based hierarchies, in order to both leverage external sources of information (community ontologies and/or authority files, e.g.) as well as increasing the flexibility of operating and extending its digital repository system.

Our collective long-term vision is that each research library, archive, or cultural memory institution will make openly available to the entire scholarly community not just the basic metadata about their scholarly information resources, but also a broad range of organizational, descriptive, and relationship context for those resources, with that information curated by local experts within the institution. This project is designed to take a few steps closer to that vision, by providing the software and ontologies to make it easy for institutions to implement their own separate SRSIS instances, or to share SRSIS-compatible data from their own existing systems.

These SRSIS instances will interoperate at two levels: directly, through semantic subject headings, authority records linked to ORCID and VIAF, and other shared identifiers and vocabularies; and at a deeper level by asserting semantic web equivalence relationships between resources in separate SRSIS instances (e.g., for resources sharing a DOI, ISBN, or OCLC Control Number). In addition, an organization could aggregate the RDF triples (or create a common Blacklight index) from many SRSIS instances to create a comprehensive resource for discovery and access across many institutions, perhaps specialized to a particular discipline or media type. This approach puts the responsibility for the data with the local institution that curates it while providing maximum flexibility for all kinds of reuse.

---

[1] http://librarycloud.harvard.edu

[2] http://librarylab.law.harvard.edu/blog/shelflife-alpha/

[3] http://fedoracommons.org/

[4] http://projecthydra.org/