Enhancing discovery with entity reconciliation:
Use cases from the Linked Data for Libraries (LD4L) project

Dean B. Krafft, Cornell University Library
Workshop on Reconciliation of Linked Open Data
Dec. 1-2, 2014
Linked Data for Libraries (LD4L)

- On December 5, 2013, the Andrew W. Mellon Foundation made a two-year $999K grant to Cornell, Harvard, and Stanford starting Jan ‘14
- Partners are working together to develop an ontology and linked data sources that provide relationships, metadata, and broad context for Scholarly Information Resources
- Focus is on linking to authoritative external sources and making accessible library-curated information about resources that goes beyond standard bibliographic information
Feminist, Gender, and Sexuality Studies: A Research Guide

A guide to discovering the extensive FGSS resources available in the Cornell University Library and beyond. Click on the PURPLE TABS to access each section in this guide.

Last Updated: Apr 1, 2014 | URL: http://guides.library.cornell.edu/fgss | Print Guide | RSS Updates | Email Alerts

Encyclopedias, Dictionaries & Almanacs

Encyclopedias, Dictionaries and Almanacs

(online and Olin Reference HQ 1090.3 .A45X 2003)
"R[eference guide and an introduction to the many ways in which men have defined, imagined, and experienced male identity in the social, cultural, and political contexts of the United States." [Introduction]

(olin Reference HQ 1115 .A543 1997)
"The main aim of the glossary is to identify the concepts that have structured feminist theory in the period of the second wave, from about 1963, indicating their provenance in travelling theory from this time... The glossary aims to trace the variety of ways in which these concepts have been transformed as they were critiqued and appropriated." [Introduction] Liberally cross-referenced. Extensive bibliography of works cited.

(olin Reference BT 83.55 .D53x 1996)
"The Dictionary covers standard topics of Christian theological studies, as well as topics of other religions that are important for the continuing feminist dialogue among many different religious, cultural, and racial groups [e.g., see Islam]. There are also entries on topics of particular interest to those engaged in feminist studies. There are no entries on persons..." [Preface] The Bibliography of Works Cited is comprehensive for English-language publications on feminist theologies in the U.S. context.

(olin Reference PN 98 .W64 E53x 1997)
"The mission of this encyclopedia is to map out the vast intellectual territory that has arisen mostly since the 1970s, due to the efforts of feminist literary scholars and critics in the United States and Great Britain." [Preface] Definitions of important terms, summaries of critical ideas, and essays on theoretical movements and ideas in feminist literary scholarship and thought.

(olin Reference HQ 1190 .E63x 2000)
Succinct entries on major individuals, theories, developments, and terms in feminist history and thought. Bibliographies.

(online and Olin Reference HQ 1115 .E54 2009 +)
"Gender is so significant in the ways it shapes individuals and societies that it isn't possible to cover every aspect of gender scholarship in a single venue. Our intent with this encyclopedia is to provide users with a "gender lens" on society by focusing on significant gender scholarship within commonly recognized areas of social life."
Brittain, Charles

**Positions**

- Department Chairperson, College of Arts and Sciences, College of Arts and Sciences
- Professor/Classics/Philosophy, Classics (CLASS), College of Arts and Sciences

Charles Brittain is a Professor of Classics and Philosophy, specializing in ancient philosophy.

His research is primarily concerned with Hellenistic philosophy (especially epistemology and ethics), Cicero, Augustine, and the Platonic tradition from Plato to Simplicius.

**Websites**

- Classics profile
- Philosophy profile

**Selected Publications**

1. *Philo of Larissa: the last of the academic sceptics* (2001)
   - English.
   - Text in English, with some Greek.
LD4L Use Cases

• Currently have draft set of 12 use cases in 6 clusters on LD4L public wiki: https://wiki.duraspace.org/display/ld4l/

• Users include: faculty, student, dean, researcher, librarian, bibliographer, and cataloger

• Use cases are driving the ontology development, MARC to BIBFRAME conversion, data source selection, and engineering work
1. Bibliographic + curation data, e.g. create virtual collection, tag resources
2. Bibliographic + person data, e.g. see works authored by faculty who have taught a given course
3. Leveraging external data including authorities, e.g. search with geographic data
4. Leveraging the deeper graph (via queries or patterns), e.g. identify related works or closely related authors
5. Leveraging usage data, e.g. discovery guided by community usage
6. Three-site services, e.g. cross-site search
Identifying data sources

- Bibliographic data: CUL/Harvard/SUL catalogs
- Person data: VIVO, Stanford CAP, Profiles
- Usage data: LibraryCloud, Cornell/Stanford circulation, BorrowDirect circulation
- Collections: Archival EAD, IRs, SharedShelf, Olivia, arbitrary OAI-PMH
- Annotations: Cornell CuLLR, Stanford DMS, Bloglinks, DBpedia, LibGuides
- Subjects & Authorities: FAST subject headings, OCLC Works, ORCID iDs, VIAF, LCNA, Getty, etc.
Entity Reconciliation

• Locally critical to link information across library system silos
• Essential to link across the three partners to support discovery, annotation, virtual collections: works, people, places, subjects, etc.
• Linking to web of LOD surfaces new relationships and networks
• Library role: expose our own unique entities and connect them to the rest of the world

*The more we can reconcile, the more we can discover*
The More We Can Reconcile: Loosely Connected Researchers
Strongly Connected Researchers

CSS 4830: Environmental Biophysics
Fall 2012

National Institute of Food and Agriculture
USDA

Computational Agriculture
2012-2014

Susan Riha

Cornell's supercomputers will crunch weather data to help farmers manage chemicals

New York State Water Resources Institute (NYS WRI)

Crop and Soil Sciences (CSS)

Earth and Atmospheric Sciences (EAS)

Harold Van Es
How will LD4L make these connections?

• By using ontologies commonly found in linked data
• By connecting with Cornell VIVO/Stanford CAP/Harvard Profiles information
• By using persistent, stable local identifiers (URIs)
• By linking stable local identifiers to global identifiers
• By supporting annotations with provenance
• By linking to external sources of networked relationships: Dbpedia, IMDB, OCLC
Leveraging OCLC work identifiers

• OCLC WorldCat functions as a union catalog of bibliographic identifiers shared across institutions.

• Goal is to reference common OCLC Work URIs in bibliographic resources from Cornell, Harvard, and Stanford to support common search discovery and interlinking:
  – Harvard: 82% of 13.6 million bib records can be matched to OCLC Work identifiers.
  – Stanford and Cornell have ~2.6 million records in common out of ~5.8 million in each collection.
  – Annotations and usage information can then be compared across 3 institutions.
  – If it can work for 3, it can work for many.
Likely components of the LD4L ontology

• Library resources: BIBFRAME
• Additional bibliographic types and partonomy relationships: FaBiO
• People/Organizations: VIVO-ISF (includes FOAF)
• Annotations: OpenAnnotation
• Provenance: PAV
• Virtual Collections and Structured Relationships: OAI-ORE
• Concepts: SKOS (or vocabularies such as Getty with stable URIs)
• Many identifiers: VIAF, ORCID, ISNI, OCLC Works
Challenges

• Perfection as the enemy of the good (e.g., in ontologies and reconciliation)
• Minting vs. finding identifiers
• Wider issues of entity reconciliation -- when do we mean sameAs vs. seeAlso?
• Scale
• Leveraging technologies developed for other purposes (e.g., LibGuides, DMS, CAP)
• Thinking outside the bibliographic box
Enhancing discovery and understanding with fully reconciled linked data

Ideally library descriptions of our information resources will:

- Refer to **identified** works, people, organizations, places, events, and subject headings
- Be discoverable in concert with other explicit and implicit library metadata of all types
- Join with local personal and organizational profile data to form a coherent, richer local authority picture
- Interoperate across libraries
- Interoperate with any other linked data via common, global identifiers and shared ontologies
Questions?