Linked Data for Libraries: A Project Update

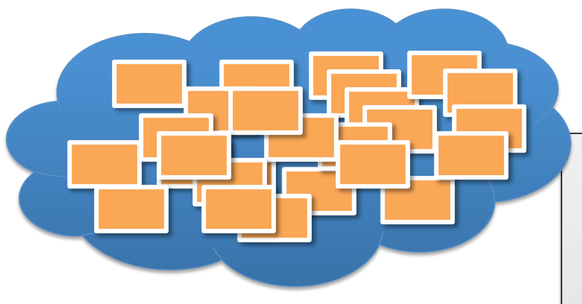
Dean B. Krafft and the LD4L Project Team

Cornell University Library
HARVARD LIBRARY
STANFORD UNIVERSITY LIBRARIES

This poster reports on the first eighteen months of the Mellon-funded two-year Linked Data for Libraries (LD4L) project [http://ld4l.org], a partnership of Cornell University Library, Stanford University Libraries, and the Harvard Library Innovation Lab. The goal of the project is to use Linked Open Data to leverage the intellectual value that librarians and other domain experts and scholars add to information resources when they describe, annotate, organize, select, and use those resources, together with the social value evident from patterns of usage. The project is producing an ontology, architecture, and set of tools that work both within and across individual institutions in an extensible network.

Vision: Create a LOD standard to exchange all that libraries know about their resources

Courtesy National Gallery of Art, Washington



42 raw use cases

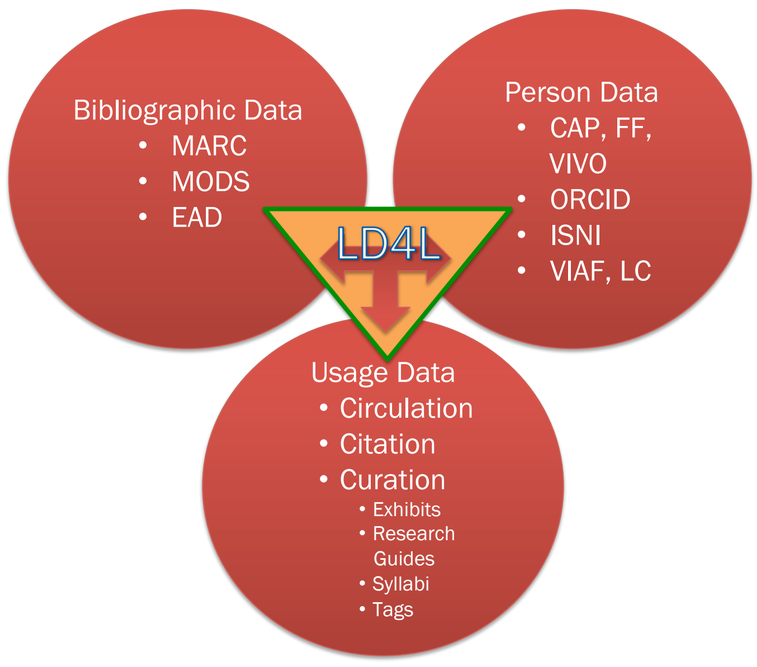
12 refined use cases in 6 clusters…

LD4L Use Case Clusters

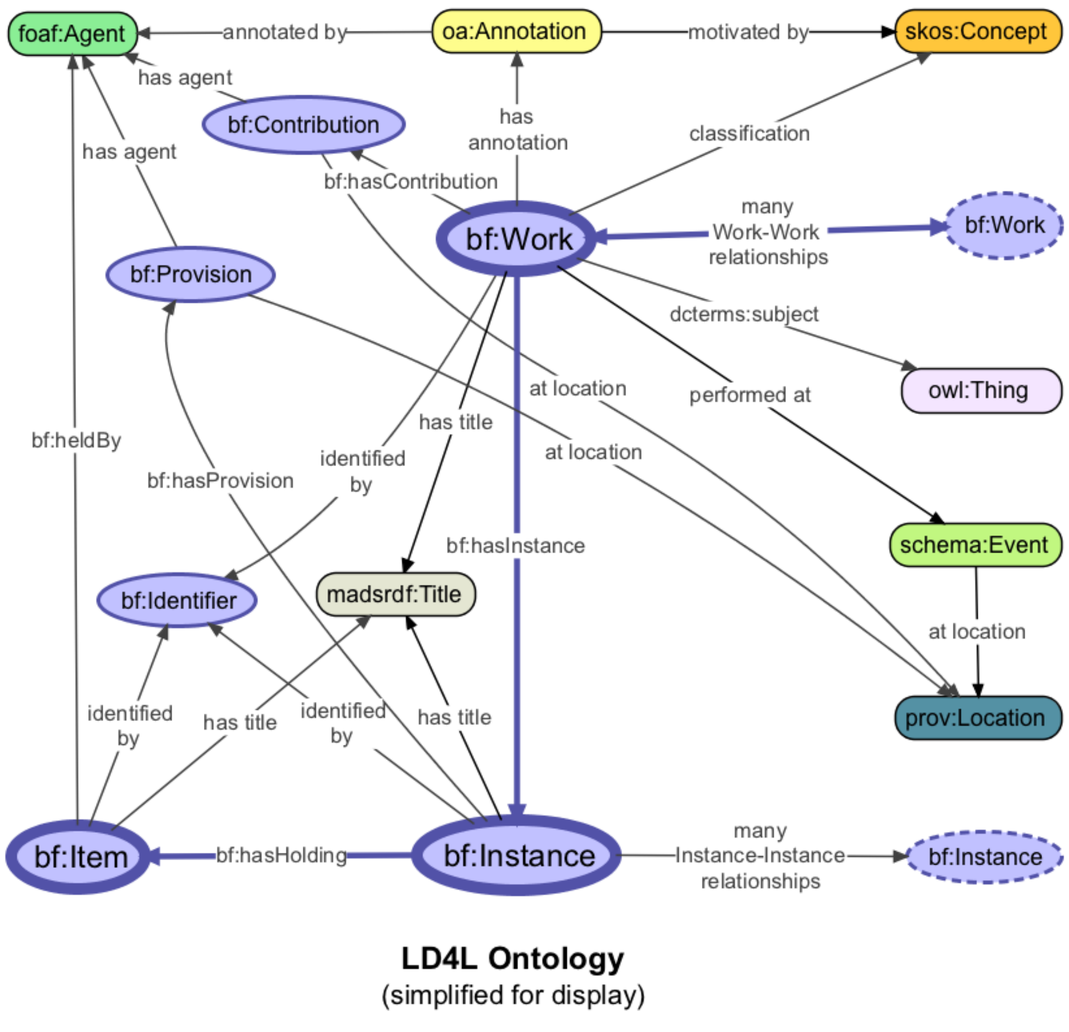
1. Bibliographic + curation data
2. Bibliographic + person data
3. Leveraging external data including authorities
4. Leveraging the deeper graph (via queries or patterns)
5. Leveraging usage data
6. Three-site services, e.g. cross-site search

+

LD4L data sources



+



LD4L Ontology
(simplified for display)

- Convert MARC data for 3 catalogs to LD4L LOD and combine
- Include non-MARC metadata, usage data (Stackscore), collection and annotation data, researcher profile data (VIVO, CAP, FF)
- Reconcile OCLC Work IDs and VIAF IDs across LOD
- Build demonstration search across LOD from 3 libraries representing 29 million scholarly resources
- Demonstrate tool to allow librarians and patrons to create and share virtual collections by tagging and optionally annotating resources
- Model non-MARC metadata from Cornell Hip Hop Flyer collection using MusicBrainz/LinkedBrainz
- Significant feedback to Library of Congress on making BIBFRAME compatible with the LOD cloud and Web, instead of just representing MARC as RDF
- Final public release of LD4L ontology work & code at https://github.com/ld4l

LD4L Workshop Outcomes

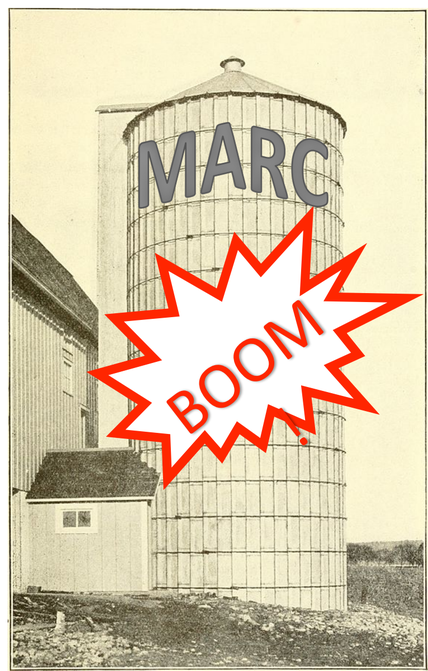
Recommendations

- Our goal should be that others outside the library community use the linked data that we produce
- We must create applications that let people do things they couldn't do before – don't talk about linked data, talk about what we will be able to do
- Local original assertions (new vs. copy cataloging) should use local URIs even when global URIs exist
- Look to LD to bring together physically/organizationally dispersed but related collections
- Libraries must create a critical mass of shared linked data to ensure efficiency and benefit all of us

Identified Challenges

- At this stage, perfection as the enemy of the good (e.g., in ontologies and reconciliation)
- When to mint vs. reference existing URIs
- Wider issues of global entity reconciliation including drift, variable adoption
- Scale – handling billions of triples
- Challenge of ontology design: Small ones don't necessarily work well together, big ones aren't finished
- Leveraging technologies developed for other purposes (e.g., LibGuides, DMS, CAP)
- Thinking outside the bibliographic and record boxes

Exploding library data silos and linking the data to the Web



Acknowledgements. The work described in this poster includes contributions from the entire LD4L project team, including from Cornell: Dean B. Krafft, Jon Corson-Rikert, Brian J. Lowe, E. Lynette Rayle, Rebecca Younes, Simeon Warner, Chew Chiat Naun, Steven Folsom, Jason Kovari, and Jim Blake; from Harvard: David Weinberger, Paul Deschner, Paolo Ciccarese, Jonathan Kennedy, and Randy Stern; and from Stanford: Tom Cramer, Philip Schreur, Rob Sanderson, Lynn McRae, Naomi Dushay, Nancy Lorimer, Darren Weber, and Joshua Greben. The project would like to thank the Andrew W. Mellon Foundation for its generous support of this research.