Agenda for LD4 Workshop 2018

Presentations and meeting notes: bit.ly/LD4notes

Plenary sessions, lunch, and coffee breaks are all held in the Bender Room (Green Library, 5th Floor).
Smaller sessions held in Bender Room, Ida Green conference room (Green Library, 1st floor), and Cubberley 205A

Day 1 (Tuesday, May 1, 2018)

Time	Session description	
8:15 - 9:15 am	*Continental breakfast*	
9:15 - 9:45 am	Plenary session: Introduction to the Workshop *Michelle Futornick, Stanford Univ.* and **Participant Introductions**	
9:45 - 11:00 am	Plenary session: **Communities** *Facilitator: Ray Denenberg, Library of Congress* Experiences and priorities of major communities implementing linked data in libraries: latest activities of the European BIBFRAME community, including work with ILS vendors; role of the Program for Cooperative Cataloging (PCC) in linked data implementation in North America; and report from the RDA (Resource Description and Access) community. • European BF Community / Leif Andresen, Royal Danish Library • PCC / Jennifer Baxmeyer, Princeton Univ., and Amber Billey, Bard College • RDA Community / Gordon Dunsire and Kathy Glennan, RSC	
11:00 - 11:30 am	*Coffee Break*	
11:30 am - 12:45 pm	**Wikidata** **Location: Bender Room** *Facilitator: Simeon Warner, Cornell Univ.* This project is creating linked open data for archival and special collection materials related to Indigenous communities in North America, using Wikidata to incorporate traditional knowledge structures as valid conceptual frameworks with a focus on provisions to respect and defend the agency and authority of individuals, families, and communities to exercise their right to not participate, or have their information used in linked data initiatives. This exploration of the application of Wikidata into linked data environments in research libraries focuses in particular on Wikidata's place as a source for identifiers. Wikidata was chosen because of its easy-to-use interface and its ability to act as a linking hub to other linked data stores. A grounded overview of methods for interacting with Wikidata, especially via the SPARQL endpoint and the Mediawiki action API. A structured conversation and activity. Bring ideas for how you might want to interact with Wikidata. What data might you want to connect, get out, or add to Wikidata?	**Ontology Collaboration and Community** **Location: Ida Green** (up to 20 people) *Facilitator: Nancy Lorimer, Stanford Univ.* The Program for Cooperative Cataloging (PCC) develops and maintains the BIBCO Standard Record (BSR) and the CONSER Standard Record (CSR) Metadata Application Profiles, which guide catalogers in creating authenticated bibliographic records according to PCC and RDA guidelines. Mapping these application profiles to BIBFRAME raises questions about ontology modeling and about mapping maintenance. Kathy Glennan will review the RDA Governance structure, summarize the RSC's communications strategies, and give brief examples of RDA ontology development. Bring ideas about governance, mechanisms, projects & other thoughts about cross-ontology /vocabulary collaboration
12:45 - 1:45 pm	*Lunch*	

1:45 - 3:15 pm	Plenary session: **Implementation Case Studies** *Facilitator: Philip Schreur, Stanford Univ.* This case study of the migration of Northwestern University Libraries' digital image repository to Hyrax/Fedora 4 will cover reconciliation with OpenRefine; mapping among Dublin Core, BibFrame, VRACore3, and schema.org; the use of multiple controlled vocabularies; extending the Hyrax metadata editor; and integration with Primo discovery environment. The Ontario Council of University Libraries (OCUL) Scholars Portal (SP) provides shared technology infrastructure and shared collections for all 21 university libraries in the province. This case study describes SP's project to model journal entitlement and collection metadata as linked data to build links among SP collections and to interact with external sources for enhanced discovery. Topics addressed include modeling decisions and use of existing ontologies; experience with the functionality of the MarkLogic Semantics platform; challenges of integrating data from various sources; and plans for integration with local library systems. This case study describes implementation of a Fedora-based digital asset management system that supports discovery across multiple metadata schemas describing library resources as varied as wine labels, historical photos, and digitized books. For several years now The National Library of Sweden has been developing a new cataloging system based on Linked Data, in order to replace the MARC21-based one in use with something more fit for the needs ahead. The time has now come to go into production. This talk will summarize our current state, the perceived road ahead, and reflect upon the opportunities and risks we've encountered.
3:15 - 3:45 pm	*Coffee Break*

3:45 - 5:00 pm	**Focus on Linked Data Editors** **Location: Bender Room** *Facilitator: Huda Khan, Cornell Univ.* Overview of challenges, successes of using LC Editor in "production cataloging" environment BIBFRAME Pilot 2, posting to a live catalog. Discussion of how profiles help/hurt. Need for good admin metadata about the packages being edited. A quick tour of the CEDAR capabilities applied in the library environment so far, what we are targeting for the sandbox environment, and a vision of interconnected annotations in a future production environment. Plans for integrating the Questioning Authority lookup service with the LD4P2 Editor Sandbox. A major initiative in the forthcoming phase of LD4P is a linked data editor "sandbox". This cloud-based linked data creation platform aims to: give catalogers p ractical experience in creating linked data descriptions; equip libraries to evaluate and choose the appropriate linked data creation tools and workflows for their environment; offer rich use cases and requirements for library system developers; and provide training materials and best practices for a wide audience. This presentation presents early procedural and architectural plans for the sandbox for feedback from attendees.	**Scalable and Shareable Linked Data Descriptions** **Location: Cubberley 205A** (up to 30 people) *Facilitator: Michelle Futornick, Stanford Univ.* What can libraries learn from the Portalwatch project in the Open Government Data domain? Challenges addressed include linking among diverse datasets, improving metadata quality through auto-generation and curation, and agreement on vocabularies and vocabulary usage. The HathiTrust Research Center uses linked data for "layered digital libraries" that support scholarly analysis in ways that traditional digital libraries cannot. With a linked open data model of worksets (coherent collections of digital objects designed to support scholarly inquiries), the HathiTrust will enable scholars to create links to objects and annotations from other digital libraries and linked data services; and to study resources at the granular level of chapters, articles, individual pages, etc. The linked data workset model supports transforming graphs on the fly and retrieving data from multiple digital library triple stores. The presentation will also describe work on adding domain-specific, computed, and extended metadata to existing workset descriptions. This presentation will focus on the idea of distributed scholarly authorities for people and places, and methods for the enrichment of resources through scholarly endeavour by using annotation as a transport and attribution mechanism. These ideas and strategies are demonstrated in Reassembling the Republic of Letters, a 3-year project to create a digital framework for multi-lateral collaboration for the study of Europe's intellectual history from 1500-1800, a period in which the advent of postal services, along with significant population movements, resulted in the emergence of one of the first large-scale social networks. This network played a critical role in the explosion of intellectual activity that led to the Enlightenment and the emergence of a European identity. Both the research corpora and the academic community are, by definition, widely distributed. The project has brought together academics, librarians and IT specialists from over 30 countries to collaborate on methods of enabling scholarship in such a distributed environment. New library systems based on linked open data will still need to provide bridges to older clients that use legacy protocols. Rather than constraining ourselves by those legacy protocols, we must design the new means of access and interchange using widely adopted open standards.	**Follow-on to Case Studies** for further discussion **Location: Ida Green** (up to 20 people) *Facilitator: Philip Schreur*

5:00 - 6:30 pm	*Reception: light hors d'oeuvres and non-alcoholic beverages* **Location: Traitel Fairweather Courtyard**

Day 2 (Wednesday, May 2, 2018)

Rooms reserved for additional sessions if needed: Lathrop 370 all day (20 people); Ida Green after 12 pm (20 people)

Time	Session description
8:15 - 9:00 am	*Continental breakfast*

9:00 - 11:30 am *including coffee break around 10:15 am*	Plenary session: **Ecosystems and Tools** *Facilitators: Sally McCallum, Library of Congress, and Dave Eichmann, Univ. of Iowa* SHARE-VDE is a research and development initiative driven specifically by the library community to facilitate the implementation of BIBFRAME in libraries. This presentation describes the project components including: conversion of over 100 million bibliographic and authority records from 12 North American institutions to BIBFRAME 2.0; reconciliation of entities within the set of converted data, creating a knowledge base of clusters; enrichment of these reconciled clusters with URIs from external sources; and the publication, supply, and management of authority and bibliographical data in RDF. Technical hurdles will be described, along with the solutions adopted, results, feedback, and evidence received from the international library community. UC-Davis is one of 3 institutions piloting a new suite of linked data services using out-of-the-box services from WikiBase and entities available from FAST, VIAF, and Wikidata to 1) reconcile names for people, organizations, concepts, places, and events against an index based on entities, returning language-tagged headings and persistent identifiers; and 2) create, share and edit entity descriptions while also allowing for the contribution of additional contextual relationships between entities, beyond those that can be found by mining structured data in bibliographic and authority data via an Editor. The partnership aims to build new tools and best practices for designing linked data environments for production-level library workflows. FOLIO developers and designers are exploring an approach to metadata management in a LOD environment that focuses on empowering the cataloger, and takes a step away from the raw RDF substrate. The internal architecture of the FOLIO project supports multiple models and approaches to metadata management, with common data models that are heavily inspired by the BIBFRAME entity model. The FOLIO design process focuses on the (librarian) user experience, involving subject matter experts collaborating with user experience specialists to produce functional and beautiful designs. One outcome of the 2017 European BIBFRAME Workshop is a specification for an ILS that supports BIBFRAME. The specification is intended to be used as a basis for evaluating vendor systems and working with vendors to develop ideal ILS components for a linked data production environment. This presentation covers the development of the specification and next steps. Since 2015, MarcEdit has included a linked data framework that has been grown in parallel with the PCC's recommendations related to linked data best practices in library data. This presentation will look at OSUL's use of MarcEdit's link data services, as well as how MarcEdit's rules files related to reconciliation could be reused outside of the application.
11:30 - 11:45 am	*Break*
11:45 am - 12:30 pm	**Caching and Ecosystems** **Location: Bender Room** *Facilitator: Simeon Warner, Cornell Univ.* This presentation describes the architecture of a robust set of authority services that serve linked data on demand, with response times sufficient for integration into an end-user application. The presentation will demonstrate on-the-fly linkage between BibLeo, a sandbox catalog prototype, and DBpedia, both indirectly via VIAF authority data for persons, and directly for works, merging related images and data from DBpedia with local catalog data. In a linked data model, libraries can describe manifestations by pointing to existing work descriptions, without having to duplicate those descriptions, but how will this model accommodate work metadata that will be expanded and improved after initial creation? This presentation will describe a model for caching entities, allowing dynamic updating.	**Application Profiles** **Location: Lathrop 370** (up to 20 people) *Facilitator: Josh Greben, Stanford Univ.* Practical implementation of BIBFRAME requires application profiles to specify how the ontology should be used in practice to create metadata. Cornell University libraries developed application profiles for cataloging audio works using BIBFRAME and related extensions; used the SHACL specification to express the application profile, including nested and other complex interactions; and developed code that allows the SHACL profiles to be used directly as configurations for the VitroLib linked data editor. Expressing application profiles in SHACL makes it possible to share the profiles, providing a single record for the decisions and definitions that have been made for particular types of content, and has the potential for reusing portions of the profiles for other content types and workflows.
12:30 - 1:30 pm	*Lunch*

1:30 - 3:00 pm	Plenary session: **Discovery** *Facilitator: Jason Kovari, Cornell Univ.* What would it be like to explore relationships in our collections through an interface that visualizes those connections? Do novel visualizations help users find things they didn't know they were looking for? Does the familiarity of a world map offer meaningful entry points into a collection? Do we actually want graph data to look like a graph? Harvard Library is creating UIs for exploring moving image materials and geospatial datasets described with BIBFRAME and related ontologies, leveraging relationships between entities in the data (collections, works, people, places, genres, materials) and between the data and external linked open data resources (Geonames, ISNI, FAST) to demonstrate queries that would not be available to end-users in a traditional library discovery environment or possible with string-based metadata. This presentation covers the use of open source javascript libraries to create the visualizations, the process by which the underlying metadata was generated, lessons learned about selecting data sources, challenges of integrating manual reconciliation with automated conversion, and the impact of choosing to reconcile or not reconcile metadata with external entities. Discovery is rarely a single step process. For many searches, library users still seek context and descriptive cues to help them understand, identify and select from the resources returned by their search, even as free-text indexing and ranking algorithms become increasingly sophisticated and powerful. This presentation discusses work at the University of Illinois at Urbana-Champaign to improve discovery in digitized special collections using three approaches: "knowledge cards" constructed with real-time queries of a local triple store and of external LOD services; sidebar descriptions and links fetched in real-time from external LOD services; and the use of Iconclass LOD services to enable multi-lingual keyword searching. These approaches are demonstrated with the Motley Collection of Theatre and Costume Design, the Portraits of Actors collection, the Kolb-Proust Archive, and Emblematica Online. The presentation also covers initial experiments to manually add unique information from these collections to Wikipedia and to create an annotatable visualization of person co-occurrence data mined from Kolb-Proust and made available for retrieval by person identifiers as JSON-LD. The next phase of LD4P will enhance the Blacklight open source search engine to harness the power of linked data. Features that will be developed include "knowledge panels" in search results to present contextual information from open web sources; browsing powered by a combination of authority files and linked data; semantic search to allow discovery of related concepts and to provide richer geographic-based browsing; and adding RDFa, or microdata, to item pages so that search engines easily find library items. This presentation will seek feedback from attendees on plans for these Blacklight enhancements.
3:00 - 3:30 pm	*Break*
3:30 - 4:30 pm	Plenary session: **LD4P2 and Building an LD4 Community** *Philip Schreur and Tom Cramer, Stanford Univ.*
4:30 - 5:00 pm	Closing / wrap-up