ALA Midwinter 2019, Seattle, WA  
Report by Tracey Snyder, Chair, MLA Cataloging and Metadata Committee (CMC)

[Link to Linked Library Data Interest Group](#) (abstracts, bios, and slides available at this link)

Saturday, January 26, 2019

Heather Pretty of Memorial University of Newfoundland gave a presentation about a practical implementation of Linked Data to manage information about Royal Newfoundland Regiment soldiers who fought in World War I, collected by the Trail of the Caribou Research Group. Data points collected by the research group include name, hometown, date of enlistment, date of death, country of burial, cemetery name, and much more. It was decided to use Linked Data for this project because it would enable searching across multiple sources to answer complex questions. For example, a researcher could gather names of and information about soldiers from the same specified hometown who were buried at the same specified cemetery. Heather reviewed Tim Berners-Lee’s four basic rules of Linked Data (Use URIs as names for things; Use HTTP URIs so people can look up those names; When someone looks up a URI, provide useful information; Include links to other URIs, so they can discover more things) and gave an example of a series of RDF triples using URIs for the subject, predicates, and objects. In the example, statements were made about a single subject (a person, John Pretty, with a URI from the Trail of the Caribou Research Group), using predicates from different ontologies (such as "is memorialized at" with a URI from the Muninn Project) and pointing to objects from different sources (such as memorial location, with a URI from DBpedia). This system of RDF statements using URIs, combined with SPARQL (to query the data), RDFS, SKOS, and OWL (to define ontologies and set reasoning logic for the data), allows a computer to infer relationships beyond those explicitly stated in the triples and enrich the original dataset with this new information. For example, after stating in a triple that the John Pretty represented by a URI from the Trail of the Caribou Research Group is the same as the John Pretty represented by a URI from the Muninn Project, one can construct a SPARQL query to retrieve birth, death, and burial information about John Pretty from endpoints such as the Muninn Project and then update the local triplestore with this information. Heather demonstrated using the command line program Apache Jena to conduct the SPARQL query and the web-based Apache Jena Fuseki to update the local triplestore. A possible next step for the project would be to utilize the Muninn Project’s "Graves" ontology (which accounts for cemetery, grave, tombstone, skeleton, and soldier) to update data when new discoveries are made about the identity of previously unidentified or misidentified remains. Heather gave some recommendations for learning about Linked Data (including Steven Miller’s six-week course, Robert Chavez’s series of one-month courses within Library Juice Academy, and a few books) and suggested finding a project that would benefit from Linked Data and learning by doing.

Lucas Mak of Michigan State University gave a presentation about using Linked Open Data to enhance subject access in the library's digital repository by generating subject knowledge cards. MSU’s Islandora repository uses MODS as the default standard and FAST as the default subject system. Although the FAST terms are searchable through indexing and browsable through faceting, the hierarchy of broader, narrower, and related terms could not be browsed. To improve the situation, the library created subject knowledge cards in the digital repository, similar to knowledge cards seen in Google search results, that offer users linked access to three things: the set of broader, narrower, and related subject terms; contextual information on the given topic from Wikidata and DBpedia; and external scholarly resources on the given topic, including JSTOR, PubMed Central, and the library catalog. Broader and related terms in the FAST subject hierarchy are captured from FAST authority RDFXML serialization, and narrower terms, which present a trickier problem, are obtained with a FAST API query looking for terms that name
a given subject heading as a broader term. Contextual information, such as a person's birth date and death date or a country's capital and date of founding, is obtained via a SPARQL query against Wikidata, using LCSH/LCNAF IDs. Linkage to external resources is accomplished using JSTOR topic IDs (for JSTOR), MeSH IDs (for PubMed Central), and FAST URIs (to execute a search for the equivalent LCSH or LCNAF term in the library catalog). AJAX is used for making live queries to APIs and loading subject data from the various sources to generate the knowledge card. Lucas showed an example of an entry in the digital repository for a photograph of fencing at the Auschwitz concentration camp. After each subject heading (for concepts such as concentration camps, Holocaust memorials, buildings, and electric fences), there is an "info" icon. Clicking on the icon brings up the knowledge card pertaining to that particular subject. For example, the knowledge card associated with the subject heading for buildings displays a summary and image from Wikipedia, a list of clickable narrower terms for types of buildings, and links out to results about buildings in JSTOR and the library catalog. The benefits of such a project are that it is low-cost (with no need to maintain a triplestore for cached data), flexible (with the ability to add, update, or remove Wikidata identifiers), and always up-to-date (with current data being pulled from external sources on the fly). There are limitations, too, though. For one thing, a Wikidata entry may not exist for a concept or entity represented in FAST, especially subdivided headings, compound headings, and headings qualified by nationality, ethnic group, or language. Also, there are differences in data modeling between FAST and Wikidata. For example, Wikidata has separate entries for the asparagus genus, species, and vegetable, whereas FAST collapses the genus and species. Another example illustrates a difference in how the two treat a name change; FAST contains separate entries for Michigan State University and Michigan State College, while Wikidata merely mentions Michigan State College and other older names in the single entry for Michigan State University. Information from Wikidata and DBpedia is not guaranteed to be accurate, and sometimes the two even conflict with each other. Finally, speed and reliability in generating the knowledge cards can be affected by response time from the library catalog and maintenance-related downtime on external sites such as DBpedia.

Theodore Gerontakos of the University of Washington in Seattle give a presentation about publishing static library Linked Open Data. Reporter notes are not available for this presentation. However, the slides are quite detailed, and the abstract reads as follows:

We need a model for publishing linked data when we don't have linked data services like content negotiation, a triple store with a friendly user interface, or a SPARQL endpoint. Our publishing model is a step toward establishing a practice for publishing static library linked open data. It requires a method for minting IRIs locally as hash DOIs; care must be taken to ensure all local IRIs dereference and are persistent. The raw data, once scrubbed, is mounted on a web server in multiple serializations, all constellations to the dataset landing page modeled as HTML+RDFa. Appropriate steps are taken to optimize discoverability by search engines, and human users should see the HTML landing page and not just raw data. The data is broken into multiple files based on entities in our data models. Attempts are made to bring the datasets under version control.

PCC at Large

Sunday, January 27, 2019

PCC Chair Xiaoli Li introduced the agenda, which included four items: (1) the implementation of the PCC ISBD punctuation decision; (2) searching on the revised id.loc.gov Linked Data Service webpage; (3) NACO functionality in OCLC’s Record Manager; and (4) the newly official PCC Guidelines for the Application of Relationship Designators in NACO Authority Records. For the first item, Xiaoli reviewed
the history of PCC's work examining the question of removing or limiting ISBD punctuation. In 2011, the PCC ISBD and MARC Task Group issued a report. The group reconvened in 2015 and issued a revised report in 2016. Testing took place in 2018, and PoCo (PCC Policy Committee) announced their decision. Initially, PoCo stated that beginning in spring 2019, PCC libraries could use one of three options when creating or authenticating bibliographic records: (1) continue to use full ISBD punctuation; (2) omit the terminal period in any field; or (3) omit ISBD punctuation between subfields of descriptive fields and omit the terminal period in any field (except terminal periods integral to the data, such as periods after initials). However, further discussion in the past couple of months revealed that the complexity of implementing all three options simultaneously would present more challenges than anticipated. PoCo issued a revised policy, in which terminal periods may only be omitted from descriptive fields in the second and third options. (When headings are controlled in OCLC Connexion, a terminal period is automatically added, and it will take time for OCLC to change this.) Implementation will occur in phases. In phase 1, PCC libraries may begin to use the second option (omit the terminal period in any descriptive field) as early as March 31, 2019. In phase 2, PCC libraries may begin to use the third option (omit ISBD punctuation between subfields and the terminal period of any descriptive field), but an implementation date is yet to be determined. (It will depend on factors such as creation of guidelines, training materials, and tools such as OCLC macros. Note that PCC policy statements for RDA are still frozen during the RDA Toolkit overhaul.) PCC catalogers shall not engage in "editing wars" once the options are implemented. Members of the community may ask questions on the PCC List.

Paul Frank, of the PCC Secretariat at the Library of Congress, announced that some new vocabularies have been added to id.loc.gov as a result of the LC BIBFRAME pilot. Additionally, search results can be refined using some new facets. As an example, a search for Mount Rainier retrieves results from three different sources listed in the Scheme facet--LCNAF (LC Name Authority File), LCSH (LC Subject Headings), and BFEntities Providers (a newly generated collection of provider/publisher names taken from provider statements in bibliographic metadata). Readers can learn more about BFEntities Providers in the report on the LC BIBFRAME Update Forum.

Cynthia Whitacre, Manager, Metadata Policy, OCLC, gave an updated version of her OpCo (Operations Committee) 2018 presentation on NACO functionality in WorldShare Record Manager, the intended successor to Connexion. This functionality was introduced in September 2018, and some U.S. libraries are experimenting with it. Functions include creating new authority records, editing and replacing existing authority records, duplicate record detection, and linking between records (in 5XX and 7XX fields). Users should not be deterred by an apparent validation error that encourages adding $0 to headings; $0 is not required. Besides the 5XX and 7XX linking, some differences from Connexion are that Record Manager uses "roles" and "permissions" (instead of "authorizations") and does not allow macros, constant data, or import of records. NACO libraries wanting to try Record Manager may complete and submit the form online and add a note in the comment section requesting NACO functionality. Demonstration videos are available at: https://www.youtube.com/playlist?list=PLkN3y9CSC9Dx9SPRa2c_R4wBPBVzOGy7U

Judith Cannan and Paul Frank from the Library of Congress praised the work of the PCC group that created the guidelines for using relationship designators in authority records and reviewed the document's appendix, which lists many forward-looking ideas that could not be implemented immediately but are being explored. (For example, a task group is being formed to look at pseudonyms and alternate identities.)

Program for Cooperative Cataloging (PCC) Participants Meeting
Sunday, January 27, 2019

PCC Chair Xiaoli Li presented Past Chair Lori Robare with a certificate and introduced the agenda, which included updates on PCC Strategic Directions, LD4P2 (Linked Data for Production Phase 2), and BIBFRAME profiles, as well as a presentation on the experience of an LD4P2 cohort member. Xiaoli pointed out that the 2018-2021 PCC Strategic Directions document emphasizes (1) transitioning to a Linked Data production environment and (2) ensuring the effectiveness of PCC as an organization. Related to these areas of focus, in 2018, PCC began its partnership with LD4P2 and also formed four new groups (Task Group on Engagement and Broadening the PCC Community, Task Group on Linked Data Best Practices, Task Group on Legal Status, and PCC Communication Board (one-year pilot)). Plans for 2019 include providing training on Library Reference Model, working on training, policy statements, and application profiles for the new RDA Toolkit, implementing the ISBD punctuation policy, and continuing the work of experimenting and training in the LD4P2 Linked Data editor (to be named Sinopia). There will be a PCC 25th anniversary celebration at ALA Annual.

PCC Chair-Elect Jennifer Baxmeyer introduced the Mellon-funded LD4P2 project as a continuation of work done 2016-2018 as part of LD4P Phase 1 and LD4L Labs (Linked Data for Libraries Labs). LD4P2 (2018-2020) will build a pathway for the cataloging community to begin shifting to a Linked Data implementation for resource description, and PCC’s partnership, which includes 17 cohort libraries, is a key component of this. LD4P2’s main work areas relate to the development of a cloud-based cooperative cataloging environment (to be named Sinopia), lookups/linkage to external data (such as Wikidata), enhancement of resource discovery (especially via Blacklight), and the strengthening of an international collaborative community (named LD4, which will host a conference in May 2019). PCC’s work with LD4P2 advances the strategic directions pertaining to Linked Data, identifier creation, and authority management (see SD3, SD4, and SD5). PCC activities to date include attending an in-person meeting in October 2018, sandbox creation (for Sinopia), data conversion by Casalini (for SHARE-VDE), and formation of a profiles working group. Future plans include training for cohort members.

Paul Frank and Les Hawkins from the Library of Congress demonstrated the use of the BIBFRAME Editor for interacting with customized profiles organized by format of resource (such as music audio recording). They also displayed a Work form and an appended Instance form in the Editor and reviewed the Editor’s use of lookups for vocabulary terms from id.loc.gov (such as genre terms from LCGFT) and drop-downs for standard values (such as values for elements like language, audience, and file type). Finally, they showed the use of PMO (Performed Music Ontology), an output of LD4P, for detailed medium of performance information in the profiles for notated music and music audio recordings. The namespace for PMO can be seen in the RDF data resulting from a description of a music resource.

Jacquie Samples, Head, Metadata & Discovery Strategy Department, Duke University Libraries (DUL), discussed Duke’s decision to join the PCC cohort in LD4P2. This work relates to DUL’s strategic directions to advance discovery and transform the information ecosystem, and it has synergy with other DUL initiatives such as FOLIO development, TRLN (Triangle Research Libraries Network) Shared Discovery Services (using Blacklight), and Ivy Plus Borrow Direct Shared-Index. DUL formed a Duke LD4P Core Group, which includes the AUL for technical services, the head of the resource description department, the principal cataloger, and Jacquie. Jacquie presented the work as being in phases. Phase 1 is the initiation of authority control modernization, in which entities can be matched and linked and data is retrieved from external sources such as ISNI. In Phase 2, efforts center on communication, training, and metadata creation.