

Linked Data in Library Services: Transforming the Library Catalog to Linked Data

Semantic Web in Libraries 2015
23th – 25th November 2015
Hamburg, Germany

Myung-Ja Han
Metadata Librarian
University of Illinois Library
mhan3@illinois.edu

As memory institutions, libraries have unique and rich collections that are not readily available for users outside of each institution. Semantic web and linked data technologies now allow libraries to publish their data to the web, to create relationships with other available information resources so that users outside of the library can find and use library collections, and provide additional contextual information available from the web for users inside of the library. However, best practices and technologies for this transformation have not been established yet in the community. The Semantic Web in Libraries (SWIB) conference provides a place for conversations on linked data experimentations and programs in libraries; and for librarians and developers to share codes and ideas, and form working groups to look at common issues together. Thanks to the Bibliothek & Information International (BI-International)'s Travel Grant, I attended the SWIB 2015 conference, held in Hamburg, Germany.

This year's SWIB conference is divided into four areas: Applications, Research, LODLAM, and Metadata, providing real implementation examples, actual methods to transform and work with linked data, and best practices when working with traditional library catalog records. All codes presented in this year's conference are published as open data in GitHub pages and can be found in the conference program page (<http://swib.org/swib15/programme.html>). For this report, I will reorganize the program into three categories: Linked Data in Production, Linked Data in Experimentation, and Understanding Library Metadata.

Linked Data in Production

The three keynotes delivered at the conference shared challenges and outcomes of their linked data projects, including how they transformed their data to linked data for production services.

The first keynote by Asunción Gómez Pérez titled *Maximising (Re)Usability of Library metadata using Linked Data* demonstrated what linked data can do in real time search and browse services that are currently available at the National Library of Spain (<http://datos.bne.es/>), based on 3LD (Linguistic Linked Licensed Data). By exploiting 3LD, including BabelNet (<http://linghub.lider-project.eu/datahub/babelnet>), which covers 272 languages, the beta version of the library site supports multi-lingual search and discovery services to users who want to find resources in and by specific languages. In addition, it uses the Lexicon Model for Ontologies (<http://lemon-model.net/>) to enable the representation of relationships between resources.

This keynote also provided detailed information in three areas:

1. How to transform traditional library catalog records in MACHine Readable Cataloging (MARC) format to linked data. The presentation showed step-by-step workflows and the people involved in the transformation processes as an example for other libraries that work with MARC format data. The National Library of Spain (and the research group) created its own ontology based on different entities, e.g., name and work (<http://datos.bne.es/def/ontology.html#/def>), with detailed descriptions about each property, and made it available for other libraries. They also capture all these processes in provenance metadata using PROV-O (<http://www.w3.org/TR/prov-o/>) to document the workflow.
2. How to license linked data. Not all linked data is ‘Open,’ but we tend to forget to add rights information into the linked data. The presentation showed how many data in the linked data cloud have a license (or rights statement), and stressed the importance of having linked data licenses using standard languages, such as Open Digital Rights (<https://www.w3.org/community/odrl/>).
3. How to integrate other linked data sources with library linked data. The experimentation also included real linked data services (in my opinion), e.g., exposing available information resources on the web, such as geographic data, to local users, that improve the overall user experience.

The second keynote titled *The Digital Cavemen of Linked Lascaux* addressed a refreshing and also important topic librarians must examine; why it is necessary to have a long-term sustainable plan for linked data and the semantic web. Ruben Verborgh said, in this presentation, that we have to prepare for changes in three areas; data models, technology and interfaces, and should establish shared meanings including URIs and ontologies, in effort to make things easier and simpler when working with linked data.

The third keynote was *Data-Transformation on historical data using the RDF Data Cube Vocabulary* by Sebastian Bayerl and Michael Granitzer, which showcased the data transformation process from messy data in excel spreadsheets to linked data. It dealt with issues surrounding languages, hierarchical structure, and inconsistency of data from digitized sources, clearly demonstrating how messy data can be and the challenges that we may encounter and have to work with.

Linked Data in Experimentation

Many libraries, if not all, have been experimenting with linked data options in library services, and knowing how others are doing is always helpful to understand the trends and new developments that also make local experimentation better and more productive. There were several presentations in this year’s SWIB 2015 that shared linked data experiments.

The talk on *Linked Data for Libraries: Experiments between Cornell, Harvard and Stanford* provided actual library linked data experimentation work being carried out by three

institutions using the Library of Congress' ontology BibFrame (<http://bibframe.org/>) and other ontologies and services including VIVO (<http://vivoweb.org/>). The group created various use cases, and developed their own ontology (on top of BibFrame) and tools for libraries to create their records as linked data. Although there are not a lot of outcomes to share at this stage, it will be interesting to follow the processes, discussions and outcomes during the next couple of years.

Modeling and exchanging annotations for Europeana projects outlined how Europeana is experimenting with an annotation service on its site. It is known that user control for annotation services, implementing different types of annotations, and allowing for the capture and use of annotations are all difficult tasks to engineer. This talk explained the importance of annotation in user engagement and services, and discussed data modelling and use cases to help cultural heritage institutions understand and implement annotation services.

Other presentations, including *ALIADA, an Open Source Solution to Easily Publish Linked Data of Libraries and Museums*, *Metadata Records & RDF: Validation, record scope, state, and the statement-centric model*, and *A RESTful JSON-LD Architecture for Unraveling Hidden References to Research Data*, provided more technically focused experiments and solutions for linked data transformation and services that would aid developers in their local implementation projects.

Understanding Metadata

As a Metadata Librarian, I found two presentations especially interesting: Karen Coyle's *Mistakes Have Been Made* and Valentine Charles's *Evaluation of metadata enrichment practices in digital libraries: steps towards better data enrichments*.

Library professionals are all familiar with the concept of FRBR and accept it as something we have to understand and use in data creation and discovery services without knowing exactly how the concept was created and what it is for. Karen Coyle's talk was provoking in that sense – she reminded audiences that FRBR describes a conceptual model that was developed to understand how users use data, and not a technology that we suppose to use in data creation. She assured that we, as librarians, have to have a new mindset to look at and understand library data, and create a data model that really meets users' needs in combination with the right technology. Although the question of 'How To' is still not clear, it was an impactful talk that made me rethink what FRBR is and how we should work with library data. Because it will be hard to develop a good (and new) data model that works for library catalog data and meet users' needs without understanding how it was developed and for what purpose.

While Karen's talk asked us to rethink the way we look at library data, Valentine's talk provided us step-by-step guidelines for enriching library data with linked data sources. Oftentimes, we focus only on the 'enrichment' part, rather than why we do it and how. She asked us to identify why we enrich the data and for what user services we would accomplish

with the enriched data. While discussing lessons learned from Europeana's data enrichment experiences, she first asked audiences whether the semantic enrichment process really ensures better records, then warned metadata professionals that the automatic enrichment process may harm data that are perfectly fine otherwise. The talk also added that incorrectly enriched data may lead to devaluation of curated metadata and loss of trust from providers and users, ultimately causing irrelevant search results and bad user experiences. Among the proposed recommendations three recommendations, defining the goals of enrichment, testing the workflows, and have the evaluation strategy seem the advice relevant to everyone who are doing the metadata reconciliation work.

Building Networked Resources

On top of regular programs, this year's SWIB offered a new experimental session called breakout sessions that were designed to provide opportunities for all participants to meet people with the same interests and discuss specific topics. There were four topics posted before and during the conference, and I attended the RDF Application Profile Validation session. As many libraries develop their own RDF based application profiles, we need to have a good tool that checks the constraints of vocabularies so that we can have valid RDF data. Although the vocabularies and processes involved in validation seem very technically challenging, this was a really informative session since our institution is also developing an RDF based application profile to work with linked data and the semantic web.

Semantic web and linked data discussions are not always easy to follow and keep up with, as one presenter mentioned, since new technologies are constantly developed and new ontologies are created without announcements. As with last time, the SWIB conference provided more information than I can digest and look at in just two days, but it also provided me with associates whom I can contact and look to when I have questions on linked data projects. I look forward to following up with these people and projects.