

Notes from Linked Library Data Unconference

American Library Association Conference, June, 2010

A group of about 50 persons gathered for a half day at the ALA meeting in Washington, DC, to discuss linked library data. These are the notes from the discussions that took place at that meeting.

Contents

[Application Profiles](#)

[Literacy and Training](#)

[ONIX Data as Linked Data](#)

[FRBR + Linked Data](#)

[Use Cases](#)

[Authorities and Linked Data](#)

Discussion of Application Profiles

Although this discussion began with APs, it soon became clear that the topic was broader, including (or primarily) dealing with the question of how do you determine what vocabularies you can use, and when it is necessary to create properties of your own.

We talked about the importance of DEFINITION and CONTEXT in the re-use of defined elements.

1. you should only re-use an element if your use is compatible with the definition of the element
2. in determining if your use is compatible, you need to look at the context in which the element is defined and used. This is important because formal definitions may not fully express the usage of the property, so looking at the larger context can fill in information relating to the property's meaning
3. 3) is the data model of the vocabulary compatible with your usage?

You may also want to determine the possible SUSTAINABILITY of the schema before making use of it. Is the schema being currently maintained? Is there an active user group?

Then we talked about what you do if you do not find a property that fits your need. In particular, we talked about two possibilities:

1. there is a vocabulary that fits your needs (at least to use some of its data elements) but it has not been registered or defined in a linked-data compatible way
2. there is no vocabulary that has a property that meets your need

In the case of 1), it may be necessary for you to define a property under your domain that represents the un-registered property. In this case, it is best to create a separate name space under your domain that makes it clear that this is a "borrowed" property. If the property is later defined by its owner, you can create a link between the two definitions that declares them as "equivalent" (e.g. owl:sameAs)

In the case of 2), if you can find a broader property that has been defined, you can define your property as a sub-property of a defined property, thereby connecting your vocabulary to an existing one. As an example, if you need to express a journal title, you may be able to define your property as a sub-property of dc:title

```
dc:title
  journalTitle: subproperty of dc:title
```

Why Create APs?

APs have (at least) two major functions:

1. communication: an AP expresses your metadata realm, and makes it possible for you to communicate this to others
 2. validation: an AP allows you to validate that data meets your requirements
-

Literacy and Training

Notes from Unconference session on literacy and training

Where are we aiming our efforts?

- Working librarians
 - Catalogers and other staff
 - Administrators and decision-makers (effects on work practices and markets for products)
 - IT staff
- Library schools
 - Next generation
- Vendors
 - Want to be responsive to users and current clients
 - Reluctant to invest without strong interest in products

Curricula considerations

- Information ecology
 - Info seeking behavior
 - Expectations/mis-matches with current practices
- Principles vs. implementation – a need to re-evaluate our efforts
- Granularity consensus broken
- Focus on demographic and learning curve

Structures

- Understanding the content models, encoding issues, presentation (very fuzzy now)

Clarify future vision

(Central to all demographic groups)

Address the tyranny of the record

Changing our focus from records to statements

Essentials

- Foundations of RDF
- Vocabularies and vocabulary development -- Not necessarily RDA

Tools

- Management strategies for statements
- Tools to support the new environment and principles + Outputâ€™MARC as legacy exchange
- Support users â€™both staff and end users

Onix Data as Linked Data

The first issue we discussed was that in our current model we tend to create metadata that is static. Data, however, is constantly changing and we are very poor about reflecting this. Linked data gives us a chance to make use of dynamic data that more accurately reflects the real world. The web has made our users much more accepting of constant change.

This led to a discussion of the use of linked data in our ILS. Most (all?) are not currently able to make use of it. If we want to make use of linked data in a library environment it will have to be outside the ILS. Stanford is interested in making use of linked data for a digital map project. Because we will be working outside of the ILS, we cannot put the controlled headings under authority control. Linked data will allow us to do this as the headings will be dynamically updated. It was pointed out that a simple link to data would not be enough. At some point, you would need to capture the data for indexing, faster display, preservation, etc.

Is there an incentive for publishers to share their ONIX data? It's a by-product of their business but they make it for internal use. OCLC has developed a project in which they receive ONIX data from publishers to enhance catalog records in Worldcat and in return publishers receive quality "work" information that they can use to enhance their metadata. Both sides win.

Do we really want all the ONIX data a publisher creates? They start creating it for the book at a concept level and it grows until production, and then post production with reviews etc. Often, the metadata is very poor by our standards because it's not meant to do the same things. If publishers are aware, however, that people can make use of this data to sell books, they will be more motivated to curate it. When do we tap into the stream? Do we want it all and all it to grow in our discovery environments?

Will linked data change the way catalogers do their work? Can we really on linked data to create a basic descriptive record? In a world of links, we should no longer need to create unique text

strings to identify entities. This will allow, for instance, name headings to be registered (in VIAF?) for all those millions of names we cannot put through the NACO process. Our focus will shift from the hand crafting of individual bibliographic records for individual items to an evolving cluster of links that dynamically draws in data as an item evolves with time.

We ended our discussion talking about the importance of a testbed for developing ideas, some place that we can get in and play. The complete record examples in RDA would be interesting to use, also OCLC's place as an aggregator of information could be very important.

Also, we need to develop the apps that will sell the concept of linked data. It's the classic chicken and the egg. Perhaps by just exposing the data the apps will be developed? Linked data is like the web before MOSAIC. There is some practical ONIX data that would be of great use to use currently: TOCs, author bios, reviews etc. but if we could load them into our ILS how would we display them? Would they make the bibliographic display unusable?

The last issue we discussed was the persistence of data particularly when dealing with publishers - they don't commit, they go out of business or are absorbed - We need to think of preservation of all this data as well. Should we become the archivers? Could we afford this?

FRBR + Linked Data

The group introduced themselves and mentioned their interest in joining the discussion, responses included:

- E-R modeling with Parts and Aggregates, particularly with respect to:
 - Linkages at different levels or to different parts;
 - Handling "gaps" in data
- E-R modeling with respect to Serials
- Relating manifestations, e.g. the "multiple versions" "debate"
- How do we transition to another format?, that:
 - Allows multi-tiered records;
 - Eliminates data redundancies that then need to be kept in synch.
- Standards? Specifically with respect to the hierarchical nature of library data
- The transition - deconstructing current data structures to have "hooks" to which to link.
- How to bring this vision/future practices to the "simple, country cataloger"?
 - Will s/he bridge the gap, only when compelled to?
 - Will s/he actually be able to bridge the gap, when compelled to?

Further questions, thoughts, and ideas were floated:

- What would FRBR linked data look like? Would it rely on identifiers?
- How far do we go in deconstructing records into new linked records or linked data? - How do we set the parameters?, How do we set the level of detail? How do we incorporate a "mash-up" of data from different sources?
- A system that doesn't require people to learn linked data? (i.e., that's what is needed, but is

it possible to develop?)

- Should we/Can we use "Good reads" as a model for such a system?
- It should employ a variety of "fill in the box" interfaces with the models underlying it, but with the models not "in your face".
- Persistent data should be present to facilitate use of the interface *
- How get old/existing data into the new model?
 - Employ various approaches/algorithms
 - How reliable will the results be?
 - How to manage our past practices, that in cases does not support current data needs and/or elements?
 - How to manage inconsistencies in where data was "parked" in the past (e.g. VHS in 300 or 538, "widescreen" in 250, 500, etc.)
- Need better control of some elements that are currently in uncontrolled forms in notes*
- Should we use authority control as a starting point?*
- Historically, it has been "name control"
- Contrast "name control" with VIAF approach
- Form of name has been used as "identifier" vs. use of an actual identifier number
- Authority control systems still rely on string matching, rather than use of identifier numbers
- Application profiles as solutions to the "form of name" problem?
- What about when people draw "lines" between entities in different places?
- Opening linked data to "non-expert", "non-library" community(ies)?
 - By definition, linked data is open
 - Solution - we control internally what data we link to
- *"Tokens" instead of text
 - e.g. the RDA registries
 - a wonderful consequence of linked data!

* -- retrospective linkage between the last idea and some related concerns mentioned earlier in the discussion

Use Cases Discussion Group

This combined two proposed break-out topics:

- Linked Data Use Cases for Scholars (non-library uses)
- Usage of existing linked library data sources (id.loc.gov, viaf, etc)

Use Cases generally fell into two categories, which might be labeled "inward" and "outward" facing.

Conversation started around discussion of existing usage statistics:

- LC just started collecting stats and aren't seeing much use [1]
- OCLC's viaf seeing a doubling of resolution of URIs each month:
 - 30,000 "303" redirects in June
 - This begged the question "Why?"

Discussion of Value Proposition of Linked Data

- Europeana Whitepaper mentioned [2]
- Discussion of "Rationalizing Serendipity" as being the primary use case for scholarship
- Scholars finding each others works, related works, looking at interdisciplinary collaboration

- Cornell/Florida "Vivo" project to expose university faculty on the web [3]
- Driven largely by Grant-Tracking needs?
- Was also subject of a session on Sunday morning (along with id.loc.gov)

Discussion of general use case to expose our authorities for disambiguation

- NY Times has approached both LC and OCLC to discuss using linked names and subjects in their own linked data apps.

Use case for matching and controlling names in dissertations

In general, largest needs seem to be for working with Names of People and Places (Geo-coordinates?)

- IETF has a URI Scheme for Places [4]

Generic Use Case for pulling cross-references from existing vocabularies into non-ILS/non-MARC indexes

Generic Use Case for connecting Names (via, eg, Vivo) to Topics (areas of interest) using id.loc

- Note that this is in part what WC Identities is doing, though not via Linked Data

Brief discussion of Intellectual Property Applications

Use case re: Exposing Library Vocab to the web for purposes of Search Engine Optimization

- Enhance participation in programs like Google's "Rich Snippets" and Facebook's "Open Graph" [5],[6]

Quick Win: Putting RDFa in OPACs - Google has indicated they would be interested in seeing this

In Summary:

* Inward facing use cases are typically about supply chain issues, improving re-use of data and aggregating search

* Outward facing use cases are harder to discuss but allow other users to leverage linked data sources from the library world.

[1]<http://bit.ly/id-loc-gov-chart>

[2]<http://version1.europeana.eu/web/europeana-project/whitepapers>

[3]<http://vivoweb.org/>

[4]<http://tools.ietf.org/html/rfc5870>

[5]http://code.google.com/apis/customsearch/docs/snippets.html#structured_data

[6]<http://developers.facebook.com/docs/opengraph>

Authorities and Linked Data

Names:

- LC has done work on modeling MARC name authorities; looking at parsing out pieces
- LC using MADS for that and doing MADS into RDF transformation; working with Antoine Isaac in the Netherlands. Will add transliteration, language, script to MADS.
- With VIAF, there was an attempt to find existing ontologies. Not enough overlap with FOAF; SKOS a bit better, want to extend it. VIAF shows multiple preferred forms and multiple alternate forms (4xx); how do we differentiate? VIAF simplifies data - we thought punctuation wouldn't be important, but if data is displayed, it could be needed.
- Corporate bodies and conferences are difficult. A month ago Dave Reynolds published his

work on an organizational ontology

(<http://www.epimorphics.com/public/vocabulary/org.html>) which extends FOAF. String matching could be problematic because of AACR2 rules on what to include/leave out.

- Do we need a record (URI?) for the label, for the person as person, and for the person as concept in SKOS?
- Relationships - we don't have a way in authority files to show relation between, e.g. James Billington and Library of Congress. [Linked data could be useful here?]

Subjects:

- LCSH - LC is working on modeling relationships between concepts.
- How granular does this need to be? Do we need a key and separate element for everything? Do we need language in a uniform title? When we move a step away from the strings, how will we reassemble the data?
- There are choices (predicate or object?); it would help to know how people are going to use the data. What is the promise here?

General:

- We seem to spend a lot of time trying to model the existing data to the new data model. Some think we should try to model where we need to go. Model a top level, relationships, where we need to go? LC subjects - many users don't know how to use them and it's a vocabulary unique to us (libraries) - it'd be great if we could integrate more with (the world)... LC has some ideas, but thinks we need to figure out how to model what we have, before trying to do something else. A lot of hard work is involved. Both VIAF and LC linked data work has already led to going beyond "modeling MARC".
- A lot of discussion about Wikipedia/DBpedia and other sources. RDA expands the kinds of data that would appear in authorities. Could we "pull in" data to authorities from Wikipedia etc.? or, point to it? or, use it to make inferences? How do licenses fit into the picture of what we want to do? Should we be contributing data from authorities to Wikipedia?
- How do we outreach and share information? Where do we post this stuff? Google Docs? Library organization called Metro? <http://www.metro.org/> - they have licensed LibGuides