

The Semantic Web, Authority Control, and You

MCLS Linked Data Users Group Virtual Meeting, 1 May 2018


Abstract: The semantic web community figured out what libraries have known for a century and a half: access and discovery requires authority control. Library data is ideal for implementation in a linked data environment because catalogers have long relied on controlled vocabularies for subject terms and authorized files for names of persons, corporations, etc. I'll briefly highlight the work of the Program for Cooperative Cataloging (PCC) Task Group on Identity Management in NACO, which is charged with lowering barriers for participating in the Name Authority Cooperative (NACO) and thus enabling more libraries to add authority records to the Library of Congress Name Authority File.

Jennifer A. Liss is Head of the Image Cataloging team at Indiana University, Bloomington. She has served on and chaired several national standards committees. Jennifer will be speaking today about her work with the PCC Task Group on Identity Management in NACO.

The semantic web authority control and you

MCLS Linked Data Users Group
1 May 2018 | Virtual meeting






Hello!

I'm Jennifer Liss
Head, Image Cataloging, Indiana University, Bloomington
jaliss@indiana.edu

2

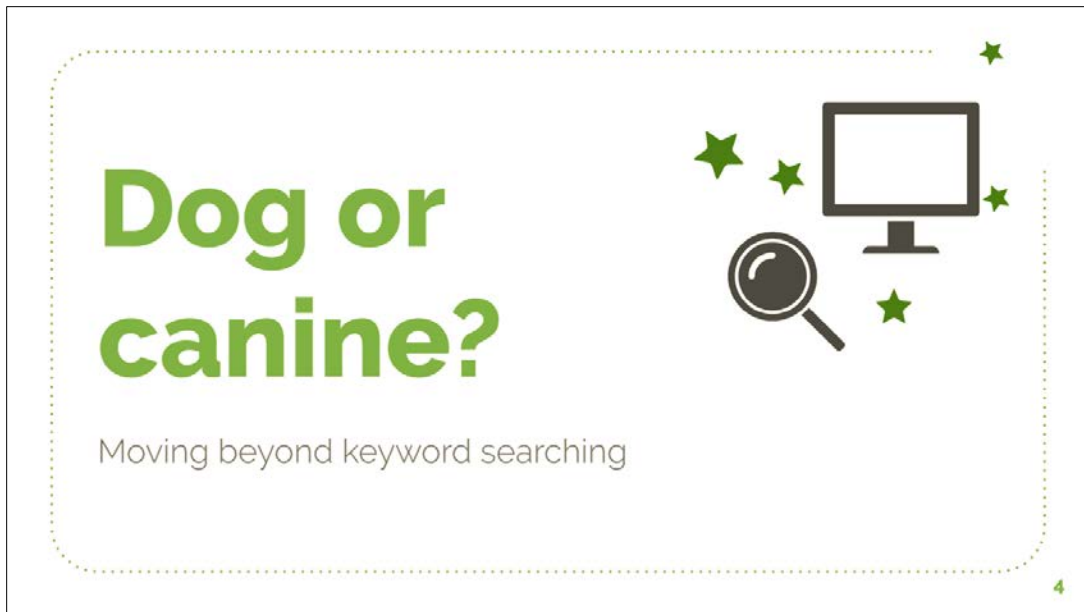
Before I get into what the Task Group is up to, I want to provide a little context.



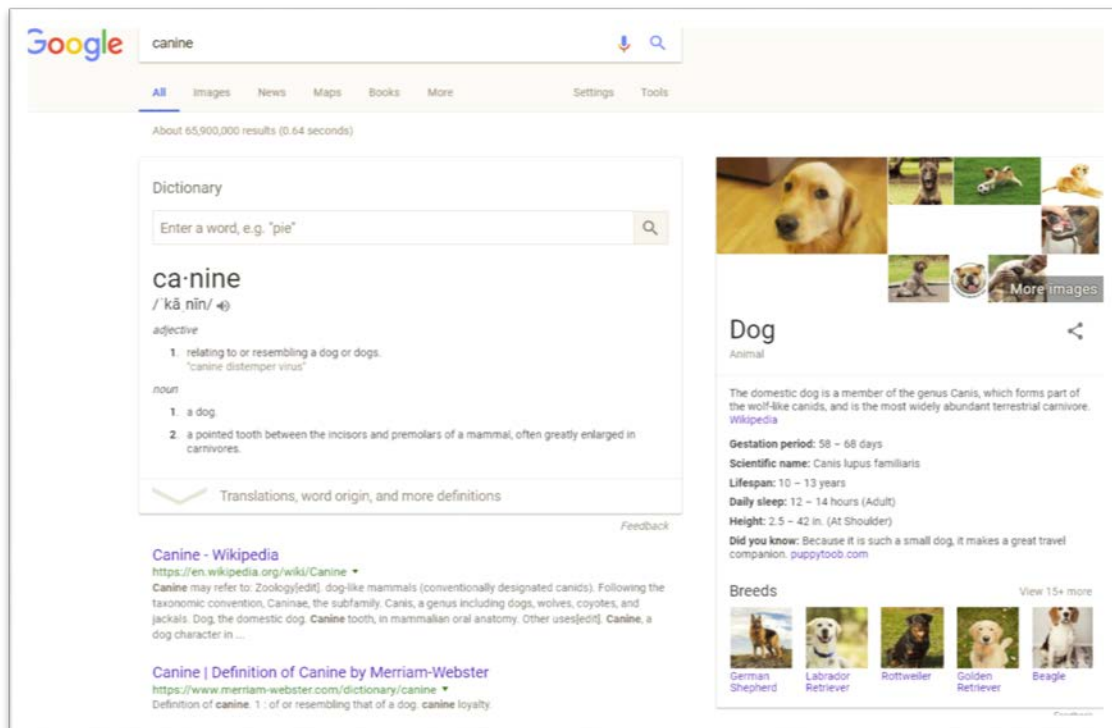
Why linked data?

Keyword search vs. semantic search

To address the question that perhaps brought everyone here today: why linked data? Why not just keep the web we have? You may have noticed that the web is little disorderly. Think back to what web searching was like 8-10 years ago.




In a keyword search environment, you needed a thesaurus on hand in order to effectively use search engines because search engines didn't know that "canine" was another word for "dog". If you searched "dog," you wouldn't find the webpage with the word "canine" on it. Happily, search engines have become increasingly aware of the semantic web in the last five or six years.



Google's algorithms now know that "canine" is another word for "dog"—that [Google Knowledge Graph](#) card on the right is powered by the semantic web, even if all of the webpage titles in the search results list contain the word "canine".

In library cataloging-speak, we would say that the web is a huge, complex system of information that is not under any sort of authority control.




**Linked data is
authority control for
the semantic web.**

5

Linked data, then, is essentially authority control for the semantic web. This is great news for libraries! It means the World Wide Web is getting serious. Now they're playing on our court and we've had a century and a half to master this game.

The not so great news: across the nation, libraries are hemorrhaging the talent needed to create metadata records that utilize controlled vocabularies and authorized forms of names.



~9,000
OCLC members with full-level cataloging authorization


~700
Libraries creating/editing LC authority records

~10%
Libraries contributing authority records to OCLC database

6

Looking at just the OCLC database, there are about 9,000 libraries with full-level cataloging authorizations. There are 700 libraries with the ability to create and edit Library of Congress name authority records. That means that the current system of policies that are in place today results in roughly 10% of those libraries doing LC name authority work for the entire database

of English-language records. This amounts to a significant capacity problem if we wish to express all of our bibliographic and authority data as linked data. So what do we do?



PCC Task Group on Identity Management in NACO

- Strings (headings/AAPs) → Things (URIs)
- Explore a "NACO Lite" option

Read the charge: bit.ly/pccidNACO

7

This is where the PCC Task Group on Identity Management in NACO comes in. In late 2016, the group was charged to do two major things in preparation for linked data: identify pathways for transitioning authority control work from a focus on creating authorized headings (i.e., "strings" of text) to minting identifiers (or URIs or "things"). The group was also charged to explore what folks are calling a "NACO Lite" workflow. The Name Authority Cooperative Program (NACO) is the program through which libraries can add or edit authority records in the LCNAF. The idea is that adopting linked data processes will greatly simplify NACO authority work, since emphasis is being shifted away from formulating the text string—which honestly, is the hardest part of learning to do NACO—to minting identifiers.





What's identity management?

A rebranding of authority control?

What is identity management? In some ways, identity management is a bit like a re-branding of authority control for a linked data context. However, identity management definitely won't work the same way that authority control works now.

Authority control


Manages access to entities by **authorizing a unique form of name** for each entity.





9

Authority control manages access to entities (personal and corporate body names) by authorizing a unique form of name for each entity.

Authority control



Rowling, J. K.





10

Authority control's emphasizes distinguishing each entity with a unique string of text, or heading.

Identity management

Process of associating a registered identifier with a single entity.



11

Identity management is the process of associating a registered identifier (or a URI) with a single entity.

Identity management



<http://id.loc.gov/authorities/names/n97108433>



12

Unlike in authority control, the differentiation of names or headings is of secondary importance in identity management.

At present, the Task Group is figuring out what a name authority record created in a NACO Lite workflow might look like. What is the absolute minimum we need in an authority record in order to disambiguate? Will policy changes trigger new MARC coding? Do the records produced in a NACO Lite program even have to be in MARC? How and where do we launch a pilot? There is lots more to sort out but the Task Group plans to communicate progress regularly.



*Metadata is a love
note to the future.*

—Jason Scott, Internet Archive

13

To wrap up I thought I'd share this quote by Jason Scott. Cataloging work has always been expensive. Cataloging is a [long tail venture](#): the benefits of cataloging compound over time and value is realized as the number of linkages grow. Linked library data is not venture for the shortsighted. Being serious about linked data means being serious about cataloging—and that is the message that we need to take to our administrators.



Contact the Task Group

John Riemer (UCLA), chair

jriemer@library.ucla.edu

14

If you'd like to get in touch with the PCC Task Group on Identity Management in NACO, you may [contact the chair](#), John Riemer.



Thanks!

Continue the conversation-

Email jaliss@indiana.edu or Twitter [@cursedstorm](https://twitter.com/cursedstorm)

15

I'm especially interested in helping advance discussions around these ideas—[reach out through email!](#) I'm happy to take questions after this session. Thank you!