What to Expect When You’re Analyzing, Transforming, and Inputting: A Linked Data Guide

Music Library Association
Thursday, February 1, 2018
Portland, Oregon
Now that we’ve heard about analyzing data and transforming data, I’ll be addressing the “inputting” part of the program session title. The LD4P, or Linked Data for Production, and LD4L, or Linked Data for Libraries, multi-institution grant projects have been covered elsewhere, including at this morning’s Cataloging and Metadata Town Hall, so I will jump right in discussing one piece of the much larger whole. In a sentence, Cornell has a very large collection of LPs and other audio discs of various genres — pop, rock, disco, soul, funk, etc. — that were used in the early formation of hip-hop, and we have selected a subset of LPs for which our catalogers will be inputting native Linked Data using an ontology that we have developed and an RDF editor that we are developing. The work of developing the editor has experienced delays due to unforeseen circumstances, so what I will show you today is a glimpse at the in-process tool rather than a finished product.
The key players in this work include, in no particular order, catalogers, cataloging coordinators, developers, metadata librarians, and ontologists. Contributions include the following:

Metadata librarians and ontologists create a demonstration ontology based on BIBFRAME, called bibliotek-o. (It’s Esperanto for library.)

Catalogers show developers the current MARC cataloging process and environment. Cataloging coordinators supply developers with documentation, templates, and screencasts.

Developers conduct user studies with catalogers and cataloging coordinators.

Developers iteratively improve the tool based on catalogers’ and cataloging coordinators’ input/feedback and the ontology.

Metadata librarian creates an application profile for audio based on catalogers’ and cataloging coordinators’ input using SHACL (Shapes Constraint Language, a W3C specification).

Also, catalogers and cataloging coordinators give input on what elements from authority records are needed for disambiguation on a search results screen when choosing among authorized names and other headings, such as a person’s field or occupation, without having to click on each one to find out.
We all have different areas of expertise; there is no single person among us who can do it all — catalog music, write an ontology, create an inputting tool and application profile — at least not to the level that is needed for this project. So it is a lengthy process requiring lots of meetings, demos, tests, recordings, collaboration, and documentation.
You may be familiar with VIVO, a platform originally developed at Cornell for sharing the scholarship of Cornell faculty members, librarians, and other researchers in a structured way, as seen here, with tabs across a person’s VIVO page for different categories of information. Several other institutions have also implemented VIVO.
Our instance has been re-engineered and re-branded as Scholars@Cornell. The software underlying VIVO and Scholars@Cornell, called Vitro, was the starting point for the development of the present Linked Data cataloging tool called...
... VitroLib. Early iterations closely resembled VIVO in look and layout, as you can see here, with information from across a bibliographic description tucked away in numerous tabs. This arrangement was not conducive to entering data pertaining to a bibliographic resource and proofreading it, so...
...the developers worked up a new prototype, which has tested better with catalogers, that generally collects data at the work level and the instance level — that's the publication in hand — on two separate screens that are linked.
Some nice features are the drop-down menu of preloaded terms from the ontology for roles or types of activity, informed by RDA relationship designators, and integrated lookups that access name entries in the name authority file and subject and genre entries in LCSH and LCGFT respectively. Thanks to the application profile created for this collection, the list of terms for a person’s or group's role in a work is limited to terms that we felt would be applicable to a musical recording, such as composer, performer, etc. Notice also that the screen where one creates a new audio instance includes fields for types of numerical identifiers that we determined to be the most relevant for these audio recordings, including issue number and matrix number.
Don't worry if you can't read this easily, but this is a small excerpt from the application profile that designates our pre-selected activity terms for the interface’s drop-down box of a person’s or group's role in an audio work. There are more that I didn't include in this slide.

Here’s an example of a partially described audio work, Grab them cakes. The tabs from left to right represent different property groups determined in the application profile, such as the titles group,...
...the agent and capture information group, and so on. All the way to the right is the View All tab, where the cataloger can proofread all of the statements that have been made about the work and do final editing.
Here in the View All tab, you can see the title statements, activity statements, and more, including a link to an instance of this audio work.
Here is the view of the audio instance, which has its own property groups determined in the application profile, including measurements, sound characteristics, and so on.
<table>
<thead>
<tr>
<th>Nick Cappadona</th>
<th>Jason Kovari</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roswitha Clark</td>
<td>Margaret Nichols</td>
</tr>
<tr>
<td>Steven Folsom</td>
<td>Lynette Rayle</td>
</tr>
<tr>
<td>Beth Kelly</td>
<td>Tracey Snyder</td>
</tr>
<tr>
<td>Huda Khan</td>
<td>Rebecca Younes</td>
</tr>
</tbody>
</table>

In summary, work continues on the application profile for music recordings and on all facets of the VitroLib tool, and we look forward to doing full inputting for the selected LPs in the near future. I want to acknowledge the work that has been done and is being done by various people at Cornell, listed here selectively and in alphabetical order, and mention that...
Huda and Lynette have presented on their tool development work and Jason, Steven, and Rebecca have presented on their ontology work, most recently at SWIB in December 2017. Their presentations are online and may hold the answers to technical or philosophical questions that you might have, but feel free to ask me and I’ll try to answer them or route questions to others as appropriate.

- SWIB (Semantic Web in Libraries)
  
  http://swib.org/swib17/programme.html

  Huda Khan & Lynette Rayle
  Steven Folsom & Jason Kovari & Rebecca Younes

- DLF (Digital Library Federation)
  
  https://osf.io/view/forum2017/

  Huda Khan