2015-09-24 Hydra Connect unconference meeting notes:
Best Practices for Descriptive Metadata

Date
24 Sep 2015

Attendees

- Jacob Glenn
- Carolyn Hansen
- Karen Estlund
- Unknown User (escowles@ucsd.edu)
- ... more people I don't know (please add your name)

Goals

- Review results of Hydra community survey on descriptive metadata
- Discuss minimal requirements for descriptive metadata in Hydra applications
- Create action items for engaging the broader Hydra community on this topic

Action items

- [ ] Distribute results of metadata survey more broadly within Hydra community
- [ ] Choose a venue for sharing use cases around descriptive metadata
- [ ] Send recommendations from this session to Hydra email lists and request input
- [ ] Add authorities recommendations to the priorities backlog for the Metadata WG (on Confluence)

Discussion

(N.B. My writeup makes it appear as though the discussion was dominated by 2 or 3 people. In fact I just didn't get everyone's name. Please add your name to the list of attendees if you were involved in this discussion.)

Background: Survey of Hydra/Fedora community about descriptive MD practice - ~25 respondents

28% favor RDF

Need to make sure there is no conflict w/ Hydra In A Box

No Standards - the hill you go to die on

Fedora 3 or 4? RDF or XML? This shouldn't be about requiring people to use a particular format or encoding - just say what elements are required

Esme & Karen had a goal to include metadata recommendations with every Hydra installation, at least basic vocabs & fields - is this crazy?

Yale wanted every object to have 5-7 baseline properties. They were trying to roll a bunch of little repos into one, dealing with a variety of standards and lots of dirty metadata. They required MODS initially, this worked for 6 months then blew up (catalogers revolted).

Karen: We could pick from DPLA application profile and implementers could use MODS or whatever they prefer. Work with DPLA to determine which fields should be included.

Yale was trying to define minimal metadata for discovery. Recommend allowing people to pick their own vocabularies.

Karen: But we need guidance on how to pick from among multiple vocabs in Hydra. New adopters are looking from guidance from the Hydra community on which vocabularies & which elements to use. One thing folks like about ContentDM: built-in vocabs. Need a good set of defaults.

The survey results indicated a variety of use cases. We need a good way of sharing these use cases - Github? Previous feedback indicates a preference to have use cases in Confluence instead. A lot of people are doing similar things, but working in isolation. Pick a venue to collaborate.

Another outcome of the survey: Lots of local field usage. Do we want to worry about weird local usage? If so, where do you host URIs? Does anyone else want to use your weird local fields? We thus come back to the issue of a linked data vocabulary hosting service...

Karen: We want people to use URIs for metadata; people need their local fields; therefore we need a service for them to host their URIs.

Esme: Duraspace is already hosting some vocabularies and exposing as HTML, shouldn't be too hard to add more.

Q: What were Yale's five fields? Title, author, ISO date, subject, format, content type (optional). They did some user studies to determine the most useful content types (a local field).
Q: How does Yale handle related objects? This is more like structural metadata, it's done with locally created tools. There is some bleed between descriptive and structural metadata.

Q: Hydra in a box... any ideas what the metadata requirements will be? It's early to say... they're just starting to think about it. This question is very closely tied to what DPLA is doing. The Stanford metadata unit is evolving, filling some openings - working on consistency among digital collections. Thinking about sharing their own metadata templates - are other institutions doing this? (Oregon used to but has largely stopped, mainly just to clean up public web pages.)

Emory does this - site just launched:

http://metadata.emory.edu

Penn State would like to publish their metadata info for other librarians.

Karen: Since there's time, let's talk metadata elements! Which ones do we want to see? Suggestions from around the room: Title, identifier, rights, rightsholder, content type, ISO date, genre (optional), author/creator and place (if applicable).

Q: Language? (Yale included this is their extended 12-element set.)

U. Alberta used a subset of DCterms in migrating their IR to Hydra: Type, language, title, ID, author, creator, subject, keyword, license. IR content doesn't always fit a standard publishing model - publication date can mean just about anything, for example. And intellectually, not everything has a title or a label.

Q: How important is content type?

Q: If we decide on any recommendations in this session, should we send them to the Hydra email lists? (Yes)

Should there be recommendations about how to do authority control in Hydra? (Suggestion: roll this into the recommendations deliverable.)

Esme: Is this a thing for the Applied Linked Data group? Some people seem to think it's out of scope?

Should we tell everyone to suck it up and use linked authorities? There are many smaller institutions without the resources to do this.

General consensus is that this stuff is a better fit for the metadata WG. The foregoing recommendations to be added to metadata priorities backlog on Confluence.

"Questioning authority" gem doesn't include the work at Oregon on giving catalogers a better interface to use linked authorities. (Respondents to the survey largely did not use the gem.)

Karen: Using the QA gem presents a huge burden for full-time catalogers. Not so bad for lighter authorities work. Trey Terrell worked on an improved interface before leaving Oregon, it's not really finished. It begins to address the cataloging interface problem - hopefully someone can pick it up.

Q: Do we want to recommend vocabularies or just elements? "It should have a title" or "it should have a DC:title"? They're very different requirements.

Esme: If the main goal is interop, as long as someone (DPLA?) can ingest it (e.g. MODS) then it should be fine. The task is to figure out a minimal field set, aiming for a bit less than Dublin Core.