MODS and RDF Call 2015-08-24

Time: 9am PDT / Noon EDT

Call-In Info: 712-775-7035 (Access Code: 960009)

Homework Reminder: Fill out survey! http://goo.gl/forms/GzHtKI2tXk (Results: MODS Survey #1 Results:)

Moderator: Steven Anderson (Boston Public Library)

Primary Notetaker: Rebecca Fraimow (raw etherpad notes link: http://etherpad.wikimedia.org/p/RDF-MODS-20150824)

Attendees:

- Steven Anderson (BPL)
- Danny Pucci (BPL)
- Christina Harlow (University of Tennessee)
- Unknown User (steve@northwestern.edu) (Northwestern University)
- Simon O'Riordan (Emory University)
- Rebecca Fraimow (WGBH)
- Shawn Averkamp (NYPL)
- Sara Rubinow (NYPL)
- Kelcy Shepherd (Amherst College)
- Karen D. Miller (Northwestern University)

Notes:

1. Skipping September 7th meeting due to Labor Day Holiday. Next meeting would be 9/21st (during Hydra Connect but hopefully still works).

2. MODS Survey #1 Results:
   a. majority responders in the survey said that they preferred to do simplest mapping and more options for more complex cases (with some minor fidelity loss); no objections.
   b. majority responders said they did care if they could reconstruct MODS from the mapped RDF
      i. one person from the comments section: unnecessary to create because you can simply store the old MODS; however, at BPL, going forward, would prefer not to create two records and would like to have RDF as the primary format rather than having to keep both
   c. which mapping held most promise: most people voted themselves, Indiana University had two votes; anyone willing to volunteer reasons for that? Rebecca from WGBH: liked that they mapped to more complex and simple options, but didn't have strong feelings between Indiana and BPL.
      i. Kelsey: really liked the BPL model for complex, but for the simple a lot of people went the same way with dc:terms, so not much to choose between
      ii. Danny: question about approach moving forward; Indiana first took XML and transformed it using LoC RDF transformation; do we want to keep looking at that? LoC transformation has the blank node problem, and some of that is apparent in the transformation. BPL, trying to find an alternative; just curious if we want to keep looking at LoC
         1. Christina: LoC is building a new ontology, but that will still use blank nodes, which are valid in OWL but don't work well in Fedora 4.
         2. Does continuing to look at that make our job easier or harder?
            a. Right now it makes it harder, creates kind of a wacky output, just meant to get the ball rolling; was hoping Mods RDF would follow Bibframe and that Bibframe would be further along; LoC is hoping that an ontology for v2 will still do the Mods RDF mapping and that other institutions will create use cases; how realistic that is, we will see, but that's the hope. It's not in a state now where it can really help guide, but it might be too difficult or too much of a distraction.
            b. Rebecca: at least worth making it interoperable.
         3. Christina: Hope is that what comes out of this will influence what comes out of Mods RDF from LoC; two versions show that there are a lot of modeling changes they're considering even from the base and we'll continue to see how other people handle it in the future, but the process right now is such that it's flexible enough to embrace what comes out of the Hydra group.
   d. minting of local URIs -- majority are not against having to do that, which opens up additional options, since many of the most simple options don't support local URIs very much.
   e. freehand survey responses: 'what did you / didn't like about the models?'
      i. Summary of each comment was read.
      ii. Anybody who didn't comment want to voice support for comments? no.
   f. last section: general comments, recs, and ideas; does anybody have anything else they want to say about this section?
      i. Kelsey is interested in the idea of acceptance criteria, though not sure what that might look like, but having a general idea of what the expectations for each element are and doing the mapping to meet certain requirements.
      ii. Q: how could we do acceptance criteria?
         1. element-by-element basis, have institutions mention what elements they are not willing to give up?
         2. Danny: from the title element perspective: BPL were willing to give up non-sort and subtitle, thought about that as they were doing the work, so maybe before we make homework for a particular element, as a group say, these are the things represented here, what are we willing to lose.
            a. more complex model had minted solutions for the title to make sure they lost less going from different titleTypes, so maybe talking about those things before we do the homework next time would help us figure out what is acceptable loss
         3. Shawn: suggested the acceptance criteria, doesn't have an idea of what it would look like, likes Danny's suggestion
         4. Danny: maybe we should think about our examples and how to map them in two separate stages, in the first meeting talk about what we all bring, and in the next meeting discuss the mappings. Spend first half of each meeting talking about homework, next half of each meeting talking about the next element, what's acceptable, etc.

iii. Steve (Northwestern): about mapping back to MODS, problem with keeping separate bitstream is that then you have your metadata in two different places, hard to keep updated, so you do want to map back to MODS, but also thinking if you find it too difficult to completely map all the RDF metadata back to complete MODS, maybe there's a compromise to map just the elements that are really necessary to map to MODS. Maybe restricting ourselves too much by making reverse mapping a particular requirement.

    1. Steven: You don't want to have to maintain your metadata in two places (MODS XML and RDF). The "fidelity" should have been decided in the "acceptance criteria" test as we should be able to make MODS that represent the data points we agreed we all cared about.

    2. Kelsey: It depends on how much you want to break elements down, how granular you want to be. At Amherst, created MODS in a granular way, but that's not required for valid MODS. A MODS name could have one child element that holds the entire string. So how much of the granular detail needs to be retained, if MODS records get turned into something else in Fedora do the original MODS need to be retained; holy grail is to not have to manage multiple formats of the same thing. There are a lot of different things that mapping back to MODS could mean, could be done in a very granular, complex way, or it could be done in a more simplified way and still be valid MODS.

    3. Some of that answer comes from how useful to the user the granularity of MODS is. For example, in the 'creator' example, is it useful to parse it all out into component parts, does that add any value?

    4. End result: depends on what we're willing to give up in different MODS elements; what we're willing to lose could be granularity, or could be whole elements or attributes; maybe also goes back to the acceptance criteria of what is important to port over to begin with

iv. Are there acceptance criteria that need to be carried through across all elements?

    1. Christina: how to handle authority in elements?

    2. Steve: General requirement that it work well with Fedora 4 and the blank nodes problem is something we want to avoid, and things like that, could be something that we don't know yet, will cause problems in Fedora 4, and maybe some things with linked data that could come up as well, but we won't know what those are until we run into them.

v. general thoughts on anything else? was the survey useful?

    1. Rebecca thinks it was useful

3. Going back to the MODS: Title element discussion:

a. acceptance testing -- what is the minimal level that we're going to support for the basic example for those who just want to do the simplest option?

    i. base level: you need support for a main title

    ii. Q: should minimum also support for alternative titles, or anything else besides a main title?

        1. A: After some discussion, yes. The "alternative title" would become a dumping ground for all other title types then the main title.

    iii. what are people willing to give up and not willing to give up?

        1. BPL: willing to give up non-sort and subtitle elements, but definitely want to have support for telling between supplied, uniform, translated, and parallel titles.

        2. Karen: from a cataloger's perspective, important to keep subtitle as well, such as music titles where the main title is the same for everything (i.e. 'Fugue'); lot of name-title combinations that are exactly the same

            a. BPL: suggesting concatenating subtitle with title; ideally you don't lose any values at all, just give up granularity of how they are split

                i. Kelsey agrees with this approach -- ideally don't lose any content that's in the record, having it concatenated is a different story.

            b. Danny: depends on if you're doing anything else in the display that requires that breakdown.

        3. Steven (BPL): is anybody doing anything with the nonSort, subtitle, partNumber, partName elements that actually requires the breakout? Any actual use cases?

            a. finding aids example: specify a sorting title and non-sorting title so that everything doesn't sort under 'Guide to,' etc. Could be useful to display it with non-sort and store it dropping off non-sort.

            b. Steven (BPL): MODS-RDF v.2 thought about what to do with non-sort, and it sounds like normally you would want to not drop Guide, so would you need to have support for two values, one that is the normal title and then one that is a non-sort version of the title?

            c. Q: Would it be unusable without this, or is it just a nice to have?

                i. In the case of EAD, it's nice to have; don't want to drop it off the title completely, because then people will think it's the whole papers available online, not just the Guide, but it comes down to, it's a nice to have thing. In the MODS example of non-sort, even more nice to have rather than having to have it.

            d. Should then the more complex mapping have requirement that we have support for a sortable title, or is that just a nice to have that we should not worry about?

                i. is there a problem with using a delimiter to break out the non-sort?

                    1. In another use case, could get messy very quickly.

            e. Sounds like we would have to have a title and then a sortable string version of the title in the end.

        4. Other distinctions from MODS that people want to keep? Amherst had display label; would this be painful to lose?

            a. Amherst could live without that.

        5. BPL would also like to keep the URI to the location where the uniform title came from (some institutions only use the label); objection to having support for the URI value also available?

            a. No objections raised.

b. Steven (BPL): Should we start talking about what that looks like on the call, or start a share document with two columns, take a group effort in having the minimal and an example of the complex w/all the elements we want to support in the complex? Or is there a better way to continue from this point?

    i. Someone: are you suggesting we do this independently and come back?

        1. Steven (BPL): No, now suggesting a collaborative document where we work out a minimum and a maximum way of doing things, more collaborative work instead of each institution doing it. For minimal, dc:title and dc:alternative work fine, but we need to work out how to do the more complex example; we know the elements we need to support and what the minimum example is that we want to be able to build off of for that.

    ii. Christina: would prefer that we all work together collaboratively, build a framework to handle elements beyond that; general agreement on this.

    iii. General approach for the future, have people come up with what they want to keep and what they want to give up (acceptance level), try their own mapping. At the next meeting, then move onto the group document from that discussion and seeing what people tried to do.

    c. Any objection to using dc:title and dc:alternative for the basic mapping?
        i. No objections.
        ii. Steven (BPL) will start a document to list out examples of the more complex, and we can collaboratively start fleshing out how that would look.
    d. Christina: with the long break, should we start work on the next element as part of our homework?
        i. Steven (BPL): Sure. Anyone have a suggestion for what element we should focus on?
            1. next element on the top-level elements list would be Name, but this one is likely to be particularly thorny, so do we want to skip it for now?
            2. skip over name and go to the next non-authority element
            3. type of resource, language, and note are three easy ones, so maybe try to hit all those? avoids ones with authorities and ones with lots of child elements and lots of questions about what to pull out
            4. or should we pick one with authorities and use it as an example to how to work with authorities on other elements?
        ii. In the end, we decided to do "name" for the next element after all. Similar to the approach of how we did title except with the addition of what is acceptable fidelity loss.

4. Next meeting: September 21st at 9:00 AM PST / Noon EST.