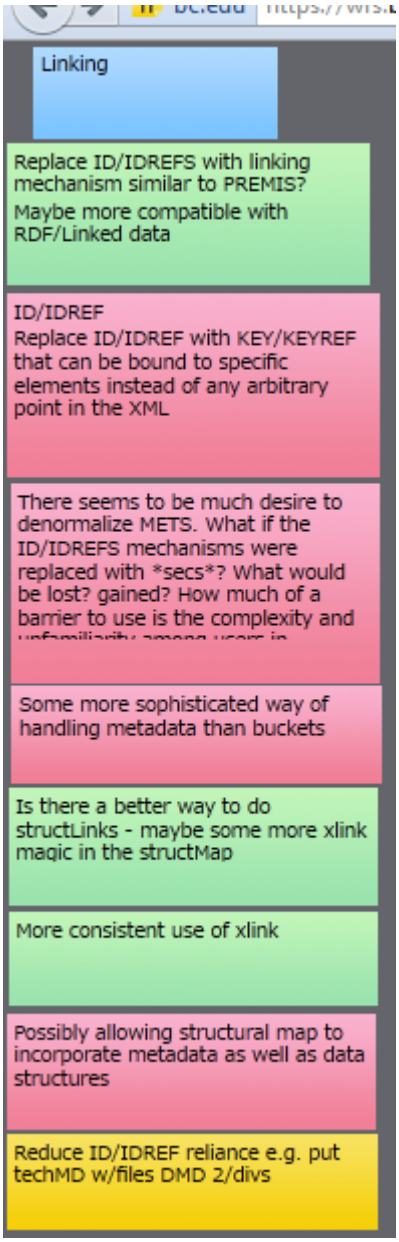
Thursday, February 9, 2012 Mets Board Teleconference

Attending: Terry Catapano, Thomas Habing, Jukka Kervinen, Betsy McKelvey (recording), Sébastien Peyrard, Leah Prescott, Jenn Riley, Tobias Steinke, Brian Tingle

Contents

Linking (From Affinity Diagram)



ID/IDREF

ID/IDREF are a common theme

- KEY/KEYREF were discussed as a possible alternative to tighten up the kinds of linking that are possible. However, there was general consensus that KEY/KEYREF option is not worth pursuing

- Linked Sections as Children

  Terry suggested that the desire to get rid of ID/IDREf is a desire for **de-normalization.** A possible solution is to allow for both normalized and de-normalized versions of METS. In the de-normalized version, users could represent data that would have been in a linked section as children of the structMap divs.

- PREMIS Model

  Tom raised the possibility of making linking in METS similar to PREMIS linking. PREMIS has a wrapper which is called *identifier*. There are four entities (agent, rights, object, event) and each one has an *identifier* element that wraps two sub-elements: *identifier type* and *value*. *Identifier* is repeatable – so every section within PREMIS can have multiple *identifiers* with different *types.* When you need to link things together, for an example, an *event* has to be linked to an *agent*, *event* would have an *identifier* with two sub-elements that refer back to the *agent*. This mechanism is used it consistently throughout PREMIS schema (which, for better or worse also supports ID/IDREF linking.)

  Effectively this mechanism is replacing a series of attribute values with a series of child elements giving you the ability to type the links. Also, you can

have alternate identifiers and use anyone of those identifiers to create the relationship.

PREMIS type linking addresses issue of ID/IDREFS – vagueness of the purpose of destination metadata. It would be easy to migrate existing instances – you could automatically create identifier elements.

Quite often in PREMIS, the identifier type is local. Embedding PREMIS in METS is common – advantages to unifying the linking mechanisms.

- XLink Discussion

  Another alternative in diagram is to more fully embrace XLink. Pet peeve is XLink supports sophisticated capabilities but here we use it in a very limited way . . . so what's the point of using it. We only use a couple of the attributes – how much does it complicate the schema to import it?

  On the xlink Wikipedia page were listed as one of the only implementers . . .

  We've talked about wanting to allow more extensibility – so if we built extensibility points into a new schema, that's where people could mix and match other elements from other schemas. For example, we could systematically allow arbitrary attributes from other namespaces on almost any element in the next version METS schema. Then people could use xml:id and XLink to do all kinds of creative linking. It's probably opening Pandora's box a bit, so we'd want to provide guidelines.

  Opening more doors to extensibility is a good idea – it focuses what METS is for. Within the core of what we want to do we must not invent other wheels – when things meet our needs we should bring them in rather than make our own versions in the METS namespace.

  A concern was expressed that using XLINK makes it more complicated than its worth to get two attributes – href and structLink. And we're interdependent with other schemas when we report it. For this case does complexity bring any value?

  Agreement that we should make use of stuff from other namespaces when appropriate, but we need to do a complexity analysis – for our limited use of XLink may not be appropriate.

  Also, the decision to use another namespace involves analysis of take-up in other communities. A decision on XLink may go in direction we shouldn't be importing, but the general approach should be to use to reuse when appropriate. *If* we were going to fully embrace XLINK it would be a good approach.

Tom could envision replacing structlinks with something based on XLink that would allow us not to only have hierarchical structures, but to have arbitrary graphs of relationships between objects and you could implement that by fully embracing XLink standard. But since there's not been much uptake of xlink we may not be buying anything by doing in that. Not any built in xlink processors out there in the world and so people are still reliant on traditional xml parsing with xpath and so forth.

Action Items:
Explore multiple options before making really big change. Not a strong consensus on what to do, so examples will help.

1. Tom will mock up a METS example using PREMIS linking mechanisms
2. Betsy will create an example that fully embraces XLink
3. Someone else should do a KEY/KEYREF example.

More Compatibility with RDF/Linked Data

This goal is compatible with our evolutionary approach. We should be taking steps now to be friendly to that process. A canonical procedure for turning METS into triples is desirable.

- GRDDL possibility

  GRDDL is a technique for obtaining RDF data from XML documents. It would be interesting to create a Gridl mapping of METS into RDF. GRDDL allows you to embed an attribute into an xml document that points to an xslt stylesheet. The purpose of xslt is transform xml into rdf. There other mappings as well.

  To put it another way, GRDDL allows an instance to say what it authorizes in an RDF representation of itself. It indicates what the authorized graph is to derive from a given document.

  If you are an author of a particular METS document you could embed a GRDDL saying how you want your METS to be transformed. As the METS board we could choose to embed GRDDL into a schema document to define our official mapping of METS into RDF.

- ID/IDREF as a Barrier to RDF mapping

  Sébastien suggested that a problem with mapping METS to RDF is that we have internal ID/IDREF links and in RDF those links are actually typed.

This takes us back to the PREMIS discussion. In PREMIS you can declare link entities, for example, between an event and an agent. You can have the agent fire the linking identifier type and linking identifier value and then you can have linking agent role indicating that this agent it plays a role of issuer towards this event or role of performer, etc. You can also use the mechanism of identifier type and value for the relationships between objects to define structural relationships, saying, for example, it is *part of*, *has parts,* or is a *derivation* indicating that this object is derived from this other one.

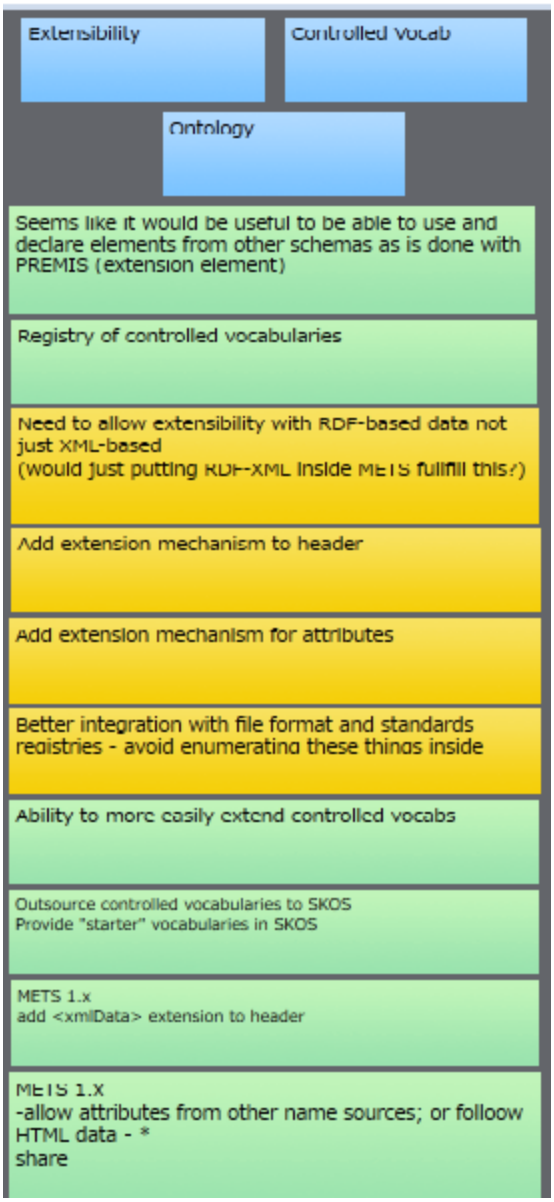
Tom agreed with Sébastien, referring back to the PREMIS presentation he gave at our Fall board meeting. What Sébastien is saying gets at how Illinois is using PREMIS to model relationships between compound objects and it may make very good sense to adopt a PREMIS-like linking mechanism for METS for 2.0

Controlled Vocabularies and Linking Relationships and RDF Friendliness

Use of controlled vocabularies to control relationships using PREMIS-like linking would be helpful. Maybe we could use controlled vocabularies to make METS more RDF friendly. For example, we could declare endorsed METS Board controlled vocabularies for attributes such file group *use*, the div *type* or the structural map *type*. Then a particular implementer could choose to use something different or to re-use it or to re-use something that is more specific but that would be sub-classed to the METS board endorsed vocabulary. Doing this sort of thing moves us towards being more compatible with linked data.

We want to be sure that METS structural info can be represented in RDF. Controlled vocabulary might not get us 100% of the way there, but would help us because we can type the link. Vocabulary is helpful tool, but will not solve all the problems.

Controlled Vocabulary Column



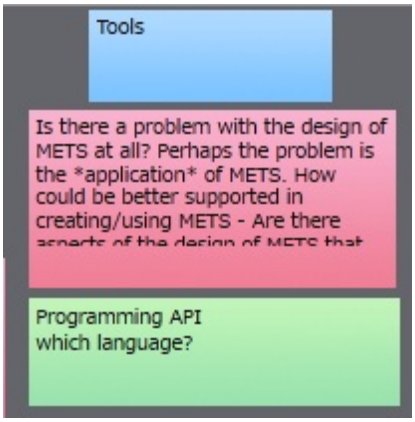
Concentrating work on the controlled vocabulary column is an easy place to start. The items in this column are more concrete than the items in some of the other columns. If we focused here we could accomplish more than in other columns where the options are vaguer.

The difficulty with controlled vocabs is that we would have to choose what values to adopt. There are already some enumerations in the schema dtd. We could look through all the values/vocabs in the METS registered profiles. Or we could re-use already published id.loc.gov vocabularies.

Brian asked whether our goal is to try to modularize vocabularies so they are outsourced from the schema? Terry indicated that, a good *starting point* is to express enumerations formally. For example, taking *use* attribute values and express them as a controlled vocabulary would be a good thing to accomplish.

Tom suggested thinking about MODS subject terms would help. Attributes allow users to point to the vocabulary they are using.

Tools Column



There was a sense that the tools column should be set aside for the time being. However, there is considerable interest in using schematron to make METS profiles more actionable in terms of validation.

- Sébastien wondered if as far as METS profiles are concerned we could delegate all the testing to other mechanisms whose jobs are to validate and automatically test a METS instance. Perhaps we could use METS profiles for the documentation purposes and then delegate all the automated testing work to another tool, for example, schematron. This is an idea that came out of experiences at BNF. Maybe that's a way for us to go with METS profiles
- Tom noted that Sébastien had proposed this in a January 14th this e-mail and Tom thought it was a good idea. The 2.0 profile has a placeholder for xpath validation you can turn into schematron.
- Terry would love to see us do work in the area of delegating rules checking to schematron. We could start with making available certain rules that are widely applicable like ID/IDREF checking. We could easily publish a schematron with rules that could be imported by local schematron instances. There are a number of generic rules that could be expressed and promulgated.
- Sébastien noted that if you use only schematron to express formal actionable rules, you could make a very simple xslt that would turn your METS profile into a machine actionable one. Another possibility, would to provide a validation service on the METS website which would allow users to validate a METS instance against a profile. This could be used to support exchange as discussed in the last meeting.
- Brian: this is a good idea, but running something on LC's site might be hard. Regardless of what we decide on this, we should establish a METS git hub group for stuff like this. Sébastien could provide some content and act as a tester. Tom could create aregistry of already developed schematrons. We could provide provide a service to build your own schematron from scratch.
- The METS profile listing (in the registry right now) should indicate which profiles have the validation capabilities – Tom could possibly accommodate this in new style sheet he's working on.
- This discussion is leading to making profiles more actionable in terms of validation. The schematron would be part of the "profile package" which included – (1) a description and (2) a schematron part. A METS profile

should perhaps have an appendix that contains a schematron – this might be easier than embedded profile.

- Possibilities – add another appendix to profile to hold not just an example but also include a processing or validation code and it can be referenced by individual items within the profile. Rather than taking each schematron role and embedding it inside constraints describe a constraint and then have a pointer to the actual rule in the appendix.
- Brian can grab *METS* on github (https://github.com/mets), people should send him their accounts.

Action Items

1. Establish github presence – send Brian your accounts
2. PREMIS linking – Tom will send his proposal
3. Linking – Betsy will work on xlink example
4. Controlled vocabularies
   - Terry: Start some work here, for example a SKOS expression of enumerations for fileSec USE? Do this by mining profiles. We could also be adventurous and go out on a limb and define div types.
   - Tom: I think that would be a useful task if someone wanted to take that on. How is it going to have a direct impact to METS 2.0? I think developing SKOS ontologies for controlled vocabs currently in use and mining profiles is a good idea, but I consider this a separate path from METS 2.0 profile work.
   - Jenn I'm unclear how a METS 2.0 instance doc would reference those linked data controlled vocabs and I think a little work is needed exactly what that would mean.
   - Terry: one simple way be for div type, anyone can already put a uri in the value of type, so people could already start using those values in a document and having uri there might make it easier to generate rdf graph from struct map.
     We would want to come up with uris for all controlled vocabs so we could de-reference the uri and get info back. That's the linked data principle at hand. It does involve a service component so you need to make it possible to dereference these. You can do that by having a static SKOS document. You don't need to go do the full treatment like the LC does.

   - Sebastien: I may be able to start some listing of all the places in the schema such as div type, etc, and come back with something on the wiki on this so it can fire up the work on the controlled vocabs.

Updates

PREMIS 3.0

Sébastien: Work continues on PREMIS 3.0. The main difference in PREMIS 3.0 is that you will be able to pull environment out of an object so that an environment relating to an object may be described separately and the description can be shared. It's a big change so it becomes a new entity.

Sébastien is unsure how this fits in with the PREMIS in METS document. For the moment we still don't have the big picture – it's complicated. Once we have a definite evolution of the data dictionary the question of where environment would go in METS it is one he will think about –it could be related through techMd, digiProv, or behaviorSec: a lot of possibilities to think about!

BNF updated METS PROFILES

Sébastien will send these to Tom.