

Thursday, December 1, 2011 Mets Board Teleconference

Attending Board Members: Brian Tingle, Betsy McKelvey (recording), Jenn Riley, Jukka Kervinen, Nancy Hoebelheinrich, Richard Gartner, Robin Wendler, Sébastien Peyrard, Terry Catapano, Thomas Habing, Tobias Steinke

Contents:

1. [Swedish National Archives METS 1.10 Change Requests](#)
2. [containerMD Schema \(expressing technical metadata for container files\)](#)
3. [Database approach / RDF database implementation example posted on the wiki](#)
4. [For the next call . . .](#)

Swedish National Archives METS 1.10 Change Requests

The Swedish National Archives recently submitted four change requests.

The requests are posted at https://www.socialtext.net/mim2006/mets_1_10_change_requests The requests were discussed at the November meeting in Baltimore. One was provisionally approved and three were deferred for further discussion.

Provisionally Approved Request: [Add EAC-CPF in the value list for MDTYPE](#)

Tom will contact the Swedish National Archive for a use case and an indication of how time sensitive this request is. The next steps would be for Tom to (1) add the value to the schema, (2) generate an example and make sure it validates, (3) send notification of a new version to the listserv, (4) wait two weeks for objections, (4) then accept new version.

As we don't have any other changes pending it does seem of particular value to see if the need for this change is time sensitive.

The remaining requests were not approved.

The next step for these is for Tom to write up responses to each request with examples of how we would address the requests in the current schema. (The precedent in the past has been that if there is a reasonable way to accomplish a change request within the current profile, then it is a good signal not to change the profile.)

1. [Add repeatable element altRecordType in metsHdr](#)

Discussion:

- There are no restrictions on the type attribute in the mets root which means you could use a semi-colon separated list of types
- Another option is to link to a dmdSec from the header
- A lot of the information in the examples is provenance-like, which really isn't type info

- The type attribute in the root element is not describing a record type, it's defined to describe an object type
 - General consensus around recommending the use of a linked metadata section (dmd, amd) in the mets header or another place if more appropriate
2. [Add format attributes to mdRef and file elements](#)
- Discussion:
- The request acknowledges that these information elements are defined in PREMIS, convenience is a driver for the request
 - If we consider the proposed attributes to be core things we want to say about files on the same order, as, say, checksums, then there is no problem adding them. But we're not sure we consider them at the same level that they need to be enumerated as core standard data elements that need to be said about files. Leaning toward making them technical metadata.
 - If you look at the attributes already provided, many of them could reasonably be considered unique. Mime type is an exception, but size, created, checksum are specific to this image as opposed to something you might want to normalize so that it applies to a group of files. Another reason for putting this information elsewhere.
 - Another potential argument for not adding the proposed attributes is that there's already a lot of confusion about "if I embed PREMIS then there are two places to embed checksums." If we add more attributes, it will increase that confusion. . . more redundancy. And redundancy is already a bit of a problem.
 - We're not sure adding another way to do the same thing is a good idea. Maybe we should even think about reducing the number of attributes in METS 2.0
3. [Add status attribute to fileGrp, file, Flocat and Fcontent elements](#)
- Discussion:
- You could use the USE attribute USE vs. STATUS is just semantics.
 - The SNA profile could define how their USE attributes relate to status.
 - You could do this with metsRights or concatenate note information to USE. Eg., security_copy:no_read_access.
 - If attributes from other namespaces were allowed, some of these questions would be moot. Allowing attributes from other namespaces would reduce interoperability. It's still an ongoing discussion about whether the goal of METS is to facilitate exchange between repositories, so at this point we shouldn't introduce things that will make that harder. Although there is a great diversion between profiles already and somebody's going to have to read profiles to figure out how they are put together and then write code . . .

containerMD Schema (expressing technical metadata for container files)

Sébastien raised the issue of the technical metadata for container files (ARC, WARC, TAR, ZIP, etc.). New functionality in JHOVE2 can be used to extract container technical md, but the resulting xml output is too large to be stored in BnF information packages without modification.

METS could be used to describe container files, but using METS as a techMD extension within a METS file seemed risky and cumbersome in terms of maintenance.

To solve these problems, the BnF created a solution, containerMD (<http://bibnum.bnf.fr/containerMD-v1>), for expressing container technical metadata. The containerMD schema addresses (1) the need to aggregate information in a non-verbose fashion (to have XML files of manageable size) and (2) the need for "format specific XML buckets" to express fields that are specific to a particular format (ARC being the first one).

Noting that this is a strategically important topic now that JHOVE2 provides the ability to do container file analysis, Sébastien asked the METS Board to consider the following:

1. Is containerMD a good solution for expressing technical information for container files? Or would another system / schema be a better solution? E.g. should we recommend METS as a container characterization format? Is it, from a Board standpoint, something realistic to consider?
2. Would we go for METS as a characterization schema, should it be considered within the scope of METS 2.0 ("METS as a characterization schema")? Should it be handled with a specific Board-endorsed profile for example?

Discussion:

There was general agreement that containerMD is impressive work and is a good idea. It is a clever construction that was an effective solution for the BnF needs. There was also a sense that METS can do so many things and that it may be a good idea to have more targeted schema for specific use cases rather than trying to accommodate all different types of use cases.

It was also pointed out during the discussion that you can have objects in container files that are not files (eg. directories, symbolic links, etc.). Enabling METS to handle these would mean that METS which is already so generic would have to become more generic.

[Database approach / RDF database implementation example posted on the wiki](#)

Sébastien posted [some information](#) on the Board's wiki on the BnF METS-in-a-database implementation.

Sébastien noted that the example is very specific because it's an RDF triple store, and that he isn't sure it's the right place to start to determine if a database approach should be considered (the question of aggregating the information being the only question interesting for all databases).

However, the example can feed both the "controlled vocabularies" approach (since it requires clearly defined div TYPES, and structMap TYPES to work) and (overall) the "data model" approach: the creation of the ontologies supporting the RDF database model required to define an underlying model (very close to the conclusions of Jerome's paper).

Discussion:

The importance of the structMap and the fileSec in the example, echoes a statement at the face-to-face meeting indicating that the structMap and to a lesser extent the fileSec are where METS has a niche.

The example uses ORE concepts and there was a sense that it could be used as the beginning of both (1) an abstract data model and (2) an ontology.

The call ended before the group got to the final agenda item.

For the next call . . .

Where do we go next with METS

- Look at affinity diagram
- Ideas about abstract modeling

To make call productive:

- Look at Sebastien's [information](#) more closely and consider what it provides with respect to data modeling
- Tom likes the idea of looking at ORE. It makes sense if we agree that structMaps and fileSecs are what sets METS apart

Interoperability vs. Exchange question

- Tom will talk more about this.

METS 2.0 profile at BnF

- We didn't get to this agenda item on our call, however Sebastien won't be able to attend the January call.