

METS Editorial Board Meeting

Thursday, November 7, 2013

DLF Fall Forum

AT&T Executive Education and Conference Center

Austin, Texas

Contents

METS Editorial Board Meeting.....	1
Minutes	2
Present:	2
On Phone:	2
Board's Mission for the Upcoming Year	2
Board's Internal Procedures and Mechanisms	2
Membership/recruitment.....	2
Chairmanship	2
Teleconferences frequency/schedule/infrastructure.....	3
Backup/mirror site for schema	3
Github as a tool for wiki/documentation	4
Formal adoption of 1.10 Schema changes.....	5
Summary of changes.....	5
Summary/Actions:	5
Open Floor – Other presentations; updates on other initiatives; continued work on topic from earlier discussion.....	6
Documentation/Wiki	6
ALTO and Structural Info.....	6
METS 2.0 Presentation and Discussion (Tom)	6
Discussion:	7
Communication and Documentation Issues	8
Goal: new XSLT, XSL-FO for METS Profiles– Review of progress, next steps.....	8
Action item:.....	9
Open and Wrap up.....	9

Minutes

Present:

Tom Habing, Jukka Kervinen, Jean-Philippe Moreux, Betsy Post (recording), Brian Tingle, Nate Trail, Frederick Zarndt (guest)

On Phone:

Nancy Hoebelheinrich, Tobias Steinke, Robin Wendler

Board's Mission for the Upcoming Year

1. METS 1.0 maintenance
2. Focus on documentation for 1.10 including updating Primer
3. Continue METS 2.0 data modeling exploration and development

Board's Internal Procedures and Mechanisms

Membership/recruitment

- Review new member recruitment procedures
- Put out a call for new members fill seats vacated by Patrick Yott and Sébastien Peyard
- Applications due by end of January; subcommittee makes recommendations to Board by end of February.
- In recruiting, consider importance of cross representation with other groups (Jukka, Nate, Jean-Philippe members of ALTO Board; Marcus/PREMIS; Betsy/MODS)
- Recruit broadly, putting out a call to METS, PREMIS, MODS, lists. Send out individual recruitment e-mails.
- Talk with vendors and implementer; look for people from museums. Eg. Singapore National Library, Smithsonian, Archives of American Art, ExLibris, Duraspace, U Wisconsin.

Chairmanship

- No formal mechanisms around co-chairs – consider new members.

- Betsy and Tom will each serve another year

Teleconferences frequency/schedule/infrastructure

- Six weeks is a good schedule.
- Betsy will schedule for next year.
- Like Ready Talk.
- Recordings – after a call, delete the last call.

Backup/mirror site for schema

Discussion triggered by recent federal government shut down.

- Concerns
 - (1) Availability of documentation and specification. IA had complete archive of LOC pages.
 - (2) With LC servers down, direct link for schema validation caused error messages on ingest. Schema validation is vital function.
- Discussion

Best practice is to maintain a local catalog of local schemas. Modern xml allows you to refer from local catalog and this is a best practice.

LC wants to put schemas into GitHub, but uris would vary. Would still need to tell people what to put to in their own instances. Github doesn't want people to validate schemas directly from GitHub.

For the long term LC is considering – moving documentation and schemas off LC site – so work load can be distributed.

Internet Archive mirror site was used during shut down.

Brian created a local catalog, but hasn't tested it -- OASIS test for catalogs should work for testing.

- Action Item

Add something in our documentation indicating it's a best practice to make catalogs and a reference to a resource that shows how to make/implement a catalog.

Github as a tool for wiki/documentation

Discussion:

- Is it easy to get stuff out of github? Yes – you can clone – but you are locked into wiki language. Just by cloning you've exported.
- But, how would you migrate issues? We don't know how they work internally. We think they are in a proprietary format. It looks like there is an API to grab issues so we may not be totally locked in. (Widely used alternative: JIRA issue tracker.)
- Our issues are in the wiki repo – not against the schema. We should put schema issues with the schema.

Goals:

- Documentation – complete migration from social text.
- Get handle on github. Be happy we're not totally locked in.

How we're using it now: 4 projects

- METS 2.0 wiki
 - They also have Github web pages.
 - You can set up c-names – get metsstandards.org
 - We aren't using separate wiki for this
- METS Schema History
 - Tom will merge 1.10 branch back in once it is approved/finalized.
- Wiki project – couldn't figure out a way to have a wiki without repo – wiki project is where issues are.
- Terry's schematron stuff.

Issues Handling

- We still want community to send issues/change requests to the mailing list. One of us (Tom) will add issues to GitHub.
- Tom and Betsy will review open issues prior to each meeting.

- When you do a commit you can also indicate that a particular issue should be closed.

Formal adoption of 1.10 Schema changes

Tom walked group through changes for 1.10. For change details see:

<https://github.com/mets/METS-schema/blob/Version-1.10/mets.xsd>

Summary of changes

- Dropped copyright replaced with Creative Commons
- Added 'LIDO' as potential metadata scheme to MDTYPE enumeration
- Added xsd:anyAttribute to top level elements (avoiding elements where this would create an xlink conflict) – this will accommodate a lot of potential change requests – a lot of change requests are around local practices.
 - Our documentation should provide best practices for the use of anyAttribute. Carefully describe costs and benefits, especially what might be lost in interoperability when anyAttribute is used.
 - Encourage people to make profiles. With anyAttribute, we're allowing local practice, but not encouraging people to go overboard with it.
 - If there's an existing attribute use it, rather any anyAttribute.
 - Encourage people to register their local attributes by registering profiles – thereby gives interoperability to people who care about. This will also help the development of METS. We can add them to schema or address the need in 2.0 as appropriate.

Summary/Actions:

- Update METS profile description on LC site – noting you should document your use of anyAttribute.
- Agreed to move forward with 1.10 schema changes. Tom will publish version to github and test validation once more.
- Brian will locate his collection of sample METS documents that are included with profiles and send them out to board.
- Brian – goal stability of v. 1 schema.

Open Floor – Other presentations; updates on other initiatives; continued work on topic from earlier discussion

Documentation/Wiki

- Change SocialText link on LC site to github wiki. Add link on github wiki to SocialText. Make pdfs of old stuff on SocialText wiki. Goal for next year is to complete migration. (Betsy will talk to Leah)
- Use of which mark up on wiki? People using different things. Does this matter?

ALTO and Structural Info

Developments at ALTO Board meeting: Need to allow for adding structural components in ALTO documents. For example, there is a need to point out elements such as headlines, etc. In most implementations this is handled by using METS and structMap. However, some implementations use ALTO without METS which can be a problem because ALTO doesn't understand document except on page level. Tools like FineReader don't create METS. Some of the ALTO changes might impact METS 2.0

- ALTO Board is proposing new section (similar to the styles section/element) s – it would be called tags and would have label, description, title.
- Textblock element would reference the tags by ID. In example, tag has an extension element ; example contains a MODS record.

Discussion: Tom saw synergies between ALTO need for structural metadata and "Open Annotation Collaboration" as described on openannotation.org. Betsy raised possibility of re-use of some basic TEI tags.

METS 2.0 Presentation and Discussion (Tom)

Reviewed history of our METS 2.0 work, noting the *Reimagining METS* white paper (April 2011) and the affinity diagram (November 2011)

PowerPoint Presentation and Review of Illustrative Document (see appendix A for Powerpoint)

Discussion:

- First bullet or PowerPoint (maintain core aspects of METS1) point is a fair summary
- Second bullet – align ourselves more closely with PREMIS 3.0 (esp concept of intellectual entities). There are pros and cons:
 - there is an official PREMIS vocabulary and it is all posted to loc.gov
 - significant new thought is being given to PREMIS – so if you adopt some of it then you are getting some of that stuff for free
 - What is missing in PREMIS that we need METS at all? Should we just add to PREMIS.
 - In any alignment with PREMIS we must look at restrictions in PREMIS that are confined to PREMIS. For example, PREMIS saying non-preservation uses out of scope (Robin)
 - For METS 2.0, more consideration should be given to interoperability from the get-go. Profiles are not workable solution for interoperability. Other schemas/standards are things starting to appear in multi-media space, so maybe range of use cases for METS narrows.
 - IIIF work may help us gain traction – presentation system must do something with file format
 - Perhaps lobby PREMIS to allow us to embed METS stuff into schema.
 - Tom's concern of his mock up – it is too expressive – there are too many implementation decisions.
 - Nate – put up principles and see who responds. Do this so we know what to people who care; so we know what to support and what not to support.
 - Point of caution: avoid trap of optimizing schema for humans who are looking at the xml document vs. optimizing for implementation. For example why is fileGrp so nested. Why not just a flat inventory of files.
 - XML is inherently hierarchical which is limiting. We have hierarchies rather than full graphs. Property graphs considered. See for example: <https://github.com/tinkerpop/gremlin/wiki/Defining-a-Property-Graph>
 - Brian: to some extent we are completing model and serialization.; step back to model.
 - Have we done in current years a survey of what features people are using? There definitely should be some sort of survey and we should also try to get as many sample documents as we can. We need verification of the goals. We also need to ask what features implementers are using and not using in METS. We should also ask what needs are not being met in current METS. How do we focus between modeling and technology? How do we broaden our technology base to include vendors and other folks who are doing large scale work. Broaden beyond research library group. Get specific feedback from people who are using METS at scale. An advisory group to work on 2.0 could include people not necessarily on board.

- Is 2.0 effort internally driven?

Communication and Documentation Issues

Goal: New edition of METS Primer, incorporating latest version of schema METS Profile

- Add new elements and attributes and examples added since 1.6
- Correct errors identified since last release.
- Nancy has a list of errors or that needs to be changed. Some are in social text, some have been sent to Leah.
- Update tables at back
- One group could do writing
- Another group editing graphics that appear at the beginning of each section. They were done on a MAC in graphics.
- Master source is a word document.
- Process includes checking to language to make sure Primer and documentation in the schema use similar language
- Divvy up tasks and see who has time to do contribute
- Nancy will contact Leah about forming a group.

Goal: new XSLT, XSL-FO for METS Profiles– Review of progress, next steps

Tom started doing a version 2 xslt for METS profiles. The 2.0 profiles are simply xml. There are not html or newer versions. It would be nice if there were a human friendly formatted version of the profiles. This might be a good task for a student learning xml. Recruit library school intern. We could ask DLF to sponsor a trip to its forum for a student who created a stylesheet for the new version of the profile. Or it could be an assignment in a library school class.

Tom will look to see how far he got and we will check to see how far Jenn Riley got when she worked on this. Question: do we want pdf or xhtml? Existing standard transform doesn't show examples well. Look at Illinois examples, look at minutes to see what else was missing.

Action item: Take existing xslt style sheet and adopt for new version. Post volunteer opportunity on list. Betsy ask digital library and xslt instructors at Simmons. Perhaps Jerry could suggest a volunteer.

Open and Wrap up

Next Year's Proposed Meeting in Europe

- Nancy contacted Jose. If we are going to do a program (as opposed to a simple open meeting) we will need to submit a proposal.
- Nancy will spend some time looking at the METS 2.0 schema proposed by Tom as well as Brian's proposal about a graphing model. She will come up with questions.
- Related Standards to think about: *ISAD(G): General International Standard Archival Description ISAIDG* –hierarchical description in archives is codified.
- Archive eXchange Format something they're starting to talk about in science community. Nancy will send on link (<http://www.openaxf.org/>)
- Small group to work on 2.0.
- Nancy will keep an eye for deadline for proposals for joint JCDL and **Theory and Practice** of Digital Libraries conference. We'll see where we are at that point. Last year proposals were due march 23rd and were extended until April.
- Betsy check with Marcus re space for Board meeting at BL.
- Or we could put in a proposal to conference organizers saying we'd like to have an open editorial board meeting. At an old JCDL we did a METS implementation fair. People did lightning talks about how they implemented METS.

METS 2.0

Ideas and Principles

METS Affinity Diagram

- <http://linoit.com/groups/METS%20Board/canvases/METS%20Board>

Guiding Principles

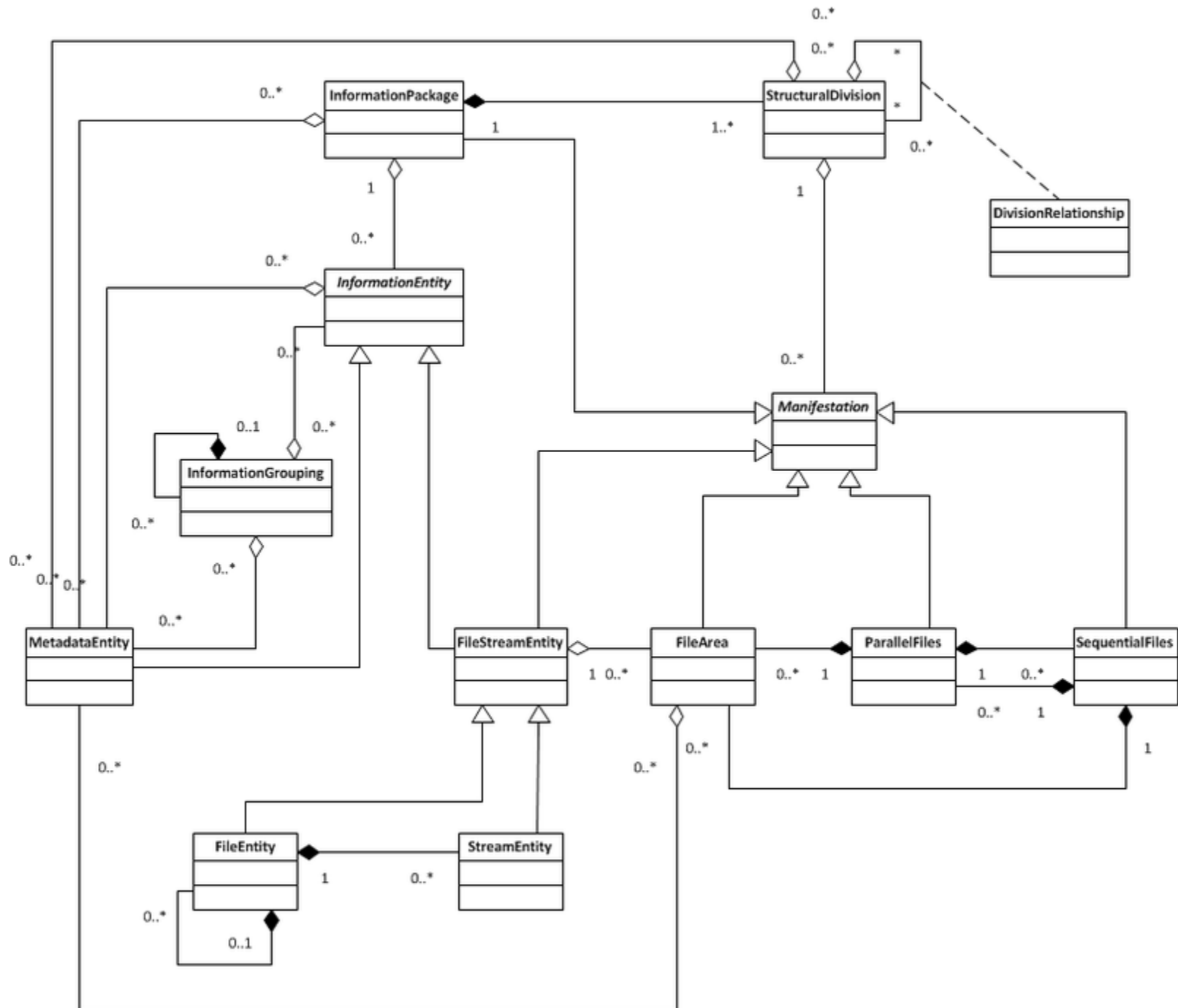
- Maintain Core Aspects of METS 1
 - Structured Inventory of Files (fileSec and fileGroup)
 - Hierarchical Structures (structMap)
 - Accommodate non-hierarchical structures (structLink)
 - Links to different kinds of metadata
- Reuse PREMIS 3.0 entities as possible
 - PREMIS File ⇔ METS File
 - PREMIS Representation ⇔ METS Div or structMap
 - PREMIS Bitstream ⇔ METS Area or nested METS Files
 - PREMIS Intellectual Entity ⇔ METS Div or structMap

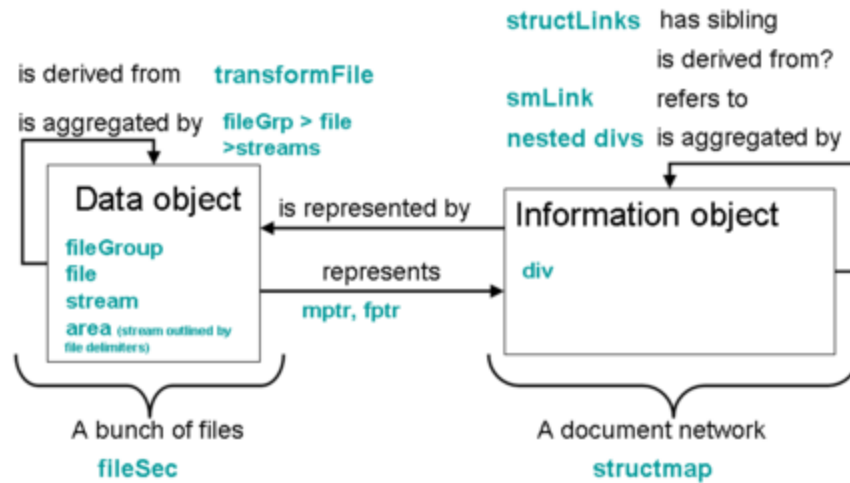
More Principles

- More PREMIS-isms
 - No attributes in XML Schema
 - Two-part identifiers
 - Reuse PREMIS events, agents, and rights
- Simplify schema
 - One Container Element <objectGroup>
 - Do not “hard-code” controlled vocabularies
 - Extension points throughout - <xs:any>, <xs:anyAttributes>

More Principles?

- id.loc.gov for vocabs
- Not backward compatible
- More interoperable
 - More constrained
- Digitized, Cultural Heritage Materials
- Disallow embedded binary
- Maintained explicit containers





METS terminology

- objectInventory
- objectRelationships