

## Mets Board Teleconference

Thursday, June 14, 2012

**Attending:** Terry Catapano, Markus Enders, Richard Gartner, Thomas Habing, Jukka Kervinen, Betsy McKelvey (recording), Sébastien Peyrard, Tobias Steinke, Brian Tingle, Robin Wendler

### DLF Fall Forum/METS Board Face-to-Face

We discussed possible meeting times/dates for Denver this fall. [We requested Friday, November 2, which is before the Forum begins.]

### METS Profile Listserv

We need to update the list of recipients.

### Controlled Vocabulary Mining Work

See the page on the board wiki: [https://www.socialtext.net/p-mets-board/formal\\_vocabulary](https://www.socialtext.net/p-mets-board/formal_vocabulary)

Step 1. Sébastien mined the registered profiles and recorded all registered values.

- His results are recorded in a spreadsheet. The spreadsheet is attached to the METS Board private wiki: [https://www.socialtext.net/p-mets-board/formal\\_vocabulary](https://www.socialtext.net/p-mets-board/formal_vocabulary)
- He had to look beyond the controlled vocab section of the profiles because in some cases controlled vocabulary was part of the profile but not mentioned in the controlled vocabulary section.

Step 2. Analyze the results of the data mining

- Sébastien presented a powerpoint which included observations from and analysis of the data mining exercise. The powerpoint is attached to the METS Board private wiki: [https://www.socialtext.net/p-mets-board/formal\\_vocabulary](https://www.socialtext.net/p-mets-board/formal_vocabulary)
- There are 3 places where vocabulary is used or could be considered:
  1. Already existing controlled vocabularies in embedded schemas
  2. Existing enumerations in the schema
  3. Uncontrolled fields in METS which can benefit from a controlled vocabulary
- The Powerpoint provides analysis for #2 and #3. #3 was the focus of the data mining exercise. The predominant uncontrolled attributes are USE and TYPE (for mets, structMap and div). The Powerpoint provides significant detail and notes that the div TYPE is the most complex/difficult.

### Data Modeling

Tom created a draft data model that he hopes can serve as a starting point for METS 2.0 modelling. [https://www.socialtext.net/p-mets-board/data\\_model](https://www.socialtext.net/p-mets-board/data_model). The model is illustrated using a UML class diagram. Below the diagram is a link to a Word document with some description of each of the top-level entities. The model is mostly based on the current XML Schema with some added levels of abstraction

and some name changes.

Comments on the model:

- Jerry McDonough commented that the structural division could be another kind of InformationEntity, further compacting the model.
- Sébastien commented that the file “stuff” could be compacted as well: there are (1) streams which are not independent and (2) files that are independent
- Sébastien commented that perhaps we should avoid use of the term manifestation (FRBR term) and use data object (OAIS term) instead?

## Wrap up

Some discussion of whether the vocabulary and data model could be presented at DLF. Any presentation at this point would need to indicate this is ongoing work for discussion. We haven't for example, proved that the old model can be fully represented with the new model. [Note: deadline for DLF submissions has passed.]