

PREMIS Editorial Committee Conference Call Notes

4 October 2007

In attendance: Rebecca Guenther, Steve Bordwell, Olaf Brandt, Priscilla Caplan, Brian Lavoie, Bill Leonard, Zhiwu Xie, Gerard Clifton (notes).

Apologies: Rory McLeod. [Yaniv Levi no longer able to participate.]

1. EC Membership

Several people have been suggested to replace Yaniv Levi (who is unable to continue on the Editorial Committee due to other commitments):

- Karsten Huth [Federal Archives Germany] – not yet implementing
- Markus Enders [SUB-Goettingen] – developing software in digitisation field and implementing PREMIS through MathArc work.
- Susan Thomas [Oxford University Libraries] – project manager – Paradigm, Cairo
- Representative form National Library of New Zealand

The Committee recommended that Markus Enders be approached to replace Yaniv, as his implementation experience and programmer's perspective can provide valuable input to discussion.

ACTION: Rebecca Guenther to contact Sally McCallum (LC) regarding making a formal approach to Markus Enders to join the PREMIS Editorial Committee.

2. Rights

Priscilla noted that the Data Dictionary excerpt on Rights has been updated: the several agents have been replaced by a single agent and the examples have been revised. Priscilla has received a suggestion to make the Rights proposal more consistent with the CDL Copyright proposal, but believes this requires more discussion, as there are some issues in doing this.

ACTION: EC members to review the revised Rights DD excerpt and examples, for discussion at a future call.

3. Controlled vocabularies

A number of new proposals have been made regarding implementation of semantic units for specifying (or controlling/registering) the controlled vocabularies used by other semantic units:

1. Rebecca's revised and simplified proposal
(<http://pec.lib.uchicago.edu:8888/pec/uploads/1/vocab-identifier-application.1.doc>):
Defines a single semantic unit (vocabularyIdentifier), which:
 - a. points to the controlled vocabulary in use, if more than one is available
 - b. is not used if the vocabulary is locally controlled, provided by PREMIS or addressed by XML signatures.
 - c. does not itself have a data constraint requiring its own definition by a controlled vocabulary, avoiding the 'infinite regression' problem. The form of the identifier should be self-evident (e.g. URI, textual)
2. Zhiwu's proposal
(<http://pec.lib.uchicago.edu:8888/pec/uploads/1/controlledVocabulary.doc> and schema examples): Defines a container unit coupled with a base 'PremisTerms' registry to allow implementers to register their vocabularies, providing for greater control and a mechanism to validate schemas which use such vocabularies.
3. Gerard's proposal (<http://pec.lib.uchicago.edu:8888/pec/29>): Proposes a container which points to a controlled vocabulary definition and the semantic units to which it is applied, rather than needing per-unit containers each pointing to the same vocabulary definition.

Discussion:

- Proposal 3: This seems similar to a 'business rules' entity approach, which we have previously decided not to pursue. This proposal was rejected.
- Discussion of Proposals 1 & 2 covered a number of points:
 - Proposal 2 includes a common schema which allows XML validation of the controlled vocabulary referenced, however this would require registration and maintenance of control lists. (E.g. MPEG7 has a central registry which is maintained by the group)
 - Repositories could refer to the registered vocabularies. They would be free to change the vocabularies to use locally, as long as they did not want to share them – would require registration.
 - This entails a maintenance task that the PEC does not really want to take on over time. We want to allow definition of the controlled vocabularies in use, but we don't want to have to maintain the vocabularies.
 - (In future, RDF may offer another mechanism to share vocabularies – maintenance can be left to the individual agencies.)
 - Proposal 1 offers a simpler architecture – allows declaration of the vocabulary in use (e.g. in XML attributes), but does not allow validation.
 - Proposal 2 is more explicit – allows declaration of the vocabulary identifier and the type of identifier, plus allows validation.

- There was a query as to whether the declaration of controlled vocabularies was seen as core metadata. There was agreement that a repository needs to know the source of a controlled vocabulary (=core), but whether a repository wants to control a set of values (i.e. as per Proposal 2) is an implementation decision and would not be core.
- There was also agreement that we want to allow users to indicate what values they are using for vocabularies, and to have choice, but we do not want to control the vocabulary values or the type of vocabularies in use.
- If we wanted to control all the vocabularies, this would require a significant extension to every unit that needed such control, which would markedly disrupt the DD.

Decision:

- It was suggested that we say that controlled vocabularies should be indicated, but we will not include a specific semantic unit for it in the DD.
- We would include a section in the DD covering controlled vocabularies, but not change the table part. We could also provide examples of how controlled vocabularies can be declared in various implementation scenarios: e.g. 'if you use METS, do it this way (e.g. profiles); if you use XML do it this other way (e.g. attributes). There can be a place for declaration of controlled vocabularies included in the schema, but not in the DD.
- Include a section in the introduction, or a special topic – we need to include this somewhere, and (even though we would probably prefer not to), we should re-do the examples.

ACTION: Rebecca Guenther, Zhiwu Xie and Gerard Clifton to draft a 'special topics' section on Controlled Vocabularies.

4. Other business

Olaf on leave for 2 weeks.

Next call: 18 october.