

## PREMIS Editorial Committee Conference Call Notes

29 November 2007

In attendance: Rebecca Guenther, Priscilla Caplan, Angela Dappert, Olaf Brandt, Brian Lavoie, Steve Bordwell, Markus Enders, Bill Leonard (notes).

Apologies: Zhiwu Xie, Gerard Clifton.

### 2. Object characteristics at Representation level

An email discussion continued this topic from the last meeting.

<http://listserv.loc.gov/cgi-bin/wa?A2=ind0711&L=premis-ec&T=0&X=2550D842725B67FFBB&P=2609>

<http://listserv.loc.gov/cgi-bin/wa?A2=ind0711&L=premis-ec&T=0&X=30F10D52462E21E456&P=2048>

The concern was expressed that it would not be possible to distinguish between files and representations if all object characteristics were available for use by representations. The two sides of the discussion are:

- according to the original PREMIS data model a representation is intangible; it cannot have the characteristics of a file; it might be a set of files, each of which has its own object characteristics; the PREMIS data model is sound and consistent.
- it is difficult to implement the theoretical model in practice. All of the files comprising a representation might be packaged up into one file, for example as a .tar file. The .tar file, therefore, has its own object characteristics, which cannot be recorded as characteristics of a representation. It is necessary to keep track of the provenance of the representation.

Implementors have found various solutions to recording the characteristics of the file, or folder, comprising the representation. Markus mentioned that METS has a pointing mechanism between the larger file and the component files. It was suggested that users of premis could employ the METS pointing mechanism. Rebecca suggested that PREMIS already has accommodation for this kind of relationship in the relationshipSubType unit. Brian questioned whether a linkingIntellectualEntityIdentifier from a file would indicate a representation.

Back in 2004, the PREMIS authoring committee had discussions concerning the metadata packaged in AIPs for exchange. It was assumed that information would be duplicated for files and representations before transmission.

Olaf said there might be two different levels of representations, making it difficult to plan for. They need to know the background of the exchange; the XML stream might contain original file names. They have a practical problem of how to retain provenance information.

Priscilla suggested recording provenance information when you receive the AIP, and associating it with the object. Olaf said that would not work when multiple representations are received in an AIP.

Steve said it is clear that we need to clarify how to record object characteristics for the bundle (representation). Priscilla said she is not convinced that we don't need a mechanism for stating that a file is also a representation. An identity relationship might be needed, e.g. these files are part of a representation.

Rebecca suggested looking again at the "is root" relationship subtype.

The discussion returned to the original question of how to distinguish between a representation and a file. It was stated that it is desirable to maintain the conceptual distinction between files and representations.

Action: Olaf and Priscilla will work together examining the capability in the relationship subtype with the aim of preparing an example, terminology, and possibly revisions which clarify the data dictionary.

### 1. Extensibility

Rebecca presented her paper ["Extensibility implementation for PREMIS"](#) Nov. 27, 2007 which is a response to questions raised in an earlier discussion about replacing mandatory semantic units. This paper includes a list of semantic units which could be extensible, and suggests that any mandated PREMIS semantic units should have semantically equivalent ones in the extension schema. Concerns were expressed about interoperability, particularly regarding mandatory elements. It should be clear from the profile that an institution is using a replacement element from an extension, instead of the mandated PREMIS element.

A concern was expressed that some institutions will want all of the relevant metadata together, whichever schema it comes from, and other institutions will want to use pointers between clusters elements from particular schemas. This flexibility can be accommodated in the schema.

It was agreed that overall, extensibility is mainly a schema issue and no new semantic units (as suggested in earlier discussions) are required.

### Issue: Which Semantic Units could be replaced by extensions?

The group was asked whether we agree the list of semantic units given in the proposal is complete. The discussion covered concerns about validation of extension units, if we opened up extensibility to all units. There were other concerns about not having guidelines or a standard for evaluating potential candidates for use as extensions. It was suggested that the semantic units which contain data inherent for preservation, e.g. event information, must be PREMIS units.

Other business or processing semantic units, e.g. signature and rights, could be candidates for

extension. There was concern about misuse, i.e. plugging in inappropriate schemas in which the replacement elements are not semantically equivalent to the replaced PREMIS semantic units. It was agreed to allow extensibility in signatureInformation and rights. There were concerns about allowing extensibility for format, hardware and software. [Note from notetaker: it was unclear whether these three units were to be on the list.]

There are two possible situations in which extensions could be used with PREMIS:

- extension units are used to replace PREMIS units, one-to-one
- more granularity is desired than that which is provided by PREMIS semantic units

In this situation, lax processing is required.

Action: Rebecca will revise the extensibility document and the schema sub-group will work on changing the schemas accordingly.

**Next call:** 6 December 2007.