

PREMIS Editorial Committee Conference Call Notes

1 November 2007

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

Apologies: Priscilla Caplan, Markus Enders, Zhiwu Xie, (Rory McLeod).

1. Change #11 – Rationale for dateCreatedByApplication

Issue: The rationale is not clear for recording the date an object was created by the creating application (dateCreatedByApplication) – is there any need for it? (see <http://listserv.loc.gov/cgi-bin/wa?A2=ind0604&L=pig&T=0&P=1038>)

Proposal: Steve's proposal notes that, although there may be limited cases where dateCreatedByApplication, the semantic unit is optional, so implementers may choose not to record it if they prefer. This requires no change to the DD. (See http://pec.lib.uchicago.edu:8888/pec/uploads/1/11_dateCreatedByApplication.doc)

Decision: EC agreed no change to DD required.

2. Change #18 – Can both formatDesignation and formatRegistry be used?

Issue: Question as to whether both formatDesignation (formatName) and formatRegistry can be used together, and whether there is ambiguity in allowing formatRegistry to be repeatable. (See <http://listserv.loc.gov/cgi-bin/wa?A2=ind0602&L=pig&T=0&P=548>)

Proposal: Steve's proposal notes that both formatDesignation and formatRegistry can be used (at least one of them is required, but using one does not exclude use of the other), and suggests collapsing the two containers into a single repeatable format container, thereby providing a more explicit link between formatName and formatRegistry information. (see http://pec.lib.uchicago.edu:8888/pec/uploads/1/18_formatDesignation_and_formatRegistry.doc)

Discussion: Covered several points:

- DD should explicitly say that 'You may use both formatDesignation and formatRegistry'.

- It would seem likely that, in practice, running tools for file identification may produce more than one identity per file – e.g. a tool may 'know' the registryKey for a format & version in a number of registries, or, where a format is a sub-type of another format (or even contains another embedded format), a tool may identify a file as being both formats. Some implementers may also choose to run more than one tool. It may be useful to more explicitly associate a (repeatable) formatName with the corresponding registryKey information. If two tools or registries return different results (as demonstrated by differing formatNames or versions), this discrepancy can be more easily detected.
- However, it was noted that such ambiguity should be resolved through the workflow design. Within an implementation there should be clear policy decisions about which tools or registries provide authoritative information used for identification, and where there is ambiguity or conflicting data detected, there should be a mechanism for determining which formatName has precedence. There should still only be one formatName (not repeatable) decided upon and recorded.
- To reduce ambiguity where several formatRegistries are recorded, formatRegistryRole should be used to indicate if a registry is being used for a particular purpose –e.g. specification, format identity & validation profile. There may only be one of the registries that is chosen as the authoritative source for identifying formats (which should be indicated by the formatRegistryRole).

Decision: The suggestion to collapse formatDesignation and formatRegistry into a single format container (with repeatable formatName) was not upheld – the current containers should stand, with additional explanation.

ACTION: Olaf Brandt to draft several sentences for usage notes indicating:

- that both formatDesignation and formatRegistry may be used together
- how more than one registry can be used and can be reconciled with one formatDesignation, noting that workflow and/or policy should be in place to resolve ambiguity to produce a single formatDesignation.

3. Change #28 - preservationLevel and significantProperties

preservationLevel

Issue: Ambiguity in the use of preservationLevel.

Revised proposal: (See http://pec.lib.uchicago.edu:8888/pec/uploads/Preservation_Level.1.doc)
 Previous discussion (see <http://pec.lib.uchicago.edu:8888/pec/uploads/ec-callnotes-20070816.doc>) had suggested inclusion of a 'Role' unit to indicate whether a particular preservationLevel represented what was intended or what was currently implemented. The revised proposal includes an optional 'preservationLevelRole' semantic unit for this purpose, as well as optional 'preservationLevelRationale' and 'preservationLevelDateAssigned' units, as per the previous proposal. If more than one preservationLevelRole can be specified, the preservationLevel container should be repeatable. The previous dataConstraint units have been removed, in line with the decision not to specify a separate semantic unit for controlled vocabularies.

Discussion: The approach appears fine – the new units are optional and so may be ignored if preferred (and so does not require further PIG discussion). Current preservationLevel values recorded could be shifted to preservationLevelValue by a style sheet, if needed.

Decision: This approach was agreed, with some modification to usage notes to exclude specifics on use of controlled vocabularies and PREMIS versions, which will be dealt with elsewhere.

ACTION: Gerard Clifton to revise preservationLevel semantic unit DD pages as noted above and move to pages for collation.

significantProperties

Issue: More guidance and/or structure required on use of significantProperties.

Revised proposal: (see http://pec.lib.uchicago.edu:8888/pec/uploads/significant_properties.doc). Significant properties moves out from objectCharacteristics (applicable at Representation level). A repeatable container includes a significantPropertiesValue and an optional significantPropertiesType, which can describe the 'facet' or attribute in which the significant property described applies. This provides flexibility for representing significant properties (until a better way comes along, at least). The previous dataConstraint container has been removed, in line with the decision to not include a specific semantic unit for controlled vocabularies.

Discussion: Covered several points:

- It is clear that significantProperties must be moved out of objectCharacteristics (even if objectCharacteristics does become applicable at the Representation level): significant

properties are not tied to the format of an object, and, even at an intellectual level, a cataloguer may discover and wish to describe an aspect of the intellectual item that needs to be retained (regardless of any file or format information).

- It was suggested that describing significant properties may imply an 'intention' to preserve that would tie it more closely to preservationLevel, as previously suggested.
- Although significant properties do indicate which properties are 'intended' to be preserved across time through preservation action, it may be preferable (as noted by several PIG respondents in previous discussion) to keep significantProperties and preservationLevel as independent units, and not bind significantProperties to preservation intention.
- As well as indicating the properties which should be (and are intended to be) preserved, significantProperties can also serve as measures of preservation success. For example, if, after some transformation as part of a preservation action, the significant properties described are not retained in the resulting object, the transformation method may have failed or be unacceptable as a preservation method.
- It was suggested that significantProperties should also be made extensible, in case, as research into significant properties continues, other superior methods or structures for describing them are proposed (even for specific material types).

Decision: Proposal for significant properties agreed, with modification to exclude specifics on controlled vocabularies and PREMIS versions. Additional usage notes on using significant properties as measures of preservation success to be included. Extensibility of significantProperties semantic unit to be included in the area discussing extensibility in general (not within significantProperties unit itself).

ACTION: Gerard Clifton to revise significantProperties semantic unit DD pages as noted above and move to pages for collation.

4. Applying ObjectCharacteristics at Representation level - some discussion

Olaf noted several usage scenarios where it is desirable to be able to record objectCharacteristics against Representations, such as exchange of data in a networked environment. An object (which could have its own name by which it is known within a repository) could be composed of several files or perhaps an XML stream which relates a bundle of files (or maybe just refers to their network location in a repository). In the exchange scenario it is important to control what was or wasn't exchanged. There are some 'file-like' characteristics in these situations that it would be advantageous to record and control:

- a representation (e.g. all the files comprising the parts of an object, including all the hierarchies – e.g. volumes, pages etc) could have a name by which it is known collectively – equivalent to an 'original name'
- a collected bundle of files may be expected to have a discrete combined size and also fixity.

However, others noted that it is still not clear how objectCharacteristics apply to a Representation, particularly if there is still some ambiguity about what a Representation is.

A Representation may only refer to one level of a hierarchy [Notetaker's note: I'm unclear on what was intended on this point.]

It may be that only some of the objectCharacteristic semantic units are applicable at Representation level.

Olaf noted that it would be helpful if (some) semantic units from objectCharacteristics were optionally available for use with Representations (as covered in the usage scenarios above).

ACTION: Olaf Brandt to go through the semantic units of objectCharacteristics and see:

- what is applicable at a Representation level
- what specific changes would be required to use these at a Representation level.

5. Other business

- Approved changes (in final DD form) should be linked to the 'Summary of decisions' page (off the main wiki page). Brian will pick up and collate this material.

Next call: 15 November

Call times will revert to regular times (to be confirmed).

[Gerard will be on leave 14-30 November, so no variation to call times due to differences in daylight saving required yet.]