PREMIS Implementation Fair Workshop at iPRES 2014
October 6th 2014, Melbourne, Australia

Notes by Michelle Lindlar and Peter McKinney

Agenda
09:00-09:15 Introduction
09:15-09:45 Introduction to PREMIS
09:45-10:30 Work since last iPRES in Oct. 2013
10:30-10:50 BREAK
10:50-11.15 Conformance to PREMIS
11:15-12:35 Implementations and their challenges
12:35-13:00 Questions and Discussion

Note: Conformance was not discussed as we focussed on implementations as being most relevant to the participants.

Participants

Dave	Allen	State Library of Queensland
Werner	Bailer	Joanneum Research
David	Bromage	Queensland State Archives
Esther	Carey	National Archives of Australia
Denise	de Vries	Flinders University
Carmel	Denholm	Department of Education Tasmania
Joanna	Fleming	State Library of New South Wales
Nobuaki	Honda	National Diet Library Japan
Peter	Jenkins	State Library of South Australia
Dana	Kahabka	State Library of New South Wales
Leo	Konstantelos	University of Melbourne
Karel	Koucky	National Archives of the Czech Republic
Chi Wa	Lai	Government of Macau
Angeletta	Leggio	Australian National Data Service
Michelle	Lindlar	German National Library of Science and Technology
Peter	McKinney	National Library of New Zealand
Deborah	Schrader	Melbourne Cricket Club Library
Christopher	Paroz	SAE Institute
Beth	Robertson	State Library of South Australia
Scott	Wajon	State Library of New South Wales
Eld	Zierau	National Library of Denmark

Participant expectations:

Participants were invited to introduce themselves and their motivation for joining the workshop. Most participants described no working knowledge of PREMIS but wanting to learn more in order to
1) be able to implement the standard in their institution
2) gain better personal understanding of requirements for preservation metadata
3) be able to better explain the needs for PREMIS within their institution

Questions

Q: Can I dump technical metadata elsewhere in the system and just include a pointer towards that place in PREMIS?
A: Yes.

Q: Are events preservation events only or really any events related to the object?
A: Per definition events can be related to creation, modification and access. Most events related to the object should be able to be mapped to those groups. The institution needs to define whether an event is indeed a preservation event to them or not.

Q: We run fixity once a month, should we keep the info?
A: It is recommended to capture information about any events that touch an object, but it is up to the institution to define how this is realized. It would be possible, for example, to only capture the info if something goes wrong.

Q: Can we capture even more information in PREMIS, like descriptive metadata?
A: In general extensive descriptive metadata should be captured elsewhere. The specific requirements of the collection should be looked at to decide what to capture where. It would be helpful for the PREMIS committee to have some concrete examples from the user community.

Q: Are there any recommendation and guidance as to what to include in the extensions to ensure long-term understandability by e.g. including fixed vocabulary or standards?
A: No recommendations and guidance are available. It is implied that standards should be used for the extensions – however, this is of course hard in cases such as the eventOutcomeDetails. It would be helpful for the PREMIS committee to have some concrete examples from the user community.

Q: Where can I dump my extensive ffmpeg output? Does it go into eventOutcome or elsewhere?
A: It is important to differentiate between tool output and problem reporting by tools. EventOutcome should capture if the event ran ok or not and include error messages.

Q; How do large institutions like national libraries deal with schema changes – e.g. in the case of the upcoming PREMIS v3?
A: At first check how you can implement this in your system. If you have a vendor, talk to them early on. It is also important to think about what to do with the preserved objects – to change the preservation metadata for those, you could consider a tool-based approach or running them through the entire system again.

Q: Why don't you directly describe the policies in PREMIS (in relation to Eld's presentation on preservation level)?
A: Because the policy may change regularly.

Q: Has PREMIS looked at incorporating the SCAPE controlled vocabulary for policies?
A: Not that we know. But will recommend the Committee to look into the relevance of it.

Q: Has the environment extension been tested? Will it ensure preservation and renderability? Who is preserving all these environments?
A: That's a general digital preservation questions – it is good if we can point towards registries for a lot of this.

Implementation examples:

Eld Zierau presented the PREMIS implementation at the Royal National Library of Denmark.

Scott Wajon (State Library of New South Wales) brought in an example of a metadata file the institution received from a service provider. The file included PREMIS and MIX metadata. It was used to look at what kind of information could be captured from external processes. The file was interesting in that only event metadata was codified in PREMIS semantic units (why had the vendor made that decision). In particular, the file included extensive information about a deskewing event. It was discussed how this information could be relevant depending on whether it was preformed on a master or on a derivative file. Explicit information about the software/agent which was used to perform the event should be included.

Michelle Lindlar presented work being done as part of the DURAARK project in a pre-ingest workbench for architectural 3D data. Regarding a PREMIS implementation in the workbench process, three questions were formulated:
1. If the pre-ingest workbench runs externally (e.g. as a service) with no knowledge of the preservation repository, is it still an agent or is it something else?
2. As the pre-ingest workbench is a complex system combining multiple tasks and wrapping separate tools for e.g. file format identification and metadata extraction, is it a series of agents or something else?
3. Within DURAARK, "a building / structure" is considered an intellectual entity. Representations of the entity always stand in temporal / spatial relationships and dependencies – i.e., scans from different years or plans describing pre-/post-refurbishing states. These representations should therefore be rather positioned at an IE level, calling for a nested IE structure. Is this possible within PREMIS and are there known reference implementations for this?

It was discussed how a pre-ingest workbench can be described using the environment entity in PREMIS v3. The DURAARK workbench can therefore be seen as a nice use case for this new entity, where an external system is described which produces a SIP (and therefore generates a PREMIS file) to be deposited into an institution's digital preservation system. The environment entity also allows for the detailed description of the different agents involved within the pre-ingest process.

The nested structure is possible in theory, however, no reference implementation is known.

Action Items

- Put all slides from the event onto PREMIS website.
- Put on website (and send to participants) sample METS showing PREMIS used for newspaper digitisation work [NLNZ].
- Investigate SCAPE controlled vocabularies.
- Editorial Committee to investiage enriching advice on container extensions (in particular eventOutcomeDetail).