

PREMIS Implementation Fair Workshop

Peter McKinney
National Library of New Zealand
Cnr Molesworth & Aitken St
Wellington
New Zealand
peter.mckinney@dia.govt.nz

Eld Zierau
The Royal Library of Denmark
Søren Kierkegaards Plads 1
DK-1016 København K
+45 91324690
elzi@kb.dk

Rebecca Guenther
Library of Congress/Consultant
101 Independence Ave SE
Washington, DC 20540 USA
+1 703 298 0157
rguenther52@gmail.com

ABSTRACT

This workshop provides an overview of the PREMIS Data Dictionary for Preservation Metadata, a standard addressing the information you need to know to preserve digital content in a repository. It includes an introduction to PREMIS and reports from the preservation community on implementation of the standard in various systems or contexts.

General Terms

infrastructure, preservation strategies and workflows, case studies and best practice, preservation strategies and workflows.

Keywords

Preservation metadata, Preservation repository implementation, Data dictionary

1. INTRODUCTION

The PREMIS Implementation Fair Workshop is one of a series of events organized by the PREMIS Editorial Committee [1] and that has been held in conjunction with previous iPRES conferences.

At iPRES 2014, the workshop will give the audience a chance to understand the PREMIS data dictionary and give implementers, and potential implementers, of the *PREMIS Data Dictionary for Preservation Metadata* an opportunity to discuss topics of common interest and find out about latest developments.

2. OUTLINE OF WORKSHOP CONTENT

2.1 Overview of the PREMIS Data Dictionary

The *PREMIS Data Dictionary for Preservation Metadata* [2] is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability. Developed by an international team of experts, PREMIS is implemented in digital preservation projects around the world, and support for PREMIS is incorporated into a number of commercial and open-source digital preservation tools and systems. This session provides an overview of the PREMIS Data Model (which was recently revised) and of the types of information specified to support the digital preservation process. Included will be a

summary of the changes in version 3.0, which includes enhanced ability to describe intellectual objects and technical environments within the PREMIS context.

2.2 PREMIS Conformance

This session describes the work of the PREMIS Conformance Working Group and its effort to clarify what it means to adequately capture the essential metadata needed to support the essential functions of a digital repository. The group is pursuing two avenues of inquiry. The first has drafted conformance levels and is exploring what metadata is required for minimum conformance to PREMIS. The second explores the relationship between preservation metadata and functionality of a preservation system. These two avenues will together allow institutions to not only be able to understand their own conformance to PREMIS, but additionally reflect on how they utilise their metadata to drive, support and record preservation functions.

2.3 Implementation reports

Implementation reports will be solicited from the PREMIS Implementers community. Included will be a report on the National Library of Australia's implementation of a Tessella solution and one on the complexities of applying PREMIS to born digital data acquired on removable media.

3. WORKSHOP SERIES

The PREMIS Implementation Fair at iPRES 2014 will be the sixth in a series and have been held in conjunction with iPRES since 2009. These events are intended to highlight PREMIS activities, discuss issues concerning implementation, and provide a forum for implementers to discuss their activities, issues, and solutions. Because this is a rapidly changing area, it is important to provide continuous updates.

4. INTENDED AUDIENCE

The workshop is designed for those involved in selecting, designing, or planning a preservation project or repository using preservation metadata. This includes digital preservation practitioners (digital librarians and archivists, digital curators, repository managers and those with a responsibility for or an interest in preservation workflows and systems) and experts of digital preservation metadata and preservation risk assessment.

5. SHORT BIOGRAPHIES OF ORGANIZERS

Peter McKinney is the Policy Analyst for the Preservation, Research and Consultancy programme at the National Library of New Zealand Te Puna Mātauranga o Aotearoa. He is a member of

iPRES 2014 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a [copy of this licence](#).

the PREMIS Editorial Committee and part of the Conformance Sub-Committee. Most recently he has been coordinating work on the NSLA Digital Preservation Technical Registry.

Eld Zierau has been a member of the PREMIS Editorial Committee since 2013. She is a digital preservation researcher and specialist, with a PhD in digital preservation. She is a computer scientist, and has worked with almost all aspects of IT in private industries for 18 years, before starting in digital preservation in 2007. She has been working with many aspects of digital preservation, and she is involved as an architect or a consultant on major initiatives such as a new digital repository including data modeling of metadata for preservation.

Rebecca Guenther is Chair of the PREMIS Editorial Committee, on which she has served since its establishment in 2006. She worked at the Library of Congress on metadata standards in the Network Development Office for 22 years and is currently an independent consultant in New York on metadata development and training; she also continues to work part-time for the Library of Congress. She was co-chair of the original PREMIS Working Group which developed the *PREMIS Data Dictionary for Preservation Metadata*.

6. PROCESS FOR SOLICITING CONTRIBUTIONS

Contributions will be solicited from the PREMIS Implementers' Group via its discussion list (pig@loc.gov). To subscribe go to: <http://listserv.loc.gov/listarch/pig.html>. The PREMIS Editorial Committee will review all requests. After workshop proposal is approved, a call will be sent for contributions to the implementation portion and deadline will be within a month.

7. WORKSHOP OUTCOMES

7.1 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

7.2 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 be recommended to the Committee.

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.

7.3 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 performed 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.

7.4 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.
- Investigate SCAPE controlled vocabularies.
- Editorial Committee to investigate enriching advice on container extensions (in particular eventOutcomeDetail).

8. REFERENCES

- [1] PREMIS Maintenance Activity, <http://www.loc.gov/standards/premis/>
- [2] PREMIS Editorial Committee. 2012. *PREMIS Data Dictionary for Preservation Metadata* (Library of Congress). <http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>