

**PREMIS Implementation Fair**  
**iPres 2012: Toronto, CA**  
**Oct. 2, 2012**

**1. Data model changes (Angela Dappert)**

Angela reviewed the issues concerning making Intellectual Entities a new level of object and taking Environment out of Object as a new entity. Comments included:

*Intellectual entities.* Describing a metadata event may be a use case for allowing intellectual entities as a PREMIS object. PREMIS allows you to give metadata about the metadata through the use of the event entity. This makes the data dictionary more self contained and the data model more compact. It was suggested that perhaps we want to carry information about the FRBR level or archival categorizations in the metadata. Eld Zierau from the Danish National Library voiced concern about deprecating linkingIntellectualEntity (although in a later message to the PIG list she felt it was not a problem if linkingObject is used instead and further clarified). There was interest in using PREMIS intellectual entity metadata for assessment before the object goes into the repository. This brought up some discussion of the proposal for semantic unit groupings and how it can be used for intellectual entities. Participants were encouraged to further discuss this on the PIG list.

*Environment.* PREMIS is not restricted to events within the repository (as in OAIS), but can include all life cycle events, either digital or non-digital, for instance creation events. It was suggested that we make it clear what sort of environment information is included, e.g. technical or business (there could be multiple environments). Some are interested in adding policy to environment (see also below).

**2. Changes for Preservation Rights metadata (Robin Wendler)**

Some may indicate policies in PREMIS rights, although not all policies belong in rights. For instance, Robin noted that Harvard had policies about storage and could put it in environment.

Priscilla Caplan mentioned that as implementers have used PREMIS Rights they have been instrumental in affecting future changes—the Rights entity has been growing organically.

**3. Repository software**

Peter van Garderen reported about Archivematica's standards based approach and implementation of PREMIS. Their processing pipeline creates AIPs and DIPs, incorporating PREMIS metadata that can be exported to Dspace, ContentDM, etc. He suggested some additional event types being used: identifier assignment and unpacking. The latest release has incorporated version 2.2.

**4. PREMIS OWL ontology**

Sebastien reviewed use cases for the PREMIS OWL ontology. Using RDF is a handy way to query your preservation data. It also allows you to seamlessly integrate data against different sources. The ontology will be put out for public review in the near future

for further development. The Library of Congress will take over its maintenance in consultation with the PREMIS Ontology Working Group.

## **5. Implementations**

**5 a.** Lynn Benson from the National Library of New Zealand reviewed their implementation, where they have substantial amounts of non-governmental born digital data. A request for more detailed provenance was submitted to the PIG list recently to enable more information about the outcome of events. They also had concerns about compliance, driven in particular because of their outside providers and the need to both be compliant among themselves and meet their responsibilities as trusted preservation repository. They would like a harder line. This brought up issues of conformance vs. compliance (the PREMIS EC put up a conformance statement about a year ago, but it was NOT called a compliance statement). See below for further discussion. NLNZ also would like to see more examples of bitstreams vs. filestreams.

**5 b.** Kyle Rimkus from University of Illinois described their new Medusa system, which is Linked Data inspired. They have dropped using PREMIS in METS and are using PREMIS only; PREMIS relationship is being used for what METS does in the StructMap.

**5 c.** Mark Evans reported about the implementation in Tessella's Safety Deposit Box (SDB). They have added extensions for preservation planning and actions, since they didn't think PREMIS had enough detail. They also extended the Intellectual Entity to include more metadata at this level, so are better aligned with what will be in version 3.0. This includes doing characterization at the Intellectual Entity level. To do active preservation and preservation planning, they want to determine what's at risk and generate a "transformational entity" so that actions can be taken in the future. They are not sure how to do this in PREMIS event, which deals more with actions already taken rather than actions in the future. In creating a preservation plan an emulation environment is launched; they are not sure how to use PREMIS to detail this.

## **6. Preservation Health Check**

### **6 a. Introduction to the pilot**

Titia van der Werf introduced the Preservation Health Check (PHC) Pilot, which is a joint initiative of OCLC Research and the Open Planets Foundation (OPF). The main research questions being considered are: what does my preservation metadata tell me about digital preservation risks to which my digital collections are exposed? How effective are standards, tools and practices in addressing risk control? The pilot will collect preservation metadata recorded by repositories and analyze it as to risk assessment. This work will involve mapping the SPOT risk model with PREMIS and then apply this as a protocol to analyze real life metadata and report back results in the form of Health Check Reports. The pilot will involve 3 pilot sites (not all 3 have been identified yet) and the duration of the pilot is set to 1 year. More information is available at: <http://www.oclc.org/research/activities/phc.html>

### **6 b. Standards, Risk analysis, Common sense and Evidence**

Bram van der Werf explained that the use of the term “risk assessment” in the context of digital preservation is confusing and that what the DP-community really talks about is “quality assurance”: the ongoing monitoring and improvement of the preservation process itself, which is all about treating and handling digital objects to stop or slow down their deterioration, obsolescence and loss of quality (renderability, functionality, etc.). Quality assurance requires continuous monitoring (PHC-dashboard concept) as opposed to one-off risk assessments or certifications. The PHC-dashboard visualizes the data gathered by sensors and triggers preventive/corrective actions. The pilot is looking at PREMIS to identify such sensors. For tool-building, the PREMIS standardization process should not be a moving target.

### **6 c. Presentation by the Bibliothèque Nationale de France (BnF), a prospective PHC Pilot Site**

Sébastien Peyrard shortly presented BnF’s expectations of the PHC-pilot and why it is interested in becoming a pilot site. He explained how the need for regular monitoring (the last assessment they did was in 2007) was becoming more urgent, in particular to evaluate the quality and usefulness of the metadata and how it might be improved. He suggested that external evaluation might trigger internal improvements and be helpful in securing activities and staff effort to improve the preservation metadata. He envisioned how the PHC would help to formulate functional requirements for preservation metadata and how the dashboard would improve the curator’s data management interface.

### **6 d. Presenting the SPOT Model**

Priscilla Caplan gave a short introduction to the SPOT Model which is a lightweight framework for addressing threats to the digital content of repositories. It explicitly does not address other threats found in more encompassing risk-assessment frameworks, such as governance, financial or legal threats to the sustainability of repositories. The threats identified in the SPOT Model have been carefully chosen on the basis of criteria such as appropriate and comparable levels of granularity and comprehensiveness. The SPOT approach was guided by the desire for practical application in the field. The Model is published in: <http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html>

### **6 e. Preservation metadata as an evidence base for risk assessment**

Brian Lavoie explained how mapping SPOT to PREMIS might lead to identifying preservation metadata that supports evidence-based risk assessment. The mapping exercise would lead to a protocol, ideally a standardized and widely-applicable protocol for threat-control, addressing both types of threats: those that have already happened and those that can potentially happen. The focus of the mapping will be on actionable intelligence and automated analysis. Brian gave some practical examples and concluded that the mapping exercise could highlight gaps and inform repositories (the gap between necessary metadata and metadata actually recorded) and advance the further development of threat models and of the PREMIS Data Dictionary.

### **6 f. Plenary Discussion**

The presentations led to practical questions and several more elaborate discussion threads. Below a short summary is given of the 4 major discussion threads.

The relationship with other threat models: TRAC and DRAMBORA are pretty high-level and they are not linked to evidence information that is maintained in the repository. The community needs regular assessments and the potential for automating these and making it a routine. The PHC-pilot is an appealing project because it is trying to do this. It would be useful to position the PHC-exercise in the bigger picture of risk assessment and certification.

The impact question: it would be interesting to understand how the PHC-results would impact our organisation and operations. How would this inform preservation managers in terms of preservation strategies or facilitate the process of becoming a trusted repository? Will the pilot provide input to justify the development of health check tools? How would these fit in existing tools and workflows? Where in the metadata cycle would this health check take place?

The relationship with preservation policies: repository mission and policies are essential for establishing context for risk-assessment. What is the relationship between policies, preservation metadata and risk factors? How will the SPOT-PREMIS mapping take policies into account?

Digital lifecycle and control over metadata: repository managers do not necessarily have full-control over the preservation metadata they maintain. Metadata might be produced by other stakeholders in the digital lifecycle. The PHC-pilot might need to take the digital life-cycle into account when mapping the metadata and the threats.

## **7. PREMIS Conformance/Compliance**

Further discussion about PREMIS conformance suggested that the PIG community discuss the need to be more stringent. A new type of conformance might be evidentiary conformance, where choices made on using aspects of PREMIS may be tied to what you gain in the digital preservation process.

## **8. Preservation Level Type**

Eld Zierau introduced her proposal to define different categories/aspects of preservation level. An institution would define values of high, medium or low to describe to what extent they are doing things like bit safety, confidentiality, availability and functional preservation strategy . This discussion again brought up the issue of where you keep policy information; PREMIS early on said that business rules (i.e. repository policy) was not in scope, but implementers are increasingly finding the need to record some policy decisions, which are related to the preservation metadata.