### PREMIS Implementation Fair (PIF) 2013, 5 Sept 2013, 14:30 to 19:30, Lisbon

Chair: Sébastien Peyrard

Minutes: Angela Dappert

Online materials: Eld Zierau

Participants:

|  |  |  |  |
| --- | --- | --- | --- |
| Mark Jordan | Simon Fraser University | Canada | mjordan@sfu.ca |
| Stina Degerstedt | National Library of Sweden | Sweden | stina.degerstedt@kb.se |
| Eld Zierau | The Royal Library | Denmark | elzi@kb.dk |
| Tamara Leuenberger | University of Bern | Switzerland | tamara.leuenberger@ub.unibe.ch |
| Anna Henry | Tate | United Kingdom | anna.henry@tate.org.uk |
| Kakha NADIRADZE | AFRD | Georgia | foodsafetyge@gmail.com |
| Courtney C. Mumma | Artefactual Systems, Inc. | Canada | courtney@artefactual.com |
| Tsutomu Shimura | National Diet Library, Japan | Japan | t-shimur@ndl.go.jp |
| Angela Dappert | Digital Preservation Coalition | United Kingdom | angela@dpconline.org |
| Rui Castro | KEEP SOLUTIONS | Portugal | rcastro@keep.pt |
| Thomas Bähr | TIB | Germany | thomas.baehr@tib.uni-hannover.de |
| Peyrard | National Library of France | France | sebastien.peyrard@bnf.fr |
| Juha Lehtonen | CSC - IT Center for Science Ltd. | Finland | juha.lehtonen@csc.fi |
| Helena Patrício | National Library Portugal | Portugal | hpatricio@bnportugal.pt |
| Beth Delaney | Audiovisual Archive Consultant | France | bdelaney515@gmail.com |
| Janet Delve | University of Portsmouth | United Kingdom | Janet.Delve@port.ac.uk |
| Inge Hofsink | National Library of the Netherlands | Netherlands | inge\_hofsink@hotmail.com |
| Hélder Silva | KEEP SOLUTIONS, LDA | Portugal | hsilva@keep.pt |
| David Allen | state library of queensland | Australia | dave.allen@slq.qld.gov.au |
| Pauline Sinclair | Tessella | United Kingdom | pauline.sinclair@tessella.com |
| Kathryn Cassidy | TCD / DRI | Ireland | kcassidy@tchpc.tcd.ie |
| Eduardo Pablo Giordanino | Universidad de Buenos Aires | Argentina | egiordanino@sisbi.uba.ar |
| Maria Patricia Prada | Universidad de Buenos Aires | Argentina | pprada@sisbi.uba.ar |
| Titia van der Werf | OCLC | Netherlands | titia.vanderwerf@oclc.org |
| Eliska Pavlaskova | Charles University in Prague | Czech Republic | eliska.pavlaskova@ruk.cuni.cz |
| Tomasz Parkola | IBCH PAS - PSNC | Poland | [aniao@man.poznan.pl](mailto:aniao@man.poznan.pl) |
| Rafael Antonio |  | Portugal | rafael.antonio@sapo.pt; |
| Walter Allasia | Eurix | Italy | [allasia@eurix.it](mailto:allasia@eurix.it) |
| David Anderson | University of Portsmouth | United Kingdom | David.Anderson@port.ac.uk |
| Angela Di Iorio |  | Italy | angeladiiorio@gmail.com |

|  |  |  |
| --- | --- | --- |
| **Part 1: A View from the Editorial Committee** | | |
| 14:30-14:45 | **Introduction** A brief introduction to the workshop. | Sébastien Peyrard |
| 14:45-14:55 | **Update on PREMIS activities**.  A brief overview of PREMIS activities since the last PREMIS Implementation Fair in Oct. 2012 will be given, including changes to the Data Dictionary as part of the development of PREMIS version 3.0. | Sébastien Peyrard |
| 14:55-15:35 | **Changes in the PREMIS Data Model**.  The next major version of the PREMIS Data Dictionary will be released by the end of the year 2013; an update on the main evolutions of the PREMIS Data Model and Data Dictionary will be given. Notably, the revised data model will consider Intellectual Entities differently. Additionally, this new version provides a better way to describe Environments separately from Objects and allow software and hardware registries to be linked to. This modeling work, and the new features that it allows, will be described. | Angela Dappert |
| 15:35-15:55 | **Changes for Preservation Policy Metadata**.  The new version of PREMIS also allows the preservation policy applied to preserved digital objects to be recorded in more detail by updating the preservationLevel semantic container. | Eld Zierau |
| 15:55-16:25 | **PREMIS OWL ontology**.  A revised stable version of the OWL ontology was published in June 2013, and all the PREMIS controlled vocabularies were published at the same time. The ontology allows users to express PREMIS using RDF, and to work easily with Linked Open Data, at a time when those technologies are being used more and more in registries (e.g. UDFR, PRONOM). Along with the existing PREMIS XML schema, this provides an alternate PREMIS endorsed serialization format, and can be leveraged to address problems of distributed preservation metadata across several preservation repositories or format registries and allow PREMIS metadata to be queried easily. | Sébastien Peyrard |
| 16:25-16:45 | Break |  |
| **Part 2: A View from the Field** | | |
| 16:45-17:05 | **Preservation Health Check Report**. The Open Planets Foundation and OCLC Research are conducting a pilot that runs through 2013-2014. The activity involves the National Library of France as a pilot site that provides the preservation metadata from their operational repository and deposit systems. The project consists of a quality analysis of the real-life preservation metadata (METS/PREMIS) used by the pilot site, and intends to demonstrate the value of preservation metadata in mitigating risks by aligning the PREMIS Data Dictionary to risk factors. An update will be given on the current state of the project with particular emphasis on the initial outcomes. | Titia Van der Werf, OCLC Research |
| 17:05-17:30 | **Implementations** | |
| 17:05-17:25 | National Library of Sweden’s implementation of PREMIS. The experiences of a newcomer in implementing PREMIS. | Stina Degerstedt, National Library of Sweden |
| 17:25-17:45 | PREMIS usage in PSNC-developed dArceo long-term preservation services. The works were conducted in scope of an ongoing Polish national R&D SYNAT project, but it is important to note that dArceo is used in production mode by several institutions in Poland. | Tomasz Parkoła, Poznan Supercomputing and Networking Center (PSNC) |
| 17:45-18:05 | Royal Library of Denmark's implementation of PREMIS. An experienced practitioner’s perspective. | Eld Zierau, The Royal library of Denmark |
| 18:05-18:25 | Archivematica’s implementation of the PREMIS 2.2 rights section, and describe how values in the PREMIS rights entity are used to automatically apply access restrictions on DIPs uploaded from Archivematica to public access systems. | Courtney Mumma, Artefactual |
| 18:25-18:45 | Implementation of PREMIS at the University of Rome. In particular, the implementation of their metadata content models and how data is modelled for providing a connection to other contextual information, and consequently how PREMIS semantics are supporting the information collecting for unleashing their preservation strategies into the repository. | Angela Di Iorio, Sapienza Università di Roma |
| 18:45-19:05 | Preservation of audiovisual digital contents: how to deal with multimedia metadata. The talk will give a summary of the PrestoPRIME and Presto4U experiences and the activity in MPEG Multimedia Preservation AhG. In particular, mapping MPEG metadata not only to MPEG-21 but also to PREMIS and W3C-Prov to support interoperability. If there’s time, then Walter will include a practical exercise on extracting technical metadata from files automatically, how to manage this information in several formats, the missing parts, and what can be done. | Walter Allasia, EURIX |
| 19:05-19:30 | **Questions and Discussion** |  |

The event started with a short introduction of the participants. The interesting difference to last year was that there were many relatively inexperienced users who wanted to use the event to learn about PREMIS rather than as a user exchange. Some mentioned that this goal was achieved for them, even though this was not a tutorial.

### Update on PREMIS activities, Sébastien

Sébastien announced the coming PREMIS 3.0, the data model changes for IntellectualEntities and Environments work, better handling of preservation policies, events additions for detail and extensions, and the OWL ontology. He stated that the conformance working group is discussing what conformance means given that many different non-interoperable solutions count as conformant. The group is gathering real-life examples of PREMIS use to see the variety of use, to use them to understand whether there are different levels of conformance, and to answer questions such as, whether it is sufficient to be conformant only for exchange, what conformance means for the implementation in tools, what conformance means if semantic units are implemented within the METS container, and how OAIS impacts the notion of conformance. Participants were invited to contribute their PREMIS samples.

### Data Model update, Angela

Angela introduced the motivations for changing the implementation of IntellectualEntities and Environments for PREMIS 3.0, elaborated the requirements, gaps and proposed solutions. The main message was that an Environment was going to use the existing Object infrastructure with the addition of being able to specify the type and subtype of the Object and with a richer set of relationships between Objects.

**The user feedback was:**

* There was a request for clarification that Objects were indeed on the same level as their Environment descriptions now.
* There was a question on whether the vocabulary for possible relationships was prescriptive.
* There was approval of the fact that the proposed solution lets us capture networks of environment components.
* There was a question whether the proposed solution addresses how to model a preservation plan and we said that we had excluded this from the scope of this working group. This is further work to be done.
* People found the simple remark that the use of environments does not become compulsory. This observation is an easy way of removing anxieties.
* We discussed that registries should be used but are currently inadequate. The creation of registries is facilitated by the provision of the environment model offered in the proposal.
* We encouraged people to read our journal article and to provide feedback on the proposal.

### Preservation Policy, Eld

Eld explained that preservation strategies and policies can be expressed at a logical or functional level, but also on a bit preservation level. Logical we can have migration, emulation or technology preservation. We want to express that at a specific preservation level. This asks for an addition to the preservation level semantic unit: preservation level type. Eld illustrated several different possible uses of the new preservation level type in order to achieve integrity, confidentiality, availability, etc.

Eld emphasised that preservation level types can be specified in one’s own vocabulary. Strategies associated with the vocabulary can change and need to be adjusted because of technological and policy changes. By specifying preservation level types, one can change the policy without having to change the preservation metadata.

**The user feedback was:**

* The example had 4 level types. Can institutions choose for themselves how many they want to use? Yes, a shared vocabulary can grow over time and is application and organisation specific.
* How do you assign preservation level types to a category of Objects? Denmark attaches it to every single Object at a bit level. With the proposed change to Intellectual Entities in version 3 we can now also attach preservation levels at that level to describe a whole category of Objects.
* We clarified that preservation levels (and significant properties) were exceptions that implement business requirements. In general business requirements would not be contained in an interchange format, but they are actually a very important form of provenance information that explains why the Objects were preserved in the way they were. Preservation level information for bit preservation may give you enough information to roll back faulty preservation actions.
* Is the preservation level meaningful to others? Yes it is if you refer to the preservation policy. But it does differ between institutions.
* Should it be restricted to purely factual assessments? This depends on the intended use in the organisation. It may have to include a value assessment.

### PREMIS Ontology, Sébastien

Sébastien explained that the ontology has been developed since 2010 and has arrived at a public, final version for PREMIS 2.2. Initially, the XML schema was literally mapped to RDF but it was not idiomatically expressed. Therefore, it was refactored to match the RDF philosophy. Sébastien gave a brief overview over RDF and showed a graphical representation of the classes and subclasses and their relationships. Properties can be ambiguous if they have the same names; therefore we introduced distinguishable property names.

The purpose of the ontology is to have a ready-to-use RDF implementation that can be shared; you can use it as a data management interface; you can use it for distributed digital preservation across different repositories; and it can be used to link to other non-preservation databases such as library catalogues or Wikidata if you have RDF on both sites.

Sébastien explained that when the data dictionary states “Value should be taken from a controlled vocabulary” those can be found in id.loc.gov and are PREMIS endorsed but can be extended through organisation specific vocabulary. Sébastien demo’ed the id.loc.gov site and the PREMIS ontology site.

Next steps will involve including the Environment and Intellectual Entity changes, and technical registries and aligning with Prov-O.

**The user feedback was:**

* How do you go from XML to RDF? You use style sheets in a pretty straight-forward way. This approach is used at the BnF.
* What is the main advantage of using RDF? RDF is complementary to XML. XML and RDF are used for different purposes. XML is used for validation and to support knowledge management. RDF is good to support linking and sharing across the internet; it supports data management, is good for querying. One cannot recommend one implementation over the other. The choice has to be with the organisation and depends on their objectives. If you keep both you need a well-defined workflow that keeps both synchronised. There are of course other options as well, such as relational databases. This is true for any metadata and is not PREMIS specific.
* Can everything that is expressed in XML also be expressed in RDF? Yes, but some shortcuts were introduced to be idiomatic and to avoid unnecessary, intermediary relationships.

### The preservation metadata health check, Titia van der Werf, OCLC and OPF

The preservation metadata health check was performed to help preservation managers to establish the health of their collection metadata. It should be monitored automatically based on objective data. The project is trying to determine what health indicators exist and whether preservation metadata is useful. The goal is to track intentional and unintentional change using a dashboard with sensors, thresholds and triggers. The research is performed both top-down and bottom-up. They were working with PREMIS and the SPOT risk-assessment methodology, which both provide properties of successful preservation.

The BnF run a trusted digital repository and have volunteered to run a pilot site.

They mapped PREMIS semantic units to the SPOT model. SPOT defines 6 basic properties for risks associated with it. For example “persistence” is associated with storage medium decay. They performed a mapping of which semantic unit in PREMIS addresses this issue?

Do the PREMIS semantic units address the threats identified by the SPOT property? Some gaps have been identified in terms of understandability. Many are already being addressed by the changes brought in with version 3 and the environment work and with respect to coverage of policies in which PREMIS makes additional provisions explicitly through preservation level types and significant properties.

Remaining shortcomings listed by the group were that:

* PREMIS is designed with the explicit assumption that the repository is a self-contained system and all digital preservation processes are performed in-house. For example, identifiers are created by the repository and no external identifiers are usable. They should be assignable by third-parties. It was clarified in the subsequent discussion that this is a misunderstanding.
* PREMIS does not require explicit encoding of all mandatory field. This may be inappropriate for a third-party. Again, it was clarified that there was a requirement for explicit recording within the repository and a requirement for explicit implementation for exchange.
* Threat assessment varies whether you look at a digital Object, a Collection, or a Repository. It applies at different levels. Collections share environments, for example, but identifiers may need to be captured at the individual Object level.

**The user feedback was:**

* APARSEN have developed in depth best practice recommendations on metadata needed to ensure authenticity. This should be used by the OCLC working group.

### National Library of Sweden’s implementation of PREMIS. The experiences of a newcomer in implementing PREMIS. Stina Degerstedt, National Library of Sweden

For 5 years they have been developing Mimer, a platform for ingest and digital preservation. It can accommodate any form of content. They are learning about PREMIS, METS, MODS. The work is pursued in collaboration with the Swedish National Archive who also use METS and PREMIS. They also have a digitised archive of audio-visual materials and a webarchive. All archives overlap in some respects, but are independent from each other. Extranet.kb.se has a document about how they use PREMIS, explaining also their use of PREMIS in METS and example. They appreciate feedback! The following issues were discussed in depth:

* The data model: The data model was developed using the terminology of PREMIS, which was helpful. Each Object has PREMIS data in the METS; Agents are used
* PREMIS relationship to METS: They distribute
  + IEs to METS dmdSec and the
  + files to amdsec in the METS section.
  + More complex structures are expressed in METS,
  + PREMIS used for preservation, e.g. to describe file formats.
* The redundancy in PREMIS was irritating. Events that belong to all files are not recorded repeatedly.

Right now the focus is on ingest rather than on preservation actions.

The Events are not finalised. Recorded are validation, ensuring that all files are there, metadata has been collected, checksums comparison, etc. (see slides). They are taken from id.loc.gov with their own vocab added.

For identifiers they feel that the PREMIS data dictionary should stress the importance of persistent unique identifiers more and explain how to use global identifiers. For identifying objects and intellectual entities they are using URN:NBN . For other types of entities they are using UUID or creating their own identifier values.

They are not using PREMIS rights.

A big issue is still to get organisational buy-in for the need of digital preservation metadata. They are running some tutorials now.

Next steps are testing, improvements to finish details, preparing for preservation planning and preservation actions. And, to do more work on sharing preservation metadata through RDF triples, having approached the LOD people at the library working on the national catalogue.

**The user feedback was:**

* Using RDF triples for collaboration on content so that preservation metadata can be exposed is interesting.

### Archivematica’s implementation of the PREMIS 2.2 rights section, and describe how values in the PREMIS rights entity are used to automatically apply access restrictions on DIPs uploaded from Archivematica to public access systems. Courtney Mumma

Artefactual’s products are Archivematica (free and open-source digital preservation) and ICA Atom (Access system – access to memory). Courtney gave a brief overview over Archivematica which implements the OAIS functionality and fills some additional gaps. They create AIPs and DIPs: normalise, produce high-quality output, use BagIt for packaging, creating METS with the 15 DC fields and PREMIS data. Archivematica is not a digital repository or an access system.

All information about their PREMIS use is on their wiki.

You start by ingesting metadata as csv files that can contain DC and any other metadata. DC and PREMIS rights can also be entered via a dashboard interface. A resulting bag contains the ingested digital Objects, the preservation version of the Objects, the submission documentation, logs that are not covered in PREMIS, METS file (dmdSec contains DC, administrative metadata contains the techmd for the PREMIS Object , digiprovMD for PREMIS Events and PREMIS Agents, rightsmd containing PREMIS rights.

There are 3 Agent groups.

* Users can be multiple for an Object since there can be multiple passes in processing an Object.
* The system
* The repository

PREMIS events are already shared with the community; new Events are “image capture” for forensic image capture and “registration” for accessioning which assigns an accession number to an Object. “quaranteen”, “unquaranteen” are not yet used. Other familiar ones can be found in the slides and on the website. She showed an example of all PREMIS events used that can be seen at their website.

Rights are “copyright”, “licenses”, “donor” and “policy”. Acts are user determined. Grants and restrictions are “allow” and “disallow”, but also “conditional” requiring some form of authentication or any other action. The have a dashboard template in which you can enter rights using dropdown menus. You can add several acts for each rights basis, as well as Agents.

With the Archivists Tool Kit paid for by Rockefeller Archive Centre they have made PREMIS rights actionable. Rights are entered into Archivematica during ingest and get uploaded to the ATK and can be configured on the dashboard in the administration tab. PREMIS rights info entered before normalisation can be used to populate the ATK database. Notes entered can be used to populate conditions of use. You can specify “Restrictions Apply” with the value options [yes from the dashboard/no/based on PREMIS -so that you don’t have to specify it in the dashboard].

You can validate PREMIS in METS, but this is not working for Rights in version 2.2 in the PREMIS in METS validator (not updated for 2.2).

The METS files are indexed by a search engine and everything in PREMIS is searchable by the dashboard and returns the metadata and the relevant files.

### PREMIS usage in PSNC-developed dArceo long-term preservation services. Tomasz Parkoła, Poznan Supercomputing and Networking Center (PSNC)

The work was conducted in the scope of an on-going Polish national R&D SYNAT project, but it is important to note that dArceo is used in production mode by several institutions in Poland.

In 2002, the first digital library in Poland was based on dLibra. There are more than 100 digital libraries in Poland providing 1.7 m digital Objects, mostly also available in Europeana. dArceo was started 2-3 years ago to help the institutions with long-term preservation as a collaborative effort of institutions. They also provide several digital preservation tools: dLibra, dMuseion, dLab and dArceo for long-term preservation. The tools are based on their experience in participation in collaborations across Europe and learning from others. He displayed their digitisation workflow.

They use METS and PREMIS for Files, Events and Objects. Characterisation is captured in PREMIS, but they don’t use it to the full extent. There is technical metadata. There is an extensible service for adding migration services, that is used in production mode.

On the Object level PREMIS metadata is embedded into METS. They have their own identifiers. There are Events for transformations and modifications. On the File level, the identifier is the File name, characterisation information and associated Events.

Their experience:

* Duplication of technical metadata is possible in METS and PREMIS. The guidelines do not give clear rules. Greater interoperability is wanted and should be better supported; they would like to be offered exactly one approach.
* Potential inconsistencies also come from being able to use file format registries as well as file format designation as well as objectCharaceristicExtension which allows you to record inconsistent information. They would like to see clear guidance. Extension allows for overwriting PREMIS fields,
* Vocabulary should be enforced. It is appreciated that others have different requirements. But smaller institutions just want to be told what vocabulary to use for different field.
* The vocab starter list is not aligned with OAIS: e.g. migration should be called transformation.

**The user feedback was:**

* You can lose users if you are too prescriptive. But they want to have clear default vocabulary.
* Look not just at the data dictionary tables but also at the supporting text. A solution would be to increase cross-references, e.g. links from DD entries to the plain text!
* If there are ambiguities of use there should be a clear explanation about which choice to use under what conditions.
* It is the user’s responsibility not to store inconsistent metadata. Extension semantic units should not be used to overwrite native PREMIS semantic units. This is recorded in the data dictionary. – There should be a clear statement which fields have to be consistent.
* Inconsistencies for format information are sometimes desired to capture contradictory characterisations from different tools or to capture complimentary format information. There are guidelines for how to handle these situations in the data dictionary. Here again, increased DD internal cross-references should tackle the problem.
* Users should please let the Editorial Committee know which parts of the Data Dictionary are not clear so that we can revise that in version 3.

### Royal Library of Denmark's implementation of PREMIS. An experienced practitioner’s perspective. Eld Zierau, The Royal library of Denmark

Eld discussed the factors influencing the use of PREMIS while building a new digital library infrastructure at The Royal Library of Denmark. Preservation, dissemination and management are different but all need to curate and share the metadata. They use articles on the LoC website for how to decide the mix between METS and PREMIS. They oriented themselves on the Australian way of splitting out the PREMIS data across the METS fields (descriptive, technical, and technical for different content types). Rights metadata is not yet determined. For file format information the PREMIS metadata overwrites MIX metadata if there should be contradictions between the two of them.

Eld showed an example using MODS as descriptive metadata and techMD consisting of the Object identifier and the MIX. As provenance metadata she showed preservationLevel examples, Events example and Agent example. She will share those with the workshop materials.

At iPres 2012 Eld explained why they use WARC. Files need to be linked to their IDs. File names go with the file system and should not be used. Putting the identifier into the File would require changing the File. They decided to wrap the File itself and its identifier in WARC. The identifier needs to be at a logical level so that people can come back in 50 years and require an Object using its ID and to request a certain version you want (original, newest x years ago). They use Intellectual Entity identifiers for this.

Eld showed a demo. In the WARC header file they define the vocabulary. Instead of using id.loc.gov they keep a local vocabulary id.kb.dk as their own domainwith their own local additions. They have a date and a harvester to harvest id.loc.gov. They use type kbinternal to specify identifier types.

**The user feedback was:**

* Why are you using WARC instead of BagIt to encapsulate the Files? They looked at the 10 different most used packaging formats and chose WARC based on different preservation criteria derived from their requirements. This is documented in the iPres 2012 paper.
* Sweden also mints their own vocabulary locally. It would be nice if people did not create their own vocabularies but shared it all with each other so communities can reuse the same vocabulary. ID.loc.gov was to support a common vocabulary. What is the state of this? Will this become available as a resource?
* Default vocabulary: PREMIS XML does not let you explicitly say that you use a certain set of vocabulary. Like in MODS: Authority = to identify the list from which you have taken it.

### Implementation of PREMIS at the University of Rome. Angela Di Iorio, Sapienza Università di Roma

This talk focused on the implementation of their preservation metadata content models and how data is modelled for providing a connection to other contextual information. This allows for linking preservation strategies to the repository. Their implementation is not yet published. They are building a digital library infrastructure based on standards. She showed screenshots of the prototype comprising many different content types, such as text, images, archives, maps, for a very large number of organisations of different types (60 libraries, 20 museums). The goal is to define a shared preservation metadata model and a shared pre-ingest workflow for creating a consistent SIP. They have an agreement with the Italian Supercomputer Centre that serves all Italian universities as technological partner. The Centre would eventually also like to make the infrastructure available to other universities. The current system is geographically distributed, with parts in Rome. The archival management and dissemination service to the system is located in Bologna. The dark repository is located in Rome.

The metadata definition was determined by investigation of OAIS: provenance, context, fixity, reference, access rights.

PREMIS conformance was the top requirement for every Object in the repository. The workflow is collection aware. The creation of the collection is essential to trigger creation of metadata that is shared and inherited by all Objects (license, identifier system used, copyright). In addition you can have resources and Object specific copyrights information.

Persistent identifiers have been applied from the Object up to the organisational level of the university. The identifier of the organisation is important in order to manage the Agents who have responsibilities, as well as the Rights and Events.

Angela recommends alignment between the PREMIS and the PROV ontologies for real interoperability.

The Object semantic units have now been implemented in the prototype. Semantic units for the other Entity types are in draft form.

Also: a digital Object must belong to a well-defined resource described by metadata, e.g. a personal database. There is also a workflow that applies to all content-types, including their Google project. When an external organisation donates material the original and original file name are preserved. They import all metadata and assign it to a collection in order to trigger the workflow. After that they create message digests and other Object metadata. They map the source metadata to the PREMIS model (normalisation) and then start the automatic processes. Descriptive metadata is encoded in MODS and DC. The container format for SIPs is METS. It results in the automatic construction of relationships between Objects, resources and collections. The Events during the workflow process were listed, as were the roles of Agents. The METS Object is an Object and a Representation of a resource.

**The user feedback was:**

* Ingested databases contain metadata which is mapped to MODS; the original records from the DB are also stored and maintained as CSV files.
* How do you determine what is a collection? This depends on the organisation. It must exist in an organisational unit of the university. The identification of the collection is the starting point for managing all the different collections.
* Do you collect metadata AND content? Yes.
* The password to access the prototype online can be given to workshop participants.

### Preservation of audiovisual digital contents: how to deal with multimedia metadata. Walter Allasia, Eurix together with Werner Bailer , Joanneum

The talk gave a summary of the PrestoPRIME and Presto4U experiences and the activity in MPEG Multimedia Preservation AhG. In particular, mapping MPEG metadata not only to MPEG-21 but also to PREMIS and W3C-Prov to support interoperability. Walter talked about his work on extracting technical metadata from audio-visual files automatically, how to manage this information in several formats, the missing parts, and what can be done.

Walter gave an overview over PrestoSpace: every bit of information it is associated with time. You can discover camera movement or zoom automatically. Lots of information is recorded on the on millisecond level. Data carriers and their players are becoming obsolete. Analogue tapes and film need to be digitised and translated to files. Additionally, you can apply digital actions even if there are no files.

Decisions have to be made about what file format should be used. JPEG2000 is very vulnerable with small bit flips. New compression mechanisms have increased resilience. Identifying the best format is not easy.

In PrestoPrime, the data model based on METS and PREMIS, a DC subset for descriptive metadata, MPEG-7 Audiovisual description profile, MPEG-21 MCO (Media Contract Ontology) GPLve. All OAIS services are based on web services.

The PrestoSpace model uses P\_Meta and MPEG-7 part 9 –AVDP. This is a quite new profile for the temporal composition for face recognition, speech, etc.

Purely technical checks can detect stream/format compliance, storage errors, transfer errors, - but many errors cannot be detected automatically. Severe visual defects need to be detected so that one can improve the best migration paths.

Mxf can provide only generic information. Several tools can go inside the wrapper to go inside the codecs and extract information. One has to decide where to put this information as an MXFDump. They moved to MPEG to get a solution for where to store this information. The MPEG-21 family of MPEG standards tries to address how to document, how tow to determine use, how to enrich, and how to handle rights.

The question he tried to answer is how one can profile this in PREMIS and how to use PREMIS for this. The usual answer is that PREMIS can be the high-level layer. Bottom layer is used through extensions.

### Several issues were brought up repeatedly:

* Some people wanted to have a strictly recommended and sharable vocabulary while others were creating local copies of the PREMIS vocabulary and extended it locally. We pointed them toward id.loc.gov. It would be very good if users could add and tag their own vocabulary there, so that they can share vocabulary amongst each other.
* Many people commented that certain things were not explained in the Data Dictionary. It turned out that people only use the tables and not the remaining text in the dictionary. It would be beneficial if we presented the information in an easily understandable hyper-linked way that gives a complete overview of existing information sources, including the PREMIS in METS use. There did not seem a need to have much more documentation, but rather that users were not finding or noticing it.
* There was a repeated request to form an AV user community and to use AV examples explicitly in the Data Dictionary. Possible partners of the community might be Beth Delaney, Eurix, Norway, Tate, Vicky Phillips, DAITSS. Things are much more complicated on the technical metadata side than for text or images, where there is a common consensus on the metadata format (MIX or textMD). Here we have multiple ones (e.g. for video, videoMD, PBCore, EBUCore, P\_META, MPEG-7…). In such a context, a need for discussion between implementers is especially important. In such a context, PREMIS could be particularly valuable as a common high level layer.
* Users request solutions to expressing preservation plans.
* Users are requesting a definition of quality in metadata.

My impression was that we are dealing with a new generation of users. The big memory institutions have dealt with PREMIS and are now using it in a business-as usual sort of way. We are now seeing smaller organisations moving towards it, as well as consortia with very varied smaller organisations. Their needs are different and we need to communicate differently with them.

### From Sébastien:

This was good. The attendance was mainly composed of mainly small to medium institutions. This seems to mean that PREMIS has become business as usual in the big institutions, and is spreading.

The audience was very active and participative, which was a very good thing!

About the questions and discussions, here are the first things that come to my mind.

Implementation questions were asked about the articulation between formatRegistry and formatDesignation (should they be consistent), and how the extensions should be implemented when some finer-grained schema partially maps some existing PREMIS semantic units (e.g. objectCharacteristicsExtension which provides a <size> field).

Those considerations are actually expressed in the Data Dictionary, but the people asking had used the Data Dictionary. This means we should check how explicit this is, and if there is sufficient cross-referencing in the Data Dictionary so that implementers can find a topic easily.

There were also questions about the new preservationLevelType should be used, especially about the genericity of the data dictionary entries and the fact that people are free to define their own set of values.

There was also some discussion about the environment evolutions. The important thing that came out of this (I think) is that we need to stress to the implementers is that it is an extra feature, that need not be implemented if one wants to describe only primary objects. You can keep on using PREMIS like before, the only difference is, you can use those environment features if you want /need to.

There was also a suggestion about having dedicated fields / or mechanisms to express the controlled vocabulary someone used. E.g. I am using this specific list of eventTypes, and I want to mention it directly in the PREMIS description that the value I use comes from vocab XXX.

It is plausible that some PREMIS controlled vocabularies will emerge in the next years that would complement the ID LC controlled vocabularies for specific values / specific communities.

There were questions about the articulation between PREMIS as XML and PREMIS as RDF. Some institutions are moving towards RDF or considering it (e.g. National Library of Sweden).