

PREMIS: Preservation Metadata Implementation Strategies Tutorial
June 21, 2007 at the Library of Congress
Report to DISC by Beth Black

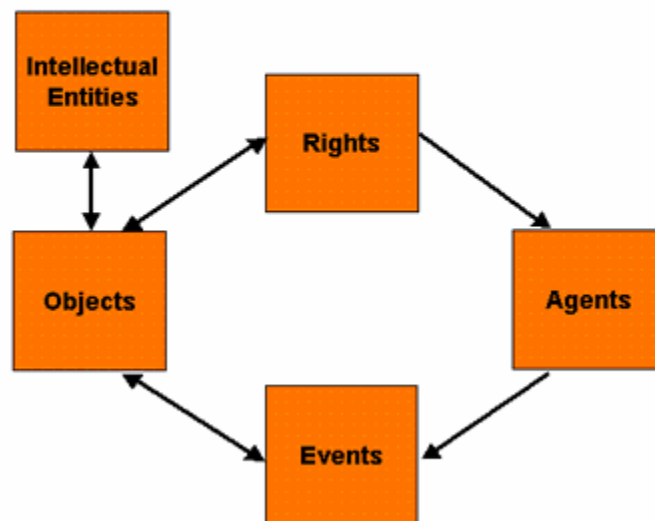
This one day workshop established the definition and scope of the data dictionary, discussed the PREMIS data model, identifiers and relationships, introduced the semantic units, discussed major implementation issues including ways of representing PREMIS in XML, and ended the day with representatives of 3 organizations sharing their experiences with PREMIS. I left the workshop with a much better understanding of what information we need about the digital objects we are promising to preserve.

The PREMIS working group was formed in June, 2003 to define implementable, core preservation metadata with recommendations for management. Their work was built on the Preservation Metadata Framework released in June 2002. The PREMIS data dictionary was released in May 2005. They are currently working on the 2nd version of the data dictionary and expect to release it 4th Quarter 2007.

What PREMIS is: common data model for organizing/thinking about preservation metadata; guidance for local implementations; standard for exchanging information packages between repositories. What PREMIS is not: out of the box solution; include all of the needed metadata (It does not include business rules, format specific technical metadata, descriptive metadata for access or non-core preservation metadata.); lifecycle management of objects outside repository; rights management.



PREMIS data model



PREMIS defines pieces of semantic units that describe properties for each of the parts of the data model (entities, the properties of the entities, and the relationships between the entities) except for the intellectual entities. They are not included in the data dictionary because they are covered well elsewhere; also because repositories ultimately

preserve objects not intellectual entities. There are 3 types of objects: file, representation and bitstream. (A bitstream in this model is slightly different from in DSpace. A DSpace bitstream can be either a file or bitstream in the PREMIS model.) The repository can select at which level to define and manage objects and the PREMIS data dictionary provides the semantic units for each. The agent object aggregates information about persons, organizations, and/or software associated with rights management and/or preservation events in the life of an object. The Rights entity aggregates information about the statements of rights between the rights holder and the repository giving the repository permission to undertake actions associated with preservation of the object. It is not a full rights expression language. The Events entity aggregates information about an action involving one or more objects.

Semantic units are properties of an entity. It is something you need to know about an object, event, agent or right in order to carry out the repositories digital preservation functions. There are two kinds of semantic units: 1) container that groups together related semantic units and 2) semantic components that are semantic units grouped under the same container. A semantic unit is NOT a metadata element. (A metadata element is an implementation decision.) The data dictionary is primarily a description of the semantic units and their appropriate uses based on the model.

The above notes outline the basics of the model. The next part of the workshop detailed the semantic units and their use with the entities in the model. For sake of brevity, I will not include that information here but I'm happy to share my notes with those who want to learn more about PREMIS.

The last part of the workshop, which outlined how organizations are using PREMIS, showed that most are using PREMIS with METS, although the model is designed to be technology agnostic. While all of the presenters described different ways of using PREMIS, they all struggled with how to incorporate PREMIS with other tools. For example, there is duplication between METS and PREMIS that can be difficult to reconcile. One presenter wished that PREMIS were more modular to make it easier to use just parts of the model. My overall impression after this section of the workshop is that PREMIS is a valuable tool but there is still need for experimentation to find the best ways to implement it. [For those interested, the presenters who described their PREMIS implementation projects were Justin Littman, Library of Congress Office of Strategic Initiatives; Michael Kaplan, ExLibris; Kate Zwaard, GPO; Tom Habing, UIUC.]