Metadata to Support Long-Term Preservation of Digital Assets: PREMIS and its Use with METS

Rebecca Guenther
Sr. Networking & Standards Specialist,
Library of Congress
rgue@loc.gov

US Workshop on Roadmap for Digital Preservation Interoperability Framework
March 31, 2010
Outline of presentation

- PREMIS background
- PREMIS community building
- METS as an information package for objects and metadata
- METS community building
- Development of PREMIS in METS guidelines
- Implementation of an exchange standard
  - PREMIS in METS toolbox
  - Controlled vocabularies
- Benefits of using PREMIS in METS
- Future PREMIS developments
PREMIS Data Dictionary

- **May 2005:** *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*

- **March 2008:** *PREMIS Data Dictionary for Preservation Metadata, version 2.0*

**Data Dictionary:**
- Comprehensive view of information needed to support digital preservation
  - Guidelines/recommendations to support creation, use, management
  - Based on deep pool of institutional experiences in setting up and managing operational capacity for digital preservation

- XML schema to support implementation
- Maintenance by PREMIS Editorial Committee; LC managing agency
- Wide implementation world-wide

http://www.loc.gov/standards/premis/v2/premis-2-0.pdf
What does PREMIS cover?

- *Information a repository uses to support the digital preservation process*
- Provides information to help manage a resource for preservation purposes
  - Technical characteristics
  - Information about actions on an object (provenance)
  - Relationships (structural and derivative)
    - Structural: indicates how compound objects are put together
    - Derivative: results of common preservation actions
  - Rights metadata associated with preservation
Scope and principles

- “Implementable, core, preservation metadata”:
  - “Core”: What most preservation repositories need to know to preserve digital materials over the long-term; not format specific
  - “Implementable”: rigorously defined; supported by usage guidelines/recommendations; emphasis on automated workflows

- “Technical neutrality”:
  - Digital archiving system: no assumptions about specific archiving technology, system/DB architectures, preservation strategy
  - Metadata management: no assumptions about whether metadata is stored locally or in external registry; recorded explicitly or known implicitly; instantiated in one metadata element or multiple elements
  - Promotes flexibility, applicability in wide range of contexts
PREMIS Data Model
PREMIS Community Building

- PREMIS Editorial Committee with Library of Congress
  - International group of PREMIS experts
  - LC, OCLC, Florida Center for Library Automation, MIT, British Library, NARA, ExLibris, Library and Archives Canada, Swedish National Archives, Fondazione Rinascimento Digitale, Saxon State Archives (Germany)
- PREMIS Implementers Group List
- PREMIS Implementation Registry
- Tutorials and Implementation Fairs
METS Background

- XML based
- Describes the structure of digital objects and associates various kinds of metadata with their components
- Uses the XML Schema facility for combining vocabularies from different Namespaces for extensibility
- Metadata is categorized into separate sections (embedded or linked)
- Records the names and locations of the files that comprise those objects (embedded or linked)
- Records a map of hyperlinks between components
- Associates executable behaviour with the components
The structure of a METS file

- **METS**
  - fileSec: file inventory
  - dmdSec: descriptive metadata
  - amdSec: administrative metadata
  - behaviorSec: behaviour metadata
  - structMap: structural map
METS Community Building

- Developed as an initiative of the Digital Library Federation
- Maintenance at the Library of Congress in conjunction with the METS Editorial Board
- METS Editorial Board:
  - Discussion list, implementation registry and wiki
  - Wide implementation, especially in cultural heritage institutions
METS with PREMIS as an OAIS Information Package

- OAIS repository functions for which METS is often used are submission or exchange (SIP), archiving (AIP), dissemination (DIP)
- A METS package is a good candidate for realization of an information object in an OAIS repository
- PREMIS satisfies need for Preservation Description Information: provenance, context, reference and fixity
- PREMIS is an elaboration and translation of OAIS information model into implementable semantic units
OAIS, METS and PREMIS

Legend

Black Arial = OAIS
Red Times New Roman = METS Primary Schema
Blue Times New Roman Italics = Extension Schema
Why do we need guidelines for using PREMIS with METS?

- Contents of each information package may vary depending on its function within a repository
- Need to determine how to include representation metadata and associate it with package components
- PREMIS data entities (objects, events, rights, agents) do not map perfectly to METS categories for representation metadata (techMD, digiProvMD, rightsMD, sourceMD)
- There are redundant elements between the two standards
- Both have extensibility mechanisms
- Flexibility of both standards requires implementation choices
- Predictability will enhance the ability for exchange with minimal human intervention
Development of *Guidelines for Using PREMIS with METS for Exchange*

- PREMIS in METS Guidelines Working Group
  - Consists of PREMIS and METS experts
  - Focuses on the METS document as a mechanism of exchange of digital objects and their metadata (SIP or DIP)
  - Facilitates communication when internal requirements and technical environments vary
- Tension between flexibility and being prescriptive to facilitate interoperability
  - Consider usage scenarios
  - If a SIP it may get unwrapped and stored in different structures
  - If a DIP it is converted from internal structures to PREMIS
  - A more liberal approach is possible for a SIP than a DIP
- Group is establishing guidelines with examples of uses and a METS profile

Implementing an Exchange Standard

- PREMIS in METS Toolbox
  - Developed by Florida Center for Library Automation
  - Converts a METS document to a stand-alone PREMIS document or a stand-alone PREMIS document to a PREMIS in METS document
  - Generates a PREMIS description using DROID for format identification and JHOVE for validation and extraction of technical metadata
  - Text, image, audio, video and software

- Controlled vocabularies
  - Many PREMIS elements recommend use of a controlled vocabulary
  - Mechanism for documenting controlled vocabularies under id.loc.gov
  - Initial offering for PREMIS will be cryptographic hash functions, preservation events, preservation roles
Benefits of using PREMIS in METS

- Packages together metadata necessary for digital preservation in a predictable format
- PREMIS provides technical and event metadata
- METS provides structural metadata
- Both standards are
  - Openly available
  - Flexible
  - Extensible
  - Maintained by an open process
- Provides an exchange standard between repositories
Future PREMIS developments

- Changes to data model under discussion
  - Provide semantic units for intellectual entities
  - Make environment and significant properties/characteristics separate entities
- Revisions to XML schema
  - Coordinate mechanism for extensibility with METS
  - Includes meta-metadata
  - Allows for more predictability for extensions
- Extensible container for Agents
- Tutorial and workshop at iPres 2010
Conclusions

- PREMIS and METS provide an important part of the infrastructure for preservation repositories in the implementation of preservation metadata.
- Agreements between exchange partners will be needed for efficiency.
- Use of standard formats allows for interoperability between repositories and facilitates building of tools.