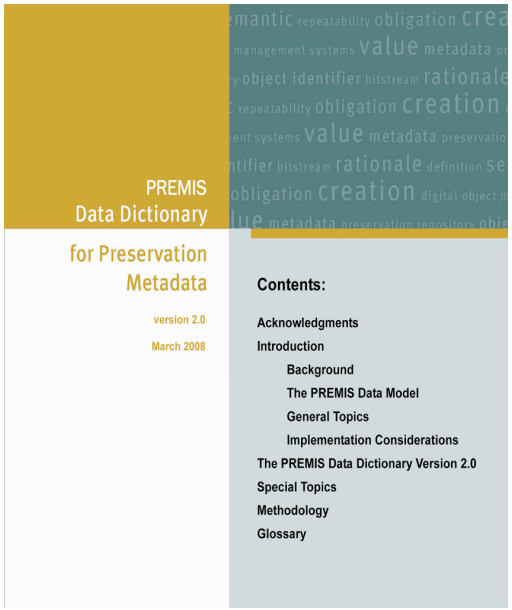
PREMIS PREservation Metadata Implementation Strategies

PREMIS version 2.0: the revision of the preservation metadata data dictionary



Rebecca Guenther, Library of Congress

DLF Spring Forum 2008

Panel: Implementing Preservation Metadata in Digital Libraries

PREMIS PREservation Metadata Implementation Strategies

Panel: Implementing Preservation Metadata: Using PREMIS with METS

- Rebecca Guenther (LC): PREMIS 2.0 revisions
- Rob Wolfe (MIT): Guidelines for using PREMIS with METS
- Nancy Hoebelheinrich (Stanford): Using PREMIS for geospatial data objects
- Ardys Kozbial (UCSD): Transferring digital objects between repositories with preservation metadata
- Tom Habing (UIUC): The ECHO Dep Generic METS Profile for Preservation and Digital Repository Interoperability

PREMIS Data Dictionary: Introduction

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

- 237-page report includes:
  - PREMIS Data Dictionary 1.0
  - Context/assumptions, data model, usage examples

- Set of XML schema to support implementation
- Commitment to keep it stable for at least a year or two

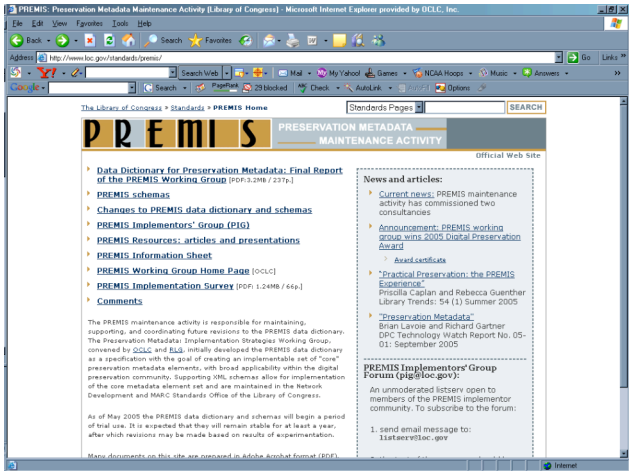
- **Data Dictionary:**
  - Comprehensive view of information needed to support digital preservation
    - Guidelines/recommendations to support creation, use, management
  - Based on deep pool of institutional experiences in setting up and managing operational capacity for digital preservation

http://www.oclc.org/research/projects/pmwg/premis-final.pdf

PREMIS guiding principles …

- "Implementable, core, preservation metadata":
  - "Preservation metadata": maintain viability, renderability, understandability, authenticity, identity in a preservation context
  - "Core": What most preservation repositories need to know to preserve digital materials over the long-term
  - "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 Maintenance Activity



- Web site:
  - Permanent Web presence, hosted by Library of Congress
  - Central destination for PREMIS-related info, announcements, resources
  - Home of the PREMIS Implementers' Group (PIG) discussion list

- PREMIS Editorial Committee:
  - Set directions/priorities for PREMIS development
  - Coordinate future revisions of Data Dictionary and XML schema
  - Membership: Library of Congress, OCLC, FCLA, General Register Office for Scotland, British Library, National Library of Australia, LANL, Library Archives Canada, Koninklijke Bibliotheek, Netherlands

http://www.loc.gov/standards/premis/

Current activities

- First revision of Data Dictionary
  - Version 2.0 was released in early April
  - Changes based on extensive discussions with implementers and on Editorial Committee
  - http://www.loc.gov/premis/v2/premis-2-0.pdf
- Draft schema to be finalized shortly
- Guidelines for using PREMIS with METS
- PREMIS Implementers' Registry
  - http://www.loc.gov/standards/premis/premis-registry.html
- PREMIS Tutorials:
  - Glasgow, Boston, Stockholm, Albuquerque, Washington, San Diego

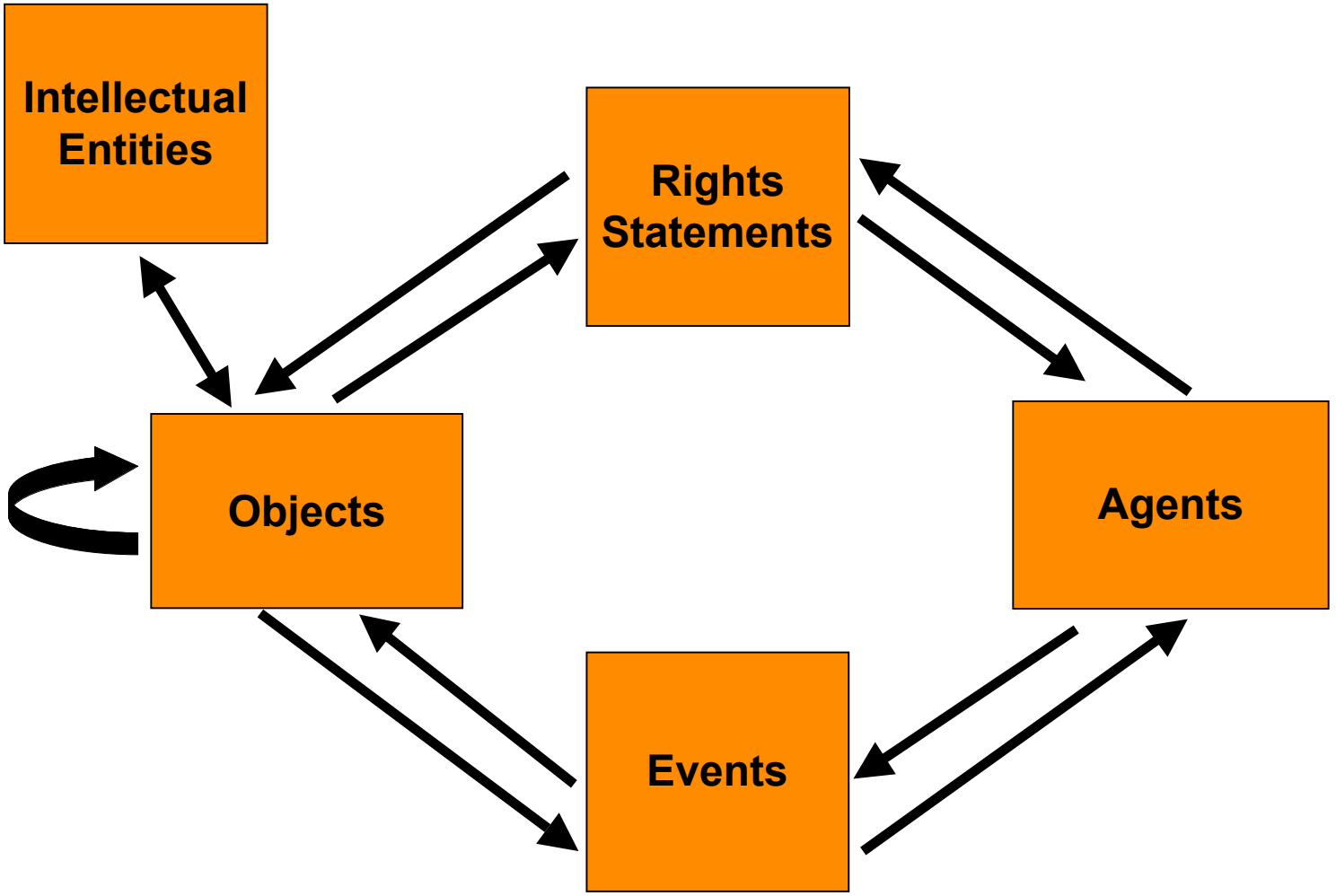
Process of the revision

- Editorial Committee reviewed feedback on the Data Dictionary from both the PIG discussion list and comments sent directly to EC members or LC
- Feedback from tutorials was useful
- Two studies:
  - Recommendations from Karen Coyle's report on rights for preservation
  - Survey report by Deborah Woodyard-Robinson about implementing the data dictionary
- Editorial Committee discussed potential changes over about a year
- Further comments were solicited on particular issues from the PREMIS Implementers' Group

PREMIS Data Model

- Establishes Entities: "things" relevant to digital preservation that are described by preservation metadata (Intellectual Entities, Objects, Events, Rights, Agents)
- PREMIS semantic units are properties of defined Entities
- Relationships between Entities documented through metadata associated with each entity
- In first version, relationships between Rights and Agents and between Events and Agents were defined as uni-directional because of narrow scope for agent metadata
- Now all relationships are defined as bi-directional
- Allows more flexibility for recording these in practice

PREMIS Data Model



Changes to Rights Entity

- Intended to support an automated process that determines if a particular preservation-related action is permissible in regard to an Object(s) within the repository
- Expanded to include a richer description for different types of rights statements
- Revision distinguishes among the following types of intellectual property rights:
  - Copyright
  - Statute
  - License
- Allows for extensibility (external schema)

Changes to Preservation Level and Significant Properties

- Preservation level
  - Related to an institution's policies and capabilities
  - Previously an unstructured unit suggesting a locally controlled vocabulary
  - Now structured to give context information: why, when and in what context the preservation level assignment is made
- Significant properties
  - Characteristics of a particular object that must be preserved for future understandability and renderability, e.g. look, feel, functionality, intellectual content
  - This is subjectively determined by the preservation context and related to an institution's preservation policies
  - Previously an unstructured unit for this information
  - Now structured to include facet (e.g. content, appearance functionality)
  - Current work in this area may influence a future revision

Extensibility

- Allows for use of externally defined structures within PREMIS
  - significantProperties
  - creatingApplication
  - environment
  - signatureInformation
  - eventOutcomeDetail
  - Rights
  - objectCharacteristics
- Consists of a bucket with the name of the container and "extension"
- May supplement or replace PREMIS metadata EXCEPT for objectCharacteristics
- objectCharacteristics may only supplement; used for format specific metadata not defined by PREMIS (so that all preservation related metadata may be kept together)

New work: Controlled vocabularies

- PREMIS defines many units that suggest controlled vocabularies
- LC is establishing a registry for controlled lists of values
- PREMIS suggested value lists will be included as starter lists
- Implementers will be encouraged to add their vocabularies
- Schemas will be generated on the fly to include enumerated values for appropriate elements to allow for validation and consistent application
- Will use Simple Knowledge Organization System (SKOS) for description; this will allow for wide sharing across communities
- Each enumerated value will include a URI (which implicitly defines source of vocabulary), label, definition, broader/narrower terms and alternative labels where applicable
- Test data to be available late summer

PREMIS 2.0 Schema changes

- One schema is used instead of five
- Many elements globally defined to allow for reuse
- Abstract object type allows for better validation of object category (representation, file, bitstream)
- Extensibility mechanism available to provide for further structure or schemas from other namespaces when needed
- Mechanism to enable use of controlled vocabularies coming
- Changes to accommodate version 2 of data dictionary
- Date/time definition to profile and extend ISO 8601; this will allow for open dates, questionable dates, basic ISO 8601 syntax (i.e. without hyphens), etc.
- Currently in draft, but will be finalized shortly