Implementing preservation metadata: PREMIS conformance and best practices

Rebecca Guenther, Chair, PREMIS Editorial Committee/Library of Congress

Evelyn McLellan, Member, PREMIS Editorial Committee/Artefactual Systems Inc.

Digital Preservation 2014
July 23, 2014

PREMIS Data Dictionary

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

- **April 2012:** version 2.2

- XML schema to support implementation

- **Data Dictionary:**
  - Comprehensive view of information needed to support digital preservation
    - Guidelines/recommendations to support creation, use, management

http://www.loc.gov/standards/premis/premis-2-2.pdf

Scope

- What PREMIS DD **is**:

  - Common data model for organizing/thinking about preservation metadata

  - Guidance for local implementations

  - Standard for exchanging information packages between repositories

  - Compatible with the OAIS reference and information model

- What PREMIS DD **is not**:

  - Out-of-the-box solution: need to instantiate as metadata elements in repository system

  - All needed metadata: excludes business rules, format-specific technical metadata, descriptive metadata for access, non-core preservation metadata

PREMIS Data Model
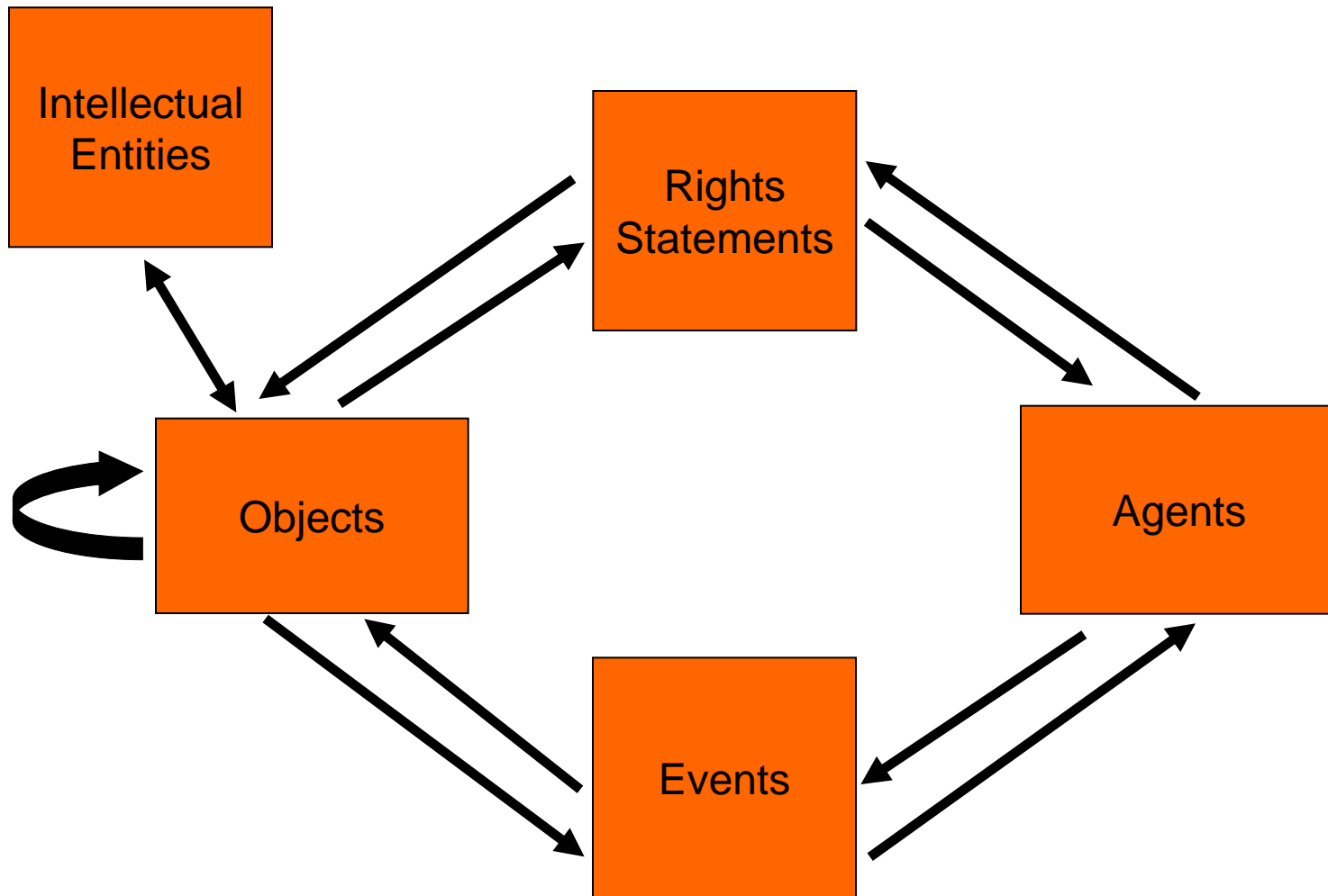
Objects entity

- Object identifier
- Preservation level
- Significant characteristics
- Object characteristics
  - fixity
  - format
  - size
  - creating application
  - inhibitors
  - object characteristics extension
- Original name

- Storage
- Environment
  - software
  - Hardware
- Digital signatures
- Relationships
- Linking event identifier
- Linking rights statement identifier

Events entity

- Event identifier
- Event type (e.g. capture, creation, validation, migration, fixity check, ingestion)
- Event dateTime
- Event detail
- Event outcome
- Event outcome detail
- Linking agent identifier
- Linking object identifier

Agents entity

- Agent Identifier
- Agent Name
- Agent Type
- Agent Note
- Agent Extension
- Linking Event Identifier
- Linking Rights Identifier

Rights entity

- Rights Statement
  - Rights Statement Identifier
  - Rights Basis
  - Copyright Information
  - License Information
  - Statute Information
  - Other Rights Information

- Rights Granted
  - act
  - restriction
  - termOfGrant
  - rightsGranted
- Linking Object Identifier
- Linking Agent Identifier
- rightsExtension

Why do we need a conformance statement?

- Technical neutrality of PREMIS
- Contexts in which conformance is important
  - Inter-repository exchange
  - Repository certification
  - Shared registries
  - Automation/reusable tools
  - Vendor support

PREMIS Conformance

- Conformant Implementation of the PREMIS Data Dictionary http://www.loc.gov/standards/premis/premis-conformance-oct2010.pdf

- What does "being conformant to PREMIS" mean?

- Conformant at which level?

  - **semantic unit:** conformant implementation of the information defined in a particular semantic unit

  - **data dictionary:** conformant implementation of all semantic units

- Conformant from what perspective?

  - **internal:** conformant implementation at semantic units and data dictionary levels

  - **external** (exchanging PREMIS descriptions): import and export

PREMIS conformance degrees of freedom

- What am I free to do now?

  - **naming**: using names that are different from the data dictionary

  - **granularity**:

    - a single metadata element can aggregate semantic units
    - information from a semantic unit can be split in multiple metadata elements

  - **level of detail**: adding more detailed information than the data dictionary

  - **explicit recording of mandatory semantic units**: need not be recorded BUT this information must be recoverable

  - **use of controlled vocabularies**: it is recommended to use controlled vocabularies

New PREMIS Conformance Working Group

- Reconsider conformance and possibility of tightening it
- Consider the benefits of conformance and advantages of more stringent levels
- Link metadata with functionality and examine different priorities
- Develop a list of core preservation activities

Draft new conformance statement

- Three levels of conformance:
  1. mapping
  2. export
  3. direct implementation
- Refinement within each level:
  A. Object entity only
  B. Object plus Events and Agents

Level 1: Conformance through mapping

- Level 1A: Object entity only
  - A repository uses one or more internal preservation metadata schemas, **elements of which can be mapped to PREMIS**. Such mapping must satisfy the principles of use at both the semantic unit and Data Dictionary levels. The repository is able to produce documentation demonstrating such mapping, at a minimum, for the **mandatory elements of the Object entity**

Level 1: Conformance through mapping con't

- Level 1B: Object, Event and Agent entities
  - A repository uses one or more internal preservation metadata schemas, **elements of which can be mapped to PREMIS**. Such mapping must satisfy the principles of use at both the semantic unit and Data Dictionary levels. The repository is able to produce documentation demonstrating such mapping, at a minimum, for all **mandatory elements in the Object entity; one or more agents; and sufficient Event metadata to document actions the repository has taken to preserve the digital objects**

Level 2: Conformance through export

- Level 2A: Object entity only
  - A repository uses one or more internal preservation metadata schemas, **elements of which can be exported as PREMIS**. Such export must satisfy the principles of use at both the semantic unit and Data Dictionary levels. The repository has established processes and tools in place to perform the exports as a routine operation, and is able to demonstrate such capability, at a minimum, for the mandatory elements of the Object entity.

- Level 2B: Object, Event and Agent entities
  - Same, except includes Object entity, Events and Agents.

Level 3: Conformance through internal implementation

- Level 3A: Object entity only
    - A repository **implements the PREMIS Data Dictionary as an internal metadata schema** in a way that satisfies the principles of use at both the semantic unit and Data Dictionary levels and **in a form that does not require any further mapping, conversion or export**. The repository implements, at a minimum, the mandatory elements of the Object entity

- Level 3B: Object, Event and Agent entities
    - Same, except includes Object entity, Events and Agents

Expanding on the conformance statement

- "…and sufficient Event metadata to document actions the repository has taken to preserve the digital objects".

- What does that mean? How can you tell?

  - A good start for Events is the Library of Congress PREMIS EventType controlled vocabulary at http://id.loc.gov/vocabulary/preservation/eventType.html

  - The PREMIS Conformance Working Group is looking both to expand that list and to identify key Events which would comprise a minimum "best practices" list

Current LoC EventType terms

- Capture
- Creation
- Deaccession
- Decompression
- Decryption
- Deletion
- Digital signature validation
- Fixity check

- Ingestion
- Message digest calculation
- Migration
- Normalization
- Replication
- Validation
- Virus check

Possible new terms

- Bitstream modification
- Dissemination
- Filename change
- File extension change

- Format identification
- Imaging
- Metadata extraction
- Others?

Ongoing work of the Conformance Committee

- Finalize conformance statement
- Re-draft conformance document
- Expand controlled vocabulary for Event types
- Work with the community to build a common body of knowledge about preservation best practices and how to support them through preservation metadata standardization

Questions for consideration

- Am I doing digital preservation if I only implement the Objects entity?
- Does implementing PREMIS mean I'm doing digital preservation?
- Should there be a set of mandatory events?
- Do we need a certification process to assure conformance and what would that process entail?