

PREMIS and Preservation Metadata Standards

**Brian Lavoie
Senior Research Scientist
Office of Research
OCLC**

**DPC Meeting on Preservation Metadata
London
September 8, 2005**

Some background ...

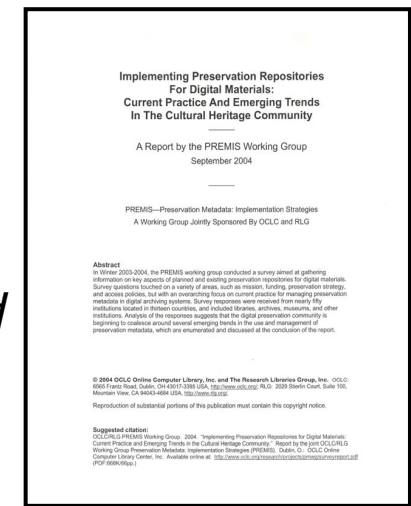
- **Pre-2002:** various preservation metadata element sets released
 - Different scopes, purposes, underlying models/assumptions
 - No international standard; little consolidation of expertise/best practice
- **June 2002:** Preservation Metadata Framework
 - Comprehensive, high-level description of types of information constituting preservation metadata
 - Consensus-based foundation for developing formal preservation metadata specifications ... but not an “off-the-shelf, ready to implement” solution
- **Post-2002:** Need implementable preservation metadata, with guidelines for application and use, relevant to a wide range of digital preservation systems and contexts
 - Motivated formation of **PREMIS Working Group**

PREMIS Working Group

- June 2003: OCLC, RLG sponsored new international working group:
 - ***PREMIS: Pre***serva***tion*** ***M***etada***ta***: ***I***mplementa***tion*** ***S***trateg***ies***
- Membership:
 - **> 30 experts from 5 countries, representing libraries, museums, archives, government agencies, and the private sector**
 - Co-Chairs: Priscilla Caplan (FCLA), Rebecca Guenther (LC)
 - Liaisons: Brian Lavoie (OCLC), Robin Dale (RLG)
- Objectives:
 - Identify and evaluate alternative strategies for encoding, storing, managing, and exchanging preservation metadata
 - Define implementable, core preservation metadata, with guidelines/recommendations for management and use

PREMIS Survey Report

- **September 2004:** *Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community*
- Survey of existing and planned digital repositories:
 - Mission, content, funding, preservation policies/strategies, and more ...
 - Use of metadata to support repository processes, functions, policies
- ~50 responses:
 - 28 libraries, 7 archives, 3 museums, and 11 other
 - 13 different countries; 45% from U.S.
 - Repositories in planning, development, and/or production stages
- Snapshot of current practices and emerging trends related to managing preservation metadata in digital archiving systems



Some findings ...

- Most repositories informed by OAIS concepts, but different interpretations of “OAIS compliance”
- Most repositories support both preservation and access
- Most do not have active preservation strategy in place; little or no basis on which to assess metadata requirements
- Recording range of metadata (e.g., rights, provenance, technical, administrative, descriptive, structural) from variety of sources
- Some common standards emerging: METS, OAIS, MIX
- Maintain multiple versions (e.g., original and migrated) with complete metadata for all versions
- Store metadata redundantly in XML/relational database AND with content

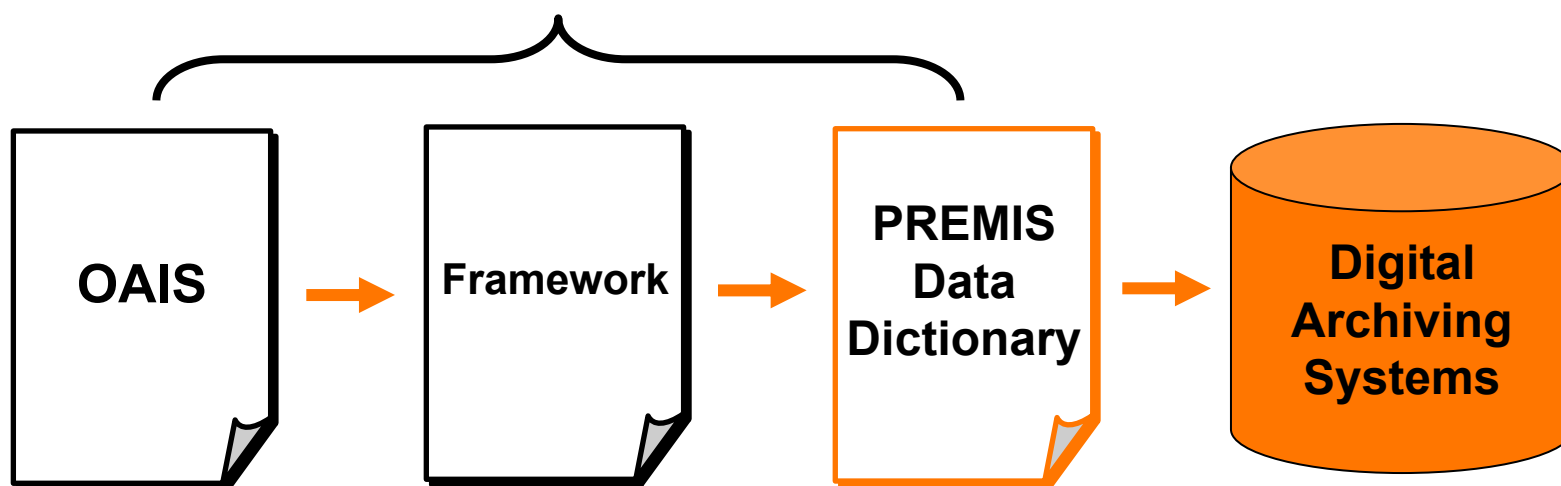
PREMIS Data Dictionary

- **May 2005:** *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*
- 237-page report includes:
 - PREMIS Data Dictionary 1.0
 - Accompanying report (context, data model, assumptions)
 - Special topics, glossary, usage examples
- **Data Dictionary:** comprehensive, practical resource for implementing preservation metadata in digital archiving systems
 - Used *Framework* as starting point
 - Detailed description of metadata elements
 - Guidelines to support implementation, use, management
 - Based on deep pool of institutional experiences in setting up and managing operational capacity for digital preservation
- Set of **XML schema** developed to support use of Data Dictionary



From theory to practice ...

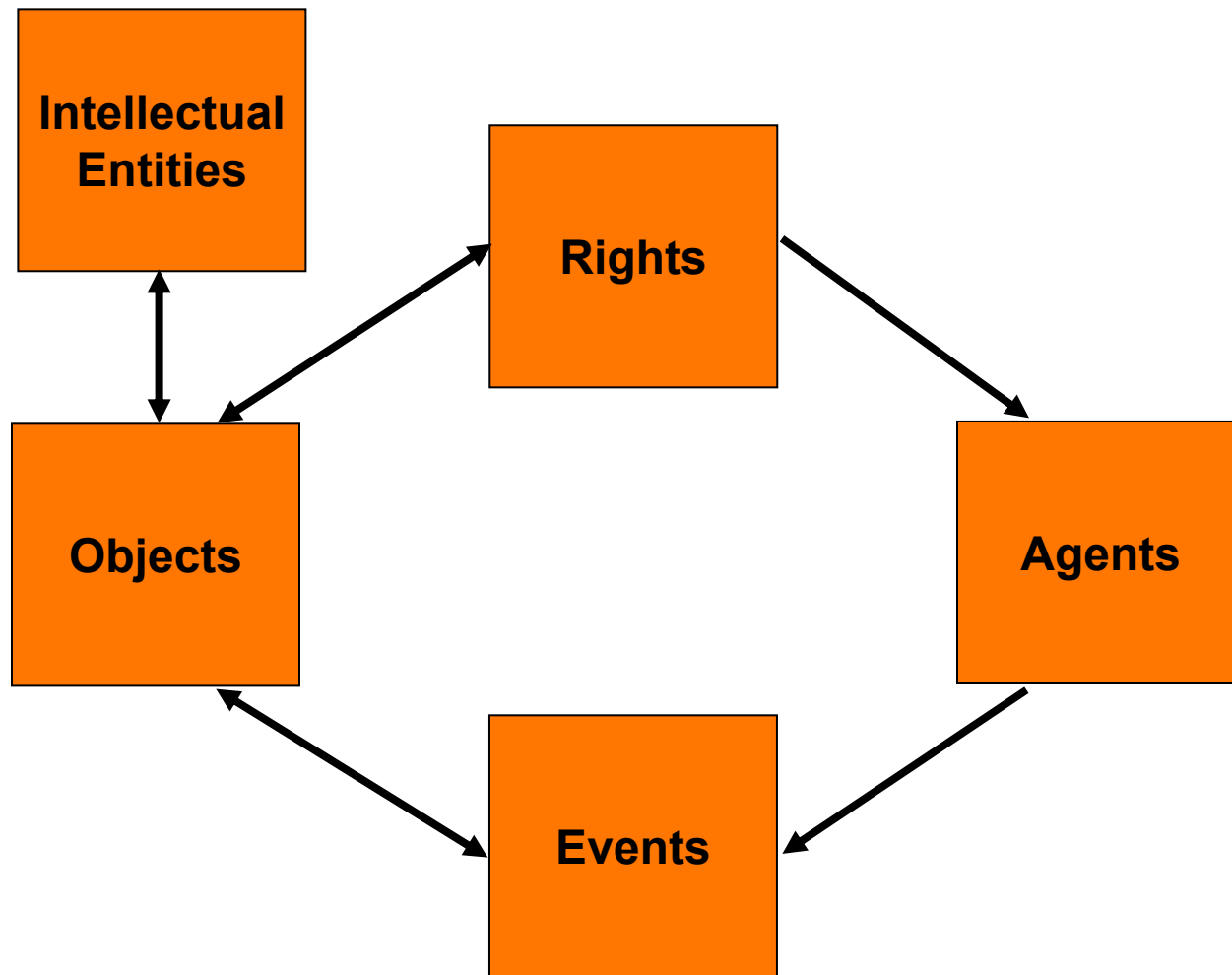
Preservation Metadata Requirements



“Implementable, core preservation metadata”

- “Preservation metadata”:
 - Maintain viability, renderability, understandability, authenticity, and identity in a preservation context
 - Spans descriptive, administrative, structural, technical metadata
 - Implementation neutral: promote flexibility/interoperability
 - Semantic units: what you need to **know** (implementation-neutral)
 - Metadata elements: how you **record** it (implementation-specific)
 - Data Dictionary defines semantic units
- “Core”:
 - What most preservation repositories need to know to preserve digital materials over the long-term
- “Implementable”:
 - Defined as rigorously as possible
 - Supported by guidelines/recommendations on creation/usage
 - Emphasis on metadata that can be automatically created and processed (avoid lengthy textual descriptions)

PREMIS data model



Sample Data Dictionary entry

Semantic unit	size		
Semantic components	None		
Definition	The size in bytes of the file or bitstream stored in the repository.		
Rationale	Size is useful for ensuring the correct number of bytes from storage have been retrieved and that an application has enough room to move or process files. It might also be used when billing for storage.		
Data constraint	Integer		
Object category	Representation	File	Bitstream
Applicability	Not applicable	Applicable	Applicable
Examples		2038927	
Repeatability		Not repeatable	Not repeatable
Obligation		Optional	Optional
Creation/ Maintenance notes	Automatically obtained by the repository.		
Usage notes	Defining this semantic unit as size in bytes makes it unnecessary to record a unit of measurement. However, for the purpose of data exchange the unit of measurement should be stated or understood by both partners.		

PREMIS Maintenance Activity

- Recognition that long-term success requires **evidence of commitment**
- Set up PREMIS Maintenance Activity to support Data Dictionary release:
 - Permanent Web presence hosted by US Library of Congress
- Maintenance Activity:
 - Centralized destination for information, announcements, and other resources related to PREMIS
 - Discussion list for PREMIS implementers
 - Coordinate future revisions of Data Dictionary and XML schema
 - OCLC, RLG, LC, FCLA currently administering Maintenance Activity; governance structure to be released soon; participation open to all

Standards and standardization

- “Standard”: Formal specification ...
 - ISO 15836 (Dublin Core), Z39.87 (MIX), ISO 14721 (OAIS)
- “Standardization”: broader perspective. Given some digital preservation requirement (e.g., metadata), develop shared view on:
 - What needs to be done
 - How it should be done
 - What skills are needed
 - What outcomes are acceptable
 - What interactions will be supported
 - What assumptions can be safely made, etc.
- Weave all of this together to develop **standardized processes**
 - Which are made concrete by a set of interlocking, mutually compatible standards: data models, schema, APIs, etc.
- Economical, sustainable digital preservation

Standardization of preservation metadata

- “Preservation metadata standard” is only first step toward developing standardized processes for creating, collecting, storing, managing, and using preservation metadata
- Shared understanding across all stakeholders on:
 - What information is needed to support digital preservation
 - Who is responsible for collecting it
- Repositories exist in networked environments:
 - Repositories link to registries
 - Repositories link to depositors and users of archived objects
 - Repositories link to each other
 - Support flow of information across repositories, stakeholders, systems
- Support development of common workflows, automated tools, third-party services applicable in multiple contexts
- Some standards emerging: OAIS, MIX, METS ... and PREMIS???

URLs

- PREMIS Working Group:
<http://www.oclc.org/research/projects/pmwg/>
- PREMIS Maintenance Activity:
<http://www.loc.gov/standards/premis/>
- *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group:*
<http://www.oclc.org/research/projects/pmwg/premis-final.pdf>
- *Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community* (survey report):
<http://www.oclc.org/research/projects/pmwg/surveyreport.pdf>