

# The PREMIS Working Group

**On behalf of PREMIS:**

Brian Lavoie

Senior Research Scientist

Office of Research

OCLC

**Presentation to DP Award Judging Panel**

**London**

**September 7, 2005**

## Digital preservation: imperative and challenge

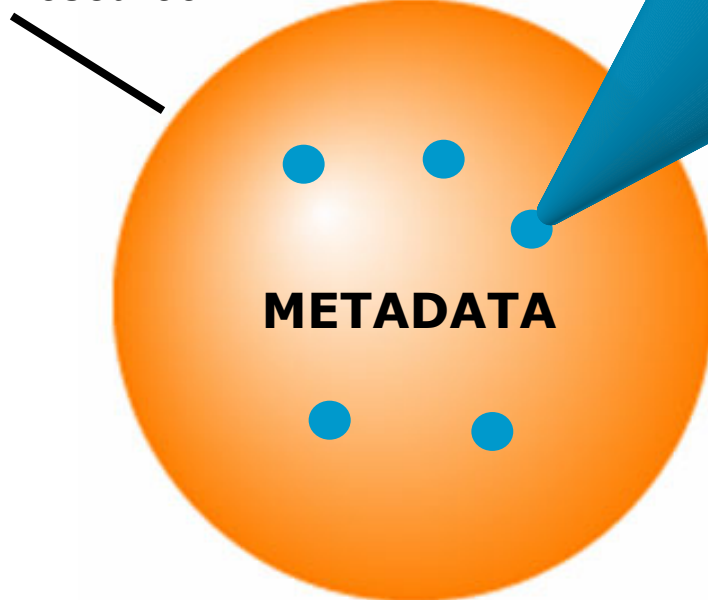
- More and more of scholarly and cultural record exists in digital form; steps must be taken to secure its long-term future
- Groups such as Digital Preservation Coalition have made significant progress in raising awareness about digital preservation imperative
- Gradual shift in focus from **articulating** problem to **solving** it ...
  - Not so much "*Why is digital preservation important*" anymore; rather, "*What must be done to achieve preservation objectives?*"
- Many practical challenges in implementing reliable, sustainable digital preservation programs
- One key challenge: preservation metadata

## PREMIS Working Group

- **A critical need ...** preservation metadata is key component of digital preservation infrastructure
- **A solution ...** *PREMIS Data Dictionary for Preservation Metadata*
  - Consensus-based, implementation-oriented, widely applicable resource
- **Momentum building ...** significant interest in PREMIS Data Dictionary; many institutions already exploring how Data Dictionary can be implemented in their environments:
  - National Archives of Scotland Digital Data Archive Project
  - US NEH National Digital Newspaper Program
  - Los Alamos National Laboratory

## Metadata and **preservation metadata**

*“Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”*

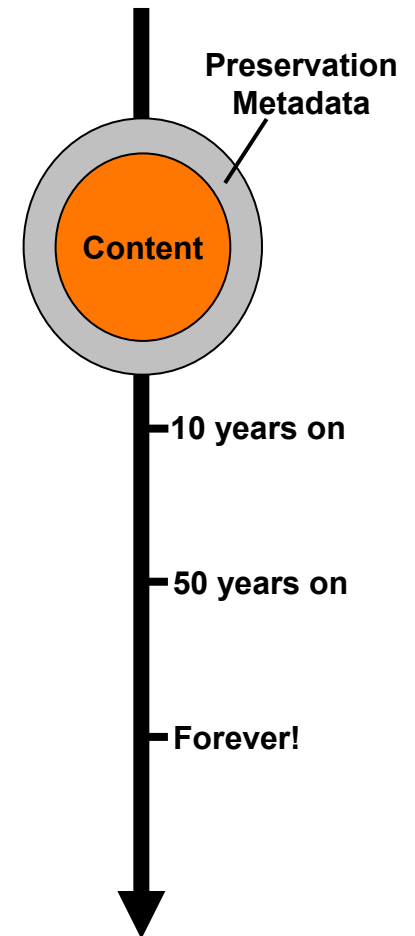


**PRESERVATION  
METADATA**

*“Metadata that supports and documents the digital preservation process”*

## Preservation metadata: Examples

- Provenance:
  - *Who has had custody/ownership of the digital object?*
- Authenticity:
  - *Is the digital object what it purports to be?*
- Preservation Activity:
  - *What has been done to preserve the digital object?*
- Technical Environment:
  - *What is needed to render and use the digital object?*
- Rights Management:
  - *What IPR must be observed?*
- **Makes digital objects self-documenting across time**



## 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
  - International working group (jointly sponsored by OCLC, RLG)
  - Comprehensive, high-level description of types of information constituting preservation metadata
  - Used OAIS reference model as starting point
  - Set of “prototype” preservation metadata elements
  - 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)
- Objective 1: Identify and evaluate alternative strategies for encoding, storing, managing, and exchanging preservation metadata
  - PREMIS Survey Report (September 2004)
  - Snapshot of current practices/emerging trends related to managing and using preservation metadata in digital archiving systems
- **Objective 2: Define implementable, core preservation metadata, with guidelines/recommendations for management and use**

## PREMIS Data Dictionary

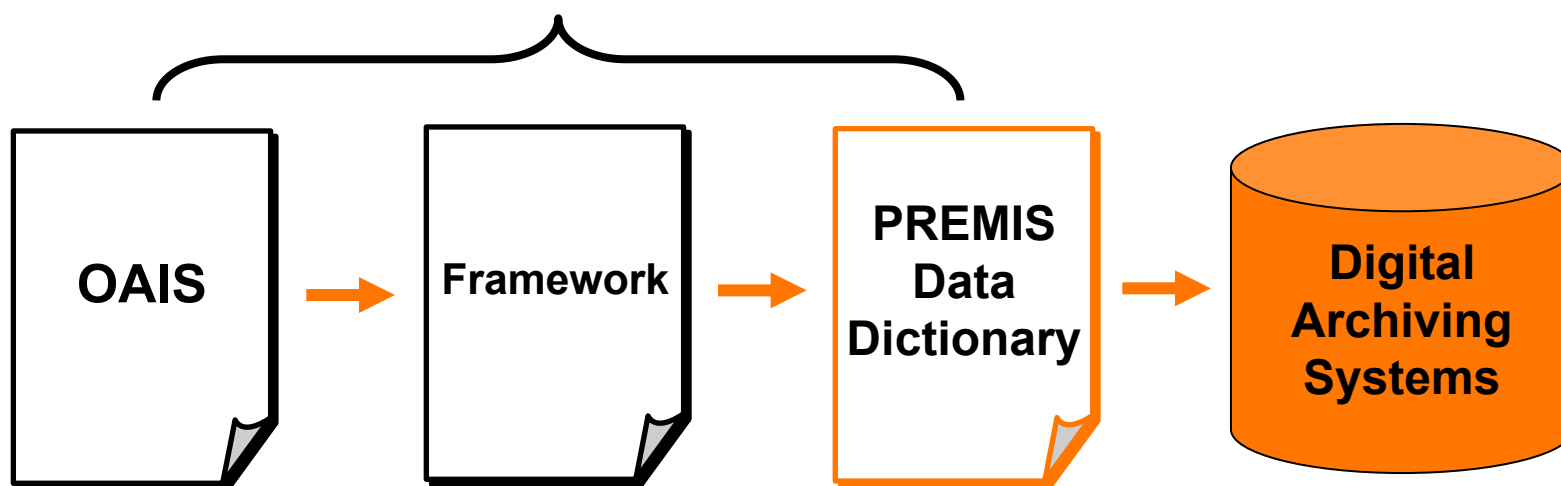
- **May 2005:** *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*
- 237-page report includes:
  - PREMIS Data Dictionary 1.0
  - Accompanying report (context, data model, assumptions)
  - Special topics, glossary, usage examples
- **Data Dictionary:** comprehensive, 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



## Benefits

- Preservation metadata is part of general infrastructure needed to support long-term preservation and use of digital materials; PREMIS Data Dictionary supplies a **critical piece** of that infrastructure
- Preservation metadata is key component of most digital preservation strategies ... so PREMIS Data Dictionary is of **immediate and demonstrable value** to existing and future digital preservation initiatives
- Preservation metadata is a shared problem; PREMIS Data Dictionary offers a **shared solution**, applicable across all sorts of institutions, digital preservation contexts, system implementations
- Effective digital preservation relies on open standards/best practices; PREMIS used **open discussion** to produce **openly available** resources that will be **openly maintained** into the future

## Special features

- PREMIS Data Dictionary is the first comprehensive technical specification for preservation metadata produced from an **international, cross-domain, consensus-building process**
- PREMIS Data Dictionary tightly **focused on implementation**:
  - Practical implementation was guiding principle in all discussions
  - Developed tools to support implementation; released with Data Dictionary
- Recognized need for **evidence of long-term commitment**:
  - Set up PREMIS Maintenance Activity to support Data Dictionary release
  - Permanent Web presence hosted by US Library of Congress
  - Central destination for PREMIS information, announcements, resources
  - 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

## Beneficiaries

- ***Any institution or organization committed to the long-term preservation of digital materials***
- As of 29 August 2005:
  - Survey Report downloaded more than 5100 times
  - Complete Data Dictionary/Final Report downloaded more than 4600 times
  - Separate components of Final Report downloaded more than 2700 times
  - 70 subscribers to the PREMIS Implementers' Group discussion list
- PREMIS Data Dictionary product of **collaboration** and **consensus**
  - PREMIS membership reflects variety of institutions, domains, countries
- PREMIS Data Dictionary reflects multiplicity of perspectives; as such, it is applicable in multiplicity of contexts
  - Digital preservation is a shared problem; this invites shared solutions

## Conclusion

- Advancing digital preservation from research topic to standardized process requires reliable infrastructure of technology, standards, and best practice
  - PREMIS Data Dictionary provides critical piece of this infrastructure
- PREMIS Data Dictionary is a building block with which effective, sustainable digital preservation strategies can be implemented
- PREMIS has contributed preservation metadata resources to the digital preservation community that are:
  - Immediately and freely available
  - Oriented toward practical implementation
  - Supported by implementation tools (XML schema)
  - Supported by a long-term commitment (PREMIS Maintenance Activity)
- Not glamorous – but essential!

## 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>