

PREMIS: Implementation and Preservation Metadata

Robin L. Dale
RLG

**International Conference on
Preservation of Digital Objects
iPRES
15-16 September 2005**

Digital preservation: progress & remaining challenges

Gradual shift in focus from **articulating** problem to **solving** it

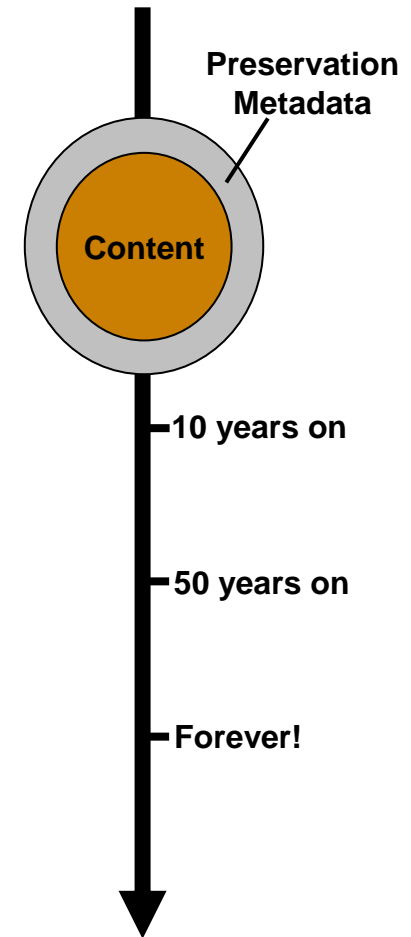
...

- 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 **implementation** challenge: preservation metadata

Preservation metadata includes:

- 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**

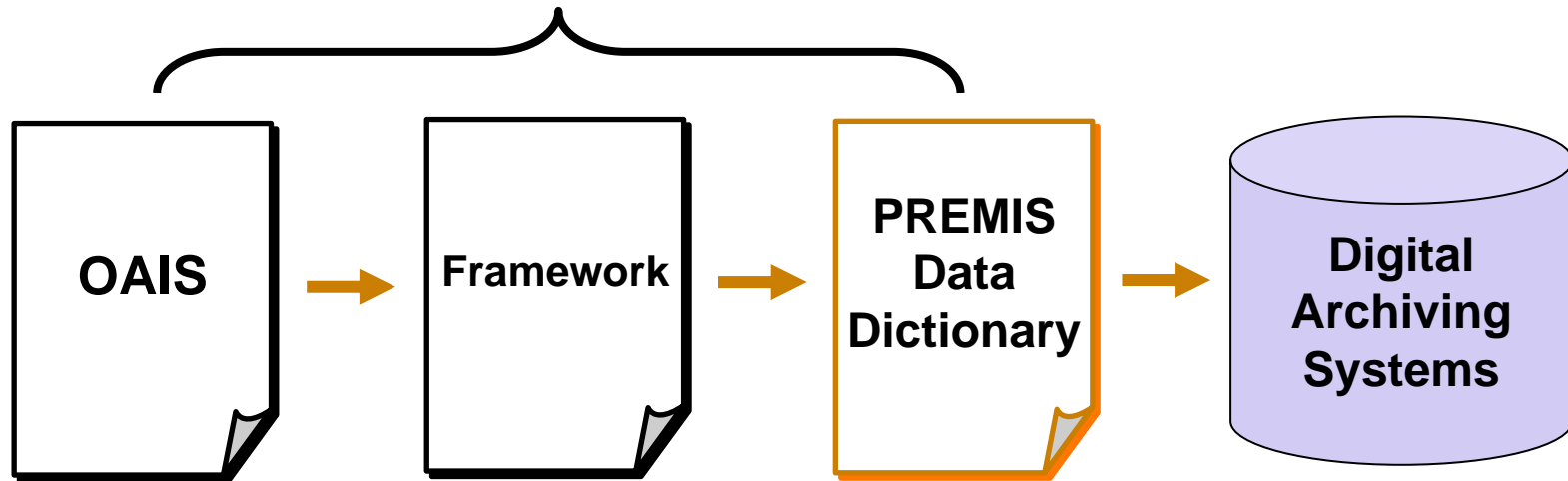


Some background ...

- **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:** Needed 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**

Need to move from theory to practice ...

Preservation Metadata Requirements



PREMIS Working Group

- June 2003: OCLC, RLG sponsored new international working group:
 - *PREMIS: Preservation Metadata: Implementation Strategies*
- Objective 1: Identify and evaluate alternative strategies for encoding, storing, managing, and exchanging preservation metadata
- Objective 2: Define implementable, core preservation metadata, with guidelines/recommendations (supporting data dictionary & rules) for management and use

Membership

- **Priscilla Caplan, FCLA (Chair)**
- **Rebecca Guenther, LC (Chair)**
- Michael Alexander, British Library
- George Barnum, GPO
- Charles Blair, U. of Chicago
- **Olaf Brandt, U. of Göttingen**
- Adam Farquhar, British Library
- David Gewirtz, Yale
- Kevin Glavash, MIT/Dspace
- Cathy Hartman, U. of N. Texas
- Helen Hodgart, British Library
- Nancy Hoebelheinrich, Stanford
- Roger Howard/Sally Hubbard, Getty Museum
- Pam Kircher, OCLC
- **John Kunze, Calif. Digital Library**
- **Brian Lavoie, OCLC liaison**
- **Robin Dale, RLG liaison**
- Vicky McCarger, LA Times
- Jerry McDonough, NYU/METS
- Evan Owens, JSTOR
- Erin Rhodes, NARA
- Madi Solomon, Walt Disney Co.
- Angela Spinazze, ATSPIN
- Gunter Waibel, RLG
- Lisa Weber, NARA
- Robin Wendler, Harvard
- Hilde van Wijngaarden, KB
- Andrew Wilson, NAA

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



Scope of data dictionary

- Implementation independent
- Descriptive metadata out of scope
- Technical metadata applying to all or most format types
- Media or hardware details are limited
- Business rules are essential for working repositories, but not covered
- Rights information for preservation actions, not access

Implementation considerations

- Schema use with specific implementations (e.g. METS)
- Machine generation of metadata
- Tools
- Role of registries (format, environment)
- Prospects for collaboration and exchanging information content
- Rights and permissions
- Emergence of best practices
- Support needed from PREMIS maintenance activity

Who is using PREMIS and why?

- *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 is a 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

PREMIS Maintenance Activity

[The Library of Congress](#) » [Standards](#) » [PREMIS Home](#)

Standards Pages

SEARCH

Official Web Site

- ▶ [Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group](#) [PDF: 3.2MB / 237p.]
- ▶ [PREMIS schemas](#)
- ▶ [PREMIS Working Group Home Page](#) [OCLC]
- ▶ [PREMIS Implementation Survey](#) [PDF: 1.24MB / 66p.]
- ▶ Governance (*coming soon*)
- ▶ More PREMIS Examples (*coming soon*)
- ▶ [Comments](#)

The PREMIS maintenance activity is responsible for maintaining, supporting, and coordinating future revisions to the PREMIS data dictionary. The Preservation Metadata: Implementation Strategies Working Group, convened by [OCLC](#) and [RLG](#), initially developed the PREMIS data dictionary as a specification with the goal of creating an implementable set of "core" preservation metadata elements, with broad applicability within the digital preservation community. Supporting XML schemas allow for implementation of the core metadata element set and are maintained in the Network Development and MARC Standards Office of the Library of Congress.

As of May 2005 the PREMIS data dictionary and schemas will begin a period of trial use. It is expected that they will remain stable for at least a year, after which revisions may be made based on results of experimentation.

Many documents on this site are prepared in Adobe Acrobat format (PDF). A free reader can be [downloaded](#) from the Adobe site for viewing.

News and articles:

- ▶ [Announcement: PREMIS Working Group Publishes Data Dictionary](#)
- ▶ [Implementing Metadata in Digital Preservation Systems: The PREMIS Activity](#), by *Brian F. Lavoie*

PREMIS Implementors' Group Forum (pig@loc.gov):

An unmoderated listserv open to members of the PREMIS implementor community. To subscribe to the forum:

1. send email message to:
listserv@loc.gov
2. the text of the message should be:
subscribe pig your name

[The Library of Congress](#) » [Standards](#) » [PREMIS Home](#)

[Contact Us](#)

May 24, 2005

Next steps

- PREMIS
 - Report, data dictionary, and XML schemas available for use
 - Draft for experimentation to remain stable for 18 months
 - **Participation opportunity: PREMIS Implementers Group Forum**
 - Revisions will be based on results of testing
- Follow-on activities
 - Testbeds for implementation and exchange
 - Tool development (esp. for automated harvest/extraction)
 - *Guidelines for Using the PREMIS Data Dictionary*
 - Community outreach
 - Maintenance activity
 - Consider formal standardization (NISO, ISO, etc)

Further questions to be addressed

- What portions of the Data Dictionary would be best served by registries?
- What portions of the Data Dictionary could be supported by automated capture and processing?
- What portions of the Data Dictionary should be supplied by the repository, and what portions would be more efficiently supplied by stakeholders external to the repository?
- What does PREMIS conformance mean?

Conclusion

- PREMIS Data Dictionary provides critical piece of reliable digital preservation infrastructure comprised of technology, standards, and best practice
- PREMIS Data Dictionary tightly focused on implementation:
 - Practical implementation was guiding principle in all discussions
 - Developed tools to support implementation; released with Data Dictionary
 - Further work with encouragement for international participation and tools development is ongoing

➤ **Unglamorous but necessary infrastructure!**

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 Metadata in Digital Preservation Systems: The PREMIS Activity" *D-Lib Magazine* (April 2004)
 - www.dlib.org/dlib/april04/lavoie/04lavoie.html
- *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>
- *RLG DigiNews* October 2004 and December 2004 issues
 - www.rlg.org/en/page.php?Page_ID=12081