

The PREMIS Data Dictionary

Michael Day
Digital Curation Centre
UKOLN, University of Bath
m.day@ukoln.ac.uk

JORUM, JISC and DCC Forum on the Long Term Curation
and Preservation of Digital Learning Objects, Wolfson
Medical School Building, Glasgow, 9 February 2006



<http://www.ukoln.ac.uk/>



JISC



Session outline

- Background
 - Metadata and preservation metadata
 - The PREMIS Working Group
- The PREMIS Data Dictionary
 - Main aims and objectives
 - Data model
 - Maintenance activities
 - Limits to scope



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



Preservation metadata

– Metadata

- Structured information about objects that supports various types of activity: discovery, retrieval, management, etc.
- Divided into descriptive, structural and administrative categories (e.g. METS)

– Preservation metadata

- "The information a repository uses to support the digital preservation process" (PREMIS)
- Cuts across all metadata categories



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS Working Group (1)

– Working Group on Preservation Metadata: Implementation Strategies

- Supported by OCLC and RLG
- Established in 2003
- International working group and advisory committee
- Chairs: Priscilla Caplan and Rebecca Guenther



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS Working Group (2)

- Building on older activity:
 - Working Group on Preservation Metadata (2000-02)
 - Preservation Metadata Framework (June 2002)
 - Based on the OAIS Information Model
- PREMIS objectives:
 - A 'core' set of preservation metadata elements (Data Dictionary)
 - Strategies for encoding, packaging, storing, managing, and exchanging metadata



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS Working Group (3)

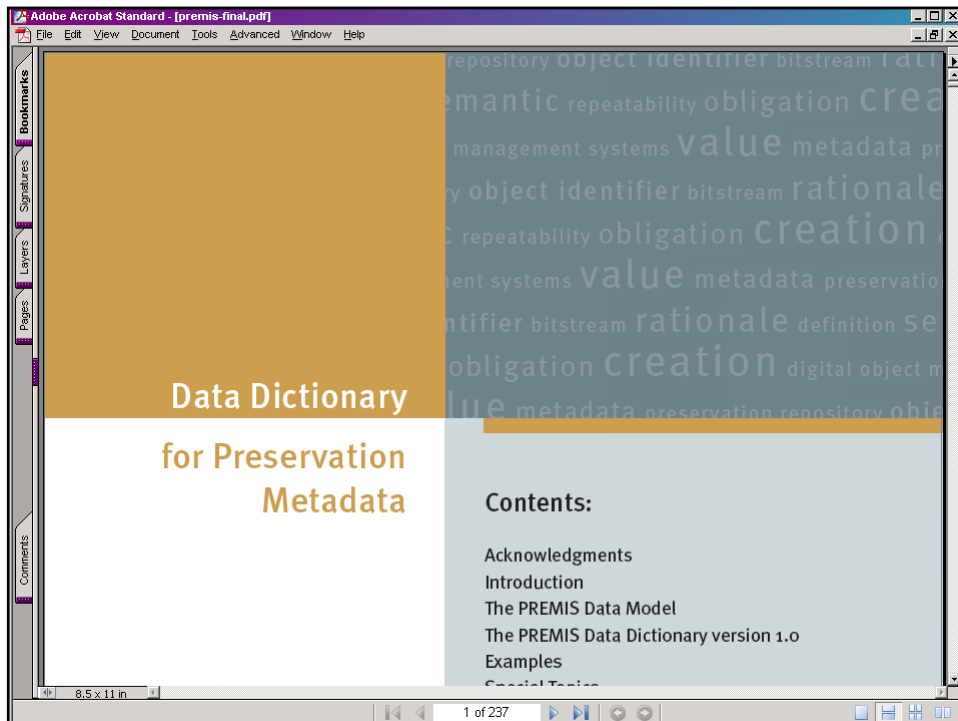
- Main PREMIS outputs:
 - Implementation Survey report (September 2004)
 - Based on ~50 responses
 - Snapshot of practice, noting trends
 - PREMIS Data Dictionary 1.0 (May 2005)
 - 237 pp.
- All WG documents are available from:
<http://www.oclc.org/research/projects/pmwg/>



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006





PREMIS data dictionary (1)

– Background:

- OAIS remains the conceptual foundation (but there are now some differences in terminology)
- The data dictionary is a translation of the OAIS-based 2002 *Framework* into a set of implementable semantic units
- Preservation metadata = "the information a repository uses to support the digital preservation process"



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS data dictionary (2)

- Core preservation metadata:
 - Data Dictionary defines metadata that supports "maintaining viability, renderability, understandability, authenticity, and identity in a preservation context."
 - Core metadata = "things that most working repositories are likely to need to know in order to support digital preservation."
 - Recognition of the need for automatic capture of metadata



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS data dictionary (3)

- The Data Dictionary is implementation independent, i.e. does not define how it should be stored
- Based on simple entity-relationship data model that defines five types of entities

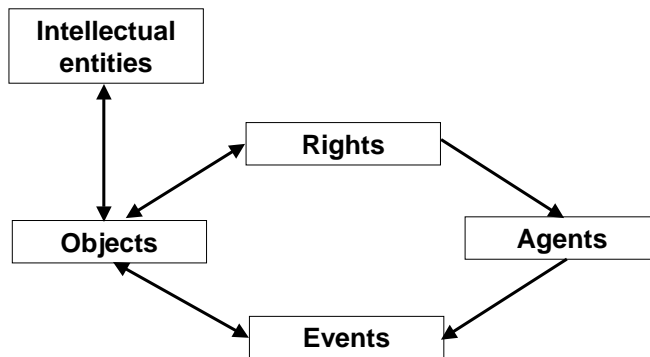


<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS data model (1)



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS data model (2)

- *Entities*:
 - Digital Object, Intellectual Entity, Event, Agent, & Rights
- *Relationships* are statements of association between instances of entities
- *Semantic Units* are the properties of an entity, and have values



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS data model (3)

- *Digital Object* = a discrete unit of information
 - Files = named and ordered sequence of bytes known by an operating system
 - Bitstream = a set of bits embedded within a file
 - Representation = the set of files needed for a "complete and reasonable" rendering of an Intellectual Entity



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS data model (4)

- *Intellectual Entity* = a coherent set of content that can be viewed as a single unit
- *Event* = an action involving at least one Object or Agent known to the repository
 - Documents actions that modify Digital Objects, records validity checks, etc.
 - Objects can be associated with any number of events



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS data model (5)

- *Agent* = persons, organisations, or programs associated with preservation events
 - Not the main focus of the data dictionary
- *Rights Statements* = assertions of rights pertaining to Objects or Agents
 - WG concentrates on rights and permissions associated with preservation activities



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS data model (6)

- *Relationships*:
 - Relationships between Objects:
 - Structural relationships, e.g. how files combine to make up an Intellectual Entity
 - Derivation relationships, e.g. resulting from format transformations or replications
 - Dependency relationships, e.g. when Objects depend on others, e.g. fonts, DTDs, etc.
 - 1:1 principle



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



PREMIS documentation

- Data Dictionary, v 1.0
 - Defines semantic units for Objects, Events, Agents and Rights
 - Implementation independent
 - Defines semantics
 - Proposed XML binding
- PREMIS Maintenance Agency
 - Library of Congress
 - <http://www.loc.gov/standards/premis/schemas.html>



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



Limits to scope (1)

- Does not focus on descriptive metadata
 - Domain specific and dealt with by many other schemes
- Does not define the specific characteristics of *Agents*
- Does not directly consider rights and permissions not directly associated with preservation actions, e.g. access or reuse



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



Limits to scope (2)

- Does not deal with technical metadata for all different types of digital file (left to format experts)
- Does not deal with the detailed documentation of media or hardware (left to media and hardware specialists)
- Does not consider in detail the business rules of a repository, e.g. roles, policies, and strategies (but this could be added to data model)



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



Some links:

- PREMIS Data Dictionary for Preservation Metadata (May 2005):
<http://www.oclc.org/research/projects/pmwg/>
- Brian Lavoie and Richard Gartner, *Preservation metadata*, DPC Technology Watch Report 05-01 (September 2005):
<http://www.dpconline.org/docs/reports/dpctw05-01.pdf>
- Michael Day, *Metadata*, DCC Digital Curation Manual Instalment (November 2005)
<http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata/>
- OAIS Reference Model (January 2002):
<http://public.ccsds.org/publications/archive/650x0b1.pdf>



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006



Acknowledgements

UKOLN is funded by the Museums, Libraries and Archives Council, the Joint Information Systems Committee (JISC) of the UK higher and further education funding councils, as well as by project funding from the JISC, the European Union, and other sources. *UKOLN* also receives support from the University of Bath, where it is based.

<http://www.ukoln.ac.uk/>



The *Digital Curation Centre* is funded by the JISC and the UK Research Councils' e-Science Core Programme.

<http://www.dcc.ac.uk/>



<http://www.ukoln.ac.uk/>

JORUM, JISC and DCC Forum, Glasgow, 9 February 2006

