Workshop A: Using metadata to support digital preservation

Michael Day
Digital Curation Centre
UKOLN, University of Bath
http://www.ukoln.ac.uk/

Driving the Long-Term Preservation of Electronic Records, London, 26-28 September 2006

Workshop outline

- Session 1: Introducing metadata (30 minutes)
- Session 2: Metadata in support of digital preservation (45 minutes)
- Coffee break (15 minutes)
- Session 3: The PREMIS Data Dictionary (45 minutes)
- Session 4: Preservation support in other metadata standards (30 minutes)

Session 1: Introducing metadata

Michael Day
UKOLN, University of Bath

Session overview

- Metadata - general overview
  - Definitions
  - Some basic questions
  - Some metadata standards

Defining metadata (1)

- Some definitions:
  - Literally, "data about data"
    - Defines the basic concept, but is (perhaps) not very meaningful
    - Refers to everything and nothing (Wendy Duff, 2004)
  - "Machine-understandable information about Web resources or other things" - Tim Berners-Lee, W3C (1997)

Defining metadata (2)

- "Structured data about resources that can be used to help support a wide range of operations - Michael Day, 2001
- "Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage" information objects - NISO, 2004
  - Hints at the many roles metadata can support

Defining metadata (3)

- Metadata is now typically defined by *function*
  - "Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics" (Dempsey & Heery, 1998)
  - Popular categorisation:
    - » Descriptive metadata
    - » Structural metadata
    - » Administrative metadata

Metadata functions

- Resource disclosure & discovery
- The retrieval and use of resources
- Resource management, including preservation
- Verification of authenticity
- Intellectual property rights management
- Commerce
- Content-rating
- Authentication and authorisation
- Personalisation and localisation of services
- …

Application areas

- "Web resources or other things," e.g.:
- Web sites, Web pages, digital images, databases, books, museum objects, archival records, collections, services, geographical locations, organisations, events, concepts, … even metadata itself

Metadata locations

- Within a resource, e.g.:
  - Title page and table of contents (books), META tags in document headers (Web pages), ID3 metadata (MP3), "file properties" (office documents), EXIF data (images)
- Directly linked to the resource, e.g.:
  - Link rel="meta" elements (Web pages)
- Independently managed in a separate database; can be linked by identifiers
  - This is the most common approach

Metadata is important

- … "is recognised as a critically important, and yet increasingly problematic and complex concept with relevance for information objects as they move through time and space" -- Gilliland-Swetland (2004)

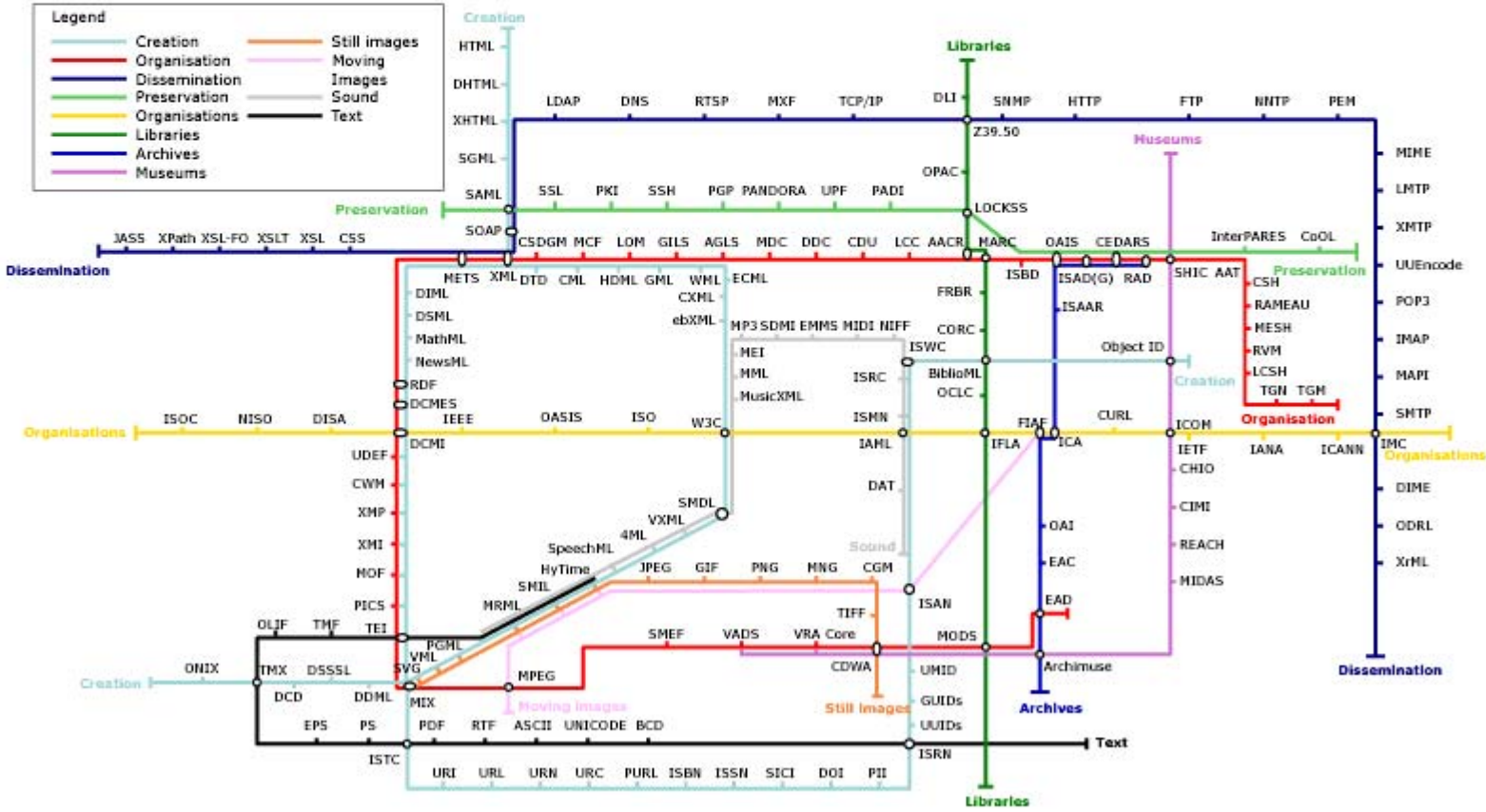
Metadata standards (1)

- But there are a large (and growing) number of metadata initiatives, formats, schemas, etc.
  - See James Turner's MetaMap for one attempt to visualise the metadata information space:
  - **http://mapageweb.umontreal.ca/turner/meta/english/**

MetaMap

[© 2004 James M. Turner, Véronique Moal, & Julie Desnoyers]

Position the mouse over an acronym to see what it stands for in a popup window.
Click on the acronym to see its definition and a link to its official site.



Metadata standards (2)

- Typically defined by "resource management communities"
  - Different traditions, perspectives, functional requirements
- Typically comprise:
  - A "conceptual model" (sometimes not explicit)
  - A set of named components ("terms", "elements" etc) and documentation on their meaning and use
  - A specification of how to represent a metadata instance in a digital format (binding)

Some examples (1)

- Bibliographic:
  - MARC (Machine-Readable Cataloguing) formats, e.g. MARC21, UNIMARC
    - Exchange format since 1960s
    - Content often based on family of related standards, e.g. the ISBD series, AACR2
  - MODS (Metadata Object Description Schema)
    - A subset of MARC
  - ONIX
    - Used by publishers and the book trade

Some examples (2)

- Archives and records:
  - ISAD(G) (General International Standard Archival Description)
  - EAD (Encoded Archival Description)
  - EAC (Encoded Archival Context)
  - Recordkeeping metadata
    - ERMS (The National Archives)
    - RKMS (Monash University)
    - ISO 23081-1:2006 (Metadata for records)
- Museum objects (and collections):
  - SPECTRUM

Some examples (3)

- Digital images:
  - VRA Core, NISO Technical Metadata for Digital Still Images
- Government information:
  - AGLS, e-GMS
- Learning objects:
  - IEEE LOM, UK LOM Core, IMS specifications
- Multimedia:
  - MPEG-7, MPEG-21 (for rights information)

Metadata implementation

- Many ways to implement metadata
  - Databases (internally)
  - Structured formats (for harvesting or exchange)
    - ISO 2709 (MARC)
    - Attribute-value pairs
    - HTML/XHTML (e.g., for header information)
    - Extensible Markup Language (XML)
      - METS, MODS, ONIX, MPEG-21 Digital Item Declaration Language (DIDL), etc. already use XML
      - Guidelines for implementing Dublin Core in XML (DCMI, 2003)
      - MARC 21 XML Schema (MARCXML)

Initial summing-up

- Metadata is ubiquitous
- Metadata enables people and software applications to do things (functions)
  - Not only about "discovery"
  - Different functions require different metadata
- There are many different standards
- Challenges remain in working across standards (interoperability), or in using standards in combination (modularity)
- XML is a current popular choice for implementation, at least to facilitate metadata harvesting (e.g. OAI-PMH) or exchange

Resource discovery and the Dublin Core Metadata Initiative

Resource discovery (1)

- A basic function of metadata
- Part of information retrieval
- Cutter's principles from "Rules for a Dictionary Catalog" (1876), slightly paraphrased:
  - To enable a person to find a [book] of which either the author, title or subject is known
  - To show what a [library] has by a given author, on a given subject, or in a given kind of literature
  - To assist in the choice of a [book] as to its edition (bibliographically) or to its character (literary or topical)

Resource discovery (2)

- A particular challenge in Web environment
  - Resource providers have moved into a **shared** network space
  - Recognition that users wish:
    - "to refer to intellectual and cultural materials flexibly and transparently without concern for institutional or national boundaries" (Dempsey, 2000)
- This is the problem that Dublin Core is designed to address (cross-domain discovery)

Resource discovery (3)

- We will now look in more detail at three standards *primarily* developed to support resource discovery
  - Dublin Core
  - The MARC formats
  - MODS

Dublin Core basics

- Perhaps the most well-known metadata initiative (there are many implementations)
- Named after a workshop held in Dublin, Ohio - a suburb of Columbus
- Mainly focused on cross-domain resource discovery
- A suite of standards (and other activities) organised as part of the Dublin Core Metadata Initiative (DCMI)

DCMI mission

- Providing simple standards to facilitate the finding, sharing and management of information, by:
  - Developing and maintaining international standards for describing resources
  - Supporting a worldwide community of users and developers
  - Promoting widespread use of Dublin Core solutions

DCMI brief history (1)

- Mid-1990s
  - Challenge of discovery on the Web
  - Search engines providing many hits, but little precision (pre Google)
  - Recognition that the traditional library approach to cataloguing could not scale to Web resources
- 1995 - first workshop
  - Hosted by OCLC at Dublin, Ohio
  - Primarily focused on Web resource discovery (document-like objects)
  - Resulted in interdisciplinary consensus on 13 metadata elements

DCMI brief history (2)

- 1996 - 2nd and 3rd workshops:
  - DC-2 (University of Warwick)
    - Recognised that DC elements would need to combine or co-exist with other types of metadata (modularity)
    - Warwick Framework devised to deal with this
  - DC-3 (Dublin, Ohio)
    - Workshop convened to deal with images (expanding beyond document-like objects)
    - Explicit focus now on cross-domain resource discovery
    - First identification of the 15 core elements

DCMI standardisation

- Dublin Core Metadata Element Set
  - Version 1.0: IETF RFC 2413 (1998)
  - Version 1.1: CEN Workshop Agreement CWA 13874 (2000), NISO Z39.85-2001, ISO 15836:2003
  - DCMI Recommendation (2004)
- DCMI Metadata Terms
  - DCMI Recommendation (latest version, 2005)
  - Specifies all metadata terms maintained by DCMI: elements, element refinements, encoding schemes, vocabulary terms
- DCMI Abstract Model
  - DCMI Recommendation (2005)

Dublin Core elements (1)

- Interdisciplinary consensus on simple element set for *resource discovery*
  - 15 elements
  - All optional
  - All repeatable
- Not intended for complex resource description
  - Initial idea of “simple document-like object”
  - Simplicity of semantics, ease of use
- Provides *basic* “semantic interoperability”
  - Across domains, across language communities
  - Does not provide detailed cataloguing rules
- A set of 15 broad “buckets”…

Dublin Core elements (2)

- Title
- Subject
- Description
- Creator
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

Dublin Core elements (3)

- ***Not*** a replacement for richer descriptive standards
- Can provide 15 “windows” into richer resource descriptions
  - disclose rich description in simple form
  - semantic cross-walks, mappings to existing data
  - export rather than create
- If metadata is language ...
  - … then DC is a “pidgin” language for use by “tourists on the Internet commons” (Thomas Baker)

Dublin Core elements (4)

- Small vocabulary, simple grammar/structure
  - Resource has Title “An Introduction to Dublin Core and the DCMI”
  - Resource has Subject “Metadata”
- Not as subtle/powerful as separate languages - but can be useful!

Extending Dublin Core

- Element refinements:
  - Narrow the meaning of a DC element
  - e.g. "date modified" v "date"
- Encoding schemes:
  - Provide additional information about a value
  - e.g. can indicate that a subject value is a Library of Congress Subject Heading
- The "Dumb-Down" principle
  - Provides rules for transforming "qualified" description into "simple" description
- the "One-to-One" rule
  - A DC description describes exactly one resource

Dublin Core Application Profiles

- In practice, metadata implementers
  - **Combine** elements from different sources (e.g. DC plus elements from other schemas, “local” elements)
  - **Refine** definitions of elements
  - **Constrain** use of elements
- Application profiles
  - If simple DC is a “pidgin”, an application profile is a “regional idiom or creole”! (Thomas Baker)
  - Element set plus policies, guidelines
  - Some DCMI Working groups developing application profiles for specific domains (government, education)

DC Application Profiles

- "Simple Dublin Core"
  - Use of the 15 properties of the DCMES
  - All optional and repeatable
  - Values represented by value strings
  - No vocabulary or syntax encoding schemes
- UK eGovernment Metadata Standard (eGMS)
  - Use of selected properties from DCMI vocabularies, additional properties
  - Guidelines on use of properties
  - Some properties mandated/recommended
  - Some vocabulary encoding schemes mandated/recommended
  - Guidance on content of value strings

Some applications of Dublin Core

- Embedded in Web pages
- Integrated resource discovery services
  - For example
    - Subject Gateways - Resource Discovery Network
    - OAI Service Providers - OAIster
    - Image services - Picture Australia

DC embedded in X/HTML

- Search crawlers can extract metadata from individual pages
- However, little or no use by the major search engines
  - Robot spamming problems
  - Lack of metadata (or quality-control)
  - Availability of better indexing tools, e.g. Google's PageRank algorithm
- But, useful in controlled environments

Session 2: Metadata in support of digital preservation

Michael Day
UKOLN, University of Bath

Wider roles of metadata

- Early recognition that metadata was not only useful for resource discovery
- Resource management
  - Managing access
  - Managing resources
  - Recording contexts (technical and other)
- Some examples:
  - Records management and archives
  - Digitisation initiatives
  - Digital preservation

Preservation metadata (1)

- Definitions:
  - All of the various types of data that allow the re-creation and interpretation of the structure and content of digital data over time (Ludäsher, Marciano and Moore, 2001)
  - "… the information a repository uses to support the digital preservation process" -- PREMIS working group (2005)
  - *All* digital preservation strategies depend, to some extent, upon the creation, capture and maintenance of appropriate metadata
  - "Preserving the right metadata is key to preserving digital objects" -- ERPANET Briefing Paper (Duff, Hofman & Troemel, 2003)

Preservation metadata (2)

- Preservation metadata fulfil a range of different roles, e.g.:
  - "… metadata accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information" (Lynch, 1999)
  - Spans the categories of administrative, structural, descriptive and technical metadata

Preservation metadata (3)

- Metadata is key to the understanding and reuse of digital information, e.g.:
  - "… it is impossible to conduct a correct analysis of a data set without knowing how the data was cleaned, calibrated, what parameters were used in the process, etc." -- Deelman, *et al*. (2004)
  - Growing emphasis on open access to research data (OECD working group)
  - The 'data deluge'

Preservation metadata (4)

- Current position:
  - Early initiatives tended to be theoretical in nature (e.g., metadata frameworks); current ones have a far more practical focus
  - Some consensus in cultural heritage domain on the *types* of metadata required
    - Major influence of the Reference Model for an Open Archival Information System (OAIS)

Excursion: the OAIS Information Model

OAIS background

- Reference Model for an Open Archival Information System (OAIS)
  - Nothing to do with the OAI (Open Archives Initiative) or OAI-PMH
  - Development led by the Consultative Committee for Space Data Systems (CCSDS)
  - Issued as CCSDS Recommendation (Blue Book) 650.0-B-1 (January 2002)
  - Also adopted as: ISO 14721:2003
  - http://public.ccsds.org/publications/archive/650x0b1.pdf

OAIS definitions (1)

- Provides definitions of terms, e.g.:
  - **OAIS -** "An archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community"
  - **Designated Community -** the community of stakeholders and users that the OAIS serves
  - **Knowledge Base -** a set of information, incorporated by a user or system, that allows that user or system to understand the received information

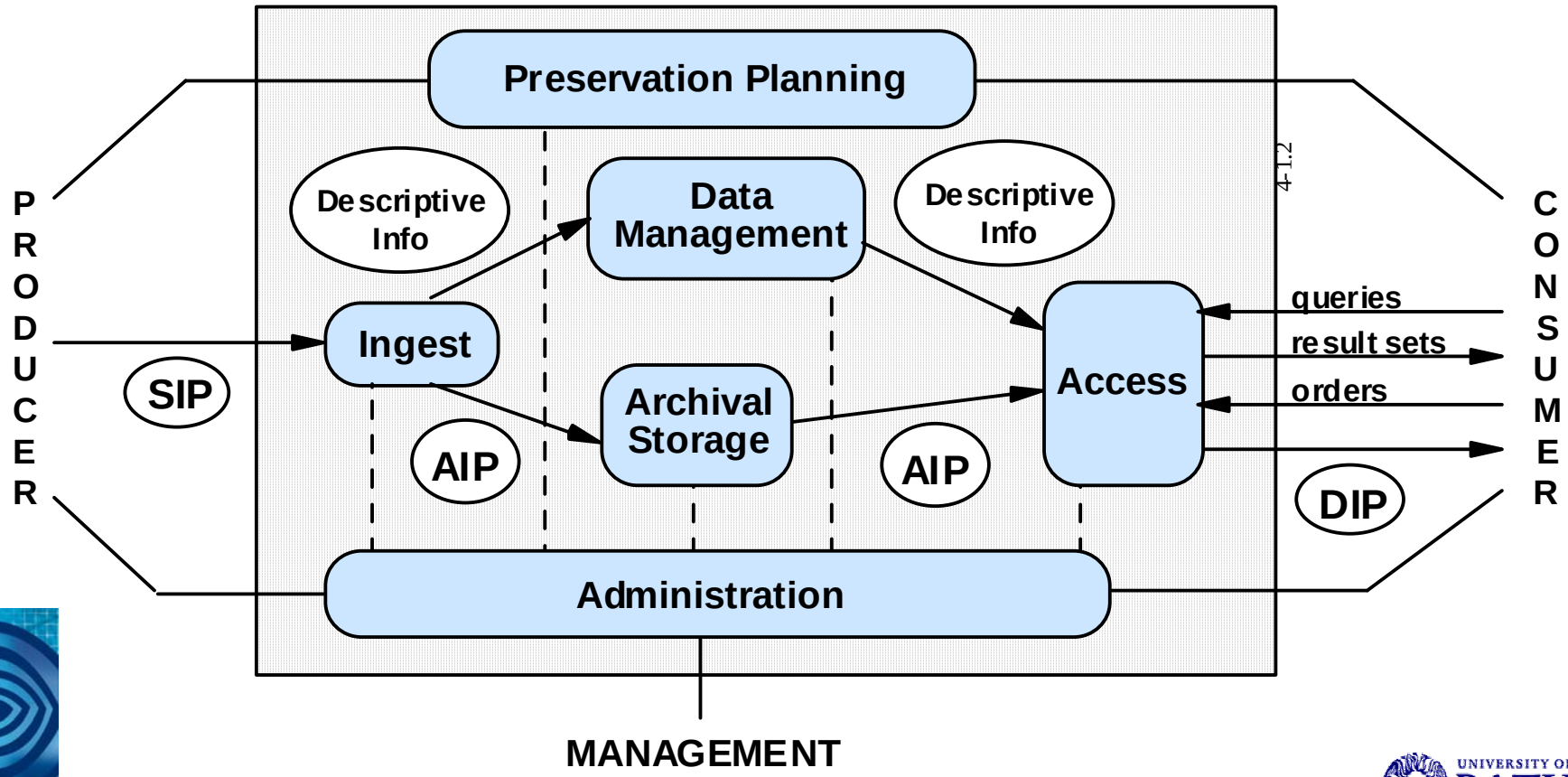
OAIS definitions (2)

- **Information Object** - Data Object + Representation Information
- **Representation Information** - any information required to render, interpret and understand digital data
- **Information Package** - Conceptual linking of Content Information + Preservation Description Information + Packaging Information (Submission, Archival and Dissemination Information Packages)
- **Preservation Description Information** - information (metadata) about Provenance, Context, Reference, Fixity information

OAIS Functional Model (1)

- Six entities
  - Ingest
  - Archival Storage
  - Data Management
  - Administration
  - Preservation Planning
  - Access
- Described using UML diagrams ...

OAIS Functional Model (2)

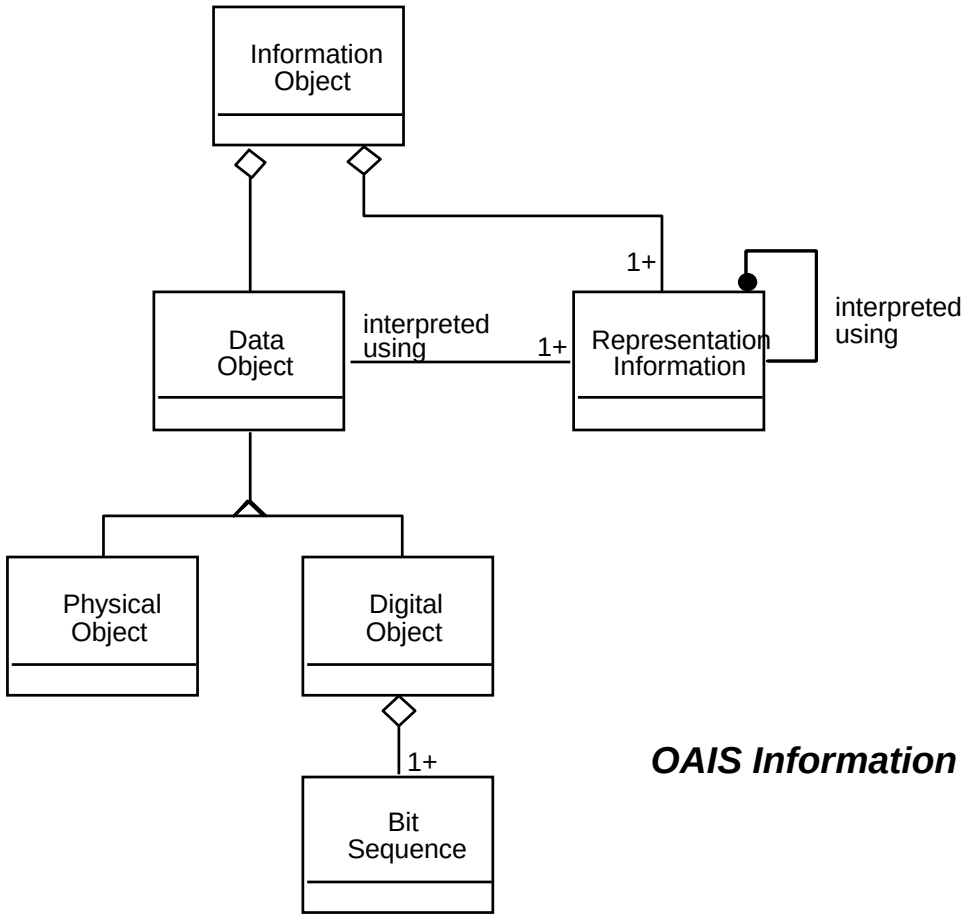


OAIS Functional Entities (Figure 4-1)

OAIS Information Objects (1)

- Information Object (basic concept):
  - Data Object (bit-stream)
  - Representation Information (permits “the full interpretation of Data Object into meaningful information”)
- Information Object Classes:
  - Content Information
  - Preservation Description Information (PDI)
  - Packaging Information
  - Descriptive Information

OAIS Information Objects (2)

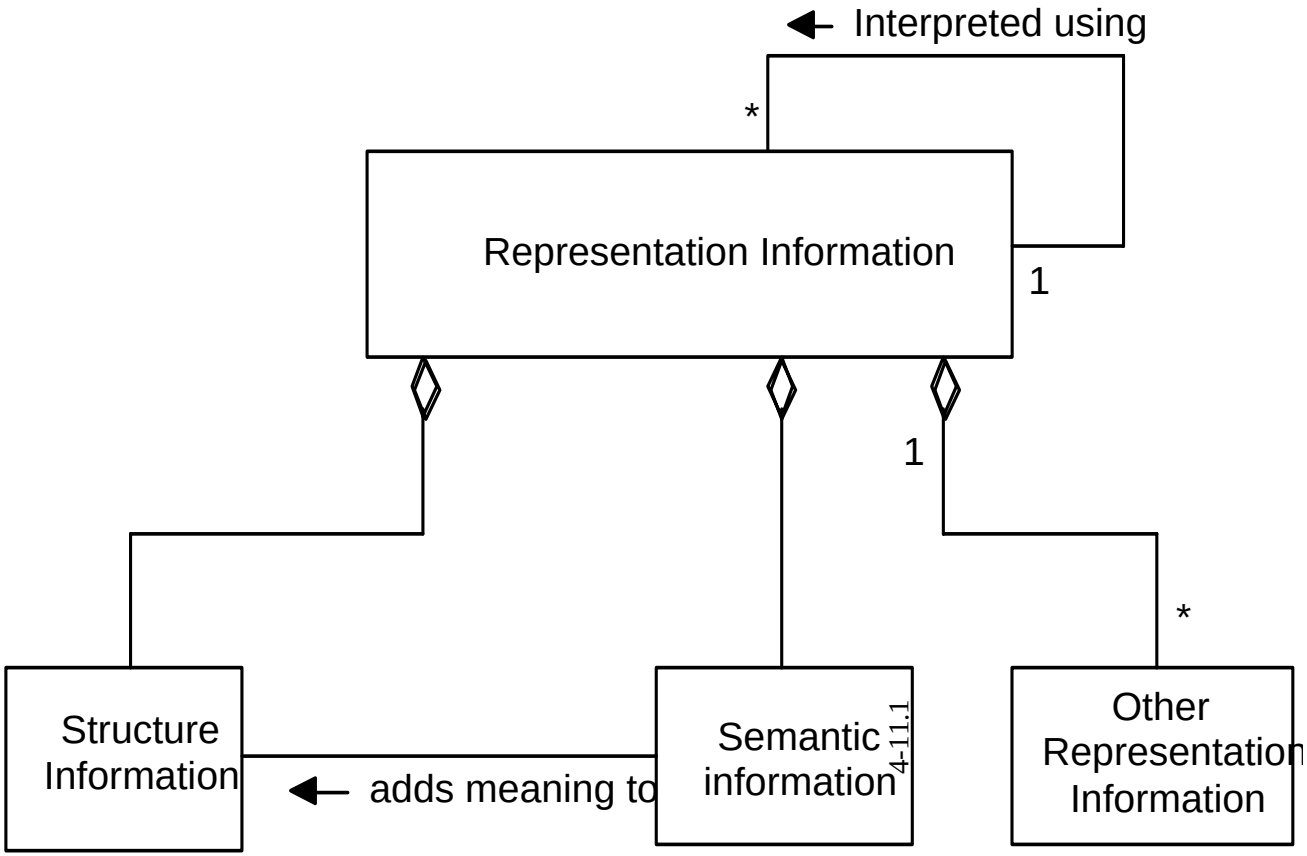


OAIS Information Object (Figure 4-10)

OAIS Information Objects (3)

- Representation Information:
  - *Any* information required to render, interpret and understand digital data (includes file formats, software, algorithms, standards, semantic information etc.)
  - Representation Information is recursive in nature
  - Essential that Representation Information itself is curated and preserved to maintain access to (render and interpret) digital data
    - e.g. Format registries (Global Digital Format Registry (GDFR), PRONOM, Digital Curation Centre Representation Information Registry)

OAIS Information Objects (4)

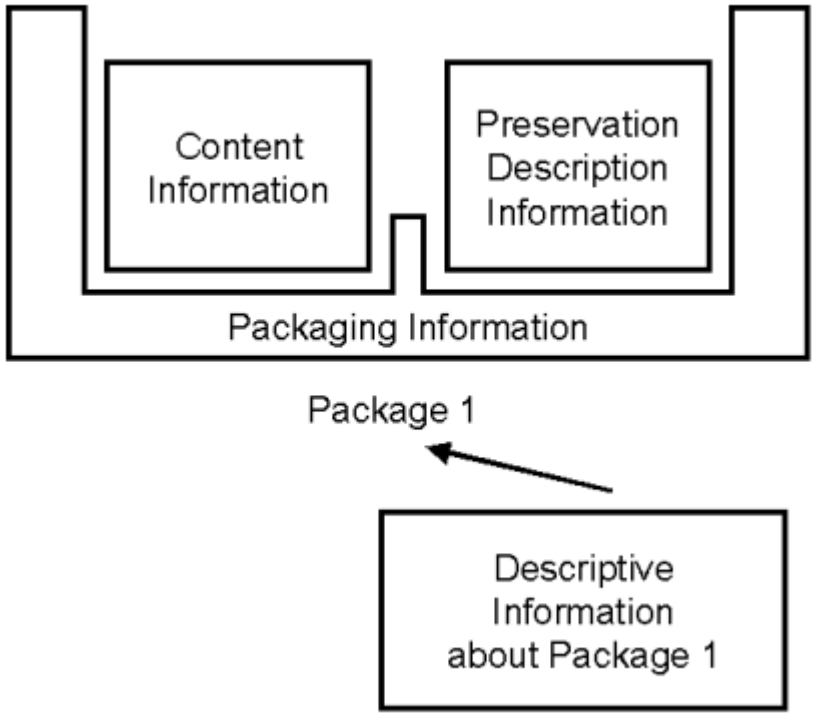


OAIS Representation Information Object (Figure 4-11)

OAIS Information Packages (1)

- Information package:
  - Container that encapsulates Content Information and PDI
  - Packages for submission (SIP), archival storage (AIP) and dissemination (DIP)
  - AIP = "... a concise way of referring to a set of information that has, in principle, all of the qualities needed for permanent, or indefinite, Long Term Preservation of a designated Information Object"

OAIS Information Packages (2)

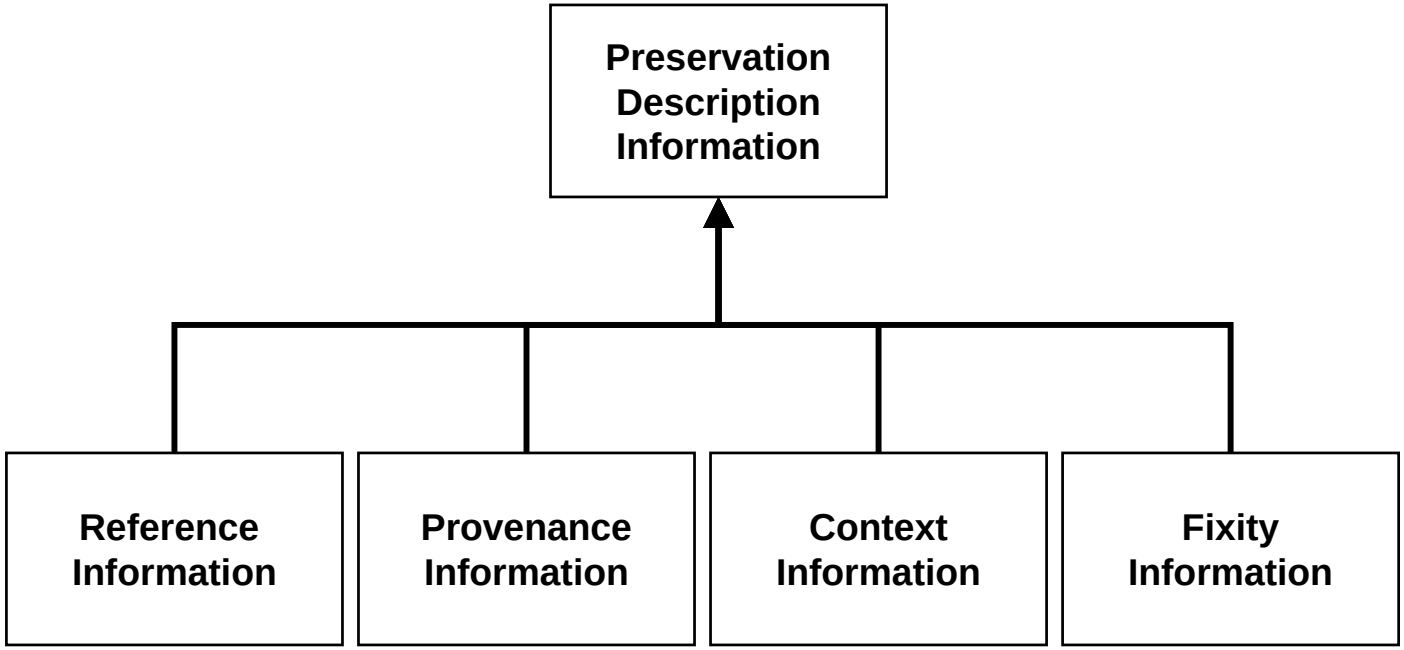


Information Package Concepts and Relationships (Figure 2-3)

OAIS Information Packages (3)

- Archival Information Package (AIP):
  - Content Information
    - Original target of preservation
    - Information Object (Data Object & Representation Information)
  - Preservation Description Information (PDI)
    - Other information (metadata) “which will allow the understanding of the Content Information over an indefinite period of time”
    - A set of Information Objects
    - In part based on categories discussed in CPA/RLG report: *Preserving Digital Information* (1996)

OAIS Information Packages (4)

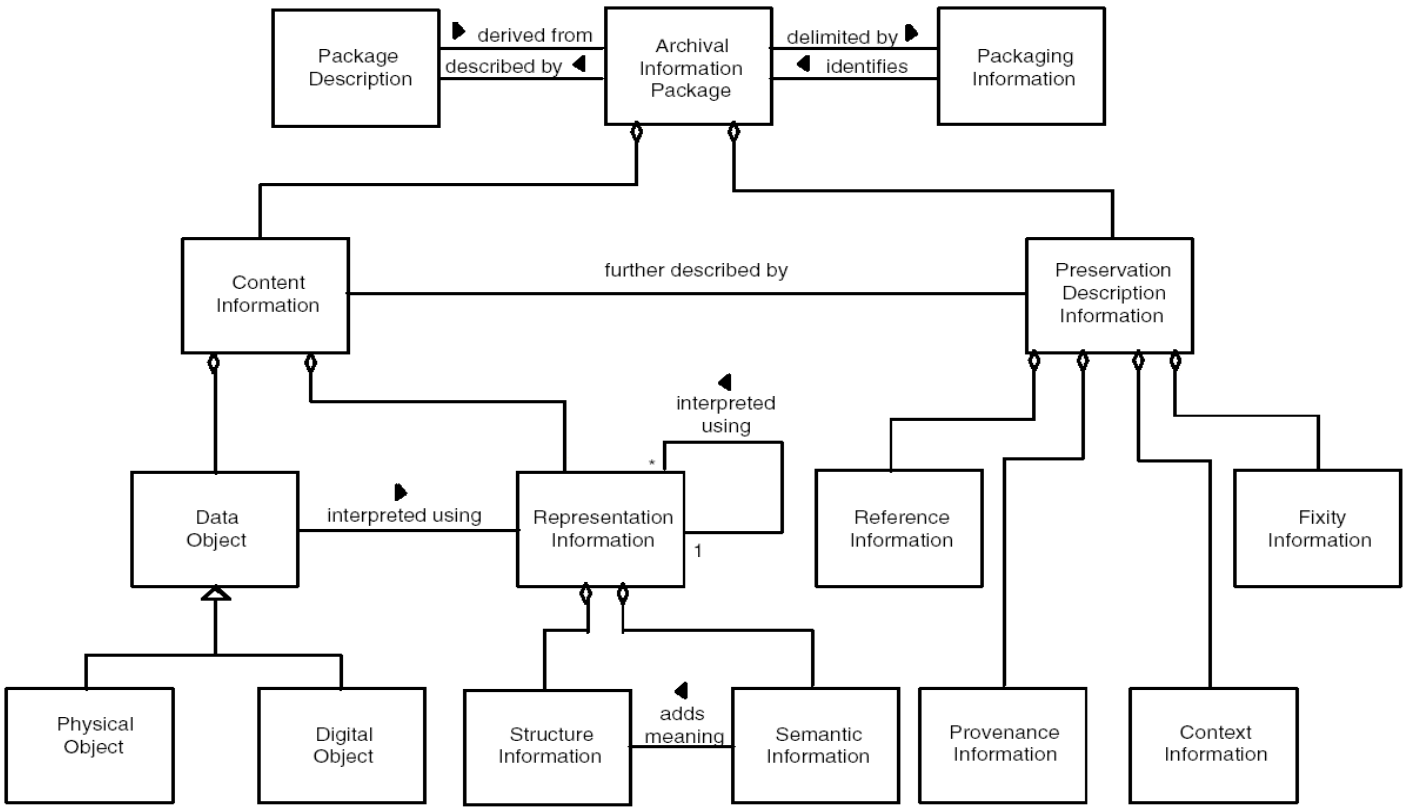


PDI Preservation Description Information (Figure 4-16)

OAIS Information Packages (5)

- **Fixity** - supporting data integrity checking mechanisms
- **Reference** - for supporting identification and location over time
- **Context** - documenting the relationship of the Content Information to its environment
- **Provenance** - documents the history of the Content Information

OAIS Information Packages (6)



OAIS Information Model

- Also defines:
  - Archival Information Units and Archival Information Collections
    - Recognises the complexity some some objects, addresses granularity
  - Information Package transformations
    - For Ingest and Access

Preservation metadata standards

Preservation metadata standards

- Two triggers:
  - An urgent practical response to the growing amount of digital content needing management:
    - National Library of Australia (1999)
    - Harvard University Library
    - National Library of New Zealand (2003)
  - Research projects
    - UK Cedars project outline specification (2000)
    - NEDLIB project (2000)

OCLC/RLG Metadata Framework

- Metadata Framework Working Group (2000 - 2002)
  - Sponsored by OCLC and RLG
  - Preservation Metadata Framework (2002)
    - Structured around the OAIS information model and the work of earlier initiatives
  - Framework was a set of recommendations, not a specification for implementation
  - Led directly to the development of the PREMIS Working Group

Coffee break

Session 3:
The PREMIS Data Dictionary

Michael Day
UKOLN, University of Bath

PREMIS Working Group (1)

- PREMIS Working Group (2003 - 2005)
  - Preservation Metadata: Implementation Strategies
  - Sponsored by OCLC and RLG
  - International working group and advisory committee
    - Primarily practical focus
    - Members from the US, the UK, the Netherlands, Germany, Australia and New Zealand
  - Chaired by Priscilla Caplan and Rebecca Guenther

PREMIS Working Group (2)

- Main objectives:
  - A 'core' set of preservation metadata elements (Data Dictionary)
  - Strategies for encoding, packaging, storing, managing, and exchanging metadata
- Outputs:
  - Implementation Survey report (Sept. 2004)
  - PREMIS Data Dictionary 1.0 (May 2005)
- All WG documents are available from: http://www.oclc.org/research/projects/pmwg/

PREMIS survey (1)

- Implementing Preservation Repositories for Digital Materials (2004)
  - Review of current practice within cultural heritage organisations
    - Based on responses to questionnaire together with follow-up interviews
    - Questions about business plans, policies, preservation strategies, as well as metadata
    - Analysis based on ~50 responses
    - Snapshot of practice, noting trends

PREMIS survey (2)

- Findings:
  - Very little current experience of digital preservation; no knowledge whether the metadata collected will be adequate
  - The OAIS model has informed the implementation of many repositories
  - METS was the most commonly-used scheme for non-descriptive metadata
  - Metadata is stored *both* in databases and together with content data objects

PREMIS survey (3)

- Trends identified:
  - Redundant storage of metadata both within databases (for ease of use) and encapsulated with data objects (self-documenting)
  - METS is commonly used for the packaging of different metadata
  - OAIS is just the starting point
  - The retention of the original versions of objects to reduce risks
  - The use of multiple preservation strategies

PREMIS data dictionary (1)

- Background:
  - OAIS remains the conceptual foundation (but 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"

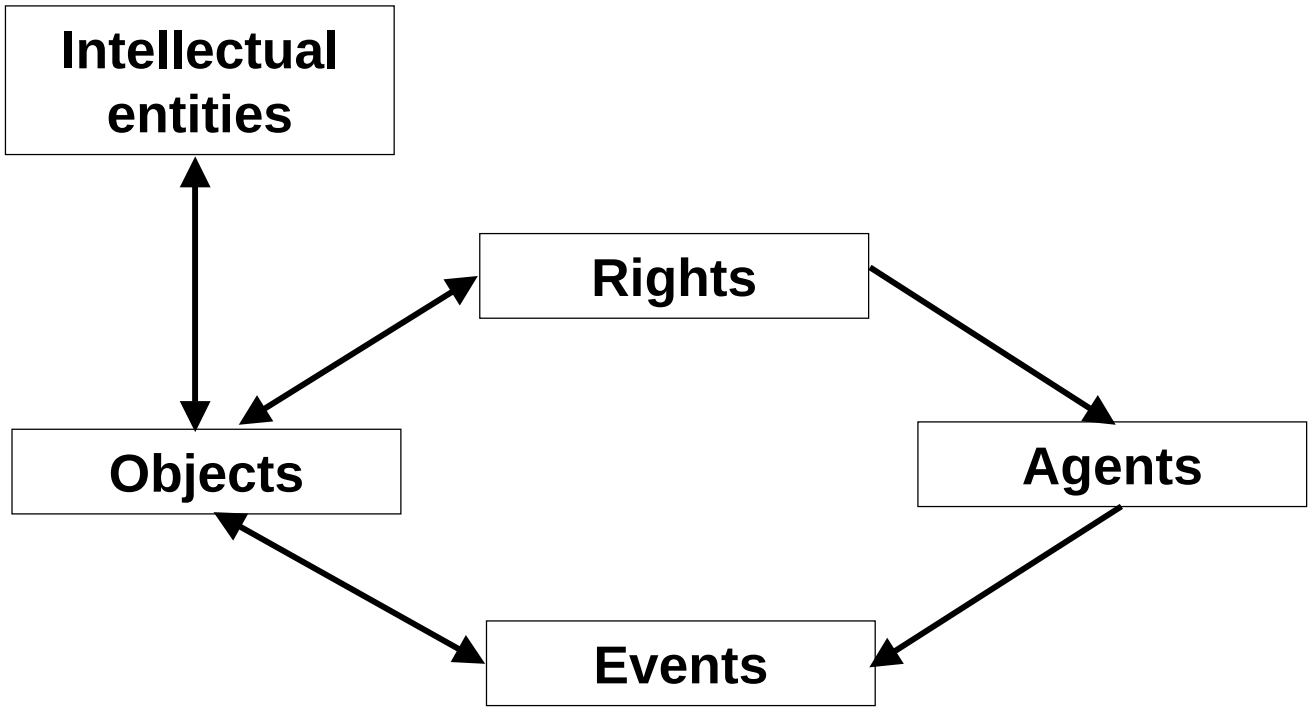
PREMIS data dictionary (2)

- Defines metadata that supports "maintaining viability, renderability, understandability, authenticity, and identity in a preservation context."
- New 'canonical' definition of preservation metadata
- 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

PREMIS data dictionary (3)

- The Data Dictionary is implementation independent, i.e. does not define how it should be stored
- Based on simple data model that defines five types of entities
- Defines semantic units for Objects, Events, Agents and Rights

PREMIS data model (1)



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

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

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

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

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

PREMIS documentation

- Data Dictionary, v 1.0
  - Defines semantic units for Objects, Events, Agents and Rights
  - Implementation independent
    - Defines semantics
- Separate proposed XML bindings (PREMIS schemas)
- PREMIS Maintenance Agency (Library of Congress)
- Editorial Committee and Implementers' Group (PIG)
  - http://www.loc.gov/standards/premis/index.html

Limits to scope

- 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
- 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)

Some issues

- The PREMIS Data Dictionary is an important contribution to the ongoing development of preservation metadata
- It is, however, implementation independent
  - Definition of semantics and a suggested XML binding
- Conformance
  - Non-PREMIS elements not conflict with or overlap with PREMIS semantic units
  - Need for more harmonisation (?)
- The exchange of Objects
  - Mandatory metadata needs to be able to be extracted and packaged with the object
- The use of controlled vocabularies

Session 4: Preservation support in other metadata standards

Michael Day
UKOLN, University of Bath

Some relevant domains

- Archives and records management
  - Focus on integrity and authenticity
  - The development of recordkeeping systems
- Digitisation initiatives
  - Focus on digitisation processes
  - Preparation of 'master' files with all appropriate metadata
- Learning object management
  - Primary focus on the management of objects (e.g. IP rights), rather than preservation
- Commercial content (e.g. television companies)
  - Not covered here

Recordkeeping metadata (1)

- Two research projects in the mid-1990s
  - Pittsburgh Project "Functional Requirements for Evidence in Recordkeeping":
    - Defined fundamental properties of records based on their role as evidence of business transactions
    - Revealed need for influence on the design of recordkeeping systems (automatic capture of metadata)
    - Business Acceptable Communications (BAC) reference model for metadata (1995)
  - University of British Columbia (UBC) project:
    - Also stressed the evidentiary value of records
    - Focus on importance of authenticity and integrity

Recordkeeping metadata (2)

- Follow-up research:
  - InterPARES project (international)
  - Australian Recordkeeping Metadata Schema (RKMS)
  - National standards:
    - National Archives of Australia - Recordkeeping Metadata Standard
    - The National Archives (UK) - Requirements for Electronic Records Management Systems - Metadata standard
  - Preservation approach (encapsulation in XML)
    - Public Record Office Victoria - Victorian Electronic Records Strategy (VERS)

Recordkeeping metadata (3)

- ISO 23081-1:2006
  - Information and documentation -- Records management processes -- Metadata for records -- Part 1: Principles
  - Developed by ISO Technical Committee TC 46, Information and documentation, Subcommittee SC11 Archives/Records Management
  - First of a family of standards
  - Builds on the framework of the ISO 15489 Records Management standard

Recordkeeping metadata (4)

- ISO 23081-1 definitions:
  - Builds on ISO 15489 definition: "data describing the context, content and structure of records and their management through time"
  - "As such, metadata are structured or semi-structured information that enables the creation, registration, classification, access, preservation and disposition of records through time and within and across domains … [and] can be used to identify, authenticate and contextualise records and the people, processes and systems that create, manage, maintain and use them and the policies that govern them"

Recordkeeping metadata (5)

- ISO 23081-1 general principles
  - Metadata capture is built into business processes
    - Defines the critical characteristics of records, must be explicit
  - Need to define roles and responsibilities
    - Records managers, information professionals, executives, unit managers, system administrators, …
  - Long-term preservation is just one of the roles fulfilled by recordkeeping metadata

Recordkeeping metadata (6)

- ISO 23081-1 types of metadata
  - About the record itself
    - Should includes metadata about structure, format and technical dependencies
  - About business rules, policies and mandates
  - About agents (people) - for accountability
  - About business activities or processes
  - About records management processes

Recordkeeping metadata (7)

- Automatic capture and sharing of metadata
  - Monash University Clever Recordkeeping Metadata (CRKM) project
    - Focus on interoperability:
      - Enabling information originating in one context to be (re)used in other ways
      - With a high degree of automation
    - Relies on standards
    - Metadata registries for storing standardised representations of schemas

Digitisation initiatives (1)

- Digitisation initiatives focus on:
  - The technical information that needs to be captured as part of the digitisation process, e.g.:
    - NISO Z39.87 Technical Metadata for Digital Still Images
  - Ways of packaging content and metadata in order to create standardised packages
    - e.g. for collecting all the individual page images that comprise a book and enabling their display in the ight order
    - Metadata Encoding & Transmission Standard (METS)
      - http://www.loc.gov/standards/mets/

METS basics (1)

- Originated in digitisation projects, i.e. Making of America II
- Now maintained by the Library of Congress
- An XML-based framework for packaging various types of metadata (and data), including:
  - METS Header
  - Descriptive Metadata - for discovery and retrieval
  - Administrative Metadata - enabling managers to administer the object (as part of a collection)
  - Structural Metadata - describing how individual components relate to one another
    - File Section, Structural Map, Structural Links, Behavior

METS basics (2)

- Implemented widely in digital library projects, e.g. Oxford Digital Library
- Supports Interoperability
  - Different metadata can be combined within a METS container, e.g. MODS, MARC in XML, DC in XML, etc.
- Supports the portability of objects
- METS can be seen as a type of Information Package (in OAIS terms), combining both data and metadata

Learning objects (1)

- Learning objects
  - Any digital resource that can be (re)used to support learning
    - "… any entity - digital or non-digital - that may be used for learning, education or training" - IEEE Learning Object Metadata (LOM) standard
  - Essentially modular
    - Includes: images (graphs, photographs), Web sites, presentations, quizzes, bibliographies, multimedia, etc.
  - There is a major focus on re-use
    - Learning objects are included in institutional repositories (and central services like JORUM)
    - Re-use is the main focus of metadata initiatives

UKOLN

UNIVERSITY OF BATH

Learning objects (2)

- But LOs reflect many of the difficulties found with other digital objects
  - Technical dependence on other resources (linear navigation, embedded content or software), complicated because of added granularity

Learning Objects (3)

- IEEE Learning Object Metadata (LOM)
  - Institute of Electrical and Electronics Engineers
  - Two main foci:
    - Resource discovery
    - Describes the structure of learning objects and management processes (e.g. rights management)
    - Used by JORUM (UK LOM Core)
  - There is now an increasing consideration of the potential role of long-term digital preservation in the learning object arena:
    - JISC report (2004)
    - JORUM watch reports

Summing up

- Metadata is perceived to be essential for the long-term management and preservation of digital objects
- There is now the beginning of consensus on what particular metadata might be required to support preservation processes (e.g., the OAIS model, PREMIS Data Dictionary) and packaging (e.g. METS)
- There is still little experience with the practical implementation of preservation metadata, but the existence of the PREMIS Data Dictionary is beginning to make a difference

Further reading

- OAIS Reference Model (2002): http://public.ccsds.org/publications/archive/650x0b1.pdf
- PREMIS Data Dictionary for Preservation Metadata (2005): http://www.oclc.org/research/projects/pmwg/
- DPC Technology Watch Report on "Preservation Metadata" by Brian Lavoie and Richard Gartner (2005): http://www.dpconline.org/docs/reports/dpctw05-01.pdf
- DCC Digital Curation Manual Instalments
  - "Metadata" by Michael Day (2005), "Preservation Metadata" by Priscilla Caplan (2006), and "Archival Metadata" by Wendy Duff and Marlene van Ballegooie (2006)
  - http://www.dcc.ac.uk/resource/curation-manual/

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/