

# Preservation Metadata

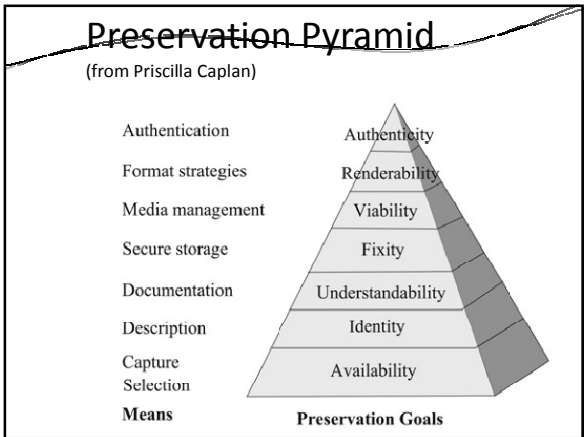
Wendy Duff

## Definitions (1)

- Metadata:
  - A relatively new term for a very old concept
  - “Metadata that are associated with either an information system or an information object for the purposes of description, administration, legal requirements, technical functionality, use and usage and preservation, play a critical role in ensuring the creation, management, preservation, discovery, use and re-use of trustworthy materials, including records” (InterPares 2 book)

## Definitions (2)

- Preservation metadata:
  - “... the information a repository uses to support the digital preservation process” (PREMIS Data Dictionary) Information that supports the viability, renderability, understandability, authenticity, fixity and identity of digital objects in a preservation context;
  - Potentially very wide scope:
    - Technical information on data structures or formats
    - Information to help better understand the content
    - Information on contexts and provenance
    - Information on preservation processes



### Preservation Metadata

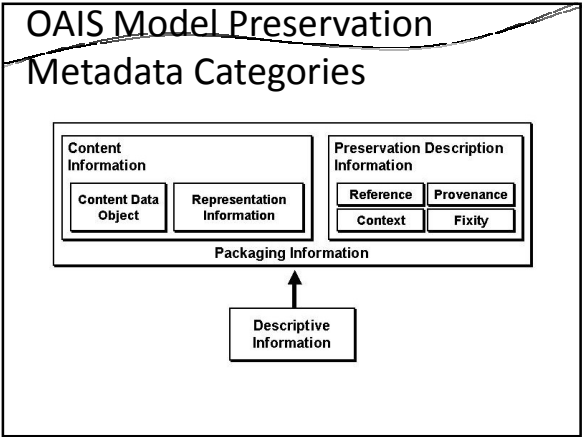
- Early initiatives tended to be theoretical in nature (e.g., metadata frameworks); current ones have a far more practical focus
- Some consensus in cultural heritage domain on the *types* of metadata required
  - A major influence on this has been the Reference Model for an Open Archival Information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

### The OAIS Information Model

- General OAIS background:
  - An ISO standard (ISO 14721:2003)
  - Development led by the Consultative Committee on Space Data Systems
  - Provides general principles, standard terminology and defines two interrelated models (functional model, information model)

### The OAIS Information Model

- Some general principles:
  - OAIS entities (Data Objects and Types of Information) are conceptually bound together
  - Information Content - Content Data Object + Representation Information
  - Information Packages - Information Content + Preservation Description Information + Packaging Information



## The OAIS Information Model

- Representation Information:
  - Representation Information - Any information required to render, interpret and understand digital data (includes file formats, software, algorithms, standards, semantic information etc.)
    - Is tightly bound with the Content Data Object
    - Provides a bridge between the bit-level information being stored in an OAIS and something that can be understood
  - Describing data structure concepts, or formats (Structure Information)
  - Providing additional information on semantics, i.e. meaning of numeric data (Semantic Information)

## Representation Information

- Team Digital Preservation and the Deadly Cryptic Conundrum
- <http://www.youtube.com/watch?v=YunghkPPF9M>

## The OAIS Information Model

- Preservation Description Information:
  - Preservation Description Information - information (metadata) about Provenance, Context, Reference, Fixity
- The additional information "needed to make the Content Information meaningful for the indefinite long-term" (p. 4-33)
  - For example, the information "needed to preserve the Content Information, to ensure that it is clearly identified, and to understand the environment in which the Content Information was created" (p. 2-6)

## OAIS Preservation Metadata Categories

- Preservation Description Information
  - Reference Information: to enumerate and describe identifiers
  - Provenance Information: to document the history of the content information (creation, modification, custody)
  - Context Information: to document the relationship of the content to its environment
  - Fixity Information: to document authentication mechanisms

**Context Information**  
documents the relationships of the Content Information to its environment (e.g., why it was created, relationships to other Content Information)

- Reason for Creation
- Is Version Of
- Has Version
- Is Replaced By
- Replaces (migration)
- Is Required By
- Requires
- Is Part Of
- Has Part
- Is Referenced By
- References
- Is Format Of
- Has Format
- Same Intellectual Content As

**Preservation Description Information**

|           |            |
|-----------|------------|
| Reference | Provenance |
| Context   | Fixity     |

## The OAIS Information Model

- Lessons from OAIS (1):
  - Data objects need to be closely coupled with additional layers of information (metadata) that will help provide meaning and context, etc.
  - These layers broadly reflect the main characteristics of digital information
  - Produces self-documenting objects

## The OAIS Information Model

- Lessons from OAIS (2):
  - It highlights the importance of preserving context and provenance (but these are quite vaguely defined)
  - OAIS works on an abstract level, defines terms but does not present a preservation metadata schemata
  - ...

## Many Different Schemas

- Cedars, PREMIS, NLA, etc.
- Deciding on a schema or elements depends on many factors, e.g., designated community, significant properties, risk, etc.

## Significant Properties of Digital Objects

Properties that affect the quality, usability, appearance, functionality, "look and feel," and behavior of digital objects

Significant Properties are defined by:  
encoding schemes  
formats  
software  
human intellectual judgments

### Why are Significant Properties Important?

For tangible objects, significant properties affect authenticity, legal admissibility, artifactual value, ease or convenience of use, aesthetic quality, etc.

For digital objects, significant properties affect all of the above, plus:  
 functionality  
 look and feel  
 behavior

### PREMIS Data Dictionary

- Background :
  - PREMIS Working Group (2003-2005)
  - An attempt to develop something that would be implementable
  - Development informed by OAIS model
  - Built upon on several initiatives that had been developing preservation metadata schemas and frameworks prior to 2003
  - Data Dictionary first published in May 2005; v. 2.0 in March 2008

### PREMIS Data Dictionary

- What PREMIS aims to do:
  - The Data Dictionary is specifically focused on defining the core metadata needed for long-term preservation
  - "... the information a repository uses to support the digital preservation process"
    - "... functions to maintain viability, renderability, understandability, authenticity, and identity in a preservation context"
  - Based on a data model

### PREMIS Data Dictionary

- PREMIS Data Model:
  - Recognises that digital preservation is as much about describing processes as well as objects
  - Five entities
    - Intellectual Entities are a set of content that is considered a single intellectual unit for purposes of management and description: e.g., a map, photograph, or database.
    - An Intellectual Entity can include other Intellectual Entities; for example, a Web site can include a Web page; a Web page can include an image.
    - Objects: a discrete unit of information in digital form
    - Events: an action that involves or impacts at least one Object or Agent associated with or known by the preservation repository.
    - Agents person, organization, or software program/system associated with Events in the life of an Object, or with Rights attached to an Object.
    - Rights assertions of one or more rights or permissions pertaining to an Object and/or Agent.

