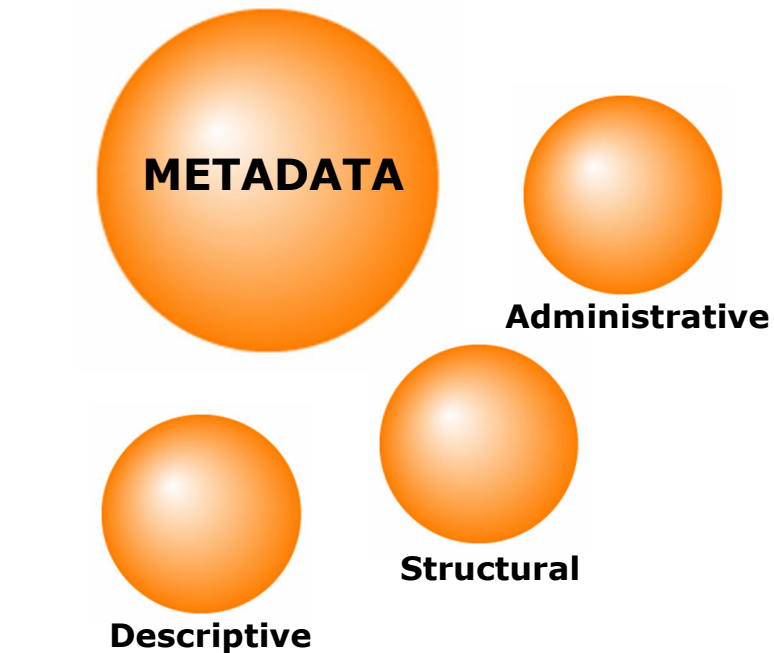


Preservation Metadata: **Setting the Scene**

Brian Lavoie
Senior Research Scientist
Office of Research
OCLC

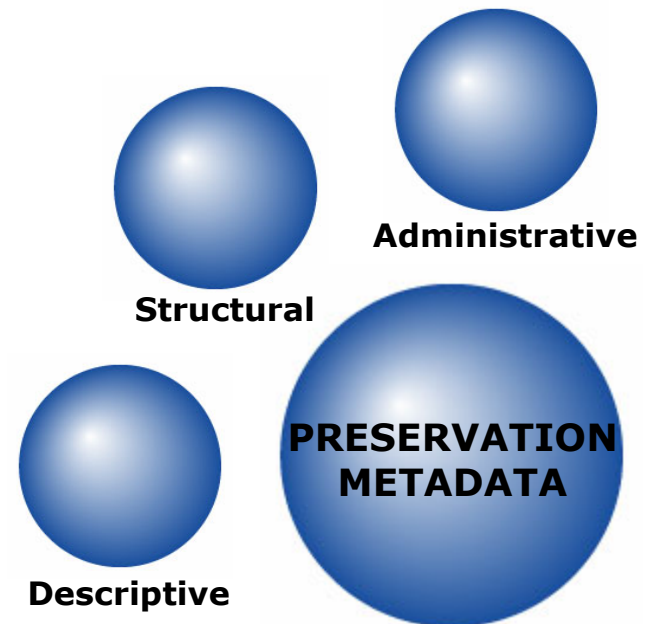
DPC Meeting on Preservation Metadata
London
September 8, 2005

Metadata and **preservation metadata**



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

“Metadata that supports and documents the digital preservation process”



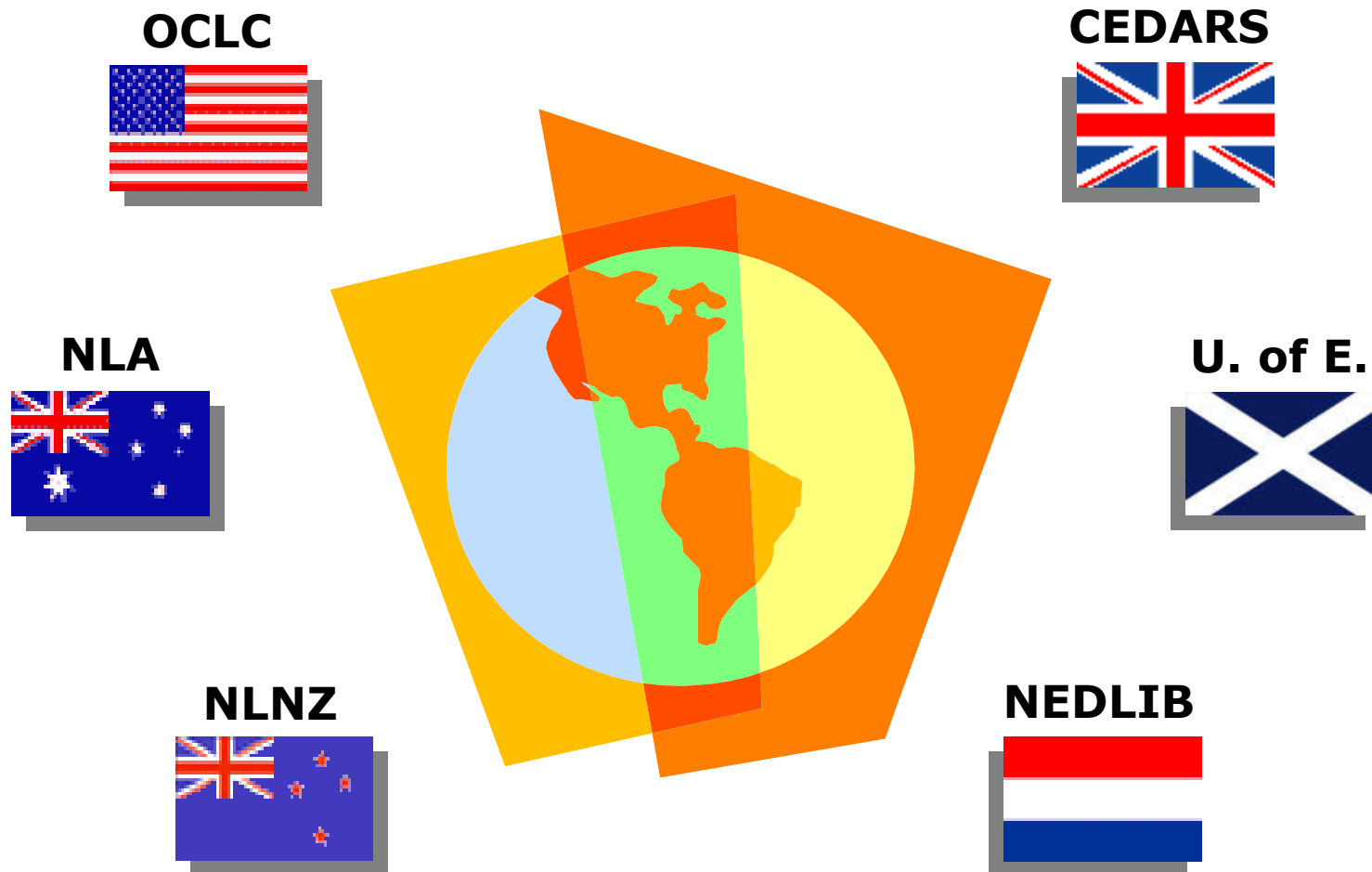
Preservation Metadata: **Examples**

- 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?*

Why is preservation metadata important?

- Digital objects are technology-dependent ...
 - Complex technological environment between content and user
 - Means to access and use archived object must be documented
 - *Technical metadata especially important*
- Digital objects are mutable ...
 - Can be easily altered, impacting look, feel, functionality
 - Changes to object must be documented/validated
 - *Provenance/authenticity metadata especially important*
- Digital objects are bound by intellectual property rights ...
 - Preservation often proceeds while copyright still in effect
 - May constrain preservation activities and access policies
 - *Rights management metadata especially important*
- **Makes digital objects self-documenting across time**

Preservation metadata around the world ...



Towards consensus ...

- March 2000: OCLC, RLG jointly sponsored international working group on preservation metadata
 - Preservation metadata framework (June 2002)
 - Comprehensive, high-level description of types of information constituting preservation metadata
 - Used OAIS information model as starting point
 - Set of “prototype” preservation metadata elements
- The Framework ...
 - Consolidated/synthesized expertise, OAIS, existing element sets
 - Served as foundation/shared departure point for schema implementations
 - But not an “off the shelf, ready to implement” solution
- Need implementable preservation metadata, with guidelines for application and use, relevant to a wide range of digital preservation systems and contexts

PREMIS Working Group

- June 2003: OCLC, RLG sponsored international working group:
 - **PREMIS: Pre**serva**tion M**etada**ta: I**mplementa**tion S**trategies
- Objectives:
 - Identify and evaluate alternative strategies for encoding, storing, managing, and exchanging preservation metadata
 - Define implementable, core preservation metadata, with guidelines/recommendations for management and use
- May 2005: Released *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*:
 - Comprehensive, practical resource for implementing preservation metadata in digital archiving systems

Preservation metadata: Perspectives

- Prospects for consensus, standards ...
 - Foundation starting to coalesce, informing current work
 - Progression from theory to practice (OAIS to PREMIS)
- Re-invention of wheels?
 - Potential overlap between other metadata initiatives
- Support internal functions AND interoperability:
 - Exchange of digital content/metadata in networked digital spaces
- Predicting the future ...
 - Hard to judge effectiveness *a priori*
 - Too much? Too little? (worse!)
 - Important to document and share practical experience

Implementation issues: **Tools**

- General consensus that:
 - 1) Metadata is key component of digital preservation process
 - 2) Preservation metadata is expensive to create and maintain
 - 3) Need to minimize human mediation
- JSTOR/Harvard Object Validation Environment (JHOVE):
 - Identify, validate, and characterize digital object formats
 - Modules for: TIFF (various versions), PDF, XML, and others
- NLNZ Preservation Metadata Extraction Tool:
 - Extracts information from digital file headers (e.g., MS Word, TIFF, WAV, bitmaps); outputs metadata in XML format
- **Surface preservation metadata tools in variety of digital repository environments (Dspace, Fedora, DAITSS)**

Implementation issues: **Economy & efficiency**

- Develop economical/efficient ways of acquiring and maintaining preservation metadata
- PRONOM File Format Registry (UK National Archives)
 - Technical metadata about specific file formats
 - Description of software needed to create, render, migrate formats
 - Metadata created once, re-used many times
- Automatic Exposure (RLG)
 - Facilitate capture of metadata specified in NISO Z39.87 (Technical Metadata for Digital Still Images)
 - Dialog with digital scanner/camera manufacturers
 - Technical metadata automatically captured when object created
- **Reduce cost/increase efficiency by leveraging opportunities for sharing and re-use, and diffusing metadata capture throughout information lifecycle**

Implementation issues: **Packaging**

- Link (physically or logically) archived digital object and all associated metadata
- OAIS Information Package
 - Conceptual structure for information moving into, through, and out of archival system
 - Digital object and its metadata, bound into single logical package
- Metadata Encoding and Transmission Standard (METS)
 - XML schema for encoding descriptive, administrative, and structural metadata associated with digital object
- **Moving digital objects and their metadata across space and time requires standard mechanisms for encoding and exchange**

Implementation issues: **Perspectives**

- Current focus on tools for format validation and technical metadata. Also need work on tools that:
 - Address other forms of preservation metadata
 - Support formal preservation metadata schemas (PREMIS)
- Preservation metadata schema should be:
 - Oriented toward automated capture/processing
 - Implementation neutral: promote flexibility & interoperability
- Division of labor:
 - Map preservation metadata requirements to appropriate stages of information lifecycle
 - Allocate responsibility for collecting metadata
- “Quality assurance”

Looking ahead ...

- Questions of “what type”, “how much” preservation metadata still unsettled ...
 - Digital preservation processes still not fully tested/understood
 - Metadata requirements shaped by local repository characteristics
- Collaboration essential:
 - Pooling expertise from variety of institutional perspectives mitigates uncertainty
 - Highlights points of convergence/divergence; helps distinguish metadata that is widely applicable vs. domain-specific
 - Helps identify best practices and encourages standards-building
- Continue to accumulate practical experience in preservation metadata

More information...

PADI Preservation Metadata Bibliography:

<http://www.nla.gov.au/padi/topics/32.html>

lavoie@oclc.org