Integrating PREMIS and METS

PREMIS Tutorial
Implementers’ Panel
June 21, 2007, 9:00-5:30
Library of Congress, Jefferson Building, Whittall Pavilion

Tom Habing, thabling@uiuc.edu
Grainger Engineering Library Information Center University of Illinois at Urbana-Champaign
Outline

• Project Background
• Hub and Spoke Architecture
• Using PREMIS in our METS Profile
• Our Java API
Quick Project Background
NDIIPP ECHODEP Project
http://ndiipp.uiuc.edu/

• Repository Evaluation
• Tools development
  – Web harvesting and archiving (OCLC’s WAW)
  –** Hub and Spoke interoperability and preservation architecture
**
• Preservation Research
  – preserving the authenticity and semantic meaning of digital resources through time.
Hub and Spoke

• Repository Interoperability Architecture with a forward-looking emphasis on preservation metadata and activities
  – A common METS-based profile
    • http://www.loc.gov/standards/mets/profiles/00000015.xml
    • Focus on preservation, not access
    • May be overlaid on top of, or inherited from, other profiles
  – A Java programming API
  – A series of scripts that use the API and METS profile for creating Information Packages which can be ‘used’ across different repositories
METS Profile in More Detail

• Descriptive Metadata
  – Primary DMD is MODS
  – Alternate DMD are encouraged
  – Provenance for DMD is required

• Technical Metadata
  – PREMIS object entities
  – MIX for images
  – Other metadata for other media types

• Digital Provenance
  – PREMIS events and agents
Using PREMIS in METS

• All linking via ID & IDREF-type attributes *not* identifier elements

• Where to Wrap the PREMIS?
  – Object in techMD
  – Event in digiprovMD
  – Rights in rightsMD
  – Agent in digiprovMD or rightsMD

• All Files at a Composition level of 0
  – No packaging, compression, or encryption
Technical Metadata for Files

• A techMD section wrapping a PREMIS object element is required for each file or bit-stream
  – Minimal required elements: fixity, size, formatDesignation
    • Duplicated in some METS attributes
  – creatingApplication and software are encouraged especially for MIME types starting with ‘application/…’
    • Plan to investigate integration with the GDFR when it becomes available
Technical Metadata for Representations

• Technical metadata can also be associated with representations
  – There is a special required techMD called the ‘primary representation’ that corresponds to the entire METS file. Used mostly for alternate identifiers for the file, but may also be used to record other technical metadata about the whole METS document
  – Each structural map may also have representation technical metadata.
Digital Provenance

• Recorded for all non-trivial changes to:
  – Descriptive Metadata (must)
    • Creation, Transformation, Modification, Deletion
  – Files and Bitstreams (should)
    • Events from PREMIS data dictionary
  – Structural Maps (may)
    • Creation, Transformation, Modification, Deletion

• PREMIS event and optional associated agents are wrapped in a digiprovMD
Java Implementation

• Partially complete and in-work
  – Support for DSpace and to lesser degree Fedora
  – Future plans for additional repository support

• Open source

• Javadopts:
  – http://dli.grainger.uiuc.edu/echodep/HnS/JavaDocs/

• Source Code:
  – http://sourceforge.net/projects/echodep
Technical Architecture (Java)