PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

PREMIS Controlled vocabularies

Rebecca Guenther

Sr. Networking & Standards Specialist,
Library of Congress

rgue@loc.gov

PREMIS Implementation Fair

San Francisco, CA

October 7, 2009

PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

Use of controlled vocabularies in PREMIS

- Many semantic units include a data constraint: "Value should be taken from a controlled vocabulary"
- Most have "suggested values"
- There is not currently a way to enforce selection of a controlled value
- LC is establishing a mechanism to develop vocabularies and to validate them

Example: messageDigestAlgorithm

Semantic unit	**1.5.2.1 messageDigestAlgorithm**		
Semantic components	None		
Definition	The specific algorithm used to construct the message digest for the digital object.		
Data constraint	Value should be taken from a controlled vocabulary.		
Object category	**Representation**	**File**	**Bitstream**
Applicability	Not applicable	Applicable	Applicable
Examples		MD5 Adler-32 HAVAL SHA-1 SHA-256 SHA-384 SHA-512 TIGER WHIRLPOOL	
Repeatability		Not repeatable	Not repeatable
Obligation		Mandatory	Mandatory

PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

Controlled vocabularies databases

- Library of Congress is establishing databases with controlled vocabulary values for standards that it maintains
- Controlled lists are represented using SKOS as well as alternative syntaxes

PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

About SKOS

- Simple Knowledge Organization System
- RDF application used to express knowledge organization systems such as classifications, thesauri, taxonomies, and the concepts within
- Allows distributed, decentralized management of KOS through Linked Data-inspired application.
- All concepts and schemes require a URI

PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

The SKOS data model (Classes)

- **ConceptSchemes** (e.g., published vocabularies, thesauri, code lists, etc.)
- **Concepts** (individual entries or terms within the larger vocabulary)
- **Collections** (logical groupings of Concepts)

PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

Advantages to using SKOS

- SKOS has a defined element set which is particularly relevant for controlled vocabularies
- Relationships between entries in a concept scheme can be expressed (broader, narrower, etc.)
- Relationships between entries in different concept schemes can be expressed (exactMatch, related)
- Having a dereferencable URI for concepts and their concept schemes enhances the ability to provide web services for consumers of these standards

PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

Reasons for developing a web service for vocabularies

- Facilitate development and maintenance process for vocabularies
- Make controlled lists openly available
- Provide comprehensive information about controlled values
- Experiment with semantic web technologies and linked data
- Expose vocabularies to wider communities

PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

Using controlled vocabularies in PREMIS

- Semantic units that specify a controlled vocabulary: realized as "concept scheme"
- Each value: realized as SKOS instance
- Implementers add their values within a concept scheme
- Mechanism to import the values into the PREMIS XML schema to enable validation
- A concept in multiple standards may be established for broad usage in a concept scheme
- There has been some interest in exploring an RDF version of PREMIS for semantic web applications

Those wishing to experiment: http://id.loc.gov

PREMIS PRESERVATION METADATA MAINTENANCE ACTIVITY

Introducing id.loc.gov

- Library of Congress Authorities & Vocabularies service: http://id.loc.gov
- Allows both human-oriented and programmatic access to LC-promulgated authorities and vocabularies.
- First offering is Library of Congress Subject Headings, but more to come: e.g. preservation events, cryptographic hash functions

Demo