Peer Reviewed

Title:
Integrating Metadata Standards to Support Long-Term Preservation of Digital Assets: Developing Best Practices for Expressing Preservation Metadata in a Container Format

Author:
Guenther, Rebecca, Library of Congress
Wolfe, Robert, Massachusetts Institute of Technology

Publication Date:
10-05-2009

Series:
iPRES 2009: the Sixth International Conference on Preservation of Digital Objects

Publication Info:
iPRES 2009: the Sixth International Conference on Preservation of Digital Objects, California Digital Library, UC Office of the President

Permalink:
http://www.escholarship.org/uc/item/0s38n5w4

Multimedia URL:
http://www.cdlib.org/services/uc3/iPres/video.html?file=ipres/Guenther&title=Rebecca%20Guenther%20and%20Robert%20Wolfe%3A%20Integrating%20Metadata%20Standards%20to%20Support%20Long-Term%20Preservation%20of%20Digital%20Assets%3A%20Developing%20Best%20Practices%20for%20Expressing%20Preservation%20Metadata%20in%20a%20Container%20Format

Abstract:
This paper explores the purpose and development of best practice guidelines for the use of preservation metadata as detailed in the PREMIS Data Dictionary for Preservation Metadata within documents conforming to the Metadata Encoding and Transmission Standard (METS). METS is an XML schema that provides a container format integrating various forms of metadata with digital objects or links to digital objects. Because of the flexibility of METS to serve many different functions within digital systems and to support many different metadata structures, integration guidelines will facilitate common practices among institutions. There is constant tension between tighter control over the METS package to support object exchange versus each implementation's unique preservation metadata requirements given the different contexts and implementation models among PREMIS implementers. The PREMIS in METS Guidelines serve primarily as a standard for submission and dissemination information packages. This paper details the issues encountered in using the standards together, and how the METS document changes as events pertaining to the lifecycle of digital assets are recorded for future preservation purposes. The

guidelines have enabled the implementation of an exchange format and creation/validation tools based on the PREMIS in METS guidelines.

Supporting material:
Presentation

Copyright Information:

iPRES 2009

THE SIXTH INTERNATIONAL CONFERENCE ON THE PRESERVATION OF DIGITAL OBJECTS

Proceedings

October 5-6, 2009
Mission Bay Conference Center
San Francisco, California

California Digital Library

Integrating Metadata Standards to Support Long-Term Preservation of Digital Assets: Developing Best Practices for Expressing Preservation Metadata in a Container Format

Rebecca Guenther
Senior Networking and Standards Specialist
Library of Congress, Network Development and MARC Standards Office
Washington, DC 20540
rgue@loc.gov

Robert Wolfe
Metadata Specialist, MIT Libraries
77 Massachusetts Avenue, Cambridge, MA 02152
rwolfe@mit.edu

Abstract

This paper explores the purpose and development of best practice guidelines for the use of preservation metadata as detailed in the *PREMIS Data Dictionary for Preservation Metadata* within documents conforming to the Metadata Encoding and Transmission Standard (METS). METS is an XML schema that provides a container format integrating various forms of metadata with digital objects or links to digital objects. Because of the flexibility of METS to serve many different functions within digital systems and to support many different metadata structures, integration guidelines will facilitate common practices among institutions. There is constant tension between tighter control over the METS package to support object exchange versus each implementation's unique preservation metadata requirements given the different contexts and implementation models among PREMIS implementers. The PREMIS in METS Guidelines serve primarily as a standard for submission and dissemination information packages. This paper details the issues encountered in using the standards together, and how the METS document changes as events pertaining to the lifecycle of digital assets are recorded for future preservation purposes. The guidelines have enabled the implementation of an exchange format and creation/validation tools based on the PREMIS in METS guidelines.
.

Introduction

The challenge and urgency of preserving born digital and digitized information has become a great concern of all institutions responsible for maintaining the wide variety of documentation of human knowledge. Although there are clear advantages of digital over analog media, digital assets risk becoming technically obsolete. Recording key pieces of information about these assets is imperative upon digital repositories that hope to preserve them over time. The *PREMIS Data Dictionary for Preservation Metadata* specifies the information that a repository needs to maintain for the long-term preservation of digital objects. PREMIS itself is a list of data elements (in the *Data Dictionary* referred to as "semantic units") with definitions, examples, creation notes and usage guidelines. It is neutral in terms of the type of system, database or encoding format that implements it. Because many institutions managing digital objects and their metadata use the Metadata Encoding and Transmission Standard (METS) in digital library applications as a container format, this standard is an obvious option as an implementation path. Since an important goal is the **exchange** of objects along with their associated metadata between repositories, many implementers of PREMIS are integrating METS with PREMIS metadata along with other information about

and links to the digital objects. For example, the project Towards Interoperable Preservation Repositories (TIPR) being undertaken by FCLA, Cornell and New York University is using PREMIS embedded in METS as part of a standard Repository Exchange Package format. Because both PREMIS and METS allow for a great deal of flexibility in their implementation, using these two digital library standards together presents issues concerning duplication and management of metadata. As an attempt to address such issues, a working group comprised of PREMIS and METS experts participated in the creation of a set of guidelines for a common exchange standard, which is now being tested in digital preservation repositories.

PREMIS Background and Principles

Many institutions of different types and environments throughout the world have adopted the *PREMIS Data Dictionary for Preservation Metadata* as they attempt to assume responsibility for preservation of their digital assets. This comprehensive specification was first issued as version 1.0 in May 2005 and was then revised as version 2.0 in March 2008. It is maintained by the PREMIS Editorial Committee and Maintenance Activity. In the years since its publication, some countries have begun to embrace PREMIS as part of their preservation infrastructure and mandated its use for certain projects. For instance Spain mandates that PREMIS be implemented in every digitization project funded by the Ministry of Culture.

The PREMIS Data Dictionary defines "preservation metadata" as *the information a repository uses to support the digital preservation process*. Specific preservation functions supported by the metadata are the maintainance of viability, renderability, understandability, authenticity, and identity of digital objects in a preservation context. Different categories of metadata may be considered "preservation metadata," including administrative (i.e. management metadata including rights and permissions), technical (i.e. technical characteristics, often format specific) and structural (i.e. information about the relationships between parts of an object). The documentation of digital provenance (the history of an object) was considered particularly important as well as the documentation of relationships, especially relationships among different objects within the preservation repository.

The PREMIS Working Group, which originally developed the PREMIS Data Dictionary, worked on the principle that the specification would be technically neutral. No assumptions are made as to the specific digital archiving system used, the database architecture, or the archiving technology. In addition the Data Dictionary does not specify details about metadata management, such as whether metadata is stored locally or in an external registry, or whether metadata units are recorded explicitly or known implicitly because of repository policies. The principle of technical neutrality allows for applicability in a wide range of contexts, regardless of the specific type of implementation used for collecting, storing, maintaining, and exchanging the PREMIS metadata. This sort of flexibility allows an institution to use the specification as a key piece of its infrastructure and to adapt it to its own needs. However, there is the disadvantage that implementers then must make their own particular local system decisions and establish local repository policies, which could affect the ability to exchange digital objects and their metadata with other institutions.

The PREMIS Working Group established a data model, which was meant to clarify the meaning and use of the semantic units in the Data Dictionary. It was not intended to prescribe an architecture for implementation, but defined the conceptual entities with which repositories would need to interact. The entities in the data model are Objects, Agents, Events, and Rights; Intellectual Entities are largely out-of-scope but links to them are defined.

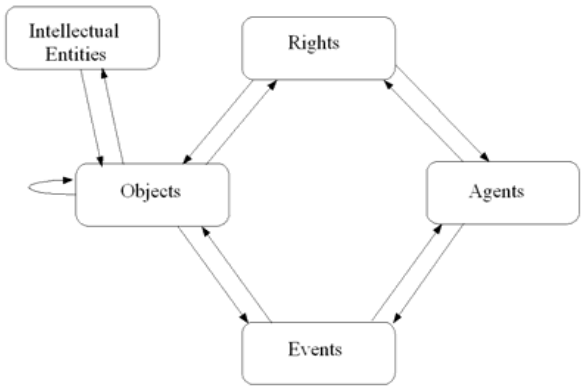

Figure 1: PREMIS Data Model

PREMIS may be implemented in a variety of ways, but, since XML is commonly used for expressing metadata, an XML schema is available to facilitate implementation. Use of the PREMIS data model is evident in the schema design, since it associates appropriate XML elements with each of the PREMIS entities (Object, Events, Agent, or Rights) to which they apply.

The PREMIS Editorial Committee provided a new feature in version 2.0 to allow implementations to include additional local metadata or to provide additional structure or granularity of metadata when PREMIS semantic units were not adequate. This extensibility mechanism is available for the following semantic units: *significantProperties*, *objectCharacteristics*, *creatingApplication*, *environment*, *signatureInformation*, *eventOutcomeDetail*, and *rights*. A container element corresponding to each of these semantic units is available

with "extension" added to the element name. This mechanism provides the flexibility to include metadata defined outside of PREMIS but to include it within the same preservation metadata description. Of particular interest is *objectCharacteristicsExtension*, which allows for including format-specific metadata, which is out of scope for PREMIS itself, into a PREMIS metadata container.

METS as an OAIS Information Package for Objects and Metadata

The Metadata Encoding and Transmission Standard (METS) defines a single document format for describing the structure of complex digital objects and associating various kinds of metadata with their components. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation. METS is expressed using the Extensible Metadata Language (XML) syntax. The content of a METS document may include: a structure map that organizes the components of a digital object; references to all the files that comprise the object; properly associated metadata about the object and its components and files; a map of the hyperlinks between the components; and an association of executable behaviors with the components. Both METS and XML have seen wide adoption, enabling repositories to share digital objects and the responsibility for their preservation. In the exchange of digital objects between repositories, METS can be understood as a very detailed packing slip. The METS document identifies the content of a digital object, represents the content's structure, and associates the contents with their metadata.

The METS document, all of the content files, and any metadata files that are referenced by and not included in the METS document may be considered a package of information for the digital object that may be exchanged or archived. These are called METS packages and they are very flexible in the way they may be used to describe and structure complex digital objects. Further definition of guidelines and restrictions for the use of METS are often needed to facilitate exchange and use of METS packages within a particular domain of information organization. One such domain is Open Archival Information Systems.

The Open Archival Information System (OAIS) standard defines an open framework for the design and function of preservation repositories. This definition contains a reference model that defines the function, organization, and interaction of content within preservation repositories. The list of repository functions that are defined by the reference model include three for which METS is widely used, namely exchange, archiving, and display or dissemination. OAIS also defines an environmental model in which the agents who exercise these functions are identified. In the environmental model, the OAIS archive occupies a central position to three kinds of agents that interact with it. These agents are Producers, Managers, and Consumers. Their roles map to the exchange, archiving, and dissemination of digital objects for which METS is well suited and widely employed.

An information model is also defined within the OAIS standard for the contents of the repository upon which the agents fulfill their functions. An Information Object is defined that is identical to the complex digital objects discussed in this paper. The information object is comprised of a data object and associated representation information. The data object is all of the contents of the information object, physical or digital. Representation information is that which is included in a METS document—the structure and associated metadata of the components of a digital object. A METS package is a good candidate for realization of an information object in an OAIS repository.

These information objects fulfill the functions of exchange, archiving, and dissemination via their deployment as Information Packages. Information Packages, as defined by the OAIS information model, are comprised of the information object, which can be expressed as a METS package, and additional representation information that is specific to preservation of the content. This additional information is called Preservation Descriptive Information. Preservation Descriptive Information is classified as four different kinds—provenance, context, reference, and fixity. These kinds of information are well defined within the PREMIS Data Dictionary, making it an ideal candidate for expressing Preservation Descriptive Information. Information Packages within the OAIS information model can be expressed as a METS package with additional preservation metadata encoded as PREMIS metadata The OAIS information model describes three kinds of packages that correspond to the aligned functions of OAIS repositories and METS and PREMIS packages. The Submission Information Package (SIP) is used for exchange. The Archival Information Package (AIP) is used for long-term storage and preservation. The Dissemination Information Package (DIP) is used to provide instructions for display and dissemination. The contents of each information package, especially the METS and PREMIS metadata, may be different for each kind of package. Guidelines for the use of PREMIS in METS are crucial for this use as information packages in an OAIS repository, providing clear instructions for the use of the two standards to express each kind of package.

One of the areas in which guidelines are needed for including PREMIS in METS documents is the way representation metadata, including preservation descriptive information, may be included and associated with package components. METS has defined several container elements for including representation metadata in the METS package. The elements categorize metadata into different kinds. The kinds identified by METS are descriptive metadata (via the dmdSec element) and

administrative metadata (via the amdSec). Administrative metadata is broken down further into four kinds:

- Technical Metadata (via the techMD element)
- Rights Metadata (via the rightsMD element)
- Digital Provenance Metadata (via the digiprovMD element)
- Source Metadata (via the sourceMD element)

The entities in the PREMIS data model (Events, Agents, Objects, Rights), do not map perfectly to the available metadata container elements in METS. The PREMIS in METS guidelines seek to provide the needed rules for the use of PREMIS in METS as a SIP, AIP, or DIP, as well as instructions for including metadata about each of the PREMIS entities in a METS document.

The Development of Guidelines for Using PREMIS with METS

The *Guidelines for Using PREMIS with METS for Exchange* was developed by a working group of PREMIS and METS experts to specify preferred implementation choices in using these two flexible standards together. It offers guidance and suggests best practices for using the PREMIS schema as METS extensions within a METS container, focusing on exchanging information packages.

The PREMIS Data Dictionary was built upon the OAIS Reference and Information Models. The OAIS information model provides a conceptual foundation in the form of a taxonomy of information objects and packages for archived objects and the structure of their associated metadata. The PREMIS Data Dictionary is an elaboration and translation of OAIS into implementable semantic units, intended for Submission Information Packages and/or Dissemination Information Packages (key concepts in OAIS). Since METS itself can be considered an information package in the OAIS sense, the *Guidelines for Using PREMIS with METS for Exchange* specifies an information package with preservation metadata.

During the attempt to establish some common usage scenarios for using PREMIS with METS it became clear that the guidelines needed to focus on the METS document as a mechanism for exchange of digital objects and their metadata, whereby a repository is either receiving a submission information package or producing a dissemination information package for exchange. Using METS within an archival information (storage) package is out of scope for the guidelines, although an institution may choose to use it internally if desired. The METS document is considered to be an exchange vehicle to facilitate communication when internal requirements and technical environments vary considerably.

In the development of the guidelines there has been an ongoing tension between allowing for flexibility and being prescriptive to facilitate interoperability. Additional usage scenarios beyond those suggested above may emerge over time, resulting in revision of the guidelines. Important considerations in using the guidelines for a particular implementation are what tools the repository is using for generating or storing METS structures, whether METS or PREMIS is primary in terms of maintenance and reliability, and whether the goal of the repository is preservation or delivery. The answers to these questions may influence implementation choices made in encoding the metadata in the METS document. Many institutions may require a more prescriptive approach to allow for an efficient processing of METS documents for exchange purposes, whereby it is predictable what form the metadata will take and what programs need to be written to process the document for a variety of purposes.

One likely exchange scenario is converting data from internal structures to PREMIS and wrapping it in METS before transmitting it to the destination. At the destination side the METS might get unwrapped. The data then might be stored in different structures, for instance a local database. Thus, the PREMIS data may become separated from its METS wrapper once it has reached its destination and later could be put together again.

Whether the METS document is an exchange object intended for display and delivery may require different encoding decisions than when it is primarily an object of preservation. The guidelines suggest that when using METS as a submission information package (SIP), a more liberal approach is possible because the trusted repository will likely have processes and internal requirements that can generate data that is not in the submission. By contrast, when using METS as a dissemination package a more restrictive approach may be required with more authoritative metadata. In the case of dissemination, the distributing repository will need to make choices as to how to output the data, which may then be stored in some other form by whoever receives it. Repositories that receive dissemination packages may have varying levels of functionality, and predictability may be important. Rather than produce two sets of guidelines for the different use cases, the working group suggests that institutions exchanging METS documents might establish profiles to document the choices made.

Implementation Issues for PREMIS in METS Guidelines

It is likely that different preservation repository implementations will result in different implementations of PREMIS in METS. The Guidelines developed for these implementations seek to balance the requirements of the individual repositories with the need for restrictions to the flexibility of METS in order to enable communication and exchange between repositories. This balance is in part achieved by restricting the focus of the guidelines to SIPs and DIPs. Individual repositories will organize preservation metadata within their systems according to their own local practices and data structures. The PREMIS in METS guidelines suggest common practices

for structuring preservation metadata for exchange. They do not offer recommendations for the structure of AIPs, nor do they offer any restrictions to local definitions of preservation metadata. The guidelines provide minimal requirements for the structure of preservation metadata that is designated for exchange.

The guidelines also balance flexibility and control by not strictly prescribing practice in key recommendations. The guidelines seek to provide recommendations in the following areas without restriction to a single practice:

- the location of PREMIS metadata within METS documents
- the use of metadata elements that occur in both PREMIS and METS
- the inclusion of PREMIS in a METS document whose purpose includes more than preservation
- the relationship of the different structural metadata mechanisms in PREMIS and METS
- the recording of PREMIS Agent entities in METS documents
- the use of identifiers to link elements in PREMIS and METS

METS Documents have two kinds of metadata containers, descriptive and administrative. Preservation metadata is widely considered to be administrative metadata. Among the types of administrative metadata defined by METS (technical, rights, provenance, and source), there is no provision for metadata whose purpose is the preservation of digital objects. Additionally, preservation metadata itself may contain technical, rights, and provenance information. This poses a difficult question when choosing a location for the inclusion of PREMIS metadata within a METS document. Should the entire preservation metadata record be contained within one of the ill-fitting administrative metadata sections in METS? Or, should the preservation record be split up and put into different sections? The PREMIS in METS guidelines are not prescriptive concerning which of these methods should be chosen. Instead, they provide instructions for each method of inclusion.

Several metadata elements occur in both the METS and PREMIS schemas. These elements share semantic definitions. The PREMIS in METS guidelines do not provide an order of preference for the use of identical elements from the two schemas. They outline a number of considerations for implementors as they choose an order of preference for metadata in both schemas. These considerations include the primary use of the metadata, the expressiveness of the two schemas, and whether or not to record the metadata in both schemas. The guidelines encourage recording metadata redundantly between METS and PREMIS when the specifics of consumption and use of the metadata are unknown.

METS documents serve many different functions within repositories. Preservation may be just one of the intended uses of a single METS document. The document may also be used to provide access to descriptive and technical metadata for its digital object. Equivalent metadata elements may be used to describe the object in two different administrative metadata sections and in two different metadata schemas. For example, the same metadata may occur in a PREMIS record included in the *digiprovMD* element and a Metadata for Images in XML (MIX) record included in the *techMD* element. The MIX metadata record may also occur within the PREMIS record, the PREMIS record contained within a METS document. Since the inclusion of an extensibility mechanism in PREMIS version 2.0, PREMIS, like METS, is a container for metadata from other schemas. This overlap of function and description is addressed in the guidelines with a list of questions repositories should consider in choosing where to record metadata and whether to store metadata redundantly.

METS records the relationships between components of digital objects in a structure map made up of nested *div* elements. These relationships are primarily content relationships, richly expressed in the METS structure map. PREMIS also contains relationship elements that can record the content relationships of the digital object. However, the METS structure map is generally a richer representation of the object's structure and its use is recommended by the guidelines. The PREMIS relationship elements are recommended for use in the preservation context, even though redundant with the METS structMap, and to map relationships among entities in the PREMIS data model (agents, events, objects, rights).

PREMIS Agents do not have a natural METS administrative metadata section. The guidelines identify two sections in which Agent metadata could appear, based upon the Agent's relationship to other PREMIS entities. If the Agent is defined in the context of a PREMIS Event, it should appear in the *digiprovMD* section. If it is defined in the context of a Right, it should appear in the *rightsMD* section. Agents could be involved in multiple events or rights statements. The guidelines recommend avoiding redundancy by placing an Agent's metadata in either a single *digiprovMD* or *rightsMD* section.

Both PREMIS and METS employ the XML ID and IDREF attribute types to link between their document sections. This is the primary mechanism by which components in the METS structure map and files in the METS file section are linked to their metadata, including their PREMIS metadata. It is also the mechanism by which PREMIS entities are related to each other. The guidelines recommend the separation of these mechanisms. METS elements referencing other METS elements and PREMIS elements other PREMIS elements.

Implementation of an Exchange Standard

PREMIS implementation tool

Successful implementation of any standard used for exchange of data across a variety of institutions is often

dependent upon adequate tools for creating and managing the objects being exchanged. The Library of Congress has contracted with the Florida Center of Library Automation (FCLA) for a set of PREMIS-related tools. One tool will convert between a METS document with embedded PREMIS metadata into a stand-alone PREMIS document. Another tool will convert a stand-alone PREMIS document into a PREMIS in METS document by segmenting the PREMIS data elements into the appropriate METS sections. Types of file formats that will be covered are text, image, audio, video and software.

In developing these tools, it is clear that explicit guidelines were needed to produce a predictable encoding without human intervention that different systems would know how to process. Where the guidelines offer choices, the tool makes implementation decisions that are followed by exchange partners.

In addition to the conversions described above, the toolkit will also provide an application to describe any single file in PREMIS-compatible terms. When a file is submitted, the DROID application is used for format identification and JHOVE for validation and extraction of technical metadata. PREMIS elements are populated if the information is available in the file and a PREMIS XML instance that can be validated against the PREMIS schema is the output. That file could subsequently be run through the PREMIS-to-PREMIS-in-METS converter if desired.

The toolkit will be made available during the fourth quarter of 2009.

Using Controlled vocabularies

Since most preservation repositories will have to deal with large quantities of data, metadata should be automatically created and used as much as possible. A number of PREMIS semantic units specify that best practice is to use a controlled vocabulary for values; the Data Dictionary generally gives starter lists which implementers may adapt to their purposes. This allows for values of PREMIS semantic units to be parsed programmatically, validated against a list of allowed values, or supplied automatically by ingest programs.

The PREMIS Maintenance Activity at the Library of Congress is establishing a mechanism to register controlled vocabularies in use with PREMIS semantic units and expose them in a way that the PREMIS schemas can include them for validation purposes. Repositories may use these or define their own, but it should be clear what the source of each controlled vocabulary is when exporting metadata for exchange. Interoperability is enhanced if common vocabularies are used and declared. In the future LC will provide a mechanism for externally created and controlled vocabularies to be registered within this service.

The mechanism for documenting PREMIS controlled vocabularies will be under the functions of the service provided by the Library of Congress under http://id.loc.gov. This will be part of the Library of Congress Authorities and Vocabularies service that enables both humans and machines to programmatically access data at the Library of Congress. This service is influenced by and implements the Linked Data movement's approach of exposing and inter-connecting data on the Web via dereferenceable URIs. It uses the Simple Knowledge Organization System (SKOS) model and markup, which defines "concept schemes" (aggregations of one or more concepts with semantic relationships defined) and "concepts" (specific units or values within a concept scheme). The main application provides resolvability to values and vocabularies by assigning URIs. Each vocabulary possesses a resolvable URI, as does each data value within it.

It is likely that initially controlled vocabularies for event type, preservation roles, and cryptographic hash functions (i.e. checksum type) will be available. Later additional PREMIS controlled vocabularies will be included. In cases where there is duplication of values (with or without the same names) in PREMIS, METS and/or MIX (NISO Metadata for Images in XML Schema), these will be brought together in a collection with relationships between names in different schemas declared.

For example, a "cryptographic hash function" is a general term for what is called in PREMIS and MIX *messageDigestAlgorithm* (under objectCharacteristics) and in METS as *checkumType* (an attribute for files). There are also minor variations in the enumerated values for cryptographic hash function within each of these schemas. The SKOS data details relationships between each list and between the values.

Algorithm	PREMIS	METS	MIX
Adler-32	*	*	*
CRC-32		*	*
Haval	*	*	*
MD5	*	*	*
MNP		*	*
SHA-1	*	*	*
SHA-256	*	*	*
SHA-384	*	*	*
SHA-512	*	*	*
TIGER	*	*	*
Whirlpool	*	*	*
unknown			*

Figure 2: Comparison of values for Cryptographic hash functions in PREMIS, METS and MIX. Note that Adler-32, CRC-32 and MNP were added to METS in April 2009.

Managing PREMIS and METS Exchange Metadata

Individual repository implementations will define their own practices for managing metadata documents within their repository systems. Some may choose to

archive, update, and redistribute the METS document, managing digital objects and their metadata through manipulation of the stored document. Other repositories may choose to disassemble METS documents, reassembling them upon demand. In both cases, changes to this metadata need to be tracked. This provenance metadata is usually reserved for the content of a digital preservation repository. The use of METS documents to exchange preservation metadata between repositories creates a need to also record the provenance of the metadata document.

A scenario may be imagined in which one repository transfers content and metadata via PREMIS in METS to a second repository with the stipulation that the version deposited will always be available for retrieval regardless of any changes that may take between deposit and
that retrieval. The repository that receives METS documents with this stipulation either has to preserve that document, or be prepared to recreate it exactly upon demand. Both strategies will rely upon provenance metadata about the METS and PREMIS metadata. PREMIS structures for capturing preservation metadata about digital objects may also be used to record and communicate the provenance of metadata. Some implementations have already used PREMIS event metadata to record transformations or modifications of the PREMIS metadata within the METS package. PREMIS Agent entities may be used to record the authority that created or modified metadata.

Sharing METS documents and their provenance metadata poses interesting questions about the proper location of this metadata within a METS document, how it may be distinguished from provenance metadata about digital objects, and how it may reference the METS document within which it is contained. These and other open questions about the long-term management of metadata within digital preservation repositories may provide challenges that will require additional guidance for using PREMIS with METS.

Conclusions

The PREMIS and METS standards can provide an important part of the infrastructure for preservation repositories in the implementation of preservation metadata (i.e. representation information) for digital objects. The development of guidelines has proven crucial for integrating these standards for use as an information package. Since both PREMIS in METS contain choices in application, balancing the need for flexibility versus control has been a challenge. The Guidelines for Using PREMIS in METS for Exchange provides guidance, but many of the recommendations remain non-prescriptive in order for repositories to make choices according to the functions of the METS package. Particular implementations will require more controlled decisions based on agreement between exchange partners.

References

PREMIS Editorial Committee. March 2008. *PREMIS Data Dictionary for Preservation Metadata. Version 2.0.*

PREMIS Maintenance Activity, Library of Congress. http://www.loc.gov/standards/premis

PREMIS Preservation Metadata Schema, version 2.0. 2008 http://www.loc.gov/standards/premis/premis.xsd

METS Editorial Board. September 2007. *<METS> Metadata Encoding and Transmission Standard: Primer and Reference Manual. Version 1.6*

METS Maintenance Activity, Library of Congress. http://www.loc.gov/standards/mets/

METS Schema, version 1.8. 2009 http://www.loc.gov/standards/mets/mets.xsd

Guidelines for Using PREMIS with METS for Exchange. Sept. 2008. http://www.loc.gov/standards/premis/guidelines-premismets.pdf

Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System*. January 2002. http://public.ccsds.org/publications/archive/650x0b1.pdf

Library of Congress. *Authorities and Vocabularies.* http://id.loc.gov

Linked Data: Connect Distributed Data Across the Web. http://linkeddata.org/

W3C Semantic Web Activity. *SKOS: Simple Knowledge Organization System Home Page* http://www.w3.org/2004/02/skos/

MIX: NISO Metadata for Images in XML Schema http://www.loc.gov/standards/mix/