Preservation Metadata (2nd edition)

Brian Lavoie and Richard Gartner

DPC Technology Watch Report 13-03 May 2013

Series editors on behalf of the DPC
Charles Beagrie Ltd.

Principal Investigator for the Series
Neil Beagrie

DigitalPreservationCoalition

DPC Technology Watch Series

Published in association with Charles Beagrie Ltd.

ISSN: ISSN 2048-7916
DOI: http://dx.doi.org/10.7207/twr13-03

Second Edition

First edition published in Great Britain in 2005 by the Digital Preservation Coalition. Second edition published in Great Britain in 2013 by the Digital Preservation Coalition.

Acknowledgements

The authors would like to thank series editor Neil Beagrie and our reviewers for many helpful comments and suggestions, which significantly improved the final version of this report.

Foreword

The Digital Preservation Coalition (DPC) is an advocate and catalyst for digital preservation, ensuring our members can deliver resilient long-term access to digital content and services. It is a not-for-profit membership organization whose primary objective is to raise awareness of the importance of the preservation of digital material and the attendant strategic, cultural and technological issues. It supports its members through knowledge exchange, capacity building, assurance, advocacy and partnership. The DPC's vision is to make our digital memory accessible tomorrow.

The *DPC Technology Watch Reports* identify, delineate, monitor and address topics that have a major bearing on ensuring our collected digital memory will be available tomorrow. They provide an advanced introduction in order to support those charged with ensuring a robust digital memory, and they are of general interest to a wide and international audience with interests in computing, information management, collections management and technology. The reports are commissioned after consultation among DPC members about shared priorities and challenges; they are commissioned from experts; and they are thoroughly scrutinized by peers before being released. The authors are asked to provide reports that are informed, current, concise and balanced; that lower the barriers to participation in digital preservation; and that they are of wide utility. The reports are a distinctive and lasting contribution to the dissemination of good practice in digital preservation.

This report was written by Richard Gartner and Brian Lavoie, specialists in digital preservation and metadata. The report is published by the DPC in association with Charles Beagrie Ltd. Neil Beagrie, Director of Consultancy at Charles Beagrie Ltd, was commissioned to act as principal investigator for, and managing editor of, this Series in 2011. He has been further supported by an Editorial Board drawn from DPC members and peer reviewers who comment on text prior to release: William Kilbride (Chair), Neil Beagrie (Managing Editor), Janet Delve (University of Portsmouth), Sarah Higgins (University of Aberystwyth), Tim Keefe (Trinity College Dublin), Andrew McHugh (University of Glasgow) and Dave Thompson (Wellcome Library).

Contents

Abstract

In the space of less than a decade, preservation metadata has evolved from a research topic to an integral part of best practice for the long-term stewardship of digital materials. The first edition of this report chronicled the evolution of preservation metadata from concept to standard, ending with the release of the *PREMIS Data Dictionary*. In second edition, this report focuses on new developments in preservation metadata made possible by the emergence of PREMIS as a *de facto* international standard. The report is intended for digital preservation practitioners interested in learning about the key developments in preservation metadata, especially as these developments concern the *PREMIS Data Dictionary*; the report will also be of interest to anyone seeking to learn more about the general topic of preservation metadata. The focus of work in preservation metadata has shifted from theory to practice; consequently, this report focuses on the key implementation topics that have emerged since the publication of the *PREMIS Data Dictionary*, including revisions of the *Data Dictionary*; community outreach; packaging (with a focus on METS), tools, PREMIS implementations in digital preservation systems, and implementation resources. The report also suggests some areas which future work in preservation metadata should address.

Executive Summary

The *PREMIS Data Dictionary for Preservation Metadata* won the International Digital Preservation Award in 2005[1]. It is remarkable to consider that only a few years before, a great deal of uncertainty had prevailed concerning what preservation metadata was and why it was important; there was certainly no agreed-upon standard for implementing it. In contrast, we now have a de facto international standard for preservation metadata – the *PREMIS Data Dictionary* – which has been implemented in digital preservation repositories worldwide, and incorporated into a variety of digital preservation systems and tools. It is no exaggeration to assert that preservation metadata, and the *PREMIS Data Dictionary* in particular, have become part of best practice underpinning responsible long-term stewardship of digital materials.

The year the PREMIS Data Dictionary won the Digital Preservation Award coincided with the publication of the first edition of our Digital Preservation Coalition *Technology Watch Report Preservation Metadata* (Lavoie and Gartner, 2005). Since the publication of that report, much has happened in the sphere of preservation metadata. The key trend characterizing these developments is a shift in focus from conceptual issues (i.e., what is preservation metadata, how should it be defined) to implementation issues (given a preservation metadata standard, how can it easily and efficiently be incorporated into real-world digital archiving systems). The emergence of the *PREMIS Data Dictionary* as the *de facto* standard for preservation metadata has served to consolidate and frame much of the current work in this area.

Our first report primarily chronicled the evolution of preservation metadata from concept to standard, ending with the release of the *PREMIS Data Dictionary*. The report noted that 'most activity to date in the area of preservation metadata has been devoted to schema development… If the [PREMIS] *Data Dictionary* does become a standard in the community, a critical gap will have been filled, and preservation metadata activities can focus energy and resources on other problems…' (Lavoie and Gartner, 2005, pp. 18–19). The second edition of the report updates the first by picking up the story of preservation metadata after the release of PREMIS. The focus here is on new developments in preservation metadata that have been made possible by the emergence and take-up of PREMIS. In this sense, the speculation in the passage above from the first report has become reality: PREMIS did fill a critical gap by becoming an accepted international standard, and preservation metadata work has focused on a range of other issues that take as a starting point the *PREMIS Data Dictionary*.

Preservation metadata is metadata that supports the distinct requirements of digital preservation: maintaining the availability, identity, persistence, renderability, understandability and authenticity of digital objects over long periods of time. Preservation metadata has moved rapidly from theory to practice. The OAIS information model conceptualized the types of information that fall within the scope of preservation metadata. More recently, the PREMIS working group defined a core set of implementable, broadly applicable preservation metadata elements, supported by a *Data Dictionary* providing guidelines and recommendations for populating and managing the elements. The *Data Dictionary* is organized around a data model consisting of five entities associated with the digital preservation process: *Intellectual Entity*, *Object*, *Event*, *Agent*, and *Rights*. Every entity is

[1] http://www.dpconline.org/newsroom/not-so-new/110-awards-2005 PREMIS (Preservation Metadata: Implementation Strategies) was an international working group formed to promote consensus-making and convergence in preservation metadata. See Section 3 for more details.

described by a set of properties called semantic units, each of which represents a discrete piece of information to be recorded as part of the metadata supporting the digital preservation process.

The emergence of the *PREMIS Data Dictionary* as the de facto international standard for preservation metadata, and its take-up in an increasing number of digital preservation systems and tools, mean that much of the significant implementation work in the area of preservation metadata has coalesced around the *Data Dictionary*. Important developments include:

Revisions of the PREMIS Data Dictionary
PREMIS 2.0 incorporated four major updates, including bi-directionality of relationships between entities in the data model; an expanded Rights entity; structured descriptions of significant properties and preservation level; and introduction of a formal mechanism to support extensibility when using the PREMIS XML schema. PREMIS 2.1 added several new semantic units for Agents, and restructured the extensibility mechanism to more closely resemble the extension schemas used in METS. PREMIS 2.2 adds new semantic units to the Rights entity, as well as several updates to the PREMIS XML schema.

Outreach
The PREMIS Editorial Committee has engaged in outreach activities aimed at raising awareness about PREMIS and preservation metadata in the digital preservation community, through tutorials, publications, and more recently, 'implementation fairs' that feature reports from the Editorial Committee and presentations from PREMIS implementers discussing issues and solutions, tools, and best practices. The PREMIS Implementers' Group listserv provides an open forum for the discussion of issues related to PREMIS and the general topic of preservation metadata.

Packaging
The most widely-used framework for storing preservation metadata and linking it to other types of metadata is METS (Metadata Encoding and Transmission Standard), an XML implementation of an OAIS Information Package. A set of guidelines has been published which provide pragmatic recommendations for using PREMIS with METS. A checklist has also been published for documenting in a METS Profile the decisions made when implementing PREMIS in METS.

PREMIS Tools
The process of implementing PREMIS in a working environment is made easier by a number of tools which can extract metadata from digital objects, and in some cases, output PREMIS XML. Examples include JHOVE (JSTOR/Harvard Object Validation Environment) and DROID (Digital Record Object Identification). The PREMIS in METS Toolbox validates the conformance of a METS document with embedded PREMIS metadata to the Library of Congress's PREMIS-in-METS Guidelines.

PREMIS Implementations
The PREMIS Maintenance Activity maintains an active registry of PREMIS implementations. The functions performed by PREMIS in each implementation vary considerably, and few use all of its features consistently. Of the five PREMIS entities, Object and Event (for provenance verification and change tracking) are the most commonly used. The metadata architectures within which PREMIS is deployed tend to be XML-based, with many using METS.

Other Implementation Resources

The PREMIS Editorial Committee has sponsored the creation of the *PREMIS OWL ontology*, which permits implementers to express the semantics of the *Data Dictionary* in RDF. A *collection of PREMIS controlled vocabularies*, represented in SKOS (Simple Knowledge Organization System) and other formats have been deployed on the Library of Congress's 'id.loc.gov' web service. The Editorial Committee has updated and expanded its definition of *conformance* to provide greater clarity on what PREMIS conformance means in practice. Finally, the TIPR (Towards Interoperable Preservation Repositories) project has developed a protocol for the *exchange of PREMIS-conformant preservation metadata* across repositories.

An area that would benefit from increased attention in the next phase of preservation metadata work is the *accumulation and consolidation of best practice.* Despite the fact that preservation metadata is now a common feature of digital preservation activities, there is little work that draws together and synthesizes the implementation experience that is rapidly accumulating in the digital preservation community. In addition, more work is needed to assess the c*osts and benefits of preservation metadata.* Estimates are needed of the costs involved in collecting and managing preservation metadata; at the same time, more evidence needs to be assembled to demonstrate the practical benefit of incurring these costs, especially in terms of how preservation metadata directly informs and supports digital preservation decision-making and workflows.

1. Background

Definition and importance of preservation metadata

Metadata – 'data about data' – is a familiar concept for information professionals, although *preservation metadata* is perhaps less so. Even though preservation metadata, in one form or another, is a standard component of most digital preservation implementations, ambiguity still surrounds its scope, and even its purpose. One reason that preservation metadata is difficult to categorize precisely is that it does not fit neatly within well-known categories such as descriptive, structural, or administrative metadata (see Glossary for definitions) (Caplan, 2003, p. 3). Instead, preservation metadata can extend across all three. Therefore, the scope of preservation metadata is best understood not so much on the basis of the detailed function of the metadata – i.e., to describe, to structure, to administer – but instead on the process, or larger purpose that the metadata is intended to support. And this is where a definition of preservation metadata begins: it is *metadata that supports the process of long-term digital preservation*.

This terse definition leaves quite a bit to unpack. The notion of 'metadata that supports digital preservation' is a telescoping one that can cover expansive (e.g., any metadata created, managed, and used by a digital repository) or narrow (metadata about file formats and rendering software stored in registries like PRONOM) swathes of information. In practice, preservation metadata is more than just technical metadata (i.e., the technical digital properties of archived objects), yet is something less than any form of metadata a digital repository finds useful to record. The true boundary lies somewhere in between.

One way to clarify the scope of preservation metadata is to focus on why it is important. Preservation metadata is important because it supports the distinct requirements of digital preservation, as opposed to other aspects of information management. Put another way, it facilitates the process of achieving the general goals of most digital preservation efforts: maintaining the availability, identity, persistence, renderability, understandability, and authenticity of digital objects over long periods of time (Vermaaten, Lavoie, and Caplan, 2012). It is helpful to invoke these goals as a way of thinking about whether a particular piece of information can be construed as falling within the scope of preservation metadata.

Fundamentally, preservation metadata establishes an informational frame of reference around a preserved digital object that remains attached to that object over time. The basic idea is that maintaining the ability to exploit the full value of a preserved digital object into the future requires preserving this frame of reference in the form of well-maintained preservation metadata. The frame of reference can be interpreted in a variety of ways, but generally speaking, it encompasses:

- The *provenance* of the object: Information describing the custodial history of the object, potentially stretching back to the object's creation, and moving forward through successive changes in physical custody and/or ownership. Provenance information includes descriptions of the actions that have been taken to preserve the object over time. Such information describes aspects of the digital preservation process used to maintain the object; it would also record any consequences of this process that alter the content, or look, feel, and functionality of the object. Related to this would be information that serves to establish and validate the object's authenticity, i.e., that the preserved object is in fact what it purports to be, and has not been altered, intentionally or unintentionally, in an undocumented way. Authenticity would include such elements as fixity and integrity.

- *Rights management* information: Information that describes any intellectual property rights currently in force that may validate (or limit) the repository's powers to preserve the object, or provide access to it. Such information would document the nature of the intellectual property rights relevant to the digital preservation process, as well as the source from which the right is conferred: for example, statutes, licenses, copyright, etc.

- The technical and interpretative *environment* associated with the object: Information that describes the technical requirements needed to access, render, and use the object. Such information would include a description of the object's file format, as well as the software applications, operating system, and hardware needed to make the object usable, given the state in which it is currently stored in the repository. In addition to technical information, 'intellectual' information may be needed to make the preserved object understandable to future users. For example, if the object consists of a set of records describing a sequence of climate observations, interpretative information might include a data dictionary describing the record structure and the meaning of each field, as well as a description of the instruments and instrument calibrations used to record the observations. Information in this category aligns with what the OAIS reference model describes as representation information – see below for more information.

These three categories of information, while broad in their definition and certainly not exhaustive in their description of all the types of information that could potentially be included in preservation metadata, nevertheless are a useful guide to thinking about its scope. This becomes clearer when looking at examples. Consider information recording the results of a virus check performed routinely upon ingest of an object into a repository. Is this within the scope of preservation metadata? Indeed it is, because it helps establish the provenance of the object. More specifically, it confirms that the object, as retained in the repository, is uncorrupted by malicious code that would impair access to its content or functionalities. This speaks to the object's authenticity. In the same way, the results of a checksum test would also be within the scope of preservation metadata, in that it establishes authenticity by validating that the object has not been altered in any way during its period of retention in the repository.

We can also use the three categories defined above to help establish that a certain piece of information falls outside the scope of preservation metadata. Consider a set of keywords or a brief abstract describing the subject content of a preserved digital object. Is this within the scope of preservation metadata? In this case, the answer is no, because such information does not directly support the long-term digital preservation process, or more specifically, it does not establish anything to do with the object's provenance, the preservation activity performed on the object, rights associated with the object, or the technical and interpretative environment needed to render and use the object. What this information does do is support *discovery* of the object: i.e., it helps make it 'findable' to potential users who would benefit from access to it. But such information is not part of the context needed to ensure that the object is preserved in a usable form over the long term, and in a strict sense, would generally not be considered part of the scope of preservation metadata. That is not to say that descriptive information supporting discovery is not important; it would however, likely be defined outside the preservation metadata schema.

In the final analysis, there is no clear, unchanging boundary between what is preservationmetadata and what is not. At a conceptual level, we can assert the general principle defining the purpose of

preservation metadata as supporting the goals of long-term digital preservation, which are to maintain the availability, identity, persistence, renderability, understandability, and authenticity of digital objects over long periods of time. This in turn leads us to some basic categories of information – provenance, preservation activity, rights, environment – that, broadly speaking, define the contours of preservation metadata. The exact nature of the metadata elements recorded under the auspices of preservation metadata will vary from schema to schema, and even from implementation to implementation within a schema. But the three categories of information described above can be used as a rough guide to the scope of preservation metadata.[2]

The fact that preservation metadata activities have coalesced around a de facto standard – the *PREMIS Data Dictionary* – means that in practice, the scope of preservation metadata has migrated from a purely conceptual question to the more practical issue of what is in or out of scope for PREMIS. With this in mind, it is useful to turn to a brief history of the development of formal preservation metadata schemas, with a special emphasis on the development of the *PREMIS Data Dictionary*.

[2] For a more detailed discussion of the conceptual underpinnings of preservation metadata, see the Preservation Metadata Framework report (OCLC and RLG, 2002). The framework is briefly described later in this report.

2. Preservation Metadata Schema Development

Engagement with preservation metadata moved quite rapidly from theory to practice. In part, this mirrored conditions in the digital preservation area itself, where efforts to develop a solid foundation of digital preservation best practices were paralleled by an immediate need to implement capacity to preserve at-risk digital materials. The movement from theory to practice in preservation metadata cannot be traced as a straight line, but rather as a series of overlapping initiatives straddling research and development, with a substantial dose of cross-fertilization at the boundaries.

2.1. Preservation metadata element sets

As the need to develop operational digital preservation capacity began to surface, a number of institutions undertook to develop preservation metadata element sets to support efforts to preserve digital materials. There is no space here to attempt an exhaustive list of these element sets, but it is useful to briefly mention a few examples of how institutions implemented preservation metadata requirements in practice, in order to convey a sense of the 'state-of-the-art' prevailing before preservation metadata consensus-building efforts began to emerge.

Early efforts to develop preservation metadata element sets were undertaken by the National Library of Australia (NLA), the CEDARS (CURL Exemplars in Digital Archives) project, the NEDLIB (Networked European Deposit Library) project, and the National Library of New Zealand (NLNZ). The NLA element set[3] was designed to support the preservation of both digitized and born-digital objects. It accommodated three levels of descriptive granularity – collection, object, and sub-object (file) – and was implementation-neutral, in the sense that no assumptions were made about the specific preservation strategy adopted by the repository. The CEDARS element set[4] was developed for use with a pilot digital archive, and was relevant to a variety of digital formats. In contrast to the NLA set, these elements were applicable at any level of description. The NEDLIB element set[5] defined a 'core' set of essential preservation metadata, with an emphasis on overcoming the problem of technological obsolescence. Elements were defined at a high level to maximize applicability across object formats and types. Finally, the NLNZ element set[6] supported the Library's ongoing efforts to develop internal digital preservation capacity. It was a starting point for implementing systems responsible for collecting and managing preservation metadata.

The earliest preservation metadata element sets – e.g., NLA, CEDARS, and NEDLIB – were largely speculative in nature, seeking to anticipate the metadata needs of programmatic digital preservation initiatives that would emerge in the future. Development of later element sets, such as NLNZ, were more closely aligned with planning and implementation of production digital archiving systems – and of course, benefitted considerably from the foundations laid by the earlier element sets.

[3] http://pandora.nla.gov.au/pan/25498/20020625-0000/www.nla.gov.au/preserve/pmeta.html
[4] http://www.webarchive.org.uk/wayback/archive/20050410120000/http://www.leeds.ac.uk/cedars/colman/metadata/metadataspec.html
[5] http://www.kb.nl/sites/default/files/docs/NEDLIBmetadata.pdf
[6] http://www.natlib.govt.nz/downloads/metaschema-revised.pdf

2.2. OAIS

The OAIS (Open Archival Information System) reference model (CCSDS, 2012) is a conceptual framework describing the environment, functional components, and information objects associated with a system responsible for the long-term preservation of digital materials. OAIS was approved as ISO Standard 14721 in 2002, but even before then, it had enjoyed widespread adoption in the digital preservation community. OAIS has exerted a great deal of influence in the development of the art and science of digital preservation, including efforts to design and implement preservation metadata. The *OAIS information model* served as the foundation for, or at least informed, the development of most preservation metadata initiatives. Indeed, one of the salient characteristics shared by these initiatives, and therefore an obvious starting point for consensus-building in preservation metadata, is the fact that each can be traced, in some form or another, back to the common antecedent of the OAIS information model.

The OAIS information model is a conceptualization of the information objects taken into, stored, and disseminated by a digital preservation repository. The core concept underlying the model is an *information package* – a combination of some piece of content that is the focus of preservation, along with its associated metadata. A key aspect of the information package concept is the fundamental link between preserved digital content and metadata, or put another way, the idea that metadata plays an essential role in preserving digital content and supporting its use over the long-term. Recall from the previous section that this is our most basic definition of preservation metadata: metadata that supports the digital preservation process.

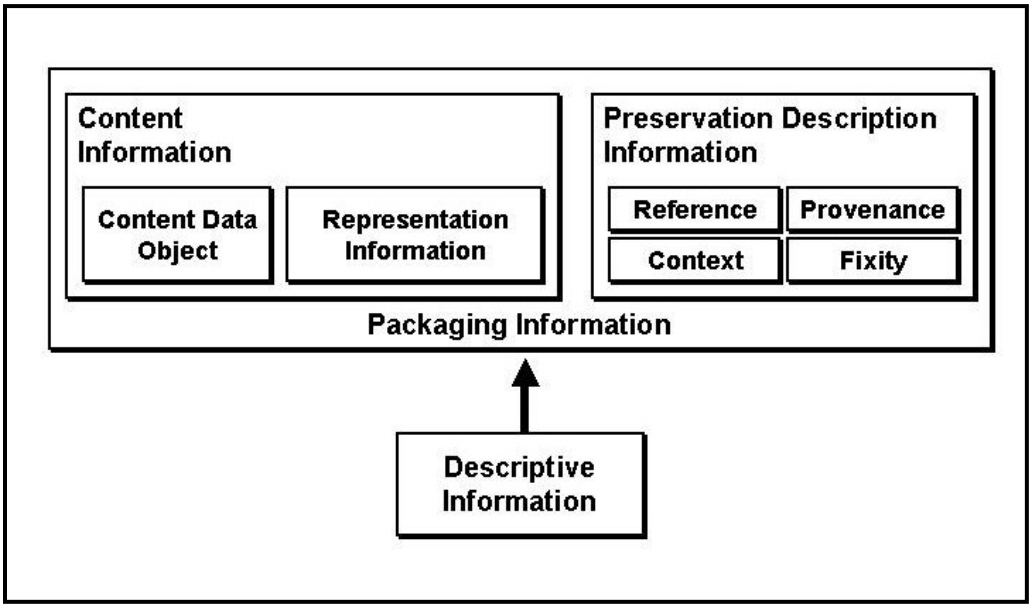
The OAIS information model (Figure 1) provides a high-level overview of the types of information associated with an archived digital object, including:

- Representation Information: information necessary to render and understand the bit sequences constituting the archived digital object.
- Preservation Description Information: information that supports and documents the preservation of the archived object, including:
  - Reference information: uniquely identifies the archived object;
  - Context information: describes the archived object's relationship(s) to other objects;
  - Provenance information: documents the history of the archived object;
  - Fixity information: validates the authenticity or integrity of the archived object.
- Packaging Information: information that binds all components of an information package into a single logical unit.
- Descriptive Information: information that supports the discovery and retrieval of the archived object by the repository's users.

With the exception of descriptive Information (for the reasons discussed above), these information types can be interpreted as the most general formal description of the metadata needed to support the long-term preservation of digital materials.[7] They serve as the starting point for most subsequent efforts to develop formal preservation metadata schema.

[7] Note that these formal categories of information easily map to the 'informal' description of the scope of preservation metadata offered in the previous section.

Figure 1: The OAIS information model[8]



2.3. A framework for preservation metadata

In 2000, OCLC and RLG jointly sponsored the creation of an international working group tasked with defining the role of metadata in the digital preservation process.[9] At the time the working group was organized, there was little or no consensus on even the most fundamental questions surrounding preservation metadata, including what types of information constituted preservation metadata, and how it could be used to support the digital preservation process. As discussed above, several institutions had developed element sets for internal use, but these reflected a wide range of assumptions, purposes, and approaches. In light of this, the working group produced a white paper (OCLC and RLG, 2001) summarizing the state of the art in preservation metadata. This provided a definition of preservation metadata, described its role in the digital preservation process, and reviewed a number of existing preservation metadata initiatives, with an emphasis on identifying points of convergence and divergence among them.

The white paper provided a foundation for the working group's next task, which was to develop a comprehensive, broadly applicable *preservation metadata framework* enumerating the types of information falling within the scope of preservation metadata. Given its extensive take-up in the digital preservation community, the working group chose OAIS as the starting point for the framework. The broad categories of information specified in the OAIS information model served as a top-level description of the types of information comprising preservation metadata. The working group then expanded each category of information, providing additional structure to articulate the OAIS information requirements in progressively greater detail and ending with a set of 'prototype' preservation metadata elements. Published in 2002, the preservation metadata framework (OCLC

[8] From Lavoie, B F (2004) *The Open Archival Information System Reference Model: Introductory Guide*. Digital Preservation Coalition *Technology Watch Series Report* 04-01, p. 12. Available at: http://www.dpconline.org/component/docman/doc_download/91-introduction-to-oais

[9] http://www.oclc.org/research/activities/pmwg/wg1.html

and RLG, 2002) was the first international consensus-driven statement on the scope of preservation metadata. It consolidated existing expertise to create a solid foundation upon which an international standard for preservation metadata could be built.

3. The PREMIS Data Dictionary

The international consensus achieved with the preservation metadata framework suggested opportunities to advance work in preservation metadata by defining a core set of implementable, broadly applicable preservation metadata elements, supported by a data dictionary providing guidelines and recommendations for populating and managing the elements. To address this task, OCLC and RLG convened a second working group: PREservation Metadata: Implementation Strategies (PREMIS).[10] PREMIS was composed of more than 30 international experts in preservation metadata, drawn from libraries, museums, archives, government agencies, and the private sector.

In 2005, the PREMIS working group published the 237-page *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group* (PREMIS, 2005). The report included the *PREMIS Data Dictionary 1.0*, a comprehensive guide to the core metadata needed to support long-term digital preservation. Subsequent to the release of the *Data Dictionary*, PREMIS released a set of XML schemas[11] to support implementation and exchange of PREMIS-conformant preservation metadata.

3.1. Introduction to the Data Dictionary

The *Data Dictionary* is organized around a data model (Figure 2) consisting of five entities associated with the digital preservation process: *Intellectual Entity* (a coherent set of content that is described as a unit: e.g., a book); *Object* (a discrete unit of information in digital form, e.g., a PDF file); *Event* (a preservation action, e.g., ingest of the PDF file into the repository); *Agent* (person, organization, or software program associated with an Event, e.g., the publisher of the PDF file who deposits it in the repository); and *Rights* (one or more permissions pertaining to an Object, e.g., permission to make copies of the PDF file for preservation purposes). With the exception of Intellectual Entity (which was deemed out of scope in that it was addressed by other metadata schemas focused on descriptive information), each entity is described by a set of properties called semantic units. Each semantic unit represents a discrete piece of information to be recorded as part of the metadata supporting the digital preservation process. A key point about the semantic units is that they are implementation-neutral – that is, no stipulations are made about how the information encompassed in a semantic unit is to be recorded in a digital archiving system. The only requirement is that this information is 'known' or recoverable in some way from the digital preservation process in which it is embedded. (This point may seem strange, but it is quite important.) In short, a PREMIS semantic unit can be recorded in any way a repository finds convenient, given the processes and architecture of its repository system, as well as its metadata management procedures. For example, a semantic unit can be recorded as a single metadata element, or broken up over multiple metadata elements if the repository prefers. Either approach is valid in the context of the *PREMIS Data Dictionary*.

The *Data Dictionary* provides a clear definition of each semantic unit, along with a rationale for its inclusion in the *Dictionary*. Data constraints – that is, restrictions on the kinds of values that can be used to populate the semantic unit – are indicated if applicable, and example implementations are shown. In addition, the semantic unit's obligation and repeatability are indicated, and notes are provided regarding the creation, maintenance, and usage of the semantic unit in a repository system. Figure 3 shows the description for one semantic unit, *preservationLevelValue*.

[10] http://www.oclc.org/research/activities/pmwg/background.html
[11] http://www.loc.gov/standards/premis/schemas.html

Figure 2: The PREMIS Data Model[12]

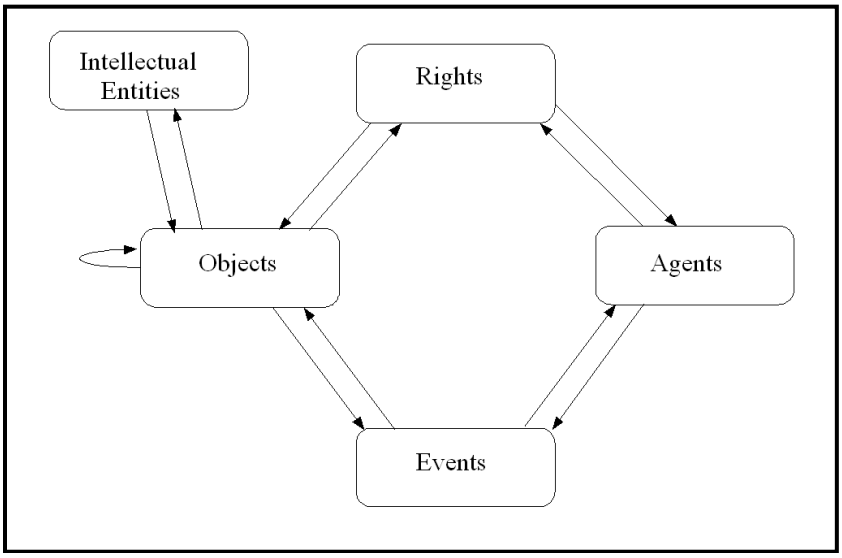


Figure 3: Example semantic unit[13]

Semantic unit	**1.3.1 preservationLevelValue**		
Semantic components	None		
Definition	A value indicating the set of preservation functions expected to be applied to the object.		
Rationale	Some preservation repositories will offer multiple preservation options depending on factors such as the value or uniqueness of the material, the 'preservability' of the format, the amount the customer is willing to pay, etc.		
Data constraint	Value should be taken from a controlled vocabulary.		
Object category	**Representation**	**File**	**Bitstream**
Applicability	Applicable	Applicable	Not applicable
Examples	bit-level full 0 1 2	bit-level full 0 fully supported with future migrations	
Repeatability	Not repeatable	Not repeatable	
Obligation	Mandatory	Mandatory	
Creation / Maintenance notes	The preservation level may be assigned by the repository or requested by the depositor and submitted as metadata.		
Usage notes	Only one *preservationLevelValue* may be recorded per *preservationLevel* container. If a further *preservationLevelValue* applies to the object in a different context, a separate *preservationLevel* container should be repeated.		

Some semantic units are defined as *containers*, which serve to group together sets of related semantic units, or *semantic components*. For example, the semantic unit *preservationLevel* is a

[12] PREMIS (2012), p. 5.
[13] PREMIS (2012), p. 34.

container that groups together four semantic components: *preservationLevelValue*, *preservationLevelRole*, *preservationLevelRationale*, and *preservationLevelDateAssigned* – each of which is a semantic unit in its own right. Containers are intended primarily as an organizational device for the *Data Dictionary*, and therefore implementation would typically occur at the level of the semantic components, although strictly speaking a repository could implement at the container level, as long as the individual pieces of information defined by the semantic components are recoverable in some way from the recorded data.

The *PREMIS Data Dictionary* provides general preservation metadata semantics, but does not offer content- or organization- specific metadata: extension schemas must be used for this purpose. For example, technical metadata for audio files might be described using the AudioMD schema.[14] Metadata elements from extension schemas may be incorporated directly within the Object entity or embedded using the PREMIS <objectCharacteristicsExtension> element. Including these extension elements may run the risk of reducing interoperability, so a careful balance needs to be struck between generality and specificity (Dappert & Enders, 2010, p.7).

Taken together, the semantic units defined in the *PREMIS Data Dictionary* represent the 'core' information needed to support digital preservation activities in most repository contexts. However, the concept of 'core' in regard to PREMIS has some looseness attached to it: not all of the semantic units are considered mandatory in all situations, and some are optional in all situations. The *Data Dictionary* attempts to strike a balance between recognizing that the intersection of metadata requirements across different repository contexts is quite broad, while at the same time acknowledging that all contexts are different in some way, and therefore their respective metadata requirements will rarely be exactly the same.

Following the release of the *PREMIS Data Dictionary*, a web presence – called the PREMIS Maintenance Activity[15] – was established to serve as a home for the *PREMIS Data Dictionary* and XML schema, as well as a central destination for news and related resources. The Maintenance Activity website is hosted and managed by the US Library of Congress, and maintains the PREMIS Implementers' Group (PIG) forum, the primary channel through which the community of PREMIS implementers interact with one another and exchange information. The Maintenance Activity was supplemented in 2006 by the formation of a PREMIS Editorial Committee. The Editorial Committee is responsible for managing the *PREMIS Data Dictionary* and associated schema, and has since overseen three significant revisions of the *Data Dictionary* (versions 2.0, 2.1, and 2.2; see below for details).

Although the PREMIS Data Dictionary is not a formal standard, in the sense of being managed by a recognized standards agency, it has achieved the status of the accepted standard for preservation metadata in the digital preservation community. Support for PREMIS is included in commercial repository solutions such as Ex Libris's Rosetta digital asset preservation system,[16] as well as open-source offerings like Archivematica[17]. The wide take-up of PREMIS can be at least partially attributed to the fact that it is the most comprehensive, detailed treatment of preservation metadata to date; moreover, it was developed through an international consensus with the goal of

[14] http://www.loc.gov/standards/amdvmd/
[15] http://www.loc.gov/standards/premis/
[16] http://www.exlibrisgroup.com/category/RosettaOverview
[17] https://www.archivematica.org/wiki/Main_Page

maximizing its applicability across as wide a range of digital preservation contexts as possible. But the widespread take-up of PREMIS can also be attributed to the fact that although it is not a formal standard, it is managed in a highly coordinated way. As mentioned above, the *Data Dictionary* is under the care of an Editorial Committee drawn from PREMIS's stakeholder community. Revisions are publically announced, and based largely on feedback from those implementing PREMIS in their local digital archiving systems. The *Data Dictionary* and its associated XML schema are carefully versioned, and there is a significant corpus of documentation supporting their use. Finally, the *PREMIS Data Dictionary* is managed under the auspices of the Library of Congress, which lends considerable prestige and credibility to the *Data Dictionary* itself, as well as the assurance that it will continue to be managed as a community resource into the future.

It should be noted that the *PREMIS Data Dictionary* must be tailored to meet the requirements of a specific repository context; it is not an off-the-shelf solution in the sense that the repository simply implements the *Data Dictionary* wholesale. Only a portion of the *Dictionary* may be relevant in some digital preservation circumstances; alternatively, the repository may find that additional information beyond what it is defined in the *Dictionary* is needed to support their requirements. For example, the *Data Dictionary* makes no provisions for documenting information about a repository's business/policy dependencies, which may be needed to support preservation decision-making. In short, each repository will need to expend some effort to adapt the *Data Dictionary* to its particular circumstances and requirements.

The value of the *PREMIS Data Dictionary* has been recognized by several awards. In 2005, the PREMIS working group earned the international Digital Preservation Award for their efforts in producing the original *Data Dictionary.*[18] In 2006, the *PREMIS Data Dictionary* was given the Society of American Archivists' Preservation Publication Award.[19] In 2012, PREMIS was shortlisted for the inaugural Decennial Digital Preservation Award for the most significant contribution to digital preservation in the last decade.[20]

Consideration of preservation metadata's brief history suggests a key shift in the trajectory of work in this area: a shift from theory to practice; from concepts to implementation; from development to use. Preservation metadata has transitioned from a research topic to one that is of interest primarily to practitioners. In this sense, there is less emphasis on issues such as defining the scope of preservation metadata, or justifying its importance. While fundamental issues regarding preservation metadata still remain (for example, how much preservation metadata is needed to support a particular digital preservation strategy and goals), the primary topic of interest is, given a preservation metadata schema, how can it be easily and efficiently implemented and maintained within a digital archiving system? Much of the impetus for this shift can be attributed to the emergence of the *PREMIS Data Dictionary* as the standard for preservation metadata, and the subsequent take-up of PREMIS in an increasing number of digital preservation solutions and tools. Much of the significant implementation work in the area of preservation metadata has coalesced around the *PREMIS Data Dictionary*.

[18] http://www.dpconline.org/newsroom/not-so-new/110-awards-2005
[19] http://www.oclc.org/research/news/2006/08-25b.html
[20] http://www.dpconline.org/advocacy/awards/2012-digital-preservation-awards

3.2. Revisions of the PREMIS Data Dictionary[21]

Upon release of the original *Data Dictionary* (version 1.0) in 2005, the PREMIS Maintenance Activity resolved to make no changes to it until a significant amount of time had passed, allowing the digital preservation community to examine the *Dictionary* and to consider potential issues that might arise in implementing it in digital archiving systems. In this way, the first revision of the *Dictionary* could be based on a critical mass of implementation experiences. In keeping with this strategy, the process of revising Version 1.0 of the *Data Dictionary* was not initiated until October 2006, following the formation of the PREMIS Editorial Committee, who would be entrusted with managing the revision.

The development of the *PREMIS* Data Dictionary 2.0 was based on feedback received from implementers through a variety of channels. The *PREMIS Data Dictionary* 2.0 incorporated four major updates:

- The structure of relationships between entities in the data model was generalized to incorporate bi-directionality in all cases, so that relationships between any two entities (for example, an Object and an Event) can be documented with metadata associated with either entity, or both.
- The Rights entity was updated and expanded to support a richer description of rights statements, including the ability to record information specific to intellectual property rights established by copyright, licence, or statute.
- The *Data Dictionary* was updated to include structured descriptions of significant properties and preservation levels, replacing single, unstructured semantic units in the previous version.
- A formal mechanism was introduced to support extensibility when using the PREMIS XML schema. A new 'extension container' permits metadata defined externally to PREMIS to be seamlessly integrated into the PREMIS schema.

Version 2.0 of the *PREMIS Data Dictionary* was published in April 2008 (PREMIS, 2008). Since that time, two additional (but relatively small-scale) revisions have been undertaken. PREMIS 2.1 (PREMIS, 2011), released in January 2011, corrected errors and provided clarifications in regard to version 2.0; in addition, PREMIS 2.1 added several new semantic units for Agents, and restructured the extensibility mechanism to more closely resemble the extension schemas used in METS. PREMIS 2.2 (PREMIS, 2012), released in July 2012, provided further amplification of the Rights entity through the addition of a number of new semantic units, as well as several updates to the PREMIS XML schema.

3.3. Outreach

In addition to managing revisions of the *Data Dictionary*, the PREMIS Editorial Committee organizes a number of outreach activities aimed at raising awareness about PREMIS, and preservation metadata generally. The Editorial Committee holds a series of tutorials on PREMIS at locations all around the world; these range in scope from a basic introduction to the *Data Dictionary* to presentations geared toward experienced implementers. The tutorials are often supplemented by presentations from representatives from projects that are implementing PREMIS in their digital

[21] Adapted in part from Lavoie(2008).

repositories, to provide a real-world context to the discussion. Materials from all of the past tutorials are available to download from the PREMIS Maintenance Activity website.[22]

Another useful resource on the *PREMIS Data Dictionary* and its implementation is Priscilla Caplan's report *Understanding PREMIS* (Caplan, 2009), sponsored by the Library of Congress. This provides a general overview to the essentials of PREMIS, and for those new to the topic, serves as a useful first step before tackling the *Data Dictionary* itself. *Understanding PREMIS* is available for download on the PREMIS Maintenance Activity website, and has proven to be a popular resource. This resource is also available in French, German, Italian, and Spanish.[23]

In recent years, 'implementation fairs' have been added as a further form of outreach. The first PREMIS Implementation Fair was held in October 2009, and was repeated in 2010 and 2012. As their name suggests, implementation fairs focus on issues associated with implementing the *PREMIS Data Dictionary* in digital repositories. In addition to reports from the Editorial Committee, they feature presentations from PREMIS implementers discussing issues and solutions, tools, and best practices. To date, the implementation fairs have been held in conjunction with the international iPRES digital preservation conference. Details about the implementation fairs, and access to materials from past fairs, can be found on the PREMIS Maintenance Activity website.

Perhaps the most important form of outreach is the PREMIS Implementers Group forum, or PIG list.[24] The PIG list is an open e-mail forum for the discussion of issues related to the *PREMIS Data Dictionary*, the PREMIS XML schema, tools, and the general topic of preservation metadata. It is also the chief mechanism by which the Editorial Committee communicates with PREMIS implementers, and vice versa. Anyone responsible for implementing the *PREMIS Data Dictionary* in a digital archiving system should subscribe to this list.

3.4. Packaging preservation metadata: METS and PREMIS

The amount, variety, and complexity of preservation metadata produced in a working digital archive or asset management environment requires careful packaging; not only should the metadata itself be stored, but it must also be linked to any additional descriptive, administrative and structural metadata associated with the objects in the system.

A perusal of the PREMIS Implementation Registry shows[25] that the most widely used framework for performing this function is currently METS (Metadata Encoding and Transmission Standard),[26] an XML implementation of an OAIS Information Package. Since METS is the most generally used method of packaging PREMIS metadata, it is useful to describe its key features in some detail.

METS is designed to act as an OAIS SIP (Submission Information Package), DIP (Delivery Information Package), or – crucially in this context – an AIP (Archival Information Package). It allows four types of metadata to be recorded, each within its own section of a METS file:

[22] http://www.loc.gov/standards/premis/tutorials.html. If you are interested in hosting a PREMIS tutorial for your own organization, contact the PREMIS Editorial Committee.
[23] In addition to these translations of *Understanding PREMIS*, Japanese and Spanish translations of the *PREMIS Data Dictionary* (version 2.0) are available on the PREMIS Maintenance Activity website.
[24] *http:*//www.loc.gov/standards/premis/pig.html
[25] See below for more information on the Registry.
[26] http://www.loc.gov/standards/mets/

- A file inventory for all the files associated with the digital object (such as still-image files, text, video or audio files);
- A section for administrative metadata, divided into four sub-sections covering technical information about the files, rights management information, information on the source from which the object was made and digital provenance information;
- A section for descriptive metadata, including bibliographic information and any other information on the intellectual content of the item necessary for users to find it and assess its relevance to them;
- A structural map of the internal contents of the item, which indicates in a hierarchical manner how its various components relate to each other, thus allowing its constituent elements to be navigated by the user: this may encode its logical structure (such as the division of a book into chapters) or its physical structure (such as the ordering of pages).

Linking these sections is a set of XML identifiers, which can be checked for consistency using any XML validation software. These may express relationships between components of great complexity if required: they may be used, for instance, to link a section in a structural map corresponding to a time-coded section of a movie to the files in the file inventory which contain the video files for that section, another pointer to a part of the descriptive metadata section containing a synopsis of that section, and another to the part of the administrative metadata section which contains technical and rights information necessary to deliver the object and control access to it.

The flexibility built into METS may cause problems in terms of interoperability. When such varied content, handled in such a variety of ways, is allowed within a METS file, it becomes more difficult to interchange METS records. This may be mitigated to some extent by the use of METS Profiles,[27] which are XML files used to document the way in which METS is implemented within a project. These documents list, amongst other things, the content schemas used within a METS file, the system of identifiers employed, whether metadata is embedded or referenced, and how it is structured within the file. They do not allow the automatic transfer of METS files between systems, but are designed to help an implementer understand another's usage of METS and how it can map to their own.

METS thus acts as a container for metadata, but the metadata itself must be encoded in other schemas: it may be embedded directly within the METS architecture, or held externally and referenced from it – for instance by URIs or URLs. These *extension schemas* may be chosen as required for a given application, although certain ones are recommended to ensure greater interoperability: for preservation metadata, PREMIS is the preferred option.

Unfortunately, the independent development histories of METS and PREMIS have resulted in a less than wholly clean fit between them. Much of the *PREMIS Data Dictionary* would, in a METS context, likely fall into the administrative metadata section of the METS file, but not all of the four top-level entities in the PREMIS data model divide readily between the four sub-sections of the administrative metadata section. Although the PREMIS Event and Rights entities slot neatly into the METS digital provenance and rights sub-sections respectively, the Object entity could be located within either technical or digital provenance metadata, and the Agent entity could refer to digital

[27] http://www.loc.gov/standards/mets/mets-profiles.html

provenance or rights information and so would have to be located in these respective sub-sections depending on context.

Problems also arise because of some duplication between METS and PREMIS elements. Checksums, for instance, may be recorded in either METS syntax (as an attribute of the <file> element in the file inventory section) or in PREMIS syntax (as an element nested within fixity information in the Object entity). In addition to these duplications, further complications can occur when data components within PREMIS are also found in other extension schemas used within a METS document. Much of the content of the PREMIS Object entity, for instance, is technical metadata which may be replicated in such schemas as MIX (NISO Technical Metadata for Digital Still Images)[28] or textMD (Technical Metadata for Text),[29] thus creating redundancies between the two. Where duplications of these types exist, it is necessary to make unambiguous policy decisions on such issues as whether the information should be repeated, which version of it should have priority, and how a system should deal with any inconsistencies.

To facilitate the use of PREMIS with METS, a set of guidelines has been published by the PREMIS Editorial Committee which provides pragmatic recommendations for dealing with the issues highlighted above.[30] Although the guidelines specifically cover the use of METS and PREMIS in the context of providing exchangeable digital objects, and so are intended for environments in which METS is used as an OAIS SIP or DIP, they still rationalize effectively the disjunctions between the two standards in the context of an archival digital object expressed as an OAIS AIP.

Among the main recommendations made by the guidelines are:

- If using the PREMIS container element (which can be used to collocate all four PREMIS top-level entities together), locate the whole package within the METS digital provenance section; if not, guidance is given on locating the PREMIS entities, some of which is dependent on context;
- Avoid the use of the PREMIS container element if PREMIS metadata can be sensibly distributed across several METS administrative metadata sub-sections;
- Handle redundancies by giving priority to the more expressive schema, and consider the intended use (usually prioritizing METS if the object is intended for display, and PREMIS if for preservation); generally, erring towards redundancy is recommended;
- Handle structural relationships through the METS structural map rather than PREMIS's relationship elements (which are used to encode links between PREMIS components): PREMIS's relationship elements are relatively basic and the METS structural map is richer and more flexible (although using PREMIS's relationships for expressing relationships between objects and their derivatives is recommended as METS does not handle these as effectively);
- Use a METS Profile to record the decisions made when using PREMIS as an extension schema, detailing particularly how redundancies are handled and how linkages are made.

28 http://www.loc.gov/standards/mix/
29 http://www.loc.gov/standards/textMD/
30 http://www.loc.gov/standards/premis/guidelines-premismets.pdf

A recent study by Vermaaten (2010) found that as of August 2009, there were 15 registered METS Profiles that implemented PREMIS. Vermaaten developed a checklist that identifies 13 aspects of using PREMIS with METS that an institution should document in their METS profiles.[31]

The METS Profile from the ECHO DEPository project at the University of Illinois at Urbana-Champaign[32] is a good example of how the two standards can be integrated. The METS Profile[33] for this project, which researched the practicalities of digital preservation architectures including such issues as semantic archiving and automatic metadata creation, illustrates effectively the scattered embedding of PREMIS elements advocated by the PREMIS-METS guidelines. Technical metadata is encoded in PREMIS's Object element within the METS technical metadata section, PREMIS Events within METS digital provenance, PREMIS Rights within METS's rights metadata section and PREMIS Agent may occur in either METS digital provenance or METS rights depending on the association of the agent to the metadata object. This profile also highlights the potential problem of redundancy between technical metadata contained within the PREMIS Object element and those in the multiple extension schemas, such as MIX,[34] VIDEOMD[35] and the Audio Technical Metadata Extension Schema,[36] which the project also uses for technical metadata.

These potential ambiguities are resolved by detailed notation within an extensive *description rules* section of the Profile which covers such issues as redundancies between METS and PREMIS identifiers, differentiations between primary and secondary representations of the archive and the use of controlled vocabularies. Detailed notes are also given on how the project handles redundancies between PREMIS and other extension schemas, including the supplementing of the technical metadata components of the PREMIS Object entity with other schemas in secondary METS technical metadata sections. As is the case with all METS Profiles, a sample METS file illustrates the application of these decisions in practice.

Both PREMIS and METS are, of course, evolving on a regular basis, and so any guidelines for their joint use will undoubtedly change. In 2010 the METS Editorial Board published a white paper[37] analyzing some current problems with implementing the standard (including the subdivisions of administrative metadata within which PREMIS elements are scattered). There is an intention to publish an evolved version of the standard which will address some issues raised in stakeholder consultations, although no date has been set for this. Similarly, work is progressing on PREMIS 3.0, which incorporates such changes as including intellectual entities as categories of Objects and extensive changes to rights metadata.[38] A revised version of the guidelines may be necessary once these new versions have been published and have established themselves, although hopefully backwards-compatibility will ensure that applications following the current guidelines will not need to be changed.

[31] http://www.loc.gov/standards/premis/premis_mets_checklist.pdf
[32] http://www.ndiipp.illinois.edu/
[33] http://www.loc.gov/standards/mets/profiles/00000015.html
[34] http://www.loc.gov/standards/mix/
[35] http://www.loc.gov/standards/amdvmd/
[36] http://lcweb2.loc.gov/mets/Schemas/AMD.xsd
[37] https://docs.google.com/file/d/0BzMhAsKzul-tN2Y0N2RjM2EtNzFmMC00NGQ3LTkyYjctNGZlZjEwODUzNmFm/edit?authkey=CIWlms8F
[38] http://www.loc.gov/standards/premis/premis-tutorial_iPRES2011_singapore.ppt

3.5. Tools to support PREMIS implementation

The process of implementing PREMIS in a working environment is made easier by a number of tools which can extract metadata from digital objects and output PREMIS XML. Not all of these tools have been written specifically for generating PREMIS-related metadata, but may be used to produce PREMIS metadata by post-processing their output. The wide take-up of these tools, all of which are open source, has clearly facilitated the wider adoption of PREMIS.

The PREMIS Maintenance Activity maintains a webpage listing the most important tools available for use with PREMIS:[39] at the time of writing this contains entries on nine tools, in addition to pointers to others which may be used to generate METS files in conjunction with PREMIS. The majority of the tools listed are for extracting technical metadata from digital objects and converting it for encoding within the PREMIS Object entity. Others can be used for checking formats, or validating files against checksums.

A popular tool among implementers is Harvard University's JHOVE (JSTOR/Harvard Object Validation Environment).[40] JHOVE is not specifically a PREMIS tool and so requires additional processing of its output to convert it to this format: a number of further tools can carry out this function. JHOVE carries out a number of checks on a digital object to identify, validate, and produce detailed technical metadata from it. These include identifying the format of the object and checking the extent to which it conforms to its format (including both its *well-formedness*, the extent to which it obeys the syntactical rules for the format, and its *validity*, whether it meets any semantically defined rules for a valid object). Finally, JHOVE produces an extensive list of information on the object itself, which can be readily processed into PREMIS Object metadata: for a TIFF file, for instance, approximately 40 information components are reported conforming to the specifications of NISO image metadata.

Another popular tool, although with more limited functionality than JHOVE, is DROID (Digital Record Object Identification),[41] a batch-processing file format identification package which interfaces directly with PRONOM, a continuously updated registry of file-format-related technical information maintained by the UK National Archives. Unlike JHOVE, which outputs its results as a simple text list without markup, DROID can output directly into XML. The XML file includes all formats for which identifications can be made with entries in PRONOM and an indication of the quality of the identification.

Neither JHOVE nor DROID can produce PREMIS metadata directly, and so their outputs must be processed into PREMIS XML format using one of a number of other tools. One such tool is Statistics New Zealand's PREMIS Creation Tool,[42] which is a set of XSL style sheets and VBScript scripts which takes JHOVE or DROID output and produces PREMIS Object records, slightly modified from the original PREMIS schema to allow information on the software package used to generate each element to be recorded.

[39] http://www.loc.gov/standards/premis/tools.html
[40] http://hul.harvard.edu/jhove/
[41] http://www.nationalarchives.gov.uk/aboutapps/pronom/
[42] http://pigpen.lib.uchicago.edu:8888/pigpen/40

Another tool which uses JHOVE to generate PREMIS metadata is the HandS (Hub and Spoke)[43] toolset produced as one of the outputs from the ECHO DEPository project at the University of Illinois Urbana-Champaign.[44] This project examined in depth the practical issues of implementing digital repositories and developed a number of automated metadata creation and extraction tools. HandS is a suite of tools written in Java which utilizes JHOVE to generate technical metadata specific to the format of the files submitted. It has particular value when PREMIS is to be packaged within a METS environment as it generates METS files into which this metadata is slotted: these METS files conform to the registered Echodep METS Profile,[45] thus aiding interoperability with other bodies which employ the same tool and conform to the same profile. The tools may be used on the command line or through a Graphical User Interface (GUI) which enables them to be submitted to a repository, disseminated from it or migrated to another in batch mode.

A further tool of use to working environments in which PREMIS is packaged within METS is the PREMIS in METS Toolbox (PIMTOOLS)[46] from the Florida Center for Library Automation. This is designed to facilitate the conformance of METS files with embedded PREMIS metadata to the PREMIS in METS Guidelines[47] mentioned above. It does this by offering a validation tool to check conformance of a METS document with embedded PREMIS metadata to the Guidelines, and a conversion tool to generate a METS/PREMIS document conforming to the Guidelines from a PREMIS file. Documents may be validated or converted through the PIMTOOLS website, either by supplying a URI, uploading them or inputting their contents directly into a web form; alternatively, the schematron file used for validation and the style sheets for performing the conversions can be downloaded for local use offline.

[43] http://dli.grainger.uiuc.edu/echodep/hands/index.html
[44] http://www.ndiipp.illinois.edu/
[45] http://www.loc.gov/standards/mets/profiles/00000015.html
[46] http://pim.fcla.edu/
[47] http://www.loc.gov/standards/premis/guidelines-premismets.pdf

TABLE 1: Five key PREMIS tools

Name of tool	Creator	Functions	Notes
JHOVE (JSTOR/Harvard Object Validation Environment[48])	Harvard University	Identify file formats and validate files, produce detailed technical metadata	Does not produce PREMIS directly
DROID (Digital Record Object Identification)	National Archives (UK)	File format identification	Interfaces with PRONOM repository. Does not produce PREMIS directly
PREMIS Creation Tool	Statistics New Zealand	Generate PREMIS *object* entities from JHOVE/DROID output	Generate XSL stylesheets and VBScript scripts
HandS (Hub and Spoke)	University of Illinois Urbana-Champaign	Generate technical metadata: package in METS	METS files conform to ECHO DEP METS Profile
PREMIS in METS Toolbox	Florida Center for Library Automation	Validate PREMIS in METS, convert PREMIS to PREMIS in METS	Checks conformance to Library of Congress's PREMIS in METS Guidelines

These tools, and others, have undoubtedly made the implementation of PREMIS in working environments much easier, as is evidenced by their widespread adoption: approximately two-thirds of projects listed in the PREMIS Implementation Registry, for instance, record that they use JHOVE as part of their PREMIS-based preservation metadata creation strategy, and the other tools mentioned, particularly DROID, are in widespread use.

3.6. PREMIS implementations

On publication, the *PREMIS Data Dictionary* was acknowledged as an undoubtedly core component of preservation metadata strategies, although its size and complexity brought with it a fairly steep learning curve before it could be implemented. For this reason, some commentators (for instance Victoria McCargar)[49] doubted that it would become established quickly outside the research library community and indicated that the laborious compilation of the metadata it required would impede its uptake. A 2007 report commissioned by the PREMIS Maintenance Activity (two years after the release of the *Data Dictionary*) found that implementations were still very limited in number and that most were still in the planning and development stages (Woodyard-Robinson, 2007, p. 9).

In the five years since this report, however, a significant number of both users and suppliers of preservation metadata have adopted PREMIS as a core component of their preservation metadata strategy and many of the initial obstacles to implementation have been addressed. The uptake of PREMIS has particularly been facilitated by an effective support network for its users. This includes

[48] http://hul.harvard.edu/jhove/
[49] http://www.loc.gov/standards/premis/No%20Pain-No%20Metadata.pdf

the PREMIS Implementers' Group forum,[50] hosted by the PREMIS Maintenance Activity, which includes an active email discussion list and a wiki for sharing documents. The wiki is a particularly useful resource for new implementers, as it includes materials from PREMIS tutorials, a collection of examples of PREMIS usage and links to information on PREMIS tools.

The PREMIS Maintenance Activity maintains an active registry of PREMIS implementations.[51] This registry, which numbers approximately 45 projects at the time of writing, includes projects from academic libraries, national libraries, archives, government agencies, and others. Two commercial companies are also represented in the registry: Ex Libris, which incorporates support for PREMIS in its Rosetta digital asset preservation system,[52] and Artefactual Systems, which supports PREMIS in the open-source preservation system Archivematica.[53]

A small number of projects (not listed in the registry) have also examined the use of PREMIS for multimedia preservation. The Preserving Digital Public Television project (PDPT),[54] funded by the Library of Congress, recommends in its Repository Design Report use of a PREMIS file for information on creating applications and rendering environments within its AIP.[55] PrestoPRIME,[56] a European project which has created a digital preservation system for audiovisual material based on an extension to the Ex Libris system, implements PREMIS as the core of its preservation metadata.

The range of digital object types for which PREMIS is used is wide and eclectic. In regard to the implementations described in the registry, textual objects form the largest category, closely followed by images. Multimedia in the form of audio files feature in approximately 33% of the implementations in the registry and video in approximately 25%. Datasets are held in approximately 15% of the implementations, while around 10% contain complete archived websites. The applicability of PREMIS to a wide variety of media is well demonstrated by this breadth of coverage.

The PREMIS functionality realized in each implementation also varies considerably; few use all of its features consistently. Only two implementers in the registry, the National Archives of Scotland and the National Library of the Czech Republic, use it for all repository workflows. Common usages for PREMIS include authentication using fixity information (such as MD5 checksums), validating the formats of digital objects, checking format migrations (including recording conversions to new formats), provenance verification (particularly using the Event entity to provide an 'audit trail' for an object), and as a packaging mechanism for technical and administrative metadata (as an alternative to, for instance, METS). It is also used to support the semantics required by OAIS SIPs, AIPs and DIPs in the case of repositories which are aiming to be OAIS-compliant (for instance, in the case of the National Library of Sweden).

Of the four core PREMIS entities – Object, Events, Rights, and Agent – Object is the most commonly used, with most implementers choosing to encode their technical metadata in Object's structure

[50] http://www.loc.gov/standards/premis/pig.html
[51] http://www.loc.gov/standards/premis/registry/
[52] http://www.exlibrisgroup.com/category/RosettaOverview
[53] https://www.archivematica.org/wiki/Main_Page
[54] http://www.thirteen.org/ptvdigitalarchive/
[55] http://cn2.wnet.org/thirteen/ptvdigitalarchive/files/2010/03/PDPTV_ReposDesign_2010-03-19.pdf, p. 14
[56] http://www.prestocentre.org/library/resources/strategy-use-preservation-metadata-within-digital-library

rather than using more format-specific standards such as MIX or textMD. The Event entity is also heavily used, particularly for provenance verification and change tracking; Agent is commonly used in conjunction with Events.

The Rights entity is the least used of the four core entities, implemented by approximately a third of those in the registry, and only a small number of these projects explicitly mention using PREMIS to control access to their holdings. The University of California San Diego's digital asset management system, which controls access by checking multiple PREMIS elements to determine the access status of an asset,[57] is perhaps the most fully developed use of this functionality at present. Most implementers use other mechanisms for controlling access to resources.

The metadata architectures within which PREMIS is deployed tend to be mostly XML-based. Of the XML-based implementations, approximately half use METS as their overall packaging mechanism and embed PREMIS elements within its architecture. In most cases, this embedding takes the form of using the Object, Agent, and Event entities within METS's digital provenance section (as for instance, in the example file provided by the National Archives of Sweden[58] in the implementation registry). None are currently using the PREMIS Rights entity within the METS architecture: instead METS's own rights schema, METSRights[59] is more commonly used (for more details on the issues surrounding the use of PREMIS with METS, see the section above on packaging).

Other implementations find different ways to integrate PREMIS into their XML metadata architectures. In some cases – for instance, the Carolina Digital Repository and projects at the University of Illinois, Urbana-Champaign – PREMIS XML is stored natively, supplemented where necessary by schemas such as FOXML and iRODS where additional semantic components are required. In some cases this XML is converted to a relational database for ease of searching and access (an approach taken by the National Library of the Czech Republic, for instance), or stored as archival objects in XML but supplemented by relational databases where convenient (as at the University of North Texas's Portal to Texas History, which stores PREMIS events in a relational database for ease of update and addition).

The embedding of PREMIS in working digital collections, although a work in progress, is nevertheless well developed. A thriving community of implementers has established momentum in the adoption of the standard, easing the way for newcomers to introduce PREMIS into their working practices. The creation of new tools, or adaptation of existing ones, has encouraged PREMIS implementation still further.

[57] http://www.loc.gov/standards/premis/registry/premis-project_name.php?proj_ID=667

[58] http://www.loc.gov/standards/premis/registry/examples/34_SwedishDigitalRepository_Q0008791_Content_METS.xml

[59] http://cosimo.stanford.edu/sdr/metsrights.xsd

TABLE 2: Features of selected PREMIS implementations (from PREMIS Implementation Registry)

Project	Sector	Function of PREMIS	PREMIS entities used			
			Object	Agent	Event	Rights
Archivematica	Company	Multiple, including ingest, fixity check, validation, creation of normalized versions	x	x	x	
Carolina Digital Repository	Academic	Storing object information, generating thumbnails and access format objects – also authentication via checksums, preservation reports	x	x	x	
Creating a digital repository at the Swedish National Archives using PREMIS	Archive	Multiple functions logged as PREMIS events	x	x	x	
Digitaal Magazijn	Library	Preservation watch and preservation actions	x	x	x	x
Digital Data Archive (DDA) Project	Archive	All repository workflows	x	x	x	x
European project SHAMAN	Library	As predicates for OAI_ORE->LMER Metadata	x		x	
Kramerius	Library	Long-term preservation and authentication	x			x
OpenSky	Research centre	Authenticity/integrity, format validation, migration	x			
Scholars Portal Project	Consortium	Content management/integrity checking	x	x		
Statistics New Zealand Data Archive	Public agency	Fixity checks, records of changes, provenance verification	x	x	x	basic
The Portal to Texas History	Library	Metadata on files, event tracking	x	x	x	

UCSD Library Digital Asset Management System (DAMS)	Library	Digital content storage, rights management, access control	x	x	soon	x
PrestoPRIME	Library	Ingest, preservation operations, rights management	x	x	x	x

3.7. Other Implementation Resources

In addition to the PREMIS XML Schema, the PREMIS with METS resources and the registries of PREMIS-related tools and implementations, the PREMIS Editorial Committee has sponsored the creation of a number of other resources aimed at facilitating implementation. One of these resources is the *PREMIS OWL ontology*. The OWL Web Ontology Language is a Resource Description Framework (RDF)-based language for creating ontologies. The PREMIS OWL ontology permits implementers to express the semantics of the *Data Dictionary* in RDF, which is especially useful for exposing information in a web environment for machine processing. A draft PREMIS OWL ontology, conformant to version 2.1 of the *Data Dictionary*, was released in 2011; more information is available on the PREMIS OWL wiki.[60]

Guidelines for use of many of the semantic units defined in the *PREMIS Data Dictionary* call for values selected from a controlled vocabulary of terms. The *Data Dictionary* often includes lists of suggested (but not mandatory) values for these vocabularies. Use of controlled vocabularies to populate implementations of PREMIS semantic units is highly encouraged, but the original *Data Dictionary* and XML schema offered no mechanism to support declaration and validation of controlled values. This gap was later filled with the release of a *collection of PREMIS controlled vocabularies*, represented in SKOS (Simple Knowledge Organization System)[61] as well as several other formats, and deployed on the Library of Congress's 'id.loc.gov' web service.[62] Use of this service and the vocabularies deployed on it are not required as part of a PREMIS implementation, but they do facilitate the process of declaring and validating controlled values for PREMIS semantic units.

The issue of *conformance* is also key with regard to PREMIS implementations. The *PREMIS Data Dictionary* makes very few requirements of implementers in terms of how it is incorporated and used in a digital archiving system. But there are a number of use cases where a more extensive set of expectations regarding the nature of PREMIS implementation is beneficial: for example, inter-repository exchange of preservation metadata, repository audits/certification, and the use of shared registries as a metadata source. As these use cases grew in importance, the Editorial Committee updated and expanded its definition of conformance to meet the need for greater clarity on what PREMIS conformance means in practice. Released in October 2010, *Conformant Implementation of the PREMIS Data Dictionary*[63] provides a set of principles against which PREMIS implementations can be assessed. In addition, the conformance guidelines detail the scope of

[60] http://premisontologypublic.pbworks.com/w/page/45987067/FrontPage
[61] http://www.w3.org/2004/02/skos/
[62] http://id.loc.gov/
[63] http://www.loc.gov/standards/premis/premis-conformance-oct2010.pdf

flexibility available to planners in shaping their PREMIS implementations while still remaining conformant. This flexibility takes the form of 'five degrees of freedom':

- Naming: names of semantic units can be changed, as long as a semantic unit's new name does not conflict with an existing name in the *Data Dictionary*.
- Granularity: semantic units can be implemented as metadata elements encompassing greater or lesser levels of granularity than the *Data Dictionary* defines (e.g., a metadata element can include information from multiple semantic units, or the information from one semantic unit can be distributed over multiple metadata elements).
- Detail: repositories can extend the detail of the information recorded for any semantic unit.
- Recording: a repository does not have to explicitly record the information defined in a semantic unit, as long as that information is somehow recoverable from the repository system.
- Controlled vocabularies: a repository can populate semantic units any way it likes, including by the use of shared, community-wide vocabularies or locally maintained vocabularies; the repository is also free to use no controlled vocabularies at all.

The purpose of the conformance statement is to establish a set of minimum requirements that support a range of use cases where conformance is beneficial (such as the exchange of PREMIS metadata between repositories), without unduly limiting the ability of digital repositories to shape their PREMIS implementations according to their specific needs.

Finally, it is useful to mention a project that, although conducted outside the auspices of the PREMIS Editorial Committee, is still an important contribution to the resources surrounding PREMIS implementation issues. The TIPR (Towards Interoperable Preservation Repositories) project[64] explored one of the use cases for which conformance is relevant: the exchange of preservation metadata between repositories. In particular, TIPR designed and tested a model for transferring OAIS archival information packages (AIPs, that is, archived content and its associated metadata) between repositories. Long-term digital preservation strategies often require 'hand-offs' of content and metadata between various stakeholders across the digital preservation lifecycle. These stakeholders will likely be operating digital repositories that are considerably different in terms of their technical specifications and implementation details. The TIPR project designed a protocol for AIP transfers across heterogeneous systems; part of this protocol deals with the extraction and transfer of PREMIS-conformant preservation metadata across repositories. TIPR demonstrated the feasibility of such transfers, and as such is an important contribution both to digital preservation practice in general and preservation metadata management in particular.

[64] See http://wiki.fcla.edu/TIPR

4. Conclusion

In recent years we have witnessed a remarkable transition from theory to practice in preservation metadata work. In a sense, the *PREMIS Data Dictionary* represents the practical fruition of the preservation metadata concepts laid out in the widely cited OAIS reference model. From these basic concepts, preservation metadata has converged to concrete expression in the form of a de facto international standard, which is now widely implemented in digital preservation repositories around the world. PREMIS-based preservation metadata is now part of generally accepted best practice for the long-term stewardship of digital materials.

But there is still much work to do in the area of preservation metadata. As the community has settled on PREMIS as the standard for defining preservation metadata, the focus of activity has shifted to developing and improving tools, workflows, and other resources to facilitate the collection, management, and use of preservation metadata in digital archiving systems. As the discussion of recent developments above makes clear, most activity surrounding preservation metadata, and PREMIS in particular, reflects this focus. It is also clear that great progress has been made in this regard. However, two areas would benefit from increased attention as we look ahead to the next phase of work in preservation metadata.

Accumulation and consolidation of best practice: Despite the fact that preservation metadata, and particularly PREMIS-based metadata, is now a common feature of digital preservation activities, there is very little work that draws together and synthesizes the implementation experience that is rapidly accumulating in the digital preservation community. The PREMIS Implementation Registry, discussed above, is a valuable first step in this direction; however, an evidence base of detailed case studies on how preservation metadata is collected and managed within digital archiving systems would help shape consensus on a set of best practices for implementers, as well as illuminate areas of priority for technical development. Incorporation of, and support for, emerging best practices for implementing preservation metadata in major open-source and commercial digital repository solutions is a key step in encouraging community-wide adoption of these practices.

Costs and benefits: Another significant gap has to do with the costs and benefits of preservation metadata. While the importance and value of preservation metadata is generally accepted, a great deal of implementation decision-making – how much metadata to collect, who will collect it, how it is managed within the archiving system – will be predicated on the relationship of costs to perceived benefits. The metadata defined in the *PREMIS Data Dictionary* is extensive, and it is not a trivial or inexpensive task to collect and manage all of it within most digital preservation contexts. More work needs to be done to provide estimates of the costs involved in collecting and managing preservation metadata; at the same time, more evidence needs to be assembled to demonstrate the practical benefit of incurring these costs, in terms of concrete examples of how preservation metadata directly informs and supports digital preservation decision-making and workflows.

Preservation metadata is a key element of the technical infrastructure supporting digital preservation. The emergence of the *PREMIS Data Dictionary* as a de facto international standard has facilitated implementation of preservation metadata by providing a common framework within which local metadata requirements can be identified and expressed. The ecosystem of tools, controlled vocabularies, guidelines, and other resources that has subsequently sprung up around PREMIS has further lowered the barriers to implementation. The state of the art in preservation

metadata has advanced considerably since PREMIS won the Digital Preservation Award in 2005. It will no doubt continue to do so in the future, as work continues to improve the efficiency and effectiveness of collecting, managing, and using preservation metadata.

5. Glossary

administrative metadata	Metadata designed to enable the management of a digital object: usually subdivided into preservation metadata, rights management metadata, technical metadata and source metadata
AIP (Archival Information Package)	In the OAIS conceptual model, a collection (package) of content and preservation description information which is preserved in an OAIS-compliant archive
Archivematica	An open-source digital preservation system created by the Archivematica project
CEDARS (CURL Exemplars in Digital Archives)	An early UK-based project which aimed to define best practice for digital preservation in libraries
checksum	A fixed-length code generated from a digital object for the purpose of detecting errors during transmission and storage
descriptive metadata	Metadata, primarily on the intellectual content of an item, designed to allow resource discovery and assessment
DIP (Dissemination Information Package)	In the OAIS conceptual model, a collection (package) of content and preservation description information which is delivered to the end user from an OAIS-compliant archive
DROID (Digital Record Object Identification)	Software tool to perform the automated batch identification of file formats using the PRONOM registry
ECHO DEPository project	A six-year research and development project which developed web-archiving tools
ExLibris	An Israeli software company providing library management systems
FOXML	An XML schema for ingesting objects into Fedora repositories
Granularity	The size of the units into which data components are divided, usually in different levels of a hierarchy.
HandS (Hub and Spoke)	A suite of tools to generate preservation metadata in METS and PREMIS
iRODS (Integrated Rule-Oriented Data System)	A data grid software system which allows data to be stored in a unified namespace using multiple storage resources
ISO	International Organization for Standards
Java	A high-level, object-orientated program language
JHOVE (JSTOR/Harvard Object Validation Environment)	A digital object validation environment written in JAVA
METS (Metadata Encoding and Transmission Standard)	An XML schema for packaging digital object metadata

METSRights	An XML schema for rights declarations
MIX (Metadata for Images in XML Schema)	An XML schema for technical metadata for still images
NEDLIB (Networked European Deposit Library)	A project from the late 1990s which attempted to develop a common architectural framework and basic tools for building deposit systems for electronic publications
New Zealand Metadata Extractor	Software developed by the National Library of New Zealand to extract preservation metadata from a range of file formats
NISO (National Information Standards Organization)	Publishes technical standards for managing information
OAIS (Open Archival Information System)	An archive that has accepted responsibility to preserve data and make it available to designated communities. Also the conceptual model which aims to allow this
OCLC	A nonprofit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world's information and reducing library costs
OWL	Web Ontology Language, a set of languages for encoding machine-readable ontologies
PIG (PREMIS Implementors' Group)	The official user group of implementors of the *PREMIS Data Dictionary*
PREMIS (PREservation Metadata: Implementation Strategies)	An initiative responsible for producing and maintaining the *PREMIS Data Dictionary* and related resources and activities.
PREMIS in METS Toolbox (PIMTOOLS)	A set of tools, produced by the Florida Center for Library Automation, which is designed to generate PREMIS metadata within METS containers
PREMIS OWL	An implementation of the *PREMIS Data Dictionary* as an OWL ontology
Preserving Digital Public Television project (PDPT)	A US-based project which aimed to devise a digital archive for the long-term preservation of public television programmes
PrestoPRIME	A European project aimed at developing methods of digital preservation for the audiovisual content of digital broadcast archives
PRONOM	An online registry of file formats, tools and preservation services, maintained by the UK National Archives
RDF (Resource Description Framework)	A general method for modelling information that underlies linked open data
RLG	Research Libraries Group, a US-based library consortium, now part of OCLC
Rosetta	An OAIS-compliant digital preservation system produced

	by ExLibris
schematron	A validation language for XML which allows extra layers of validation beyond conformance to a schema to be tested
SIP (Submission Information Package)	In the OAIS conceptual model, a collection (package) of content and preservation description information which is submitted to an OAIS-compliant archive
SKOS (Simple Knowledge Organization System)	A family of RDF-based languages for the construction of thesauri and controlled vocabularies
structural metadata	Metadata required to described the internal structure and the component relationships of a digital object
textMD	A widely-used XML schema which encodes technical metadata for text-based digital objects
TIFF	Tagged Image File Format: a widely-used digital still-image file format
TIPR (Towards Interoperable Preservation Repositories)	A project which aimed to create a Repository eXchange Package (RXP) to allow the transfer of complex digital objects between disparate preservation repositories
VBScript	A scripting language based on Visual Basic
wiki	A website which allows users to add, edit or remove its contents via web browsers
XML (eXtensible Markup Language)	A widely-used application-independent markup language for encoding data and metadata
XSL (eXtensible Stylesheet Language)	An XML-based language for style sheets for transforming and formatting XML files

6. Further Reading

The following resources may be of interest to those interested in learning more about preservation metadata, the *PREMIS Data Dictionary*, and the use of PREMIS with METS:

PREMIS 2012, *PREMIS Data Dictionary for Preservation Metadata Version 2.2*, online at: http://www.loc.gov/standards/premis/v2/premis-2-2.pdf (last accessed 13/11/12)

The latest version of the PREMIS Data Dictionary

Lavoie, B and Gartner, R 2005, *DPC Technology Watch Report: Preservation Metadata*, online at: http://www.dpconline.org/component/docman/doc_download/88-preservation-metadata (last accessed 13/11/12)

The first edition of the current report; includes a detailed history of preservation metadata up to and including PREMIS

Caplan, P 2009, *Understanding PREMIS*, online at: http://www.loc.gov/standards/premis/understanding-premis.pdf (last accessed 13/11/12)

A gentle introduction to the PREMIS Data Dictionary *and implementation issues*

Lavoie, B 2008, 'PREMIS With A Fresh Coat of Paint: Highlights from the Revision of the *PREMIS Data Dictionary for Preservation Metadata'*, *D-Lib Magazine 14*, online at: http://www.dlib.org/dlib/may08/lavoie/05lavoie.html (last accessed 13/11/12)

A detailed description of the first major revision of the PREMIS Data Dictionary

Guenther, R 2008, 'Battle of the Buzzwords: Flexibility vs. Interoperability When Implementing PREMIS and METS', *D-Lib Magazine 14*, online at: http://www.dlib.org/dlib/july08/guenther/07guenther.html (last accessed 13/11/12)

An overview of, and guidelines for, using PREMIS with METS

Readers interested in PREMIS and related issues should visit the PREMIS Maintenance Activity website for news, events, and resources (http://www.loc.gov/standards/premis/) and sign up to the PREMIS Implementors' Group Forum (see information on PREMIS Maintenance Activity home page).

7. References

Caplan, P 2003, *Metadata Fundamentals for All Librarians*. Chicago: American Library Association.

Caplan, P 2009, *Understanding PREMIS*, online at: http://www.loc.gov/standards/premis/understanding-premis.pdf (last accessed 13/11/12)

CCSDS 2012, Reference Model for an Open Archival Information System (OAIS), online at: http://public.ccsds.org/publications/archive/650x0m2.pdf (last accessed 13/11/12)

Dappert, A & Enders, M 2010, 'Digital Preservation Metadata Standards', *NISO Information Standards Quarterly,* June 2010, online at: http://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf (last accessed 17/4/13)

Lavoie, B 2008, 'PREMIS With a Fresh Coat of Paint: Highlights from the Revision of the PREMIS Data Dictionary for Preservation Metadata', *D-Lib Magazine 14*, online at: http://www.dlib.org/dlib/may08/lavoie/05lavoie.html (last accessed 13/11/12)

Lavoie, B and Gartner, R 2005, *DPC Technology Watch Report: Preservation Metadata*, online at: http://www.dpconline.org/component/docman/doc_download/88-preservation-metadata (last accessed 13/11/12)

OCLC and RLG 2001, *Preservation Metadata for Digital Objects: A Review of the State of the Art*, online at: http://www.oclc.org/resources/research/activities/pmwg/presmeta_wp.pdf (last accessed 13/11/12)

OCLC and RLG 2002, *A Metadata Framework to Support the Preservation of Digital Objects*, online at: http://www.oclc.org/resources/research/activities/pmwg/pm_framework.pdf (last accessed 13/11/12)

PREMIS 2005, *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*, online at: http://www.loc.gov/standards/premis/v1/premis-dd_1.0_2005_May.pdf (last accessed 13/11/12)

PREMIS 2008, *PREMIS Data Dictionary for Preservation Metadata Version 2.0*, online at: http://www.loc.gov/standards/premis/v2/premis-dd-2-0.pdf (last accessed 13/11/12)

PREMIS 2011, *PREMIS Data Dictionary for Preservation Metadata Version 2.1*, online at: http://www.loc.gov/standards/premis/v2/premis-2-1.pdf (last accessed 13/11/12)

PREMIS 2012, *PREMIS Data Dictionary for Preservation Metadata Version 2.2*, online at: http://www.loc.gov/standards/premis/v2/premis-2-2.pdf (last accessed 13/11/12)

Vermaaten, S 2010, 'A Checklist and a Case for Documenting PREMIS-METS Decisions in a METS Profile', *D-Lib Magazine 16*, online at: http://dlib.org/dlib/september10/vermaaten/09vermaaten.html (last accessed 13/11/12)

Vermaaten, S, Lavoie, B and Caplan, P 2012 'Identifying Threats to Successful Digital Preservation: The SPOT Model for Risk Assessment', *D-Lib Magazine 18,* online at: http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html

Woodyard-Robinson, D 2007*, Implementing the PREMIS Data Dictionary: A Survey of Approaches*, online at: http://www.loc.gov/standards/premis/implementation-report-woodyard.pdf (last accessed 13/11/12)