Practical Preservation: The PREMIS Experience

Priscilla Caplan and Rebecca Guenther

Abstract

In 2003 the Online Computer Library Center (OCLC) and Research Libraries Group (RLG) established an international working group to develop a common, implementable core set of metadata elements for digital preservation. Most published specifications for preservation-related metadata are either implementation specific or broadly theoretical. PREMIS (Preservation Metadata: Implementation Strategies) was charged to define a set of semantic units that are implementation independent, practically oriented, and likely to be needed by most preservation repositories. The semantic units will be represented in a data dictionary and in a METS-compatible XML schema. In the course of this work, the group also developed a glossary of terms and concepts, a data model, and a typology of relationships. Existing preservation repositories were surveyed about their architectural models and metadata practices, and some attempt was made to identify best practices. This article outlines the history and methods of the PREMIS Working Group and describes its deliverables. It explains major assumptions and decisions made by the group and examines some of the more difficult issues encountered.

Introduction

In 2003 the Online Computer Library Center (OCLC) and Research Libraries Group (RLG) established an international working group to develop a common, implementable core set of metadata elements for digital preservation. Most published specifications for preservation-related metadata are either implementation specific or broadly theoretical. PREMIS

(Preservation Metadata: Implementation Strategies) was charged to define a set of metadata elements that are implementation independent, practically oriented, applicable to all types of materials, and likely to be needed by most preservation repositories. In addition, it aimed at establishing best practices for the implementation of preservation metadata.

The stated PREMIS objectives were to

- define an implementable set of "core" preservation metadata elements, with broad applicability within the digital preservation community;
- draft a data dictionary to support the core preservation metadata element set;
- examine and evaluate alternative strategies for the encoding, storage, and management of preservation metadata within a digital preservation system, as well as for the exchange of preservation metadata among systems;
- conduct pilot programs for testing the group's recommendations and best practices in a variety of systems settings;
- explore opportunities for the cooperative creation and sharing of preservation metadata.

It was intended that PREMIS would build on the earlier work of another initiative sponsored by OCLC and RLG, the Preservation Metadata Framework Working Group (OCLC, 2003). That group was convened in 2001–2002 to develop a framework outlining the types of information that should be associated with an archived digital object. Their report, *A Metadata Framework to Support the Preservation of Digital Objects* (OCLC/RLG, 2002), expanded the conceptual structure for the Open Archival Information System (OAIS) information model (Consultative Committee, 2002) and mapped preservation metadata elements to that conceptual structure. Although the framework proposed a list of metadata elements, it did not contain sufficient detail for an implementer to actually use the metadata in a preservation system without considerable further specifications. The PREMIS working group was established to take the previous group's work a step further: to develop a data dictionary of core metadata elements to be applied to archived objects, give guidance on the implementation of that metadata element set in preservation systems, and suggest best practice for populating those elements.

OCLC and RLG established the working group in 2003, chaired by Priscilla Caplan of the Florida Center for Library Automation and Rebecca Guenther of the Library of Congress. Because the charge was practical rather than theoretical, members were sought from institutions known to be running or developing preservation repository systems within the cultural heritage or information industry sectors. Conveners paid particular attention to diversity of stakeholders. The group consists of representatives from academic and national libraries, museums, and archives; governments;

and commercial enterprises in six different countries. In addition, PREMIS includes an international advisory committee of experts periodically called upon to review progress and provide feedback.

In order to accomplish as much of the charge as possible in a reasonable timeframe, PREMIS divided into two subgroups with different deliverables and strategies. The Core Elements Subgroup took responsibility for drafting the "core" preservation metadata elements and supporting data dictionary. The Implementation Strategies Subgroup was responsible for examining alternative strategies for the encoding, storage, and management of preservation metadata within digital preservation systems and for developing pilot programs to test the group's recommendations in a variety of system settings.

The work of both subgroups was conducted almost entirely by weekly conference calls, which was a challenge given that the group members were from time zones ranging from the western United States to eastern Australia. Fortunately, only one person had to get up in the middle of the night to attend! However, the sheer frequency of calls and the ambitious agenda created a sense of camaraderie among participants. Members quickly learned each others' voices and mastered use of a wiki (a Web site that allows any user to add and edit content) set up for their use by the University of Chicago. The Core Elements Subgroup also held two face-to-face meetings to expedite their work. The two meetings, one in San Diego in January 2004 and the other in Cambridge, Massachusetts, in August 2004, were highly productive and contributed to the sense of community among members.

One of the group's practices has been well received and might well be found useful by other initiatives. Every month a summary of each subgroup's activities is posted on the official Web site at http://www.oclc.org/research/projects/pmwg/. For example, the Core Elements update for September 2004 reads:

> The group spent time discussing the differences between files and bitstreams and how the semantic units applied to them. It was proposed that there was a need for a new level called "filestreams." This also related to previous discussions about embedded files. The group continued its discussion of environment elements and whether this information is dependent on file format information. It continued to define what information is needed about the environment in order to render objects for the long term. Two new participants joined the group, one from DSpace and another from the Walt Disney Company. A workplan was developed to finish the data dictionary by December in anticipation of a final PREMIS report by the end of 2004.

Because of these updates, anyone interested in the PREMIS activity could follow the group's progress, see what issues were under discussion, and simply be assured the work group was working.

IMPLEMENTATION STRATEGIES

The Implementation Strategies Subgroup was charged with examination and evaluation of alternative strategies for the encoding, storage, and management of preservation metadata within a digital preservation system. To find out how preservation repositories were actually implementing preservation metadata, the subgroup decided to survey repositories that were in operation or under development. Although their work was focused on metadata, the subgroup felt that the survey provided an opportunity to explore the state of the art in digital preservation generally, and questions were drafted to elicit information about policies, governance and funding, system architecture, and preservation strategies as well as metadata practices.

In November 2003 copies were sent by email to approximately seventy organizations thought to be active in or interested in digital preservation. The survey was also made available on the PREMIS Web site and announced on various discussion lists. By the end of March 2004, forty-eight survey responses were received from institutions developing or planning to develop a digital preservation repository. Sixteen of these respondents were contacted for more in-depth telephone interviews.

Although several institutions known to be developing digital preservation repository systems did not respond, the replies received appear to be reasonably representative of the state of the art in the winter of 2003–2004. Responses came from 13 countries and included 28 libraries, 7 archives, and 14 other types of organizations. Among the respondents were 10 national libraries and 6 national archives, showing heavy involvement in digital preservation at the national level, particularly in Europe and Canada.

Key findings are summarized in the report *Implementing Preservation Repositories for Digital Materials: Current Practices and Emerging Trends in the Cultural Heritage Community* (OCLC/RLG PREMIS Working Group, 2004), so they will not be repeated here. However, a few points are worth noting.

First, there is very little experience with digital preservation. Twenty-two respondents claimed to have a preservation repository in some stage of production (as opposed to planning, development, or alpha/beta testing). However, only half of these appeared to have implemented an active preservation strategy such as normalization, format migration, migration on demand, or emulation. This list included four national libraries/national archives and six institutions categorized as "other." None was an academic library.

This finding must color all other results, including those pertaining to metadata. Whatever practices were reported on the survey, apart from these eleven institutions the results reflect repositories not yet in production or not yet implementing active preservation strategies. We do not have enough experience to determine whether the metadata these systems record or plan to record is adequate for the purpose.

Second, those engaged in digital preservation still lack a common vocabulary and, to a large extent, a common conceptual framework. Although most respondents claimed to have been informed by the OAIS reference model and to be at least partly compliant with it, there was substantial difference of opinion as to the meaning of OAIS compliance. Although OAIS has been praised for providing a standard vocabulary for basic repository concepts, it is clear that most of these terms have not been widely adopted in the community, at least not in informal communications such as survey responses.

In relation to metadata, most respondents were recording several different types of metadata, and more than half were recording metadata in all of these categories: rights, provenance, technical, administrative, descriptive, and structural. Repositories appear to draw metadata elements from various schemes to suit their purposes. The Metadata Encoding and Transmission Standard (METS) (Library of Congress, 2005), NISO Z39.87 (Technical Metadata for Digital Still Images) (National Information Standards Organization and AIIM International, 2002), and the OCLC Digital Archive metadata set (OCLC, 2002) were the only named schemes used by more than 20 percent of respondents. Overall, thirty-three different metadata element sets or rule sets were mentioned by at least one repository. In general, the survey shows a picture of a community trying to take advantage of prior work but not at the point of developing or settling on dominant standards.

CORE ELEMENTS

Methodology

The Core Elements Subgroup began its work by attempting to define the word "core" for the purpose of developing a metadata element list and data dictionary. After much discussion the group settled on a practical definition of core: those elements that a working archive is likely to need to know in order to support the functions of ensuring viability, renderability, understandability, authenticity, and identity in a preservation context. Initially the group felt that all core elements should be considered mandatory by definition, but some flexibility crept in with the acknowledgement that some elements are more core than others, and even necessary information cannot always be provided.

The Core Elements Subgroup then started analyzing the recommendations of the earlier Preservation Metadata Framework Working Group related to Preservation Description Information. This included "digital provenance," or the documentation of events associated with the digital objects. Those members of the subgroup from institutions actively running or developing preservation repositories mapped the elements from the framework to what was used in their own systems. It became clear that the elements detailed in the previous work (which themselves had been

mapped to the OAIS information model) did not always correspond to elements implemented in practice and did not give adequate guidance on how to use them. However, the exercise was useful in providing a common denominator for diverse implementations; the group discussed each element in conference calls to see where there was commonality in usage. Elements that emerged as being widely used across implementations were considered the beginning of a core element list.

The group made the decision that the data dictionary it was developing would be wholly implementation independent. That is, the core elements would define information that a repository needed to know, regardless of how, or even whether, it was stored. For instance, for a given identifier to be usable, it is necessary to know the identifier scheme and the namespace in which it is unique. If a particular repository uses only one type of identifier scheme, say one that is internally defined and assigned, the scheme can be assumed, and the repository would have no need to record it at all. The repository would, however, need to know this information and be able to supply it when exchanging metadata with other repositories. Because of the emphasis on the need to know rather than the need to record or represent in any particular way, the group preferred to use the term "semantic unit" (meaning an atom of meaning) rather than "metadata element." The data dictionary therefore names and describes semantic units.

After drafting a preliminary data dictionary for digital provenance information, the group began to consider technical metadata, or detailed information about the physical characteristics of digital objects. The group realized that it did not have either the time or the expertise to tackle format-specific technical metadata for various types of digital files. By scoping the work to include only that metadata applicable to all (or at least most) digital formats, the group was able to limit the work to a reasonable set of semantic units and leave further development to format experts. The group compiled a list of potential semantic units based on specifications for the proposed Global Digital Format Registry (GDFR, n.d.) supplemented by data elements used in the repository systems of members' institutions. Each element on the list was then discussed at some length, and those found to be both useful and broadly applicable were added to the data dictionary.

Data Model

One of the hardest issues to tackle was the development of an acceptable abstract data model. A valid criticism of the earlier framework was that the document recommended metadata elements pertaining to many different types of things while giving no guidance as to what type of thing they applied to. For example, "Resource Description" included the subelement "Existing metadata," an example of which was "a MARC bibliographic record." Bibliographic records usually describe intellectual entities, such as books, sound recordings, and Web sites. Another element, "File de-

scription" (defined as "technical specifications of the file(s) comprising a Content Data Object"), would appear to apply to individual digital files. A third element, "Size of object," might be taken to apply to the total size of a complex object (for example, a book made up of many page images) or to a single stored file. The lack of specifics as to what level of granularity of an object the elements applied to made the document difficult to actually use in metadata implementations.

The data model was intended to accomplish three purposes. First, it would force PREMIS members to be rigorous in their thinking in the development of the data dictionary. Second, it provided a structure for arranging entries in the data dictionary. Third, it would help implementers of the data dictionary understand how to apply semantic units. The data model was not, however, meant to imply any particular implementation of the semantic units in the data dictionary.

In the PREMIS data model there are five types of entities: intellectual entities, objects, agents, rights, and events. Although it is possible these definitions will change before the final report, these entities are currently defined as follows:

- An *event* is an action that involves at least one object, agent, and/or rights entity.
- An *agent* is an actor associated with preservation events in the life of an object.
- A *right* is an assertion of one or more rights or permissions pertaining to an object.
- An *intellectual entity* is a coherent set of content that is reasonably described as a unit, for example, a particular book, map, photograph, or database.
- An *object* is one or more sequences of bits stored in the preservation repository.

There are four subtypes of the object entity: file, filestream, bitstream, and representation. The most difficult part of the development of the data model has been to appropriately identify, name, and define these subtypes. Definitions in this article are slightly less elaborate than those in the actual data model, but they communicate the concepts effectively.

Of the five entity types, file is perhaps the most intuitive, as our definition resembles that of common usage: a named ordered sequence of zero or more bytes known to an operating system and accessible by applications. Every file has a file format, defined as a specific pre-established structure of a computer file that specifies how data is organized. A file may contain zero or more bitstreams and zero or more filestreams.

A "bitstream" is defined as data within a file that cannot be transformed into a stand-alone file without the addition of file structure (headers, etc.) and/or reformatting in order to comply with some particular file format. A "filestream" is a contiguous set of bits within a file that can be transformed

into a stand-alone file conforming to some file format without adding information or reformatting the bitstream. An example of a bitstream is an image embedded within a PDF; an example of a filestream is a TIFF image within a TAR file.

A "representation" is the set of files needed to provide a complete and reasonable rendition of an intellectual entity. It can be thought of as the digital embodiment of an intellectual entity. Preservation repositories never store intellectual entities, but they may store representation objects.

As an example, the final PREMIS report is an intellectual entity. There will probably be PDF and HTML versions posted on the Web; many readers will download their own copies, but all copies will have the same authors, title, and content. If the report were archived in a preservation repository, at least one representation would be stored. This might, for example, be a single, specific PDF file. The PDF file will doubtless contain embedded graphics for tables and charts, which would be bitstreams. If the HTML version were archived, the representation might consist of three or four files—the HTML file and several GIF images. Perhaps the repository will want to bundle these files together for storage by creating a TAR file. That TAR file would then have within it three or four filestreams, which could be extracted into files at some later time.

These distinctions are important because different semantic units of metadata apply at different levels. The intellectual entity may have an ISBN or technical report number, but the representation does not. The representation may have an identifier known to the preservation repository, but the intellectual entity does not. The file will have a file name and file format, the filestream will have a file format but no file name, and the bitstream will have no file name or file format, although it may have other format characteristics such as color space.

The PREMIS data dictionary attempts to define core semantic units pertaining to all subtypes of objects and events. Intellectual entities and agents are not addressed in any detail because they have been the focus of other metadata schemes and they do not present unique requirements in the digital preservation context beyond the minimum needed to establish relationships between these and other types of entities. At the time of this writing, the group was still exploring the extent to which rights and/or permissions should be described.

Relationships are the other important part of the data model. Entities can be related to entities of different types (for example, objects can be related to agents) and to entities of the same type (for example, objects can be related to other objects). Just as there may be core semantic units generally necessary in the majority of preservation repository applications, there are core relationships that most preservation repositories will need to record.

The relationships between objects, agents, and events constitute digital

provenance. As Clifford Lynch wrote in "Authenticity and Integrity in the Digital Environment":

> Provenance, broadly speaking, is documentation about the origin, characteristics, and history of an object; its chain of custody; and its relationship to other objects. The final point is particularly important. There are two ways to think about a digital object that is created by changing the format of an older object . . . We might think about a single object the provenance of which includes a particular transformation, or we might think about multiple objects that are related through provenance documentation. Thus, provenance is not simply metadata about an object—it can also be metadata that describe the relationships between objects (2000).

Objects and Events

Most of the semantic units in the data dictionary pertain to objects and events. Semantic units related to the object entity describe characteristics relevant to preservation management. It is assumed that data content objects are held in the preservation repository and that associated metadata may be held in the repository, in external systems, or in both. Data dictionary entries for objects indicate the level at which the semantic unit is applicable: representation, file, and/or bitstream. Filestream is considered equivalent to file for the purposes of applicability.

Semantic units associated with object entities include identifiers, location information, and technical characteristics. In anticipation of the development of format registries such as the proposed GDFR, the data dictionary also contains semantics for referencing format registry entries. Similarly, it provides for basic software and hardware environment information and anticipates adding references to future environment registries.

Figures 1 and 2 provide examples of entries in the data dictionary. Figure 1 shows the definition of a "container" unit (fixity), which has no data itself but serves to group together three related semantic components (messageDigestAlgorithm, messageDigest, and messageDigestOriginator). Figure 2 shows the definition of one of these semantic components, messageDigestAlgorithm.

Events are actions that involve one or more objects and may be related to one or more agents. The PREMIS report states that whether or not a preservation repository records an event depends upon the importance of the event in the context of that repository. It recommends using the semantic units related to the Events entity when recording actions that modify objects. Other actions, such as the copying of an object for backup purposes, may be recorded in system logs or an audit trail but not necessarily as an event.

Most of the documentation about the digital provenance of objects is given in relation to events. Semantic units include event identifier, event type (for example, compression, migration, validation, etc.), event out-

Semantic unit	fixity		
Semantic components	messageDigestAlgorithm, messageDigest, messageDigestOriginator		
Definition	Information used to verify whether an object has been altered in an undocumented or unauthorized way.		
Data constraint	Container		
Object category	**Representation**	**File**	**Bitstream**
Applicability	Not applicable (see usage note)	Applicable	Applicable (see usage note)
Repeatability		Repeatable	Repeatable
Obligation		Optional	Optional
Creation/ Maintenance notes	Automatically calculated and recorded by repository.		
Usage notes	To perform a fixity check, a message digest calculated at some earlier time is compared with a message digest calculated at a later time. If the digests are the same, the object was not altered in the interim. Recommended practice is to use two or more message digests calculated by different algorithms.		
	The act of performing a fixity check and the date it occurred would be recorded as an Event. The result of the check would be recorded as the eventOutcome. Therefore, only the messageDigestAlgorithm and messageDigest need to be recorded as objectCharacteristics for future comparison.		
	Representation level: It could be argued that if a representation consists of a single file, or if all the files comprised by a representation are combined (e.g., zipped) into a single file, then a fixity check could be performed on the representation. However, in both cases the fixity check is actually being performed on a file, which in this case happens to be coincidental with a representation.		
	Bitstream level: Message digests can be computed for bitstreams although they are not as common as with files. For example, the JPX format, which is a JPEG2000 format, supports the inclusion of MD5 or SHA-1 message digests in internal metadata that was calculated on any range of bytes of the file.		
	See "Fixity, integrity, authenticity," page 4-5.		

Figure 1.   Data Dictionary Entry for Fixity

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

Figure 2.   Data Dictionary Entry for messageDigestAlgorithm

come, and event date/time. When properties of an object are the result of an event, this is considered event-related information, but in practice this may be recorded with the object or with the event. An example of a data dictionary entry for a semantic unit related to the Event entity is given in figure 3.

PREMIS Report and Further Work

The final PREMIS report will go into greater detail about the findings of the working group and will present a completed data dictionary with examples. In addition, it will include a glossary, a description of the data model, discussions of some of the more difficult or controversial semantic units, and other related information. As of this writing, the working group was still conducting work by conference calls and the data dictionary was not yet completed. The target date for completion is December 2004.[1]

Semantic unit	eventType
Semantic components	None
Definition	A categorization of the nature of the event.
Rationale	Categorizing events will aid the preservation repository in machine processing of event information, particularly in reporting.
Data constraint	Value should be taken from a controlled vocabulary.
Examples	E77 [a code used within a repository for a particular event type] Ingest
Repeatability	Not repeatable
Obligation	Mandatory
Usage notes	Each repository should define its own controlled vocabulary of eventType values. A suggested starter list for consideration (see also the Glossary for more detailed definitions): capture = the process whereby a repository actively obtains an object compression = the process of coding data to save storage space or transmission time deaccession = the process of removing an object from the inventory of a repository decompression = the process of reversing the effects of compression decryption = the process of converting encrypted data to plaintext deletion = the process of removing an object from repository storage digital signature validation = the process of determining that a decrypted digital signature matches an expected value dissemination = the process of retrieving an object from repository storage and making it available to users fixity check = the process of verifying that an object has not been changed in a given period ingestion = the process of adding objects to a preservation repository message digest calculation = the process by which a message digest ("hash") is created migration = a transformation of an object creating a version in a more contemporary format normalization = a transformation of an object creating a version more conducive to preservation replication = the process of creating a copy of an object that is, bit-wise, identical to the original validation = the process of comparing an object with a standard and noting compliance or exceptions virus check = the process of scanning a file for malicious programs The level of specificity in recording the type of event (e.g., whether the eventType indicates a transformation, a migration or a particular method of migration) is implementation specific and will depend upon how reporting and processing is done. Recommended practice is to record detailed information about the event itself in eventDetail rather than using a very granular value for eventType.

Figure 3. Data Dictionary Entry for eventType

Although the data dictionary is intended to be implementation neutral, for information to be exchanged between repositories there must be some standard representation. The implementation survey showed wide use of METS among implementers. The METS initiative intends to draft PREMIS-based XML schemas suitable for use as extension schemas for the digital provenance metadata section (digiprovMD) and technical metadata section (techMD) of a METS document. The digiprovMD will be based on the events section of the data dictionary. The new techMD section will complement the other format-specific technical metadata sections and will include general technical metadata that applies regardless of file format. It will be necessary to reconcile existing format-specific extension schema with this new general one, since some data elements that apply regardless of file format will already be included in defined-format specific technical metadata extension schema (for example, MIX, the XML binding of the NISO/AIIM standard Z39.87, Technical Metadata for Digital Still Images) (National Information Standards Organization & AIIM International, 2002).

Opportunities for developing testbeds for implementing PREMIS-compliant metadata are currently under discussion, as are trials of the exchange of preservation metadata among repositories. It is unlikely that these will actually be implemented before the group is formally disbanded, so other mechanisms for continuing this work are being considered. Mechanisms for supporting the adoption of PREMIS metadata, gathering feedback and evidence of practice, and maintaining the data dictionary over time will also be necessary. The PREMIS Web site should be consulted for the status of these and other related activities.

NOTE

1. Since this article was written, the PREMIS working group released *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group* in May 2005. It is available from the PREMIS Web site at http://www.oclc.org/research/projects/pmwg/. The Web site for PREMIS maintenance activity is http://www.loc.gov/standards/premis/.

REFERENCES

Consultative Committee for Space Data Systems. (2002). *Reference model for an Open Archival Information System (OAIS)* (CCSDS 650.0-B-1). Retrieved March 8, 2005, from http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf.

Global Digital Format Registry (GDFR). (n.d.). Home page. Retrieved March 8, 2005, from http://hul.harvard.edu/gdfr/.

Library of Congress. (2005). *Standards: Metadata encoding and transmission standard.* Retrieved March 8, 2005, from http://www.loc.gov/standards/mets/.

Lynch, C. (2000). Authenticity and integrity in the digital environment: An exploratory analysis of the central role of trust. In *Authenticity in a digital environment* (Council on Library and Information Resources report). Retrieved March 8, 2005, from http://www.clir.org/pubs/reports/pub92/lynch.html.

National Information Standards Organization & AIIM International. (2002). *Data dictionary: Technical metadata for digital still images* (NISO Z39.87). Retrieved March 8, 2005, from http://www.niso.org/standards/resources/Z39_87_trial_use.pdf.

OCLC. (2002). *OCLC digital archive system guides: Digital archive metadata elements.* Retrieved March 8, 2005, from http://www.oclc.org/support/documentation/pdf/da_metadata_elements.pdf.

———. (2003). *Preservation Metadata Framework Working Group.* Retrieved March 8, 2005, from http://www.oclc.org/research/projects/pmwg/wg1.htm.

OCLC/RLG. (2002). *Preservation metadata and the OAIS information model: A metadata framework to support the preservation of digital objects.* Retrieved March 8, 2005, from http://www.oclc.org/research/projects/pmwg/pm_framework.pdf.

OCLC/RLG PREMIS Working Group. (2004). *Implementing preservation repositories for digital materials: Current practices and emerging trends in the cultural heritage community.* Retrieved March 8, 2005, from http://www.oclc.org/research/projects/pmwg/surveyreport.pdf.

Priscilla Caplan, Assistant Director for Digital Library Services, Florida Center for Library Automation, 5830 NW 39th Avenue, Gainesville FL 32606, pcaplan@ufl.edu, and Rebecca Guenther, Senior Networking and Standards Specialist, Library of Congress, 101 Independence Ave. SE, Washington, DC 20540-4402, rgue@loc.gov. Priscilla Caplan is Assistant Director for Digital Library Services at the Florida Center for Library Automation, where she is managing a project to develop a digital preservation repository for the use of the public universities of Florida. She is the author of *Metadata Fundamentals for All Librarians* (ALA Editions, 2003) and numerous articles on digital preservation, metadata, reference linking, and standards for digital libraries. In addition to co-chairing the OCLC Working Group on Preservation Metadata: Implementation Strategies, she co-chairs the NISO/EDItEUR Joint Working Party on the Exchange of Serials Subscription Information.

Rebecca Guenther is Senior Networking and Standards Specialist in the Network Development and MARC Standards Office of the Library of Congress, in which she has worked since 1989. Previous positions included cataloger at the National Library of Medicine; cataloger in the Library of Congress' Shared Cataloging Division/German Language Section, and section head of the National Union Catalog Control Section/Catalog Management and Publication Division. Her current responsibilities include work on national and international information standards, including, among others, rotating chair of ISO 639 Joint Advisory Committee on language codes and a member of the NISO Standards Development Committee. Rebecca has worked in the area of metadata since the early 1990s, including maintaining a number of crosswalks between various metadata schemes; participating in development of XML bibliographic descriptive schemas (MODS and MARCXML); serving as chair of the Dublin Core Libraries Working Group and as a member of the Dublin Core Usage Board; serving as a co-chair of PREMIS, an OCLC/RLG working group on preservation metadata implementation strategies; and participating in the Open Ebook Forum's Metadata and Identifiers Working Group, among others. She has published articles and made presentations widely on metadata and various standards-related efforts.