Conformant Implementation of the PREMIS Data Dictionary
PREMIS Editorial Committee
April 2015

Introduction

The PREMIS Data Dictionary was designed to be as flexible as possible in its implementation. No assumptions were made regarding the nature of the digital archiving system in which the Data Dictionary would be implemented, the preservation strategy being followed or even the metadata management processes responsible for creating and maintaining preservation metadata. The "technical neutrality" built into the design of the Data Dictionary is intended to maximize the Dictionary's applicability across the broad range of digital preservation contexts in which it could potentially be implemented.

Technical neutrality does not, however, override the need to establish a set of principles for implementing the Data Dictionary in ways that ensure data consistency within and across preservation repositories. Such consistency is necessary in order to support a variety of use cases, including:

- Inter-repository data exchange
- Repository certification
- Shared registries
- Automation/reusable tools
- Vendor support

To support these and other use cases, the PREMIS Editorial Committee has developed a *conformance statement* that defines a set of principles governing a conformant implementation of the PREMIS Data Dictionary. The purpose is to define a minimum set of requirements that establish certain expectations associated with a PREMIS implementation that are needed to support a range of use cases, without unnecessarily reducing the flexibility and discretion of implementers to apply the Dictionary in ways that suit their particular needs. It is important to note that adherence to the conformance principles is not a formal requirement for implementing the PREMIS Data Dictionary (although the Editorial Committee does believe that following these principles would be good practice in nearly all implementation contexts). In other words, a repository is free to implement the Data Dictionary in whatever way it chooses in situations where conformance is not asserted. However, in situations where PREMIS conformance is asserted, implementers must be able to demonstrate adherence to the conformance principles discussed below.

The PREMIS conformance statement is divided into three parts. The first part describes a set of principles that establish baseline requirements for implementing PREMIS semantic units and the Data Dictionary in a conformant way. The second part builds on these principles by introducing graduated levels of conformance, designed to provide a means of quantifying the degree to which PREMIS has been implemented by a repository. The third part offers guidance

on aspects of implementation that help support conformance: controlled vocabularies, extensible semantic units and container formats.

I. Principles of conformance
The principles of conformance enumerate the conditions that must be met if a PREMIS implementation is to be considered conformant. The principles operate at two levels:

- *Semantic Unit*: conformant implementation of the information defined by a particular semantic unit in the Data Dictionary;
- *Data Dictionary*: conformant implementation of all semantic units in the Data Dictionary relevant to a particular digital preservation activity.

* * * * *

PRINCIPLES OF USE (SEMANTIC UNIT): A conformant implementation of a PREMIS semantic unit must follow all requirements and constraints prescribed in the Data Dictionary for that semantic unit. Specifically:

- If a metadata element shares the name of a PREMIS semantic unit, it must also share its definition. If a metadata element shares the definition of a PREMIS semantic unit but does not share its name, the repository must establish a mapping between the metadata element and its corresponding PREMIS semantic unit.

EXAMPLES

Conformant:

- A repository implements a metadata element *objectCategory* that shares the definition of PREMIS *objectCategory*.
- A repository uses a relational database system with an *Objekteigenschaften* table and establishes in the system documentation that *Objekteigenschaften* shares the definition of the PREMIS semantic unit *objectCharacteristics*.

Non conformant:

- A repository implements a metadata element *objectCategory* that records information defined in PREMIS semantic units *objectCategory* and *preservationLevel*.
- A repository uses a relational database system with an *Objekteigenschaften* table but does not establish that *Objekteigenschaften* shares the definition (that is, captures the same information) of the PREMIS semantic unit *objectCharacteristics*.

- Usage requirements specified in the Data Dictionary for a particular semantic unit must be observed. Repeatability, obligation (i.e., whether a semantic unit is mandatory), and applicability (Bitstream, File, and Representation) requirements can be made more stringent, but not more relaxed.

EXAMPLES

Conformant:

- A repository implements the PREMIS semantic unit *size* and follows the semantic unit's data constraint by requiring that size be an integer.
- A repository implements the repeatable PREMIS semantic unit *objectIdentifier*, but stipulates that its implementation will only allow one instantiation of this semantic unit per Object in the system.
- A repository implements the PREMIS semantic unit *messageDigestAlgorithm*, which is defined as being applicable at the File and Bitstream levels. However, the repository chooses to record this information only at the File level.

Non-conformant:

- A repository implements the PREMIS semantic unit *size*, and records the string value "Twenty megabytes".
- A repository records three instances per Object of the non repeatable PREMIS semantic unit *objectCategory*.
- A repository implements the PREMIS semantic unit *messageDigestAlgorithm* at the level of Representation.

An implementation of a PREMIS semantic unit that fails to observe one or more of these principles is considered non-conformant.

* * * * *

PRINCIPLES OF USE (DATA DICTIONARY): A conformant implementation of the PREMIS Data Dictionary at a minimum must:

- Include the mandatory semantic units for any Data Model Entity (Objects, Events, Agents or Rights) supported by the repository.

-

Note: A PREMIS semantic component is mandatory only if the parent container is implemented. For example, the *messageDigest* semantic component is only mandatory if its parent *fixity* is implemented.

Note: Regarding the Object Entity, mandatory semantic units are only required for the Object types (Bitstream, File, Representation and Intellectual Entity) that are supported by the repository. For example, if a repository supports only Files and Representations, conformance requires implementation of the mandatory Object semantic units that apply at the File or Representation levels; mandatory Object semantic units that apply only at the Bitstream or Intellectual Entity level would not be required for conformance.

EXAMPLES

Conformant:

- A repository chooses to record information about Objects at the File level. The repository therefore implements metadata elements that, at a minimum, cover all of the information specified in the mandatory semantic units for the Objects File type.

- A repository that is conformant in regard to Objects also wants to record information about Events; therefore, it implements metadata elements that, at a minimum, capture all of the information specified in the semantic units *eventIdentifier*, *eventType*, and *eventDateTime*.

Non-conformant:
- A repository implementing Objects contains no metadata elements that capture the information specified in the PREMIS semantic unit *objectCategory*.
- The information a repository records about Events does not include information that corresponds to the PREMIS semantic unit *eventType*.

- Be able to recover all of the information specified in the mandatory PREMIS semantic units from the repository system (regardless of its specific implementation), and associate it with its corresponding Entity.

A repository's implementation of the PREMIS Data Dictionary that fails to observe one or more of these principles is considered non-conformant.

<center>* * * * *</center>

II. Levels of conformance

The principles of Conformance defined in Section I stipulate the conditions that must be met in order to achieve conformant PREMIS-based information, both at the semantic unit and Data Dictionary levels. These conditions, even when fully met, bestow a great deal of flexibility on PREMIS implementers in terms of the choice of implementation strategy for incorporating the PREMIS Data Dictionary into the repository system. This section provides a way of quantifying the degree to which PREMIS has been implemented by defining graduated levels of conformance. In situations where PREMIS conformance is asserted, implementers must be able to demonstrate the level with which they purport to comply.

The levels are built around three ways of implementing PREMIS in any repository system:
- being able to map preservation metadata to PREMIS,
- being able to export preservation metadata as PREMIS, and
- using PREMIS as an internal schema in a way that does not require any further mapping or conversion.

Level 1 describes mapping only; that is, the repository has documentation for mapping its metadata elements to PREMIS elements. For example, the repository may have a field called "formatIdentification" which can be said to be equivalent to, and is used in the same way as, the PREMIS semantic unit *formatName*.

Level 2 describes a repository that has metadata that map to PREMIS and which can also be routinely exported, through established tools and processes, as PREMIS. For example, the internal element "formatIdentification" can be mapped to and exported as the PREMIS XML element <formatName>.

Finally, level 3 describes a repository using preservation metadata that match PREMIS semantic units exactly. In that case, the field label would be "formatName" and no mapping or conversion would be required to associate the field value with that label.

The levels are further divided into two categories: Category A is implementation of the Object entity only, and Category B is implementation of Objects, Events and Agents. This division assumes that in order to assert any level of conformance to the PREMIS Data Dictionary, a preservation repository must, at a minimum, implement the mandatory elements of the Object entity for any object types (Bitstream, File, Representation or Intellectual Entity) that are supported by the repository. This requirement supports the basic identification of digital objects that is a necessary prerequisite for undertaking any actions to preserve them. It also provides a means of associating preservation actions with objects: without the Object entity, Events and Agents would lack context and meaning. Note that the Rights entity has been excluded as a requirement. Although information about rights is important, such information can be managed using other metadata standards, including descriptive metadata standards widely used by digital repositories; adding Rights to the conformance levels may therefore impose an unnecessarily onerous requirement to duplicate existing information.

Level 1 Conformance through mapping

- Level 1A: Object entity only
A repository uses one or more internal preservation metadata schemas, elements of which can be mapped to PREMIS. Such mapping must satisfy the principles of use at both the semantic unit and Data Dictionary levels. The repository is able to produce documentation demonstrating such mapping, at a minimum, for the Object entity.

- Level 1B: Object, Event and Agent entities
A repository uses one or more internal preservation metadata schemas, elements of which can be mapped to PREMIS. Such mapping must satisfy the principles of use at both the semantic unit and Data Dictionary levels. The repository is able to produce documentation demonstrating such mapping, at a minimum, for the Object entity; one or more Agents; and sufficient Event metadata to document actions the repository has taken to preserve the digital objects.

Level 2 Conformance through export

- Level 2A: Object entity only
A repository uses one or more internal preservation metadata schemas, elements of which can be exported as PREMIS. Such export must satisfy the principles of use at both the semantic unit and Data Dictionary levels. The repository has established processes and tools in place to perform these exports as a routine operation, and is able to demonstrate such capability, at a minimum, for the Object entity.

- Level 2B: Object, Event and Agent entities

A repository uses one or more internal preservation metadata schemas, elements of which can be exported as PREMIS. Such export must satisfy the principles of use at both the semantic unit and Data Dictionary levels. The repository has established processes and tools in place to perform these exports as a routine operation, and is able to demonstrate such capability for the Object entity; one or more Agents; and sufficient Event metadata to document actions the repository has taken to preserve the digital objects.

Level 3 Conformance through internal implementation

- Level 3A: Object entity only

A repository implements the PREMIS Data Dictionary as an internal metadata schema in a way that satisfies the principles of use at both the semantic unit and Data Dictionary levels and in a form that does not require further mapping or conversion. The repository implements, at a minimum, the Object entity.

- Level 3B: Object, Event and Agent entities

A repository implements the PREMIS Data Dictionary as an internal metadata schema in a way that satisfies the principles of use at both the semantic unit and Data Dictionary levels and in a form that does not require further mapping or conversion. The repository implements, at a minimum, the Object entity; one or more Agents; and sufficient Event metadata to document actions the repository has taken to preserve the digital objects.

III. Implementation notes

The Principles of Conformance defined in Section I stipulate the conditions that must be met in order to achieve conformant PREMIS-based information, both at the semantic unit and Data Dictionary levels. This section provides supplementary guidance on aspects of implementation that help support conformance by facilitating data consistency and inter-repository data exchange.

Use of controlled vocabularies: The Data Dictionary recommends the use of controlled vocabularies for certain PREMIS semantic units (see, for example, *preservationLevelValue*). In these instances, use of controlled vocabularies is not required for conformance. However, it is strongly recommended that whenever possible controlled vocabularies, whether internally defined or based on the Library of Congress terms enumerated at http://id.loc.gov, be used in order to facilitate machine processing and inter-repository exchange.

EventType **controlled vocabulary:** Category B in all of the levels of conformance described in Section II refers to implementation of "sufficient Event metadata to document actions the repository has taken to preserve the digital objects." The PREMIS Editorial Committee recognizes that it cannot prescribe what actions a repository must take to preserve digital objects. However, it seeks to encourage an emerging body of practice on this subject by supporting and expanding on the *eventType* controlled vocabulary maintained by the Library of

Congress at http://id.loc.gov/vocabulary/preservation/eventType.html. The Committee strongly recommends that repositories use the terms enumerated and defined in this vocabulary. It further recommends that repositories select a minimum core set of terms from this vocabulary and use this core set consistently for digital objects preserved within the repository. The controlled vocabulary for Events is a direct link to the preservation actions that a repository should be undertaking. Utilising the vocabulary therefore gives implementers access to the preservation community's understanding of functions and best practices in a preservation repository. Using controlled lists of precisely specified Event types across many repositories can also help during the repository certification or audit process. Given the evolving nature of digital preservation practices, the Committee welcomes suggestions on addition of new terms to capture information on actions not already reflected in the *eventType* controlled vocabulary.

Extensibility: A repository can extend PREMIS semantic units through the use of formal metadata external to the Data Dictionary. This extensibility allows repositories to add detailed and otherwise unsupported metadata while adhering to the principles of conformance for semantic units and the Data Dictionary. The Data Dictionary supplies information on which semantic units can be extended in this manner and the principles to be observed when doing so. See "Extensibility" in the introductory section.

Container formats: The PREMIS Data Dictionary makes no stipulations on how PREMIS information is recorded or otherwise managed within a repository system. However, many repositories have found it useful to adopt some form of container format to gather together and organize various forms of metadata, record structural information and bundle metadata with its corresponding digital object. METS (Metadata Encoding and Transmission Standard) is commonly used as a container format for this purpose. In view of this, the Editorial Committee has produced a set of principles to guide implementation of the PREMIS Data Dictionary with METS: http://www.loc.gov/standards/premis/guidelines-premismets.pdf. In addition, a checklist for making and documenting implementation-specific decisions regarding use of PREMIS with METS is available: http://www.loc.gov/standards/premis/premis_mets_checklist.pdf.
The Editorial Committee strongly recommends the use of a container format to support implementation of the PREMIS Data Dictionary, especially in contexts where data is being exchanged across repositories.