Conformant Implementation of the PREMIS Data Dictionary
PREMIS Editorial Committee
October 2010

I. 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.

The importance of technical neutrality as a design principle for the Data Dictionary implies that any conformance requirements associated with the Dictionary will necessarily be lightweight. But this is not to say that conformance is unimportant in a PREMIS context; in fact, there are a number of use cases where establishing shared expectations in regard to a PREMIS implementation is of practical benefit, 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 two 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 supplements these principles with a description of the key "degrees of freedom" that are left open to PREMIS implementers once the basic conformance principles are satisfied. Put another way, the conformance statement describes both what implementers must do to achieve conformance, and what implementers are free to decide for themselves while still remaining conformant.

II. 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

The principles also apply to two general usage scenarios:

- Internal conformance: conformance from the perspective of internal use by a single repository
- External conformance: conformance from the perspective of cross-repository interactions in which PREMIS information is exchanged

Note: a "metadata element" is defined as an implementation of either 1) part of the information defined by a PREMIS semantic unit; 2) all of the information defined by a PREMIS semantic unit; or 3) information defined by multiple PREMIS semantic units.

* * * * *

PRINCIPLES OF USE (SEMANTIC UNIT): A conformant implementation of a PREMIS semantic unit must follow all requirements and constraints prescribed in the *latest version of 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*.

Note: Non-PREMIS metadata can certainly be used to supplement PREMIS metadata in the repository, but it is subject to the same rules about naming and definitions discussed above. If information is taken from another metadata schema to populate PREMIS semantic units, it must conform to the requirements and constraints associated with the corresponding PREMIS semantic unit.

- 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

(bit stream, 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.

Note: 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 controlled vocabularies be used whenever possible to facilitate machine processing and inter-repository exchange. *See Section III (Degrees of Freedom) below*.

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

* * * * *

PRINCIPLES OF USE (DATA DICTIONARY): A conformant implementation of the PREMIS Data Dictionary *at the 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.

Note: Regarding the Object Entity, mandatory semantic units are only required for the Object types (bit stream, file, representation) 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 bit stream 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 the 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 the 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.

Note: The Data Dictionary does not prescribe in any way how a repository should implement PREMIS semantic units in its system. Consequently, there are many possible variations on how a specific set of PREMIS semantic units (such as the mandatory semantic units required for a conformant implementation of the Data Dictionary) can be manifested in a repository system. This bestows a great deal of flexibility on repository managers in terms of how PREMIS information is incorporated into their repository operations. However, observance of the above Principle of Use ensures that all variations in implementation strategy encompass at the minimum the mandatory information defined in the Data Dictionary, by requiring that this information is "extractable" from the system, and regardless of the form in which it is recorded or its distribution across metadata elements, it can be mapped to its corresponding semantic units and associated with its appropriate Entity.

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

* * * * *

INTERNAL CONFORMANCE: PREMIS conformance as it relates to PREMIS-based information residing *within a repository*.

- A repository that satisfies the Principles of Use at both the semantic unit and Data Dictionary levels is considered *internally conformant*.

* * * * *

EXTERNAL CONFORMANCE: PREMIS conformance as it relates to the exchange of PREMIS-based information *between repositories*. There are two forms of external conformance: *import* and *export*.

- **Import:** A repository that is *import conformant* must be able to accept PREMIS-conformant information in the form provided by another repository, parse it, and allocate the information to its corresponding metadata elements in the local repository system, as well as associate it with the appropriate Entities.

- **Export:** A repository that is *export conformant* must be able to extract PREMIS-conformant information from its local system, and provide it to another repository in an agreed-upon form, and associate it with its appropriate Entity.

Note: In order to satisfy the Principles of Use at the Data Dictionary level, a conformant exchange of PREMIS-based information between two repositories requires that the mandatory semantic units for any Data Model Entity supported in the exchange should accompany a digital object when it is transferred from one repository to another.

III. Degrees of freedom

The Principles of Conformance defined in Section II 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 describes the major "degrees of freedom" implementers can exercise in regard to shaping their implementation strategies, while still remaining conformant.

Naming: A repository is free to implement PREMIS semantic units using names different from those defined in the Data Dictionary. (However, remember that *if* a metadata element does share the name of a PREMIS semantic unit, it must share its definition; see Principles of Use (Semantic Unit) above.)
Example:
The PREMIS semantic unit *size* is implemented as a metadata element named *sizeInBytes*.

Granularity: A repository is free to implement PREMIS semantic units at higher or lower levels of granularity than what is defined in the Data Dictionary. Put another way, a metadata element implemented by a repository can incorporate information from more than one PREMIS semantic unit, or alternatively, encompass only part of the information defined in a PREMIS semantic unit (e.g., if the information from a PREMIS semantic unit is distributed over multiple metadata elements).
Examples:
A repository implements a metadata element called *iprInformation* that records the information defined in the PREMIS semantic units *rightsBasis* and *copyrightInformation*.

A repository implements two metadata elements – *identifierNameSpace* and *nameSpaceVersion* – to record the information defined in the PREMIS semantic unit *objectIdentifierType*.

Level of Detail: A repository is free to record more detailed information for a PREMIS semantic unit than what is defined in the Data Dictionary (although the information defined in the Data Dictionary should be a subset of the more detailed information recorded by the repository).
Example:
In implementing the PREMIS semantic unit *agentName*, a repository records not only the name of the Agent, but also their title and department.

Note: A repository can also extend PREMIS semantic units through the use of formal metadata schema external to the Data Dictionary; see discussion of "Extensibility" in the Data Dictionary.

Explicit Recording of Information: A repository is not required to explicitly record in its metadata management system the information populating a particular PREMIS semantic unit that it has

implemented. However, this information must be recoverable in some way when it is needed (for example, to create an information package for exchange with another repository).
Example:
A repository implements the PREMIS semantic unit *preservationLevel*. The information pertaining to this semantic unit is contained within the general policy governing the operation of the repository, which is understood to apply to all digital objects residing in the repository. Therefore, the repository does not explicitly record this information in a metadata element associated with each archived Object. However, in preparing information packages for a set of Objects it is transferring to another institution, the repository creates metadata elements for each Object that records the information defined in the *preservationLevel* semantic unit and bundles this information with the Objects prior to transfer.

Note: 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. Many PREMIS implementations of which the Editorial Committee is aware use the METS (Metadata Encoding and Transmission Standard) as a container format to support their implementations. In view of this, the Editorial Committee produced a set of rules 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.

Use of Controlled Vocabularies: A repository is free to use (or not use) controlled vocabularies to populate PREMIS semantic units. If the repository chooses to use controlled vocabularies, it is free to use either internally defined vocabularies, or externally-defined, standardized vocabularies.
Example:
A repository implements the PREMIS semantic unit *eventType*. In creating values for this semantic unit, the repository can adopt a policy of "free text" for this semantic unit; use a locally maintained list of event type values; or use the list of event type values maintained by the Library of Congress (see http://id.loc.gov/).

Note: The PREMIS Editorial Committee strongly recommends use of controlled vocabularies where possible. See the note under Principles of Use (Semantic Unit) in Section II above.

IV. Conclusion
The PREMIS Data Dictionary was designed to maximize its applicability across a wide range of digital preservation contexts. As a consequence, the required conditions it imposes in regard to conformant implementation are lightweight, and considerable scope for flexibility and choice is reserved for implementing repositories. While this approach is suitable for a general conformance policy, it may be the case that certain specific digital preservation contexts or use cases may benefit from adherence to a stricter conformance profile adapted to the particular circumstances in which it is expected to operate. In these cases, the context-specific conformance profile can establish *additional* conditions for conformance; however, it *cannot delete or modify* the core conformance conditions that are set forth in the general PREMIS conformance statement.