# PREMIS Data Dictionary for Preservation Metadata

*Sarah Higgins – Digital Curation Centre*

## Introduction

The Repositories Support Project *Briefing Paper on 'Metadata'*[1] examines how metadata standards can be used to make it easier to find, use and manage digital objects stored in an institutional repository. Long-term preservation of these objects can also be handled using metadata standards. This Web Advisory Document introduces *The PREMIS Data Dictionary for Preservation Metadata,* v2.0[2], the current authoritative metadata standard for digital preservation, and looks at how it can be used within an Institutional Repository.

## Using a repository for preservation

Institutional repositories, which aim to preserve their digital objects, will need managerial support to put in place policies and guidance. Adequate preservation ensures that the *Essential Characteristics* – the things which give the object meaning – survive both time and technical development. Preservation metadata can help to maintain these *Essential Characteristics*:

- Viability – an object is able to persist over the long-term;
- Renderability – an object can be correctly interpreted and displayed using computing equipment;
- Understandability – an object's purpose and context can be understood when rendered;
- Authenticity – an object has not been altered, adapted or substituted;
- Identity – an object can be clearly distinguished from other objects.

## PREMIS

PREMIS was specifically developed to help with long-term preservation. It builds on the OAIS Reference Model (ISO 14721)[3], which defines a framework for a successful repository. PREMIS is maintained by the Library of Congress[4] and consists of 3 parts:

1. **The PREMIS Data Model** defines 5 preservation activities or *Entities*, and how they relate to each other:

    - *Intellectual Entity* – the digital object or the parts which make up a complete digital object, *eg* the digitised pages of a book, or the complete set of files which make up a web page. An *Intellectual Entity* may have more than one *Representation* (see below) *eg* a scholarly work may be maintained in both PDF and Microsoft Word format.
    - *Objects* – a discrete digital information unit which can be described in 3 sub-types:

    - Bitstream – the bit set embedded in a file;
    - File – a named and ordered sequence of bytes known by an operating system;
    - Representation – the file set needed to render a complete *Intellectual Entity*.
- *Events* – an audit trail concerning changes made to a digital object throughout its lifecycle. Examples of *Events* are changes in custodianship, format migrations, and the creation of new relationships between objects.
- *Agents* – persons, organisations or software responsible for preservation *Events* throughout a digital object's lifecycle.
- *Rights* – rights and permissions statements for both the digital objects and their *Agents*.

2. **The PREMIS Data Dictionary** defines elements and sub-elements (called *Semantic Units* and *Semantic Components*) to describe all of the *Entities*. The exception is the *Intellectual Entity* (the actual digital object). A repository can continue to describe their digital objects using their usual metadata standard. PREMIS includes well over 100 elements, but only a selection are mandatory.

3. **The PREMIS schema, version 2.0**[5], allows the *Entities* and their *Semantic Units* to be expressed consistently in XML.

## Using PREMIS in a repository

An established repository will have already decided which metadata standards, and which elements of these to use, to make it easier to find, use and manage digital objects, on a day-to-day basis.

PREMIS should be used in addition to existing metadata to manage preservation. A repository would need to establish which of the elements included in the *Data Dictionary* are essential to ensure adequate long-term preservation of its digital objects. There may be some cross-over between existing metadata creation and the needs of PREMIS, reducing the amount of new metadata collection required for compliancy. Mandatory PREMIS elements are already frequently collected by a repository *eg*: *objectIdentifier* is a unique identifier; *objectCategory* is the type of object being preserved. Information already collected about those who manage the digital objects, and the rights associated with them, may be sufficient to satisfy the *Agents* and *Rights Entities* of PREMIS. PREMIS assumes auto generation of most of its elements. The Preserv Project is looking at ways external service providers can help in this process[6].

Many PREMIS elements mandate the use of controlled vocabularies. Few are actually specified, so relevant existing ones may need to be identified, or new ones defined. Guidance on data content and structure will also need to be developed, as this is not included in the standard.

Help is available through supporting documentation[7] and the PREMIS Implementors' Group (PIG)[8]. The latter share experiences and feed them into the ongoing revision process. Their website includes: a wiki to share documents (the *pigpen*); an implementation registry; and a listserv.

## Conclusion

PREMIS is currently the best metadata standard for managing preservation within a repository. Using PREMIS, in conjunction with appropriate metadata for description and access, can help to ensure that digital objects remain available for the future. Implementing PREMIS can be a daunting task, but there is good documentation, an XML schema and peer support available. Furthermore, many repositories may find that they are already collecting information needed by PREMIS, making an implementation easier to achieve.

## References & further information

1 Repositories Support Project Briefing Paper on 'Metadata' (April 2008)
http://www.rsp.ac.uk/pubs/briefingpapers-docs/repoadmin-metadata.pdf

2 PREMIS Data Dictionary for Preservation Metadata, version 2.0 (March 2008)
http://www.loc.gov/standards/premis/v2/premis-2-0.pdf

3 ISO 14721:2003 Space data and information transfer systems — Open archival information system — Reference model. A publicly available copy with identical text is available at:
http://public.ccsds.org/publications/archive/650x0b1.pdf

4 PREMIS: Preservation metadata maintenance activity
http://www.loc.gov/standards/premis/

5 PREMIS: Preservation metadata schema, version 2.0 (March 2008)
http://www.loc.gov/standards/premis/premis.xsd

6 Hitchcock, S et al., Preservation metadata for institutional repositories: applying PREMIS (draft) (2007)
http://preserv.eprints.org/papers/presmeta/presmeta-paper.html

7 Introduction and supporting materials from PREMIS data dictionary for preservation metadata, version 2.0 (March 2008)
http://www.loc.gov/standards/premis/v2/premis-report-2-0.pdf

8 PREMIS Implementors' Group
http://www.loc.gov/standards/premis/pig.html

| | |
|---|---|
| ***Author:*** | Sarah Higgins |
| ***Title:*** | PREMIS Data Dictionary for Preservation Metadata |
| ***Version:*** | 2.1 |
| ***Date:*** | March 2009 |