

DATA DICTIONARY SECTION *from*

PREMIS Data Dictionary for Preservation Metadata

version 2.2

This is an excerpt from the PREMIS version 2.2 document. It includes only the Data Dictionary section. The Introduction, Special Topics, Methodology, and Glossary are in a separate excerpt. The full document and both excerpts are available online from: <http://www.loc.gov/standards/premis/>

PREMIS Editorial Committee

July 2012

THE PREMIS DATA DICTIONARY VERSION 2.2

The PREMIS Data Dictionary includes semantic units for Objects, Events, Agents, and Rights. The fifth entity in the model, the Intellectual Entity, is considered out of scope because it is well served by descriptive metadata. The template for each entry includes a place for notes about how to create or use the semantic unit. In some cases the group felt additional information, such as the reason for a semantic unit's definition or issues that arose in the group's deliberations, would be useful; for these details, see "[Special Topics](#)" page [204](#).

A semantic component always inherits the applicability of the containing semantic unit. That is, if the containing semantic unit specifies that it is applicable to files but not to representations, each of its semantic components is applicable to files and not to representations. Repeatability and obligation, however, may vary.

Each entry in the Data Dictionary offers these attributes of a semantic unit:

- **Name of the semantic unit:** Names were devised to be descriptive and unique within the Data Dictionary. Using these names for the exchange of metadata among preservation repositories will aid interoperability. These names need not be used internally within any individual preservation repository.
- **Semantic components:** The semantic components each have their own entries later in the Data Dictionary. A semantic unit that has semantic components does not have any value of its own. Only semantic units at the lowest level have values.
- **Definition:** The meaning of the semantic unit.
- **Rationale:** Why the semantic unit is needed, if this is not self-evident from the definition.
- **Data constraint:** How the value of the semantic unit should be encoded. Some common data constraints are:

Container – The semantic unit is an umbrella for two or more semantic components and has no value of its own.

None – The semantic unit can take any form of value.

Value should be taken from a controlled vocabulary – The preservation repository should establish an authority list of values that are useful and meaningful to the repository. The PREMIS Data Dictionary does not specify what this authority list should be, and it is assumed that different repositories will use different vocabularies. In general, when a value is taken from a controlled vocabulary, the source of the vocabulary should be recorded. A mechanism to record the source is provided in the PREMIS XML schemas.

- **Object category:** Whether the unit applies to a representation, file, or bitstream Object. Semantic units that apply to files also apply to filestreams (see page [7](#)).
- **Applicability:** A scope of "applicable" means it applies to that category of Object.

Object Entity

The Object entity aggregates information about a digital object held by a preservation repository and describes those characteristics relevant to preservation management.

The only mandatory semantic unit that applies to all categories of object (representation, file, and bitstream) is *objectIdentifier*.

Entity types

- **Representation:** A digital object instantiating or embodying an Intellectual Entity. A representation is the set of stored digital files and structural metadata needed to provide a complete and reasonable rendition of the Intellectual Entity.
- **File:** A named and ordered sequence of bytes that is known to an operating system.
- **Bitstream:** Contiguous or non-contiguous data within a file that has meaningful properties for preservation purposes.

Entity properties

- Can be associated with one or more rights statements.
- Can participate in one or more events.
- Links between entities may be recorded from either direction and need not be bi-directional.

Entity semantic units

- 1.1 objectIdentifier (M, R)
 - 1.1.1 objectIdentifierType (M, NR)
 - 1.1.2 objectIdentifierValue (M, NR)
- 1.2 objectCategory (M, NR)
- 1.3 preservationLevel (O, R) [representation, file]
 - 1.3.1 preservationLevelValue (M, NR) [representation, file]
 - 1.3.2 preservationLevelRole (O, NR) [representation, file]
 - 1.3.3 preservationLevelRationale (O, R) [representation, file]
 - 1.3.4 preservationLevelDateAssigned (O, NR) [representation, file]
- 1.4 significantProperties (O, R)
 - 1.4.1 significantPropertiesType (O, NR)
 - 1.4.2 significantPropertiesValue (O, NR)
 - 1.4.3 significantPropertiesExtension (O, R)
- 1.5 objectCharacteristics (M, R) [file, bitstream]
 - 1.5.1 compositionLevel (M, NR) [file, bitstream]

- 1.8.5.2 swVersion (O, NR)
- 1.8.5.3 swType (M, NR)
- 1.8.5.4 swOtherInformation (O, R)
- 1.8.5.5 swDependency (O, R)
- 1.8.6 hardware (O, R)
 - 1.8.6.1 hwName (M, NR)
 - 1.8.6.2 hwType (M, NR)
 - 1.8.6.3 hwOtherInformation (O, R)
- 1.8.7 environmentExtension (O, R)
- 1.9 signatureInformation (O, R) [file, bitstream]
 - 1.9.1 signature (O, R)
 - 1.9.1.1 signatureEncoding (M, NR) [file, bitstream]
 - 1.9.1.2 signer (O, NR) [file, bitstream]
 - 1.9.1.3 signatureMethod (M, NR) [file, bitstream]
 - 1.9.1.4 signatureValue (M, NR) [file, bitstream]
 - 1.9.1.5 signatureValidationRules (M, NR) [file, bitstream]
 - 1.9.1.6 signatureProperties (O, R) [file, bitstream]
 - 1.9.1.7 keyInformation (O, NR) [file, bitstream]
 - 1.9.2 signatureInformationExtension (O, R) [file, bitstream]
- 1.10 relationship (O, R)
 - 1.10.1 relationshipType (M, NR)
 - 1.10.2 relationshipSubType (M, NR)
 - 1.10.3 relatedObjectIdentification (M, R)
 - 1.10.3.1 relatedObjectIdentifierType (M, NR)
 - 1.10.3.2 relatedObjectIdentifierValue (M, NR)
 - 1.10.3.3 relatedObjectSequence (O, NR)
 - 1.10.4 relatedEventIdentification (O, R)
 - 1.10.4.1 relatedEventIdentifierType (M, NR)
 - 1.10.4.2 relatedEventIdentifierValue (M, NR)
 - 1.10.4.3 relatedEventSequence (O, NR)
- 1.11 linkingEventIdentifier (O, R)
 - 1.11.1 linkingEventIdentifierType (M, NR)
 - 1.11.2 linkingEventIdentifierValue (M, NR)
- 1.12 linkingIntellectualEntityIdentifier (O, R)
 - 1.12.1 linkingIntellectualEntityIdentifierType (M, NR)
 - 1.12.2 linkingIntellectualEntityIdentifierValue (M, NR)
- 1.13 linkingRightsStatementIdentifier (O, R)
 - 1.13.1 linkingRightsStatementIdentifierType (M, NR)
 - 1.13.2 linkingRightsStatementIdentifierValue (M, NR)

| | |
|--|----------------------------------------------------------------------------------------------------------------------------------------------|
| | <p>recorded.</p> <p>A persistent identifier should be used, but the particular identifier scheme is an implementation specific decision.</p> |
|--|----------------------------------------------------------------------------------------------------------------------------------------------|

THE PREMIS DATA DICTIONARY

| | | | |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|---------------------------|
| Semantic unit | 1.1.1 objectIdentifierType | | |
| Semantic components | None | | |
| Definition | A designation of the domain within which the object identifier is unique. | | |
| Rationale | Identifier values cannot be assumed to be unique across domains; the combination of <i>objectIdentifierType</i> and <i>objectIdentifierValue</i> should ensure uniqueness. | | |
| Data constraint | Value should be taken from a controlled vocabulary. | | |
| Object category | Representation | File | Bitstream |
| Applicability | Applicable | Applicable | Applicable |
| Examples | DLC DRS hdl:4263537 | DLC DRS hdl:4263537 | DLC DRS hdl:4263537 |
| Repeatability | Not repeatable | Not repeatable | Not repeatable |
| Obligation | Mandatory | Mandatory | Mandatory |
| Usage notes | The type of the identifier may be implicit within the repository as long it is can be explicitly communicated when the digital object is disseminated outside of it. | | |