

PREMIS and METS

Olaf Brandt, SUB Göttingen

brandt@sub.uni-goettingen.de

Markus Enders, SUB Göttingen

enders@sub.uni-goettingen.de

Outline

Preservation Metadata

PREMIS

Integration of PREMIS in METS

Preservation Metadata

What are preservation metadata?

Information likely needed to support digital preservation

Digital Objects

their contexts

and their relationships

PREMIS:

- International Effort
- Initially sponsored by RLG and OCLC
- Crossdomain development (libraries, archives, museums, private companies)
- PREMIS developed Metadata model and a Core Set of Preservation Metadata
- provides PREMIS XML-Schemas (serialized version of abstract model)
- development and support institutionalized in PREMIS Maintenance Activity

PREMIS Data Dictionary contains:

- Common data model for organizing preservation metadata (abstract model)
- Core Set of preservation Metadata
- Guidance for local implementations

PREMIS: Entities

- Intellectual Entity
- Object Entity
- Event Entity
- Agent Entity
- Rights Entity

PREMIS: Intellectual Entity

- Set of content that is considered a single intellectual unit for purposes of management and description (e.g., a book, a photograph, a map, a database)
- May include other Intellectual Entities (e.g. a website that includes a web page)
- Has one or more digital representations
- Not fully described in PREMIS DD, but can be linked to in metadata describing digital representation

PREMIS: Object

- Discrete unit of information in digital form
- ****Objects are what repository actually preserves****
- Three types of Object:
 - **FILE:** named and ordered sequence of bytes that is known by an operating system
 - **REPRESENTATION:** set of files, including structural metadata, that, taken together, constitute a complete rendering of an Intellectual Entity
 - **BITSTREAM:** data within a file with properties relevant for preservation purposes (but needs additional structure or reformatting to be stand-alone file)

Examples:

- chapter1.pdf (a file)
- chapter1.pdf + chapter2.pdf + chapter3.pdf (representation of a book with three chapters)
- TIFF file containing header and 2 images (2 bitstreams (images), each with own set of properties (semantic units): e.g., identifiers, technical metadata, inhibitors, ...)

PREMIS: Event

- An action that involves or impacts at least one Object or Agent associated with or known by the preservation repository
- Helps document digital provenance. Can track history of Object through the chain of Events that occur during the Objects lifecycle
- Determining which Events are in scope is up to the repository (e.g., Events which occur before ingest, or after de-accession)
- Determining which Events should be recorded, and at what level of granularity is up to the repository

Examples:

- Validation Event: use JHOVE tool to verify that chapter1.pdf is a valid PDF file
- Ingest Event: transform an OAIS SIP into an AIP (may be one Event or multiple Events)
- Migration Event: create a new version of an Object in an up-to-date format

PREMIS: Agent

- Person, organization, or software program/system associated with an Event or a Right (permission statement)
- Agents are associated only indirectly to Objects through Events or Rights
- Not defined in detail in PREMIS DD; not considered core preservation metadata beyond identification

Examples:

- Markus Enders (a person)
- Göttingen State and University Library (an organization)
- JHOVE version 1.0 (a software program)

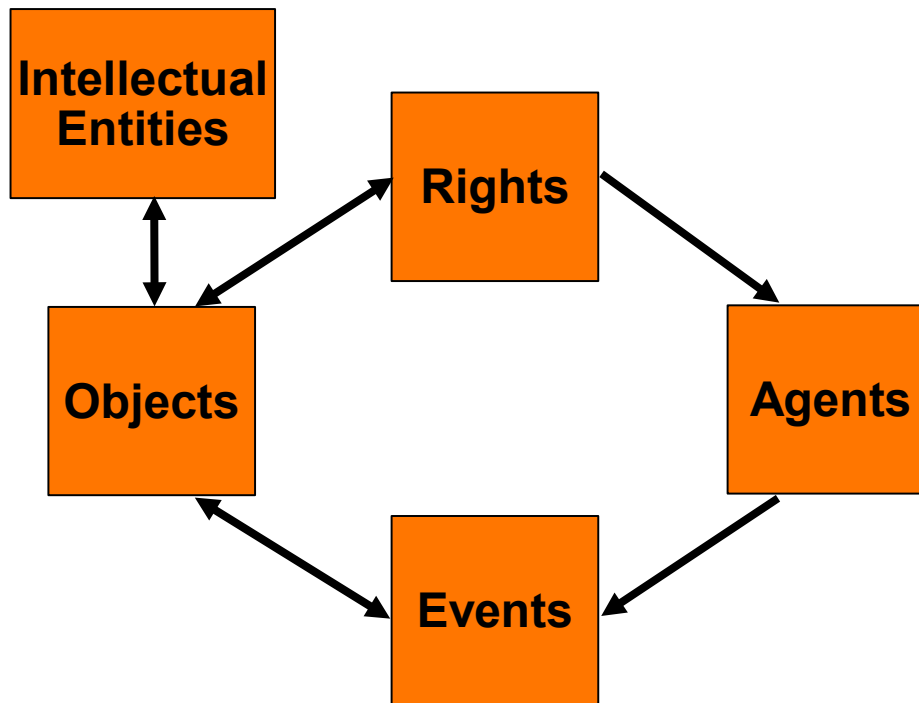
PREMIS: Rights

- An agreement with a rights holder that grants permission for the repository to undertake an action(s) associated with an Object(s) in the repository.
- Not a full rights expression language; focuses on permissions relevant for preservation.

Example:

- Priscilla Caplan grants FCLA digital archive permission to copy and migrate the Object "metadata_fundamentals.pdf" for preservation purposes.

PREMIS data model:

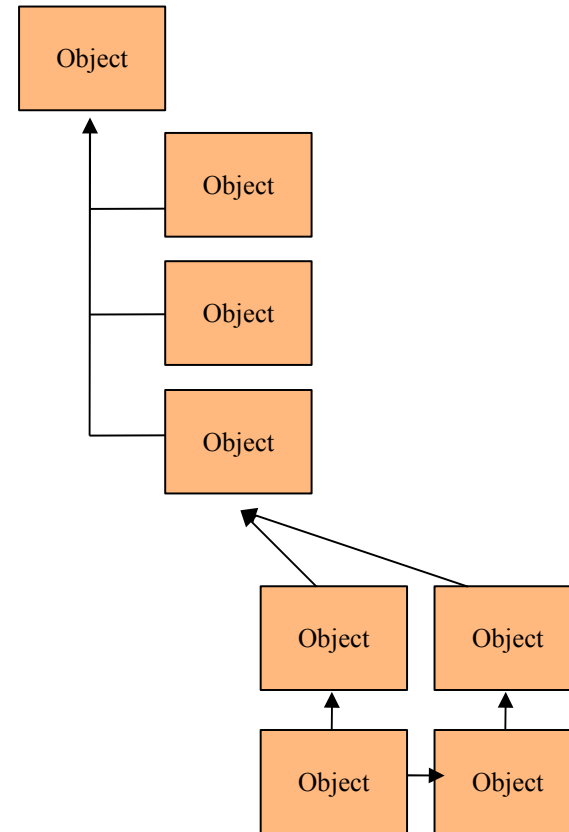


PREMIS data model: Object relationships

Relationships link objects together using unique identifiers.

Relationships can be qualified:

- structural, derivation
- isParentOf, isPartOf, hasSibling, sourceOf



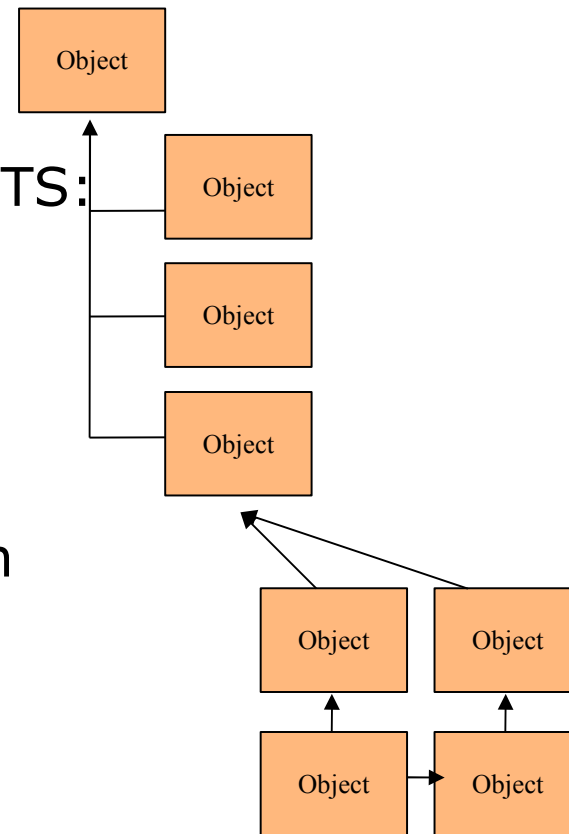
PREMIS and METS

Relationships: METS or PREMIS?

Hierarchical relationships in METS:
nested `<div>` elements

```
<div>  
  <div></div>  
</div>
```

`<mets:smlink>` to link between
`<div>` elements using XMLIDs.



PREMIS and METS

Relationships: METS or PREMIS?

PREMIS allows relationships between all kind of objects.

METS can only link between `<div>` elements.

Event- and Agent-information can be stored with a relationship in PREMIS

METS only knows xlink attributes

PREMIS can link to objects outside of current xml-file

METS can only link to `<div>` elements within the same XML-file

PREMIS and METS

Relationships: METS or PREMIS?

possible solution:

Store nested <div> elements in METS where needed and possible

Store other relationships in PREMIS extension schema
(e.g. relationships between versions or Objects and Events...)

What's the object?

<mets:div> element:

- divisions can represent monograph, movie, chapters,
- can be nested

premis:object of type „representation“

<mets:file> element:

describes a file including a “pointer” to the content
content can be embedded in METS

premis:object of type “file”

PREMIS and METS

What's the object?

<mets:stream> element:

- Bitstream embedded in a file
- Can be several streams per file

premis:object of type „bitstream“

All elements (div, file, stream) may have one or more Metadata sections attached to them (amdSec, dmdSec)

--> Use PREMIS as an extension schema to METS

PREMIS and METS

Which metadata section?

METS offers separate sections for different kind of metadata: dmdSec, amdSec (with subsections for technical, rights, digital provenance and source metadata).

Where to store PREMIS metadata?

Split PREMIS data or store them as a single unit?

There might be advantages splitting data for maintenance and reuse purposes.

```
<mets>  
  <dmdSec/>  
  <amdSec/>  
    <techMD/>  
    <rightsMD/>  
    <sourceMD/>  
    <digiProvMD/>  
  <fileSec/>  
  <structMap/>  
</mets>
```

Which metadata section?

Common up to now:

- one amdSec for a combination of PREMIS Object, Agent, Events and Rights Entity
 - use single <digiProvMD> for a complete PREMIS data set *or*
 - use one <amdSec> with <techMD> for a single PREMIS Object and one <digiProvMD> for PREMIS events and agents.
- PREMIS rights not widely used in METS up to now record them in digiprovmD or in rightsMD?

PREMIS and METS

Redundant metadata: METS and extension schemas

METS stores some metadata in attributes:

```
< mets:file MIMETYPE="image/tiff" CHECKSUMTYPE="SHA-1"  
CHECKSUM="7c9b35da4f2ebd436f..." />
```

Checksum Type
and Value



PREMIS contains similar information:

```
<premis:fixity>  
  <premis:messageDigestAlgorithm>SHA-1  
  </premis:messageDigestAlgorithm>  
  <premis:messageDigest>7c9b35da4f2ebd436f...  
  </premis:messageDigest>  
</premis:fixity>
```

PREMIS and METS

Redundant metadata: METS and extension schemas

METS stores some metadata in attributes:

```
<mets:file MIMETYPE="image/tiff" CHECKSUMTYPE="SHA-1" CHECKSUM="7c9b35da4f2ebd436f..." />
```

PREMIS contains similar information:

```
<premis:fixity>  
  <premis:messageDigestAlgorithm>SHA-1  
  </premis:messageDigestAlgorithm>  
  <premis:messageDigest>7c9b35da4f2ebd436f...  
  </premis:messageDigest>  
</premis:fixity>
```

Checksum Type
and Value



Redundant metadata in extension schemas

As METS may contain several metadata schema for different purposes, their metadata elements may overlap.

Elements may have the same semantic meaning

Some metadata schema might store additional / more detailed information.

PREMIS and METS

Redundant metadata:

PREMIS:

format independent
metadata
(for general semantic
units)

MIX:

format dependent
metadata
(for format specific
semantic units)

But there is some overlap between PREMIS and MIX
e.g. in File Type and Version or fixity information

METS and PREMIS

Redundant metadata:

MIX was recently revised; names harmonized with corresponding PREMIS semantic units

METS and PREMIS

Thank you!

METS & PREMIS

PREMIS Maintenance Activity

<http://www.loc.gov/standards/premis/>

Data Dictionary for Preservation Metadata: Final
Report of the PREMIS Working Group

<http://www.oclc.org/research/projects/pmwg/premis-f>

METS

<http://www.loc.gov/standards/mets/>