PREMIS PREservation Metadata Implementation Strategies

UNDERSTANDING AND IMPLEMENTING PREMIS
A 90 minute introduction

Karin Bredenberg
Kommunalförbundet Sydarkivera

Eld Zierau
Royal Danish Library

Micky Lindlar
TIB – Leibniz Information Centre for Science and Technology

Purpose of the Tutorial

- Introduce PREMIS
- Set a baseline of common understanding
- Introduce you to the PREMIS community

Understanding PREMIS – Entender PREMIS

ENTENDER PREMIS

Agencia de Desarrollo de Redes y Normas MARC de la Biblioteca del Congreso de EE. UU.

Autora: Priscilla Caplan

Revisión: Comité Editorial de PREMIS

Publicación original: 2009; revisión: 2017; nueva revisión: 2021

https://www.loc.gov/standards/premis/understandingPREMIS_spanish_2021.pdf

Who are we?

From the PREMIS Editorial Committee:

- Karin
- Eld
- Micky

We know that this might be true today:

- You have just started to know PREMIS
- The PREMIS data model is a bit scary
- You might have been using PREMIS in practice
- You come from
  - a library
  - an archive
  - a university
  - or something different

What do we want you to think about?

- Think about your use case!
  - I want to use PREMIS …???

- Example: I'm digitising a number of publications made at my institution and I need to preserve these for the future and making sure they are accessible 50 years from now.

Agenda

Introduction to PREMIS

Welcome
Background (brief history and rationale)
On-line presence
Benefits of implementing PREMIS

Implementation

Outline of the main entities and the data model
Your use cases

Implementation and case studies

Data Dictionary
PREMIS Conformance & interoperability

Wrap Up

Introduction to exercise (Objects, Events, Agents, Rights) at home
Answers to questions

PREMIS PREservation Metadata Implementation Strategies

Karin Bredenberg
Kommunalförbundet Sydarkivera

TODAY WE WONT TALK ABOUT SOME THINGS

Not today

- PREMIS in super detail
- PREMIS OWL in detail
- PREMIS in METS in detail
- Environments

PREMIS PREservation Metadata Implementation Strategies

Karin Bredenberg
Kommunalförbundet Sydarkivera

DIGITAL PRESERVATION METADATA - WHY IS IT NEEDED AND WHAT DOES IT LOOK LIKE?

What is digital preservation metadata?

- Digital preservation metadata =
  Metadata to ensure long-term accessibility of digital resources
- Digital objects must be self-descriptive
- Must be able to describe, manage and discover independently from the systems that were used to create them
  XML (machine and human readable)
- Often bundled with the content files in an information package

Domain

Born digital

Digitized

DP metadata supports preservation goals

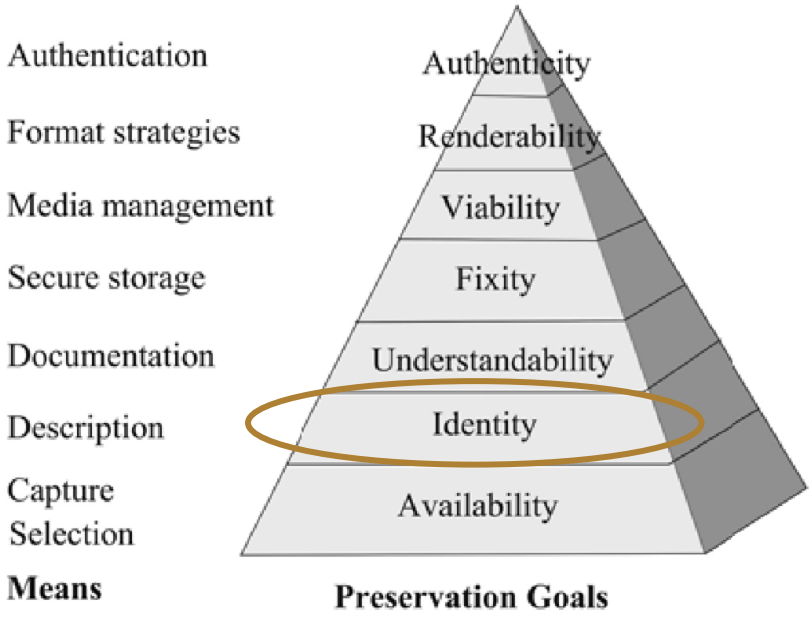
Preservation Pyramid
(from Priscilla Caplan)

Availability

- The object is in our control or in the control of a trusted, accessible repository

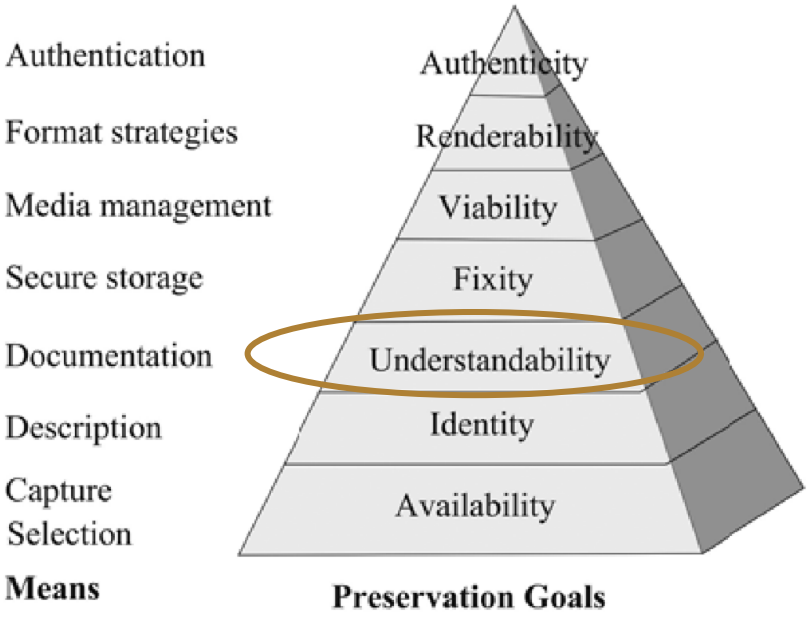
Identity

- Each relevant entity is persistently and uniquely identifiable
- The entities can be File, Work, Person, Organisation, Licence, …
- Metadata needed to be known
  - Identifier type
  - Identifier value



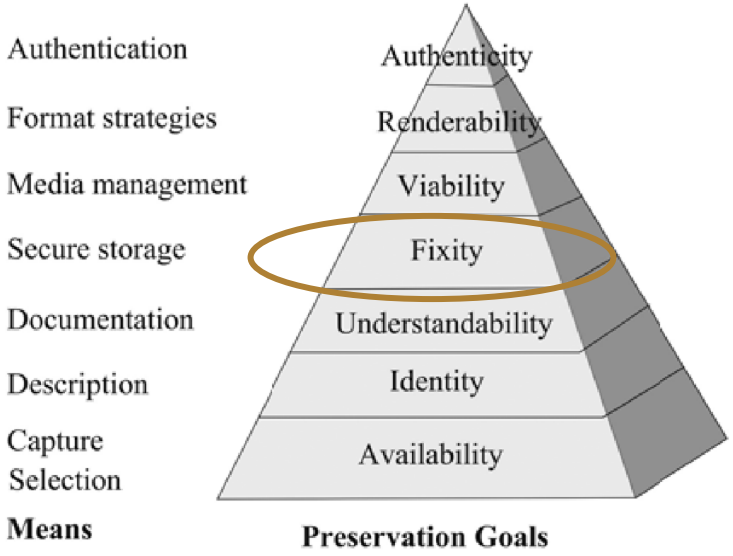
Understandability

- The object and entites is possible to understand
- Metadata needed to be known
  - Physical structure
    - Embedded files
    - File sequence
  - Logical structure
- Context needed to be known
  - Original source
  - Related items
- And much more



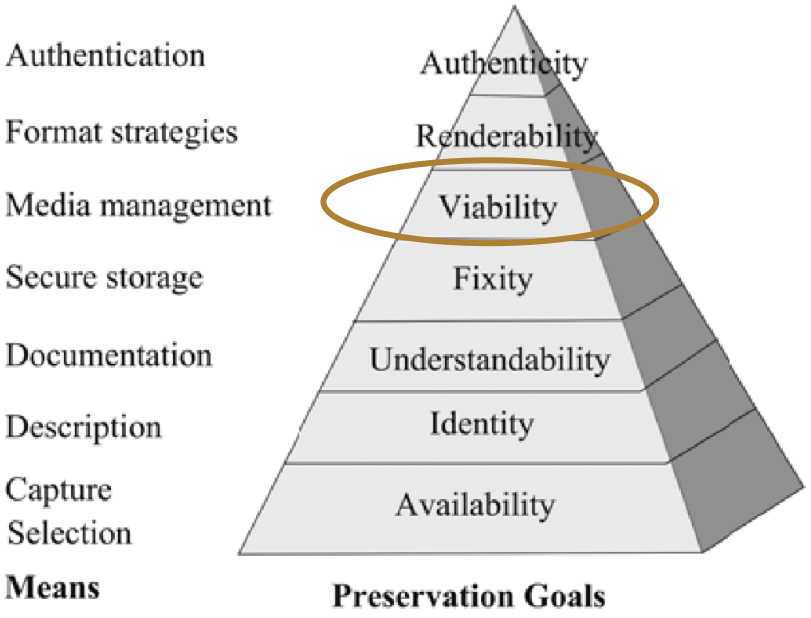
Fixity

- The object is unchanged
- Metadata needed to be known
  - Checksums
  - Message digests
  - Hash function
- Event creating them needs to give
  - Algorithms
  - Date/time
  - Originator
  - Type



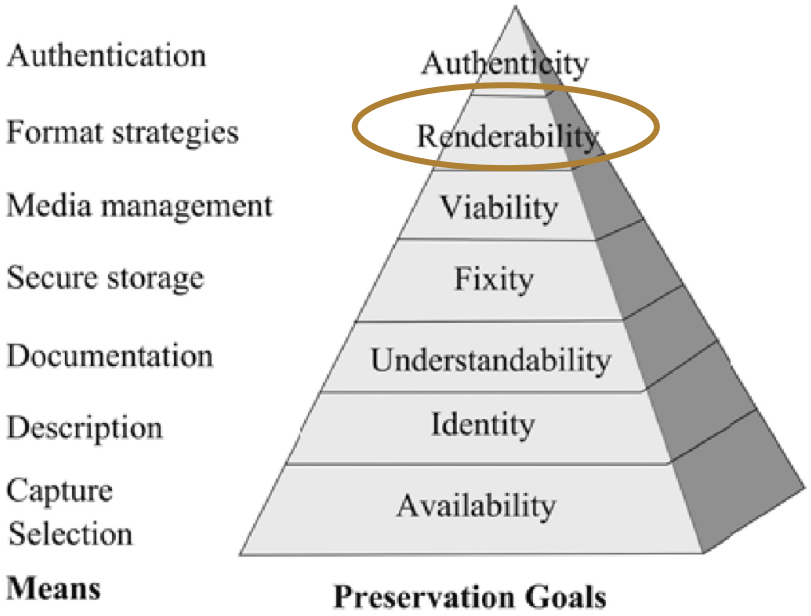
Viability

- The object is readable
- Metadata needed to be known
  - Data carrier metadata
    - Type of medium
    - Its preservation characteristics
    - Age of medium
    - Date of recording
    - Usage patterns



Renderability

- The object can be rendered or executed
- Metadata needed to be known
  - Format information
  - Rendering information
    - Software
    - Hardware
    - Other dependencies: schemas, style sheets, encodings, etc.
- Technology dependence



Authenticity

- Authenticity: The object is what it purports to be
- Integrity: The object is whole and unimpaired
- Object transformations
  - Pre-emptive preservation actions
    - Bit migration
    - Content migration
    - Replacing part of the rendering stack
  - Forensic transformation actions

Authenticity

- Metadata needed to be known
  - Structural metadata (for integrity)
  - Digital signatures
  - Access rights
  - Provenance metadata:
    - History of all actions performed on the resource (events and dates)
    - History of custodianship (agents: decision makers, administrators, tools)
  - Significant characteristics
  - Lost characteristics
  - Business rules (policy, strategy) guiding preservation actions

Rights

- Avoid Rights violations during preservation!!!
- Metadata needed to be known
  - Rights information for preservation actions during copyright / license period

PREMIS PREservation Metadata Implementation Strategies

Karin Bredenberg
Kommunalförbundet Sydarkivera

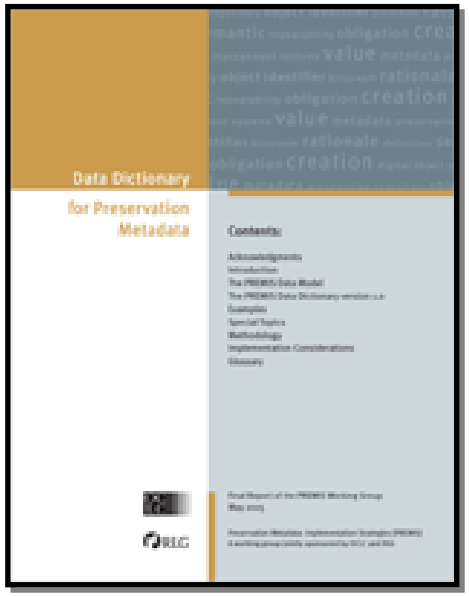
WHAT IS PREMIS?

The PREMIS standard

- International *de-facto* standard for metadata to support the preservation of digital objects and ensure their long-term usability.
  - Information you need to know for preserving digital objects

    Preservation Metadata: Implementation Strategies
- Developed by an international team of experts.
- Implemented in digital preservation projects around the world.
- Incorporated into commercial and open-source digital preservation tools and systems.

The PREMIS standard



- Data Dictionary (PREMIS 3.0)
  - http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf
  - Version 3 – major release
- XML schema v3.0
  - http://www.loc.gov/standards/premis/premis.xsd
- OWL ontology
- Supporting documentation

Activities

- The PREMIS Editorial Committee
  - Coordinates revisions and implementation of the standard
- PREMIS Implementors' Group forum (pig@loc.gov)
  - Email message to listserv@listserv.loc.gov:
    Text: subscribe pig <your name>
- Preservation metadata workshop (@iPres)

Scope

- What PREMIS DD is:
  - Common data model for organizing/thinking about preservation metadata
  - Standard for exchanging preservation metadata in information packages between repositories
  - Implementable
  - Technically neutral
  - Core metadata

Scope

- What PREMIS DD is not:
  - Out-of-the-box solution
  - All needed metadata
  - Lifecycle management of objects outside repository
    - increasing support for integration with outside
  - Rights management standard
    - strong support for rights statements

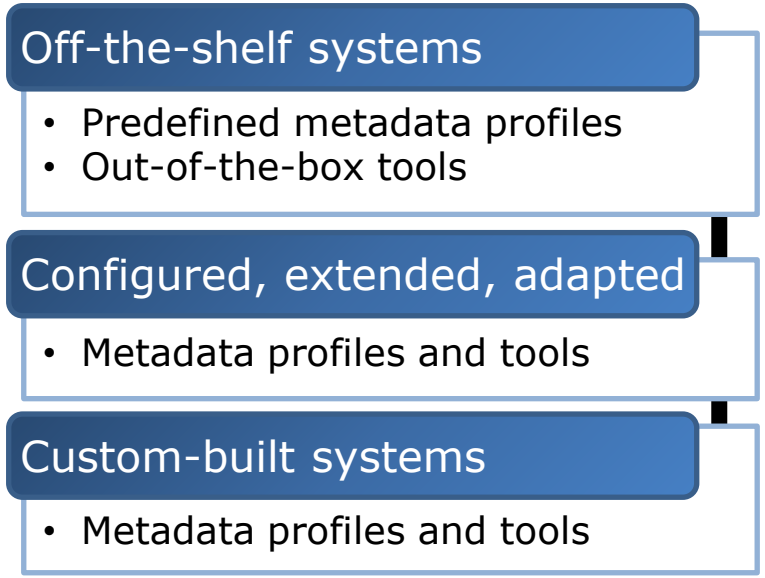
Scope

- What PREMIS DD is not:
  - It is not limited to or customized for archives and libraries.
  - It does not dictate that you need to use every feature.
    - But you should examine for yourself which features you can knowingly ignore.
  - It is not only useful if you implement metadata. You can use it to assess the metadata quality of systems you use.
  - Everyone modeling the digital landscape can and should use the high-level modeling feature.

Tailoring PREMIS to needs

- Evolving metadata
  - Increasing experience ensuring the longevity of digital objects
  - Changing future technical possibilities
  - Changing future legal framework

- Tailoring solutions
  - Varying needs
    - Content-types
    - Institutional policies
    - Intended use
  - Off-the-shelf (OS/commercial) or custom-built



PREMIS PREservation Metadata Implementation Strategies

Karin Bredenberg
Kommunalförbundet Sydarkivera

ON-LINE RESOURCES

Webpages

- Webpage
  - https://www.loc.gov/standards/premis/
- Vocabularies
  - https://id.loc.gov/vocabulary/preservation.html
- Ontology
  - https://id.loc.gov/ontologies/premis.html
- Tools at COPTR
  - https://coptr.digipres.org/index.php/PREMIS_(Preservation_Metadata_Implementation_Strategies)
- Zenodo
  - https://zenodo.org/communities/premis

E-mail list

- PREMIS Implementors' Group forum (pig@listserv.loc.gov)
  - Subscribe through the form:

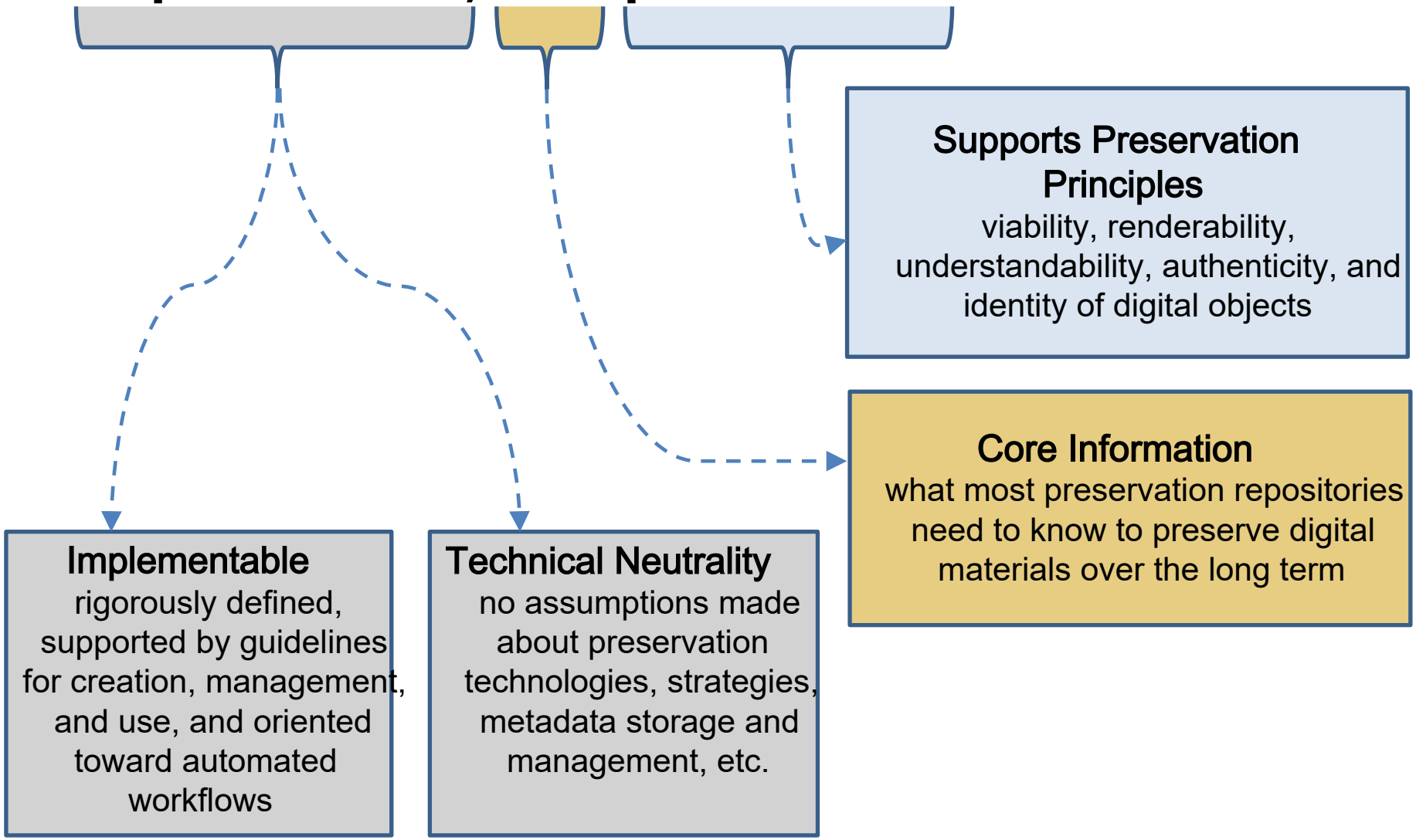
    https://listserv.loc.gov/cgi-bin/wa?A0=PIG

PREMIS PREservation Metadata Implementation Strategies

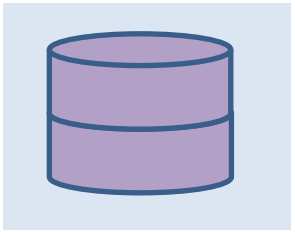
Micky Lindlar
TIB – Leibniz Information Centre for Science and Technology

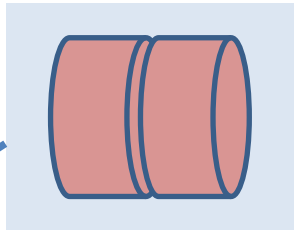
BENEFITS OF IMPLEMENTING PREMIS

Implementable, core preservation metadata



Benefits of implementing PREMIS (1)





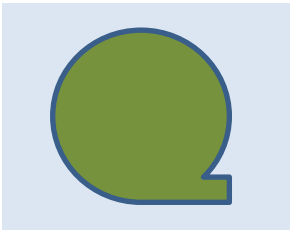
IDobj="hdl.galapagos.6754.1"

hdl="galapagos.6754.1"

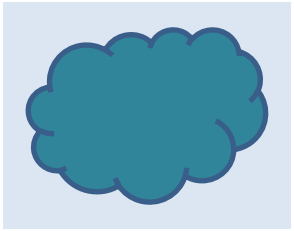
exchange
consistency

?

ID="6754.1"
GID="galapagos"



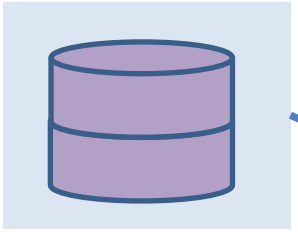
ID="1"
TYPE="hdl"
GROUPID="galapagos.6754"

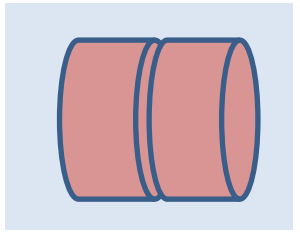


Interoperability through implementable preservation metadata

IDobj="hdl.galapagos.6754.1"

hdl="galapagos.6754.1"

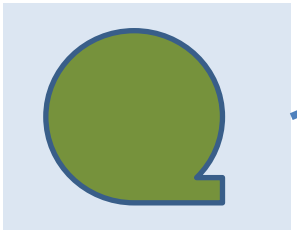




```
premis:objectIdentifier
premis:objectIdentifierType="hdl"
premis:objectIdentifierValue="galapagos.6754.1"
```

PREMIS

```
<galapagos.6754.1> a premisOwl:IntellectualEntity ;
premisOwl:identifier
<http://hdl.handle.net/galapagos.6754.1> .
<http://hdl.handle.net/galapagos.6754.1> a
http://id.loc.gov/vocabulary/identifiers/hdl .
```

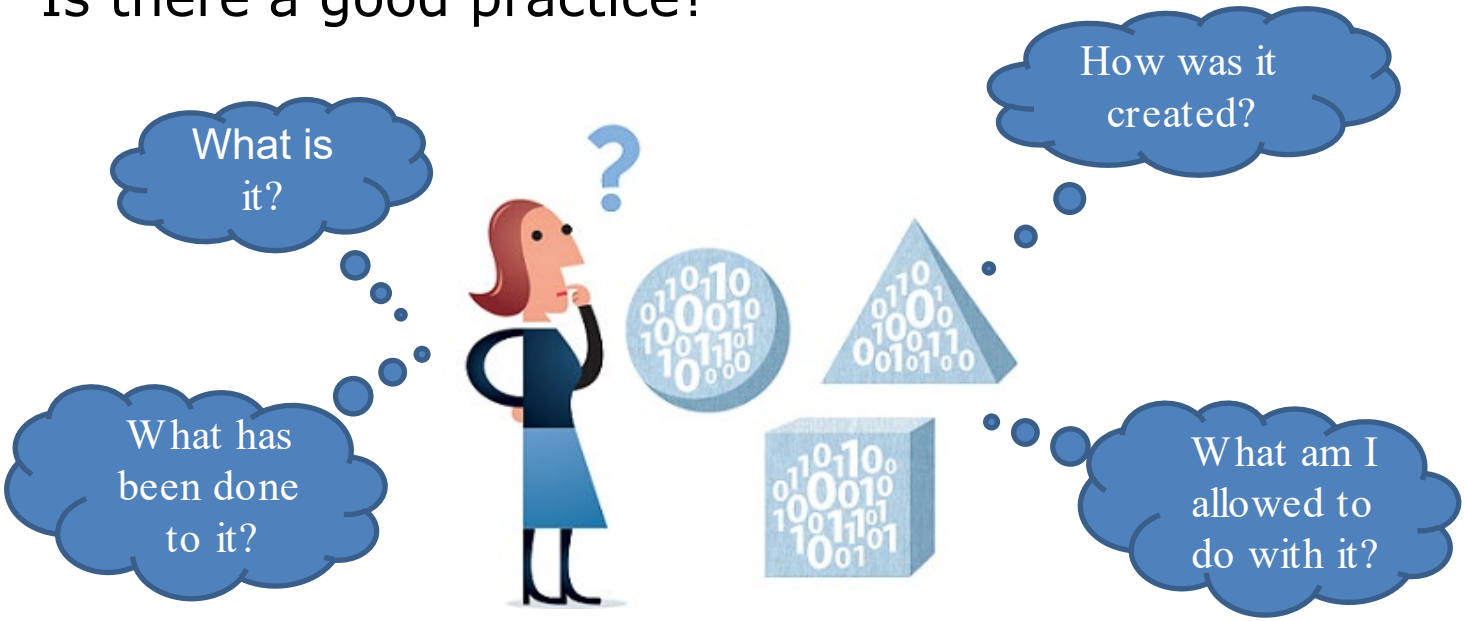


ID="1"
TYPE="hdl"
GROUPID="galapagos.6754"

ID="6754.1"
GID="galapagos"

Benefits of implementing PREMIS (2)

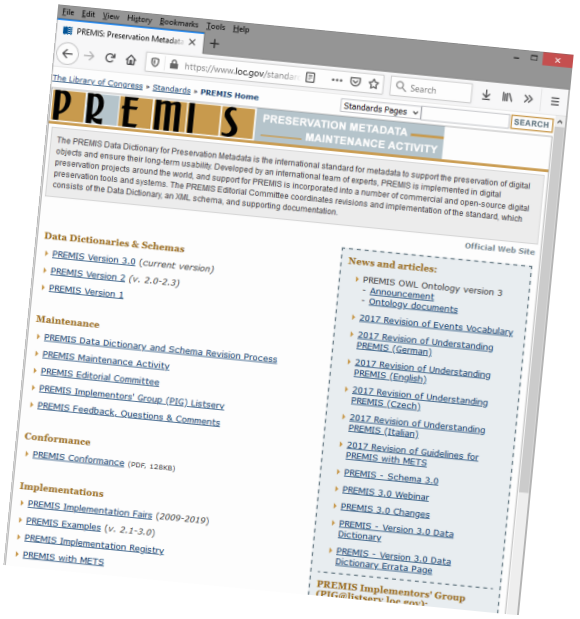
- What do I need to know about a digital object I am preserving?
  - … today
  - … in the future
- How can I best capture that information?
- Is there a good practice?



De-facto standard for preservation metadata

PREMIS = "things that most working preservation repositories are likely to need to know in order to support digital preservation"-- PREMIS Data Dictionary

- valuable resource to know what you need to capture
- a method to model the information you need
- a data model with MANY implementations in different institutions
- a community of users with good resources

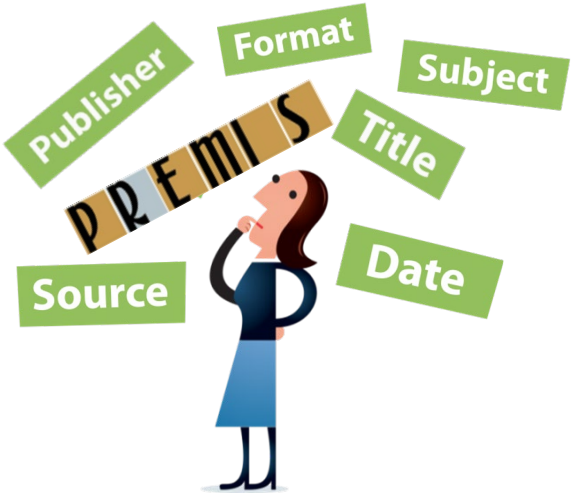


PREMIS PREservation Metadata Implementation Strategies

Eld Zierau
Royal Danish Library

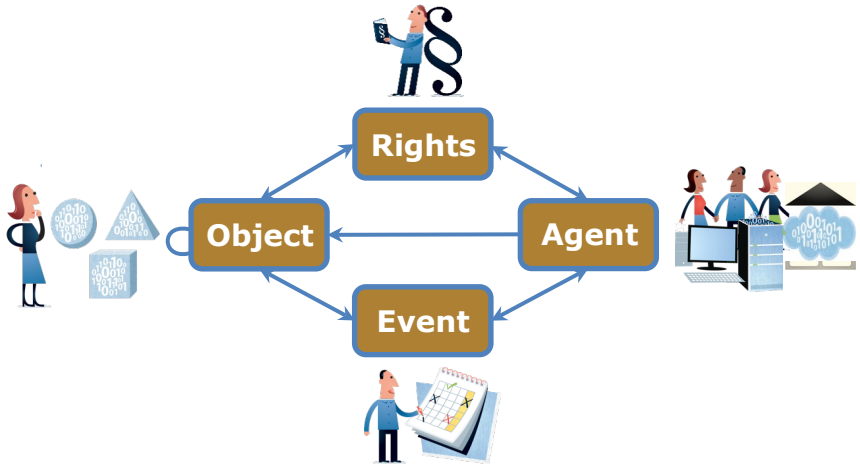
OUTLINE OF MAIN ENTITIES

THE DATA MODEL & KEY CONCEPTS



The PREMIS Data Model

The PREMIS Data Model is created by the community and experts in digital preservation



Why have a data model?

- Organizational convenience (for development and use)
- But: not a formal entity-relationship model; not sufficient to design databases

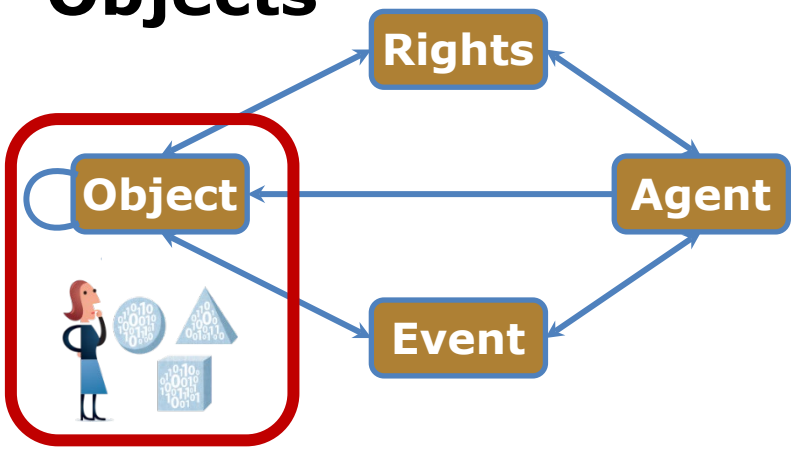
The PREMIS Data Model includes:

Rights
ing," relevant to
digital preservation

Semantic Units
detailed informat
of entities

Relationships
specifies relations
between entities

Objects



- It is the objects that you preserve
- Objects can be intellectual entities, representations or bitstreams (more later)

Implementation choices:

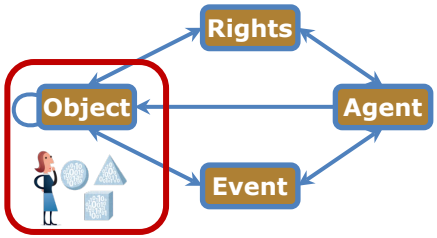
- Determine which types of Objects are in scope

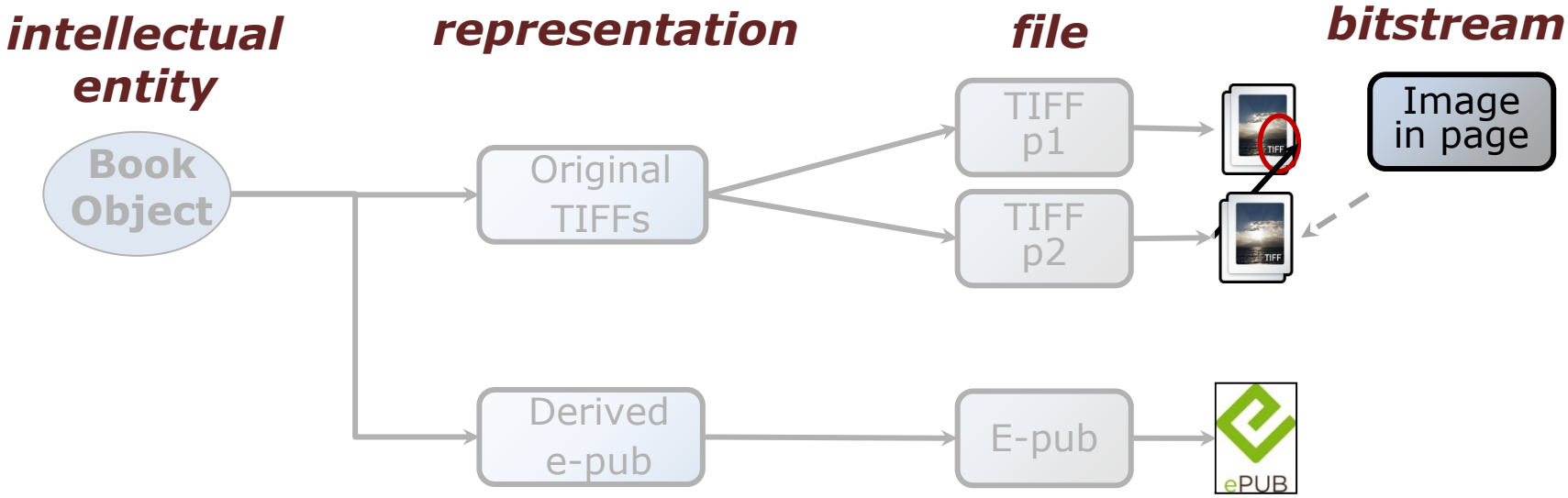
Examples:

- A PDF file
- An image within a file
- A book
- A book representation

Different types of objects

Types of objects (objectCategory)





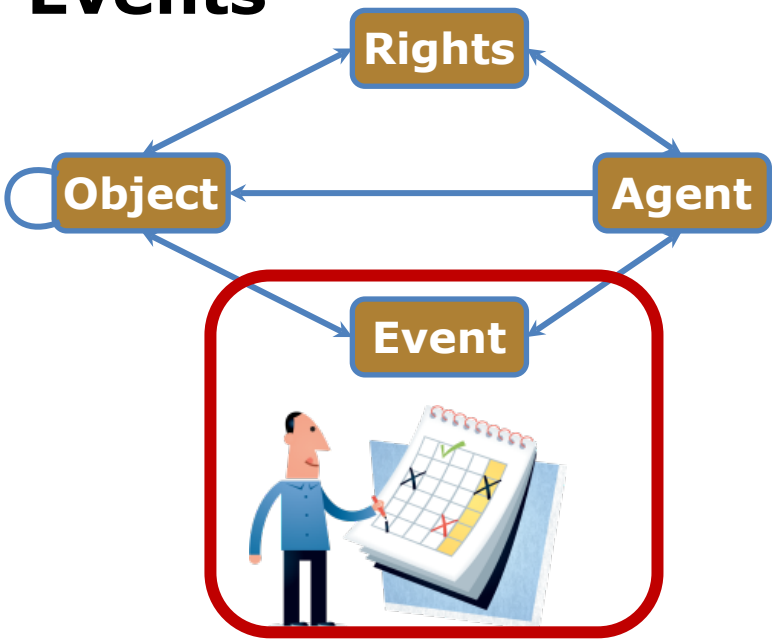
a distinct intellectual that is considered relevant in the context of digital preservation

set of objects, including structural metadata, that, taken together, constitute a complete representation

named and ordered sequence of bytes that is known by an operating system

data within a file with properties relevant for preservation purposes

Events



- An action that involves or impacts at least one Object or Agent
- Can document digital provenance, needed to track history of Object

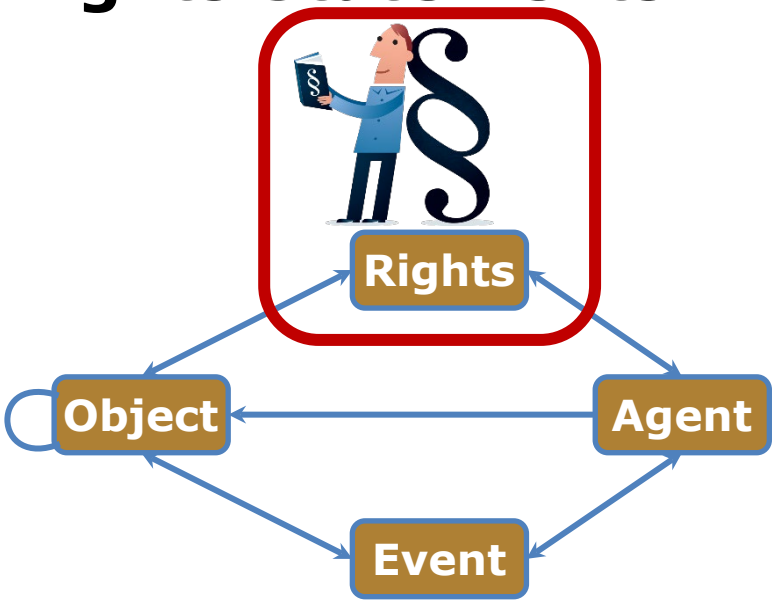
Examples:

- Validation Event e.g. using JHOVE
- Ingest Event

Implementation choices:

- Determining which Events are in scope
- Determining which Events should be recorded, as well as level of granularity

Rights Statements



- Rights to undertake an action(s) associated with an Object(s) in the repository.

Example:

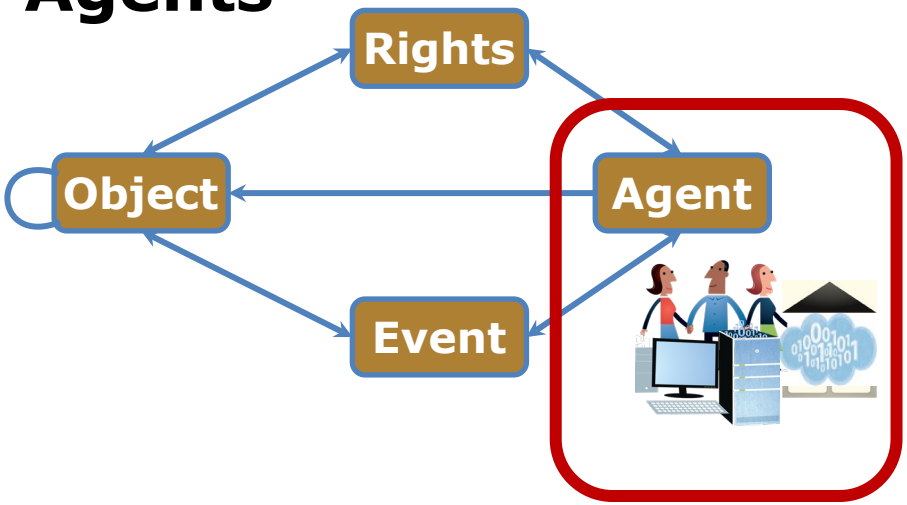
- **Helen Smith** grants **FCLA digital repository permission** to the repository in regard to make three copies of **metadata_derived.pdf** for preservation purposes.

Implementation choices :

- Can use other schemes

PREMIS form:
Agent X
grants **Permission Y**
to the repository in regard to Object **Z**.

Agents



- Person, organization, or software program/system
- Intended only to identify the agent, and to allow linking from other entity types.

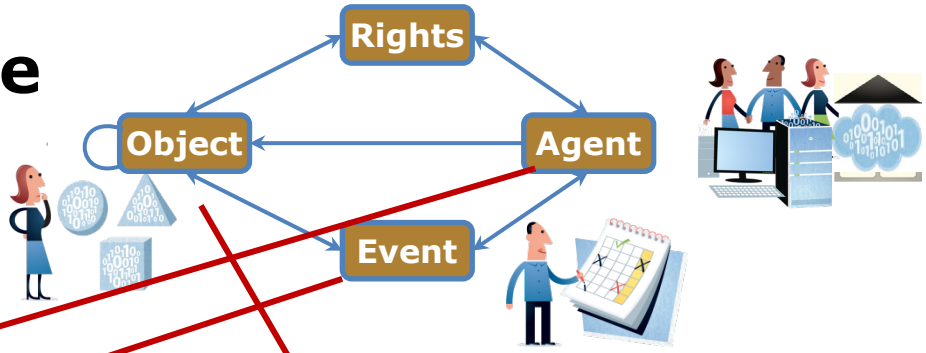
Examples:

- John Smith (a person)
- IIPC (an organization)
- JHOVE version 1.5 (a software program)

Implementation choices:

- use richer scheme that may be appropriate.

A digitization example



Identifier,
type (person),
name (Karin Bredenberg)

Identifier,
type (creation),
date/time
etc.

Identifier,
fixity,
file format,
creating application,
preservation level
etc.

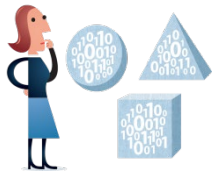
The PREMIS Data Model includes:

Entities 
"things" relevant to digital preservation

Rights

Object

Agent



intellectual entity
Representation
File
Bitstream

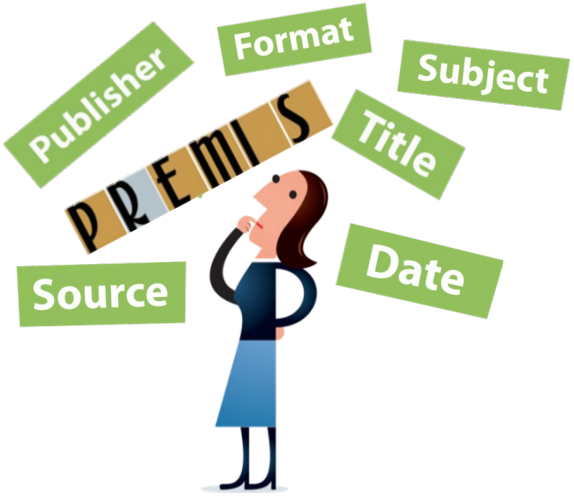
Event

menti.com

PREMIS PREservation Metadata Implementation Strategies

Eld Zierau
Royal Danish Library

DATA DICTIONARY
DESCRIPTION OF DATA MODEL



PREMIS Data model

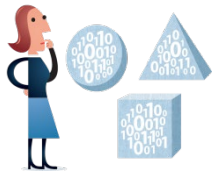
Entities
"things" relevant to digital preservation

Semantic Units
detailed information of entities

Rights

Object

Agent



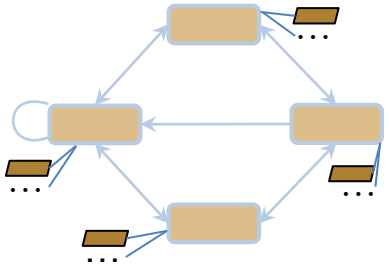
Event

- agentIdentifier
- agentIdentifierType
- agentIdentifierValue
- agentname
- agentType
- ...

Relationships ⟷
specifies relations between entities

Semantic Units

Properties of Entities



A piece of information most repositories need to know in order to carry out their digital preservation functions

- Example:

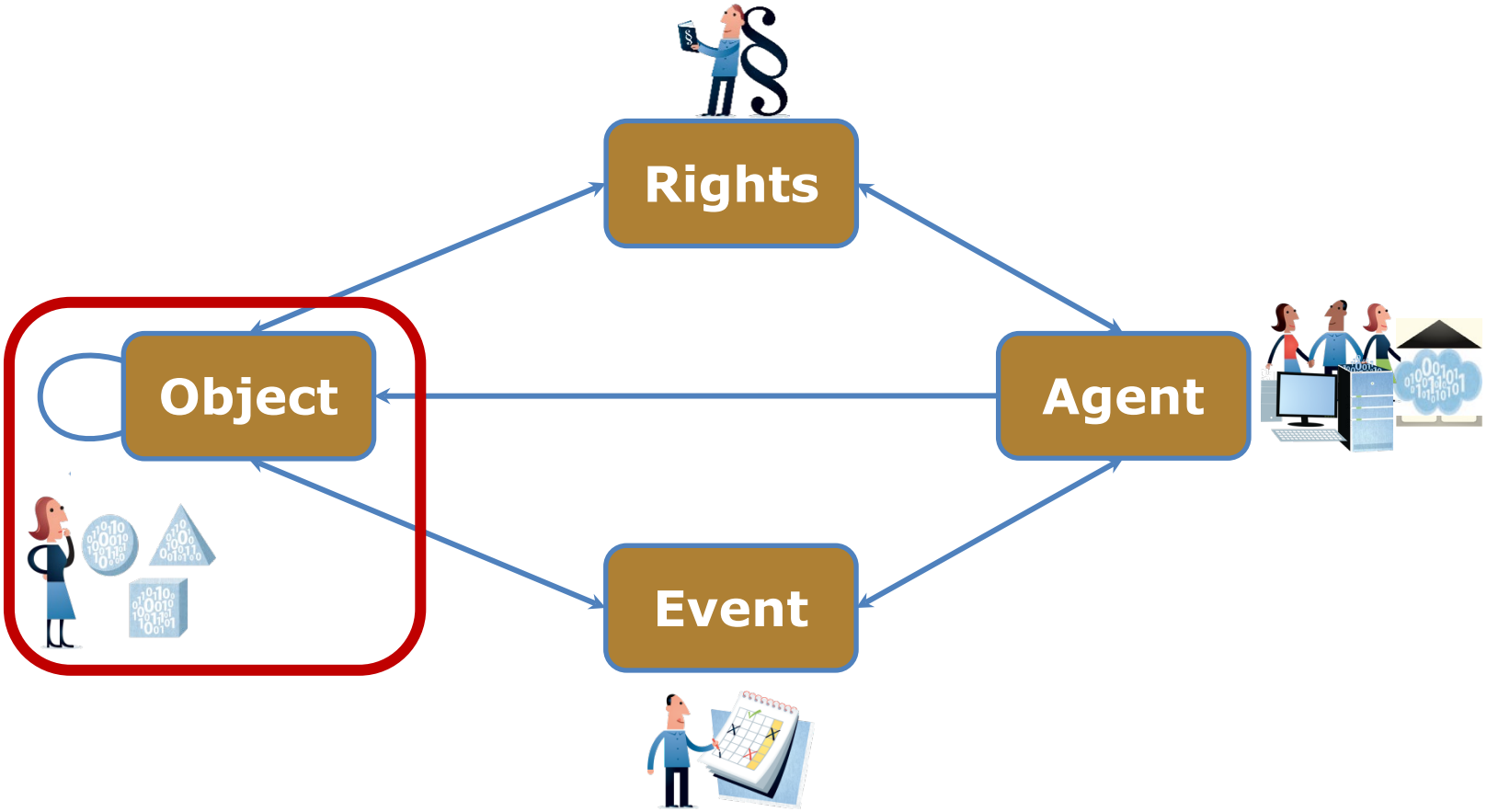
Two kinds of semantic unit

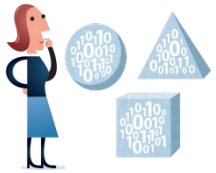
objectIdentifier [**container**] groups together related semantic units

objectIdentifierType [**semantic component**]

objectIdentifierValue [**semantic component**]

Properties of Entities - Semantic units





High level semantic units for Objects

what technical information on it?

which object is it?

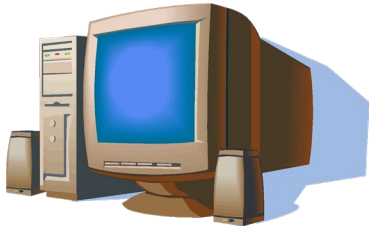
`ark:/12148/btp6k102002g/f1`

what is my preservation strategy for this object?

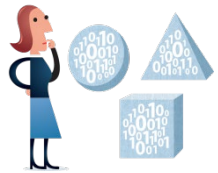
what kind of object?

which of its characteristics do I want to preserve in it?

what software or hardware should be used to handle the object?



PREMIS Object Entity – Semantic Units



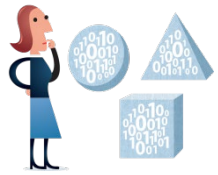
object

- objectIdentifier
- objectCategory

mandatory

- objectCharacteristics
- significantProperties
- preservationLevel
- originalName
- storage
- signatureInformation
- environmentFunction
- environmentDesignation
- environmentRegistry
- environmentExtension
- relationship
- linkingEventIdentifier
- linkingRightsStatementIdentifier

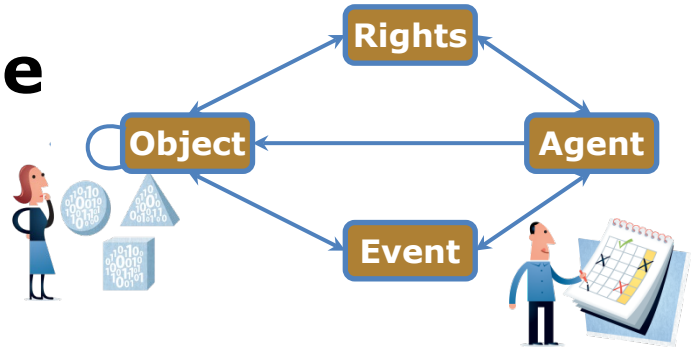
PREMIS Object Entity – Semantic Units



object

- objectIdentifier
  - objectIdentifierType
  - objectIdentifierValue
- objectCategory
- objectCharacteristics
- significantProperties
- preservationLevel
- originalName
- storage
- signatureInformation
- environmentFunction
- environmentDesignation
- environmentRegistry
- environmentExtension
- relationship
- linkingEventIdentifier
- linkingRightsStatementIdentifier

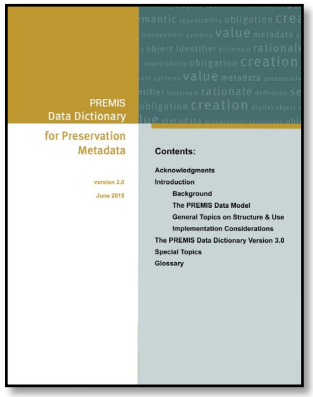
A digitization example



Identifier,
type (person),
name (Karin Bredenberg)

Identifier,
type (creation),
date/time
etc.

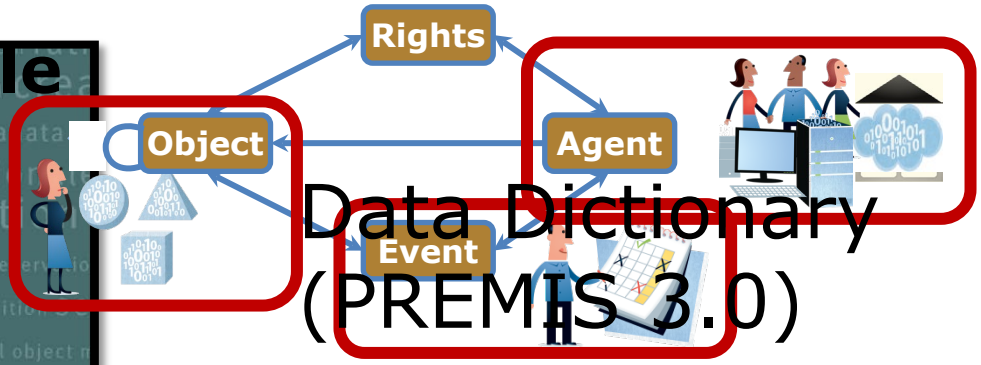
Identifier,
fixity,
file format,
creating application,
preservation level
etc.



Data Dictionary
(PREMIS 3.0)

http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf

A digitization example



Data Dictionary (PREMIS 3.0)

http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf

Agent *(digitizer)*

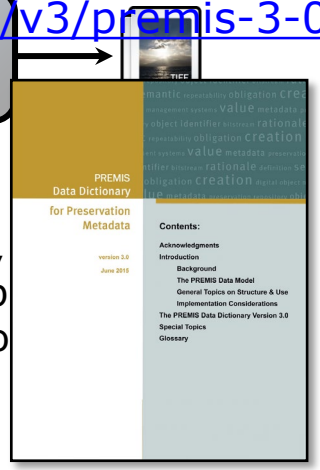
Identifier, type (person), name etc.

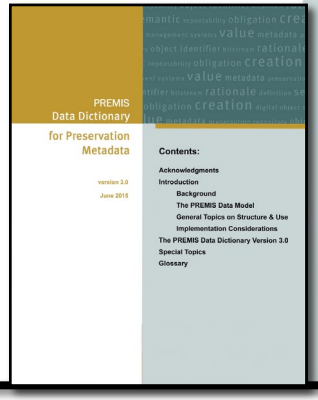
Event *(digitization)*

Identifier, type (creation), date/time etc.

Object: file

Identifier, fixity, file format, creating ap preservatio etc.





Contents:

Acknowledgments
Introduction
Background
The PREMIS Data Model
General Topics on Structure & Use
Implementation Considerations
The PREMIS Data Dictionary Version 3.0
Special Topics
Glossary

Data Dictionary (PREMIS 3.0)

http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf

DATA DICTIONARY

Entity semantic units

NB: Semantic units are applicable for Intellectual Entities, Representations, Files and Bitstreams unless otherwise indicated.

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) [Intellectual Entity, Representation, File]
- 1.3.1 preservationLevelType (O, NR) [Intellectual Entity, Representation,
- 1.3.2 preservationLevelValue (M, NR) [Intellectual Entity, Representation
- 1.3.3 preservationLevelRole (O, NR) [Intellectual Entity, Representation,
- 1.3.4 preservationLevelRationale (O, R) [Intellectual Entity, Representatio
- 1.3.5 preservationLevelDateAssigned (O, NR) [Intellectual Entity, Represe

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 (O, NR) [File, Bitstream]
- 1.5.2 fixity (O, R) [File, Bitstream]
  - 1.5.2.1 messageDigestAlgorithm (M, NR) [File, Bitstream]
  - 1.5.2.2 messageDigest (M, NR) [File, Bitstream]

objectIdentifier

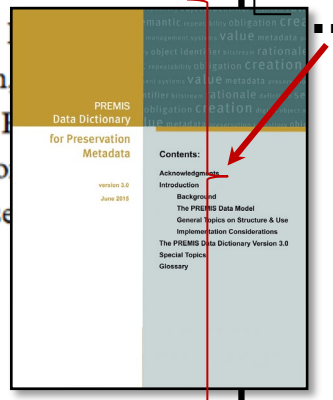
Object

objectIdentifier

objectCategory

PreservationLevel

...



DATA DICTIONARY

Entity semantic units

NB: Semantic units are applicable for Intellectual Entities, Representations, Files and Bitstreams unless otherwise indicated.

1.1 objectIdentifier (M, R)
- 1.1.1 objectIdentifierType (M, NR)
- 1.1.2 objectIdentifierValue (M, NR)

1.2 objectCategory (M, NR)

vationLevel (O, R) [Intellectual Entity, Representation, File]
- preservationLevelType (O, NR) [Intellectual Entity, Representation, File]
- preservationLevelValue (M, NR) [Intellectual Entity, Representation, File]
- ationLevelRole (O, NR) [Intellectual Entity, Representation, File]

tionLevelRationale (O, R) [Intellectual Entity, Representation, File]

tionLevelDateAssigned (O, NR) [Intellectual Entity, Representation, File]

perties (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 (O, NR) [File, Bitstream]
- 1.5.2 fixity (O, R) [File, Bitstream]
  - 1.5.2.1 messageDigestAlgorithm (M, NR) [File, Bitstream]
  - 1.5.2.2 messageDigest (M, NR) [File, Bitstream]

objectIdentifier

objectIdentifierType

objectIdentifierValue

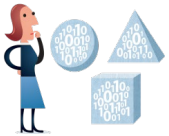
Mandatory (M)

Optional (O)

Repeatable (R)

Not Repeatable (NR)

n Strategies



Object

objectIdentifier

objectCategory

PreservationLevel

...

DATA DICTIONARY

Entity semantic units

NB: Semantic units are … Bitstreams unless other…

Semantic unit	1.1 objectIdentifier
Semantic components	1.1.1 objectIdentifierType 1.1.2 objectIdentifierValue
Definition	A designation used to identify the Object uniquely within the preservation repository system in which it is stored.

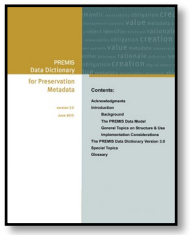
Semantic unit	1.1.1 objectIdentifierType
Semantic components	None
Definition	A designation o… unique.
Rationale	Identifier values… combination of … ensure uniquene…
Data constraint	Value should be… vocabulary is av… http://id.loc.gov/
Object category	Intellectual Entit… Representation
Applicability	Applicable
Examples	ISBN (Intellectu… Entity) DOI (Intellectua… Entity) DLC DRS

Semantic unit	1.1.2 objectIdentifierValue		
Semantic components	None		
Definition	The value of the *objectIdentifier*.		
Data constraint	None		
Object category	Intellectual Entity / Representation	File	Bitstream
Applicability	Applicable	Applicable	Applicable
Examples	0 00 221804-6 0000000312	IU2440 WAC1943.56 AMNH CD269/CD269/70/10596.PCD CDS-VDEP-200211119-24879 734	IU2440-1 IU2440-2

…tion repository must have a unique … to refer to it and to relate it to … metadata unambiguously.

XML example

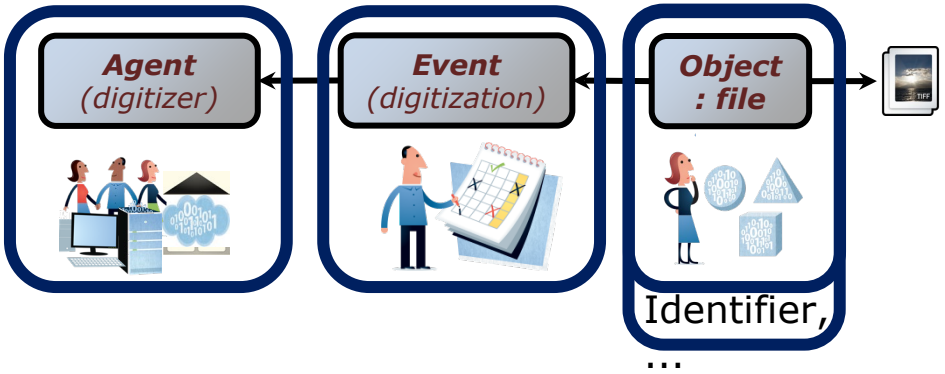
ObjectIdentifier
ObjectIdentifierType
ObjectIdentifierValue



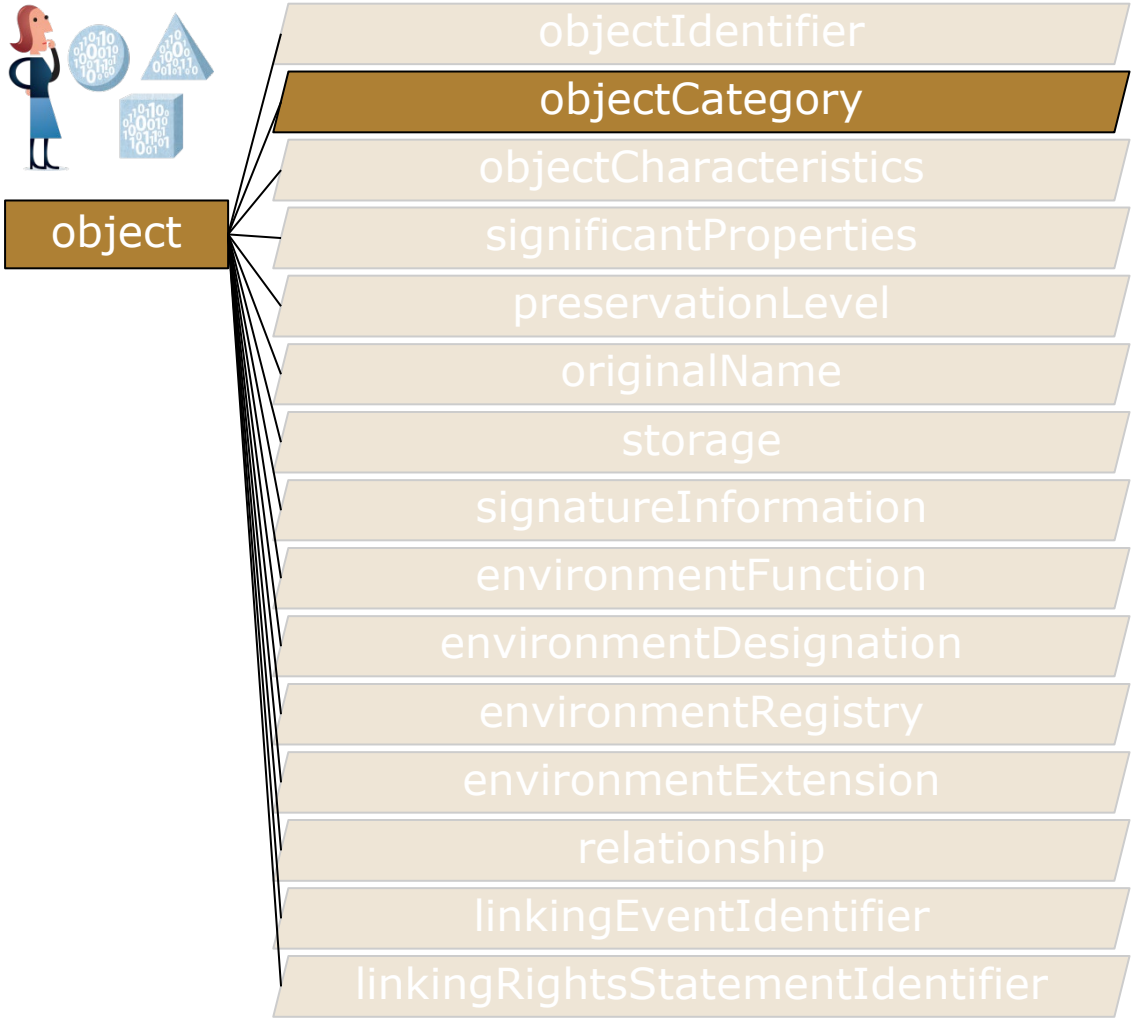
```
<premis>
  <object xsi:type="file"> ... </object>
    <objectIdentifier>
  <event> ... </event>
      <objectIdentifierType>
        UUID
  <agent> ... </agent>
      </objectIdentifierType>
  <rights> ... </rights>
      <objectIdentifierValue>
        41d10-099-1e2-9
</premis>
      </objectIdentifierValue>
    </objectIdentifier>
    ...
  </object>

  <event> ... </event>

  <agent> ... </agent>

</premis>
```

Examplified by XML using XML schema v3.0:

http://www.loc.gov/standards/premis/premis.xsd

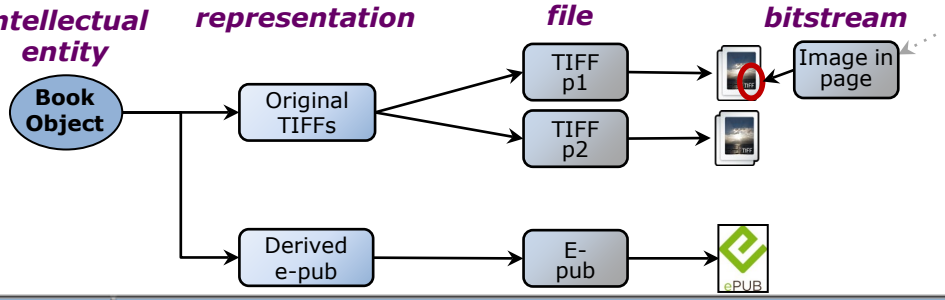


PREMIS Object Entity – Semantic Units



objectCategory

- Values:
  - intellectual entity
  - representation
  - file
  - bitstream

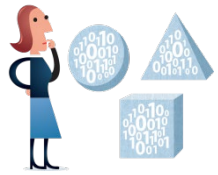


- Implemented a
so not explicitly

```
<premis>
  <object xsi:type="f
      …
  </object>
  …
</premis>
```

Semantic unit	1.1 objectIdentifier		
Semantic components	1.1.1 objectIdentifierType 1.1.2 objectIdentifierValue		
Definition	A designation used to identify the Object uniquely within the preservation repository system in which it is stored.		
Rationale	Each Object held in the preservation repository must have a unique identifier to allow other entities to refer to it and to relate it to descriptive, technical, and other metadata unambiguously.		
Data constraint	Container		
Object category	**Intellectual Entity / Representation**	**File**	**Bitstream**
Applicability	Applicable	Applicable	Applicable
Repeatability	Repeatable	Repeatable	Repeatable

PREMIS Object Entity – Semantic Units



object

- objectIdentifier
- objectCategory
- objectCharacteristics
  - specific characteristic
  - compositionLevel*
  - fixity
  - size
  - format
  - creatingApplication
  - inhibitors
  - objectCharacteristicsExtension*
- significantProperties
- preservationLevel
- originalName
- storage
- signatureInformation
- environmentFunction
- environmentDesign
- environmentRegistry
- environmentExtension
- relationship
- linkingEventIdentifier
- linkingRightsStatementIdentifier

mandatory for file or bitstream

Technical properties common to all/most file formats

Technology agnostic

Composition Level

sometimes there is more than one layer of characteristics

chapter1.pdf

Composition Level 0

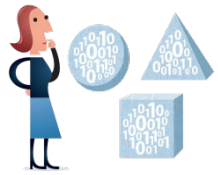
chapter1.pdf.gz

Composition Level 1

Or they may be part of other files e.g.

- Mail attachments
- Images in PDF's etc

PREMIS Object Entity – Semantic Units



object

objectIdentifier

objectCategory

objectCharacteristics

specific characteristic

compositionLevel

fixity

size

format

creatingApplication

inhibitors

objectCharacteristicsExtension*

significant[illegible]

preservation[illegible]

original[illegible]

storage

signatureInformation

environmentFunction

environmentDesign

environmentRegistry

environmentExtension

relationship

linkingEventIdentifier

linkingRightsStatementIdentifier

objectCharacteristicsExtension

Container to include external information
– e.g. for more granularity

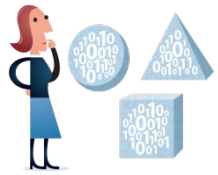
Might contain format specific metadata for a file
– e.g. technical metadata for still images (MIX)

objectCharacteristicsExtension - example

```
<premis> …
 <object xsi:type="file"> …
  <objectCharacteristics>  …
   <objectCharacteristicsExtension>
     <mix:mix xsi:schemaLocation=
         "… http://www.loc.gov/standards/mix/mix20/mix20.xsd">
       …
       <mix:BasicImageInformation>
         <mix:BasicImageCharacteristics>
          <mix:imageWidth>5894</mix:imageWidth>
          <mix:imageHeight>7768</mix:imageHeight>
          …
         </mix:BasicImageCharacteristics>
       </mix:BasicImageInformation>
       …
     <mix:mix>
   </objectCharacteristicsExtension>   …
  <objectCharacteristics>   …
 </object>  …
</premis>
```

All semantic units named … Extension works like this

PREMIS Object Entity – Semantic Units



object

objectIdentifier

objectCategory

objectCharacteristics

specific characteristic

compositionLevel

fixity

size

format

creatingApplication

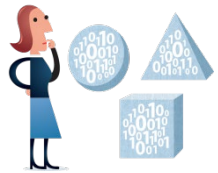
inhibitors

objectCharacteristicsExtension

OWL example:

```
<http://nri.library.ca/0912-0001.xml> a premis:File ;
    dct:format <XMLFormat> ;
    premis:fixity <0912-0001Fixity> ;
    premis:size 58465 ;
    premis:compositionLevel 0
```

PREMIS Object Entity – Semantic Units



object

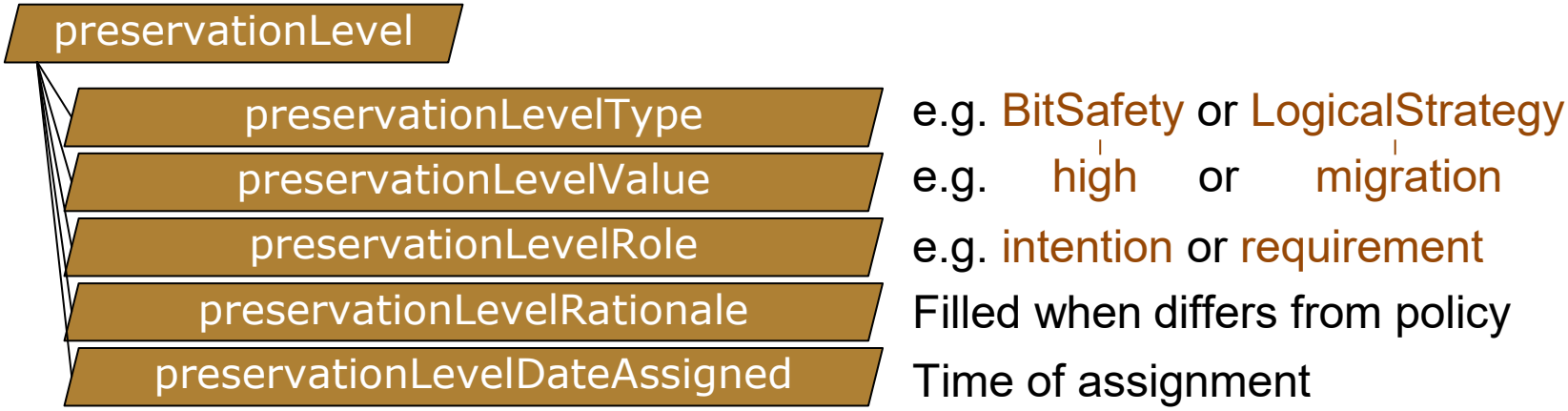
- objectIdentifier
- objectCategory
- objectCharacteristics
- significantProperties
- preservationLevel *
- originalName
- storage
- signatureInformation
- environmentFunction
- environmentDesignation
- environmentRegistry
- environmentExtension
- relationship
- linkingEventIdentifier
- linkingRightsStatementIdentifier

more general info

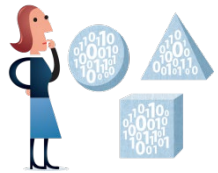
preservationLevel

What preservation treatment/strategy the repository plans for this object

- Varying preservation options dependent on factors such as value, uniqueness, preservability of format
- A business rule only relevant in a given repository



PREMIS Object Entity – Semantic Units

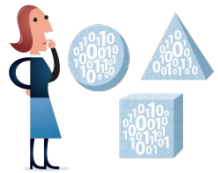


object

- objectIdentifier
- objectCategory
- objectCharacteristics
- significantProperties
- preservationLevel
- originalName
- storage
- signatureInformation
- environmentFunction
- environmentDesignation
- environmentRegistry
- environmentExtension
- relationship
- linkingEventIdentifier
- linkingRightsStatementIdentifier

more general info

PREMIS Object Entity – Semantic Units

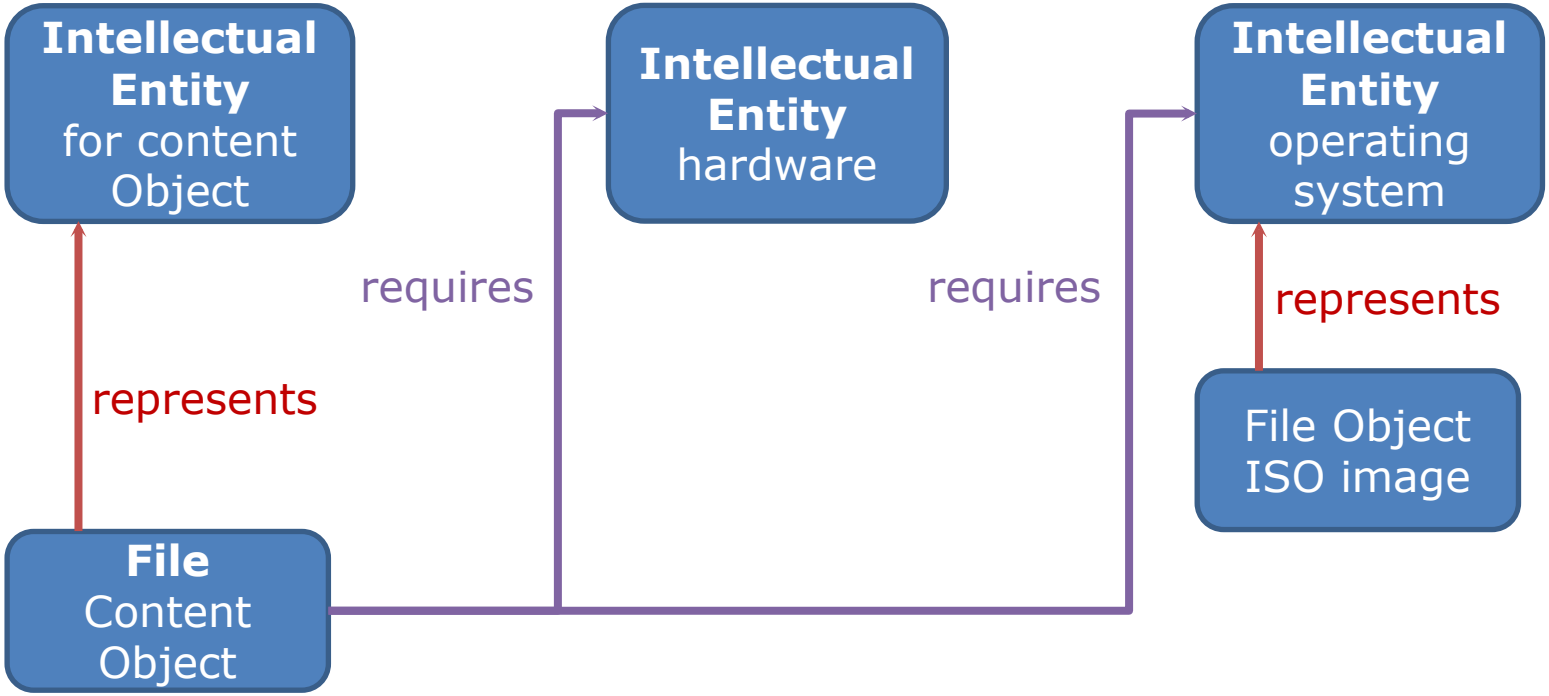


object

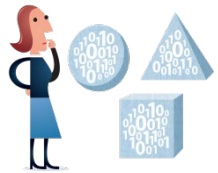
- objectIdentifier
- objectCategory
- objectCharacteristics
- significantProperties
- preservationLevel
- originalName
- storage
- signatureInformation
- environmentFunction
- environmentDesignation
- environmentRegistry
- environmentExtension
- relationship
- linkingEventIdentifier
- linkingRightsStatementIdentifier

environments

Environment example: An object and its rendering environment



PREMIS Object Entity – Semantic Units



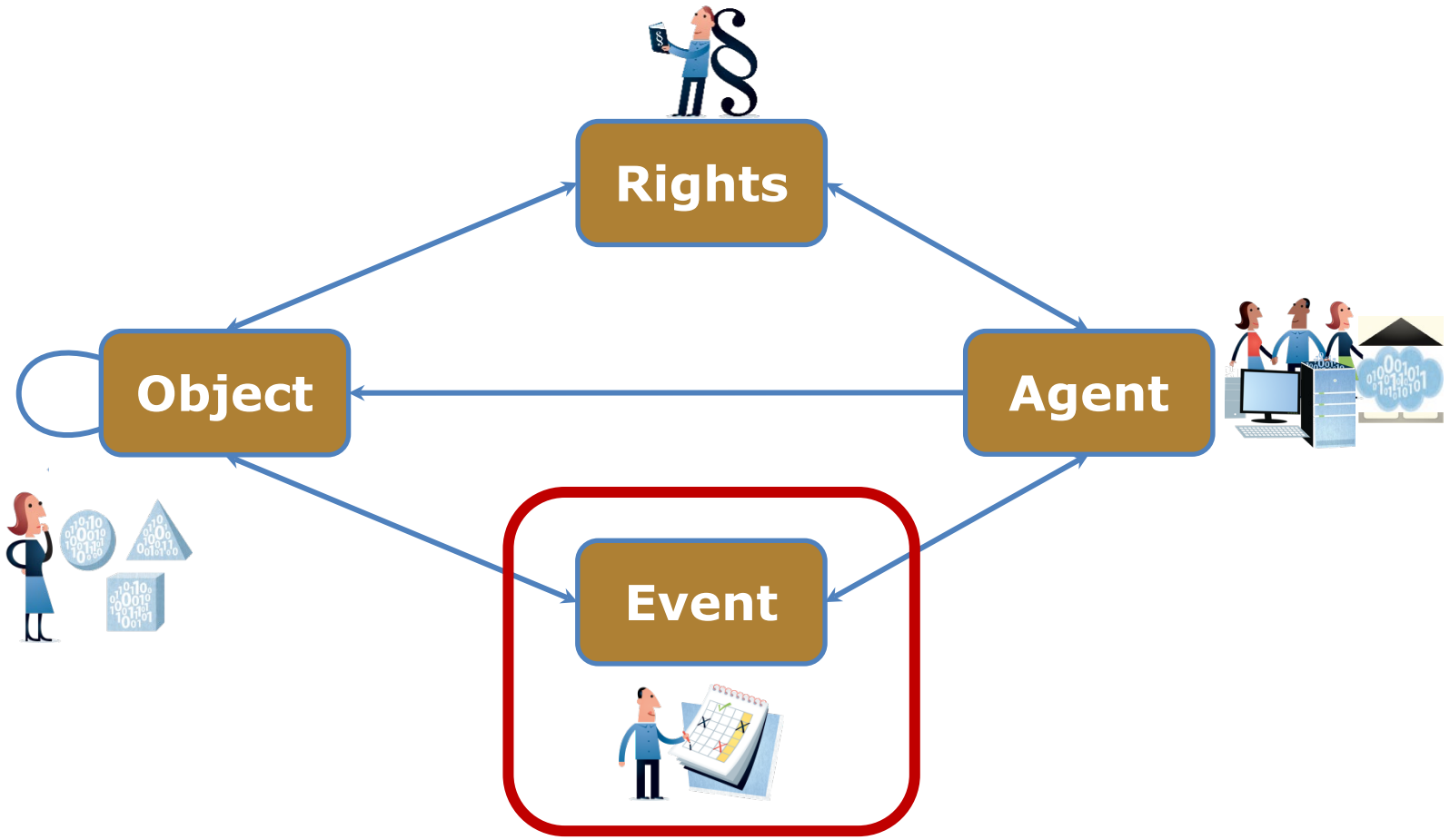
object

- objectIdentifier
- objectCategory
- objectCharacteristics
- significantProperties
- preservationLevel
- originalName
- storage
- signatureInformation
- environmentFunction
- environmentDesignation
- environmentRegistry
- environmentExtension
- relationship
- linkingEventIdentifier
- linkingRightsStatementIdentifier

Later

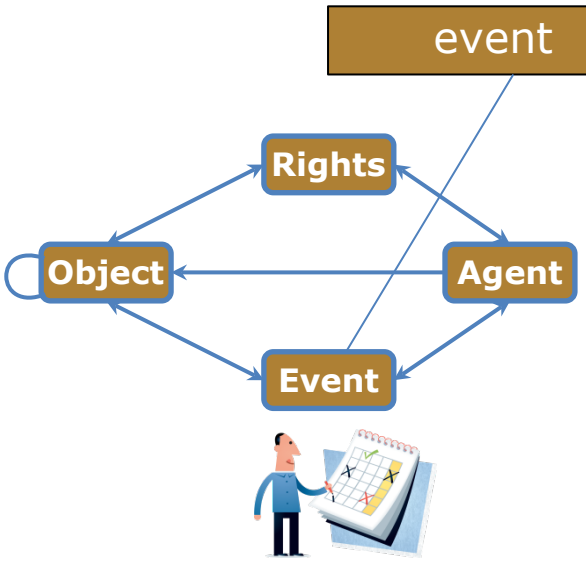
entity linking

Properties of Entities - Semantic units



PREMIS Event Entity – Semantic Units

- Must be related to one or more Objects.
- Can be related to one or more Agents.



event

- eventIdentifier
- eventType*
- eventDateTime
- eventDetailInformation
- eventOutcomeInformation
- linkingAgentIdentifier
- linkingObjectIdentifier

mandatory

eventType

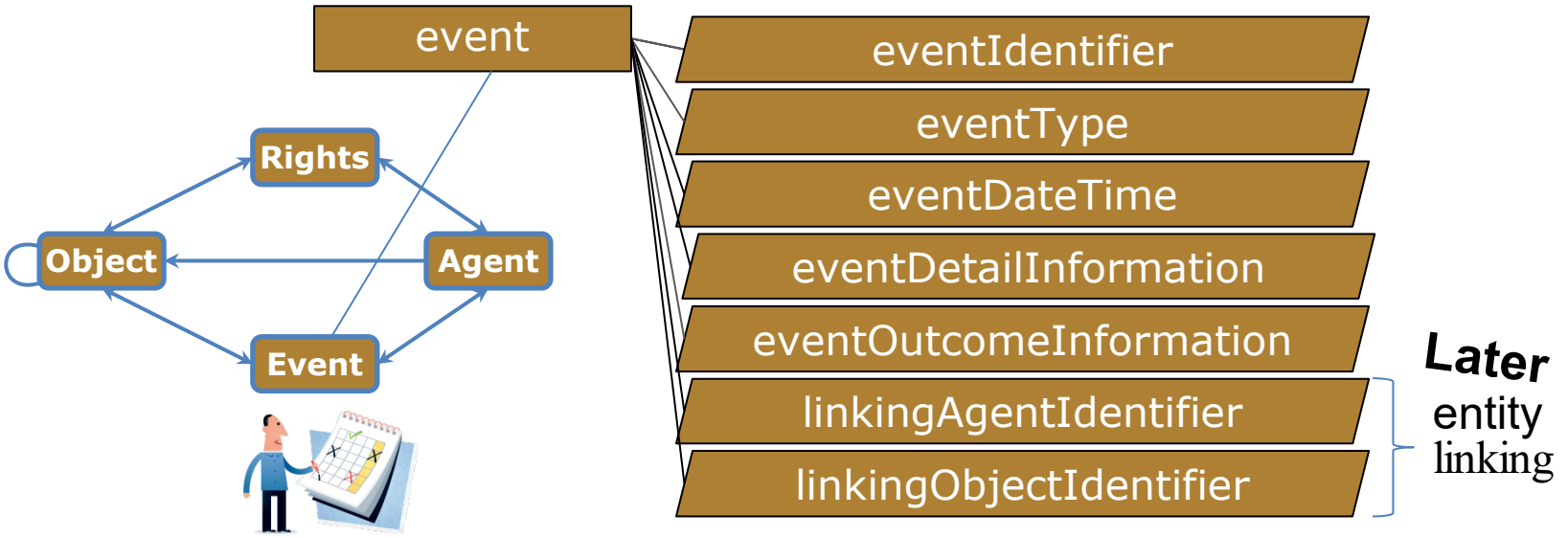
- Names the event

Ingestion **Virus check** **creation**
Validation
Compression
Message digest calculation
Decompression **…**
migration **Fixity check**

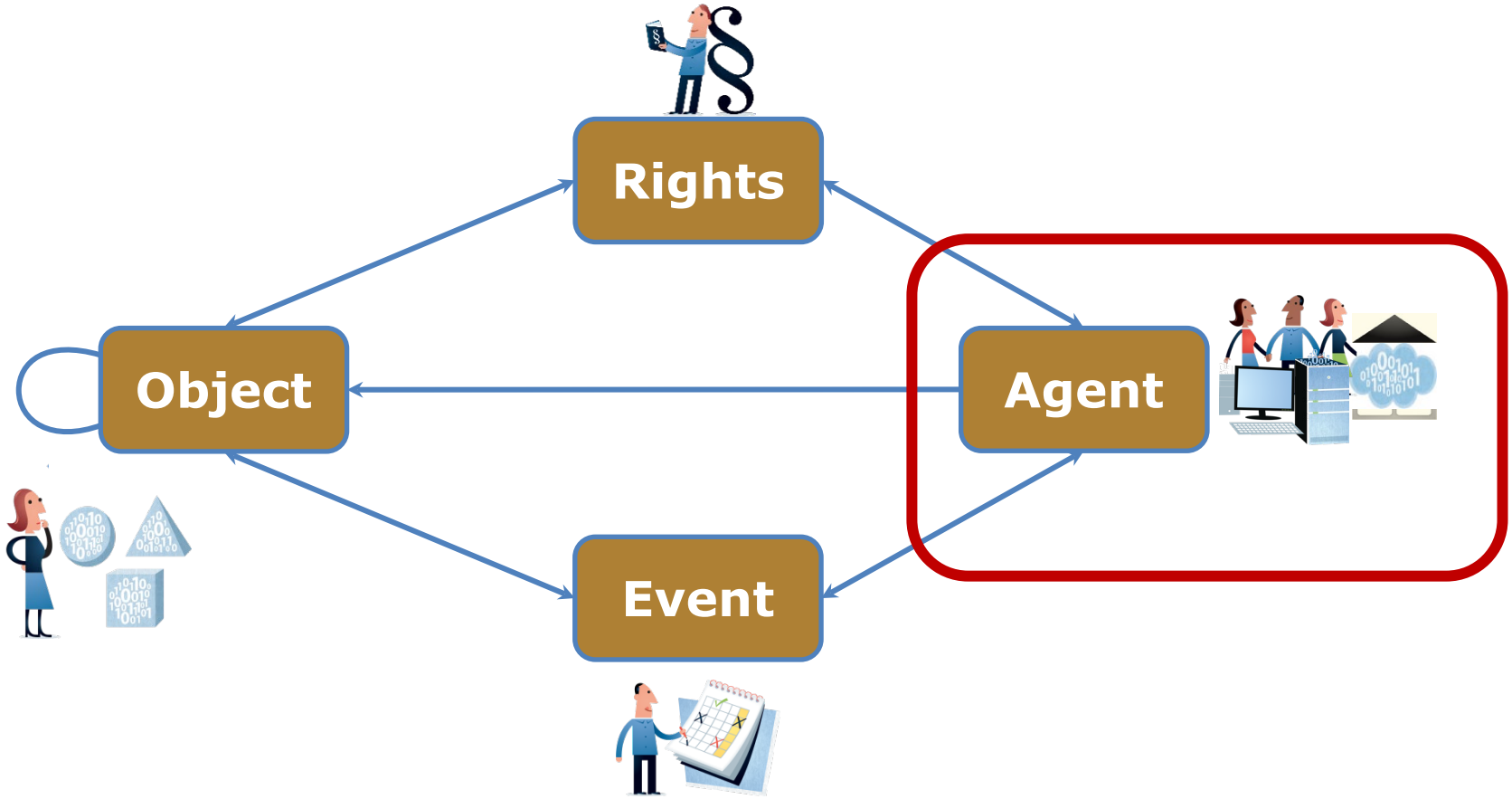
- Recommended to use a controlled vocabulary, e.g. http://id.loc.gov/vocabulary/preservation/eventType.html
- Could use coded values
- Granularity is implementation-specific

PREMIS Event Entity – Semantic Units

- Must be related to one or more Objects.
- Can be related to one or more Agents.

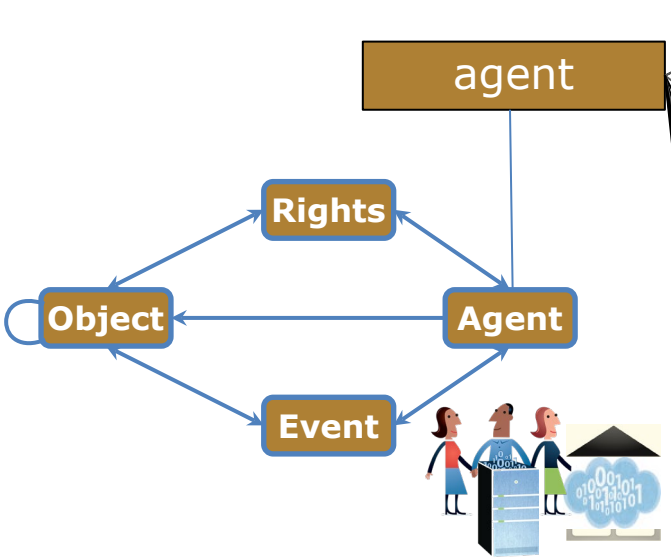


Properties of Entities - Semantic units



PREMIS Agent Entity – Semantic Units

- May hold or grant one or more Rights.
- May carry out, authorize, or compel one or more Events.
- May create or act upon one or more Objects through an Event or with respect to a Rights statement.

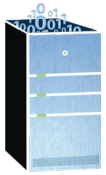
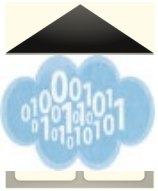


agent

- agentIdentifier
- agentName
- agentType*
- agentVersion
- agentNote
- agentExtension
- linkingEventIdentifier
- linkingRightsStatementIdentifier
- linkingEnvironmentIdentifier

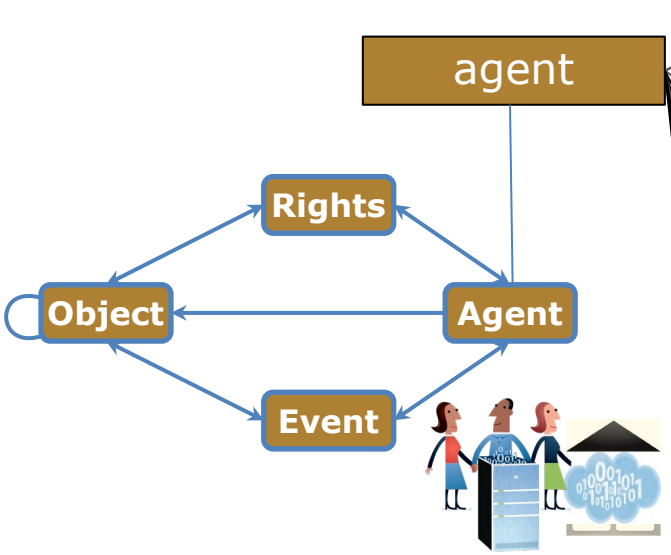
mandatory

agentType

- Can use controlled vocabulary, e.g. http://id.loc.gov/vocabulary/preservation/agentType.html
  - hardware 
  - organization 
  - person 
  - software 

PREMIS Agent Entity – Semantic Units

- May hold or grant one or more Rights.
- May carry out, authorize, or compel one or more Events.
- May create or act upon one or more Objects through an Event or with respect to a Rights statement.

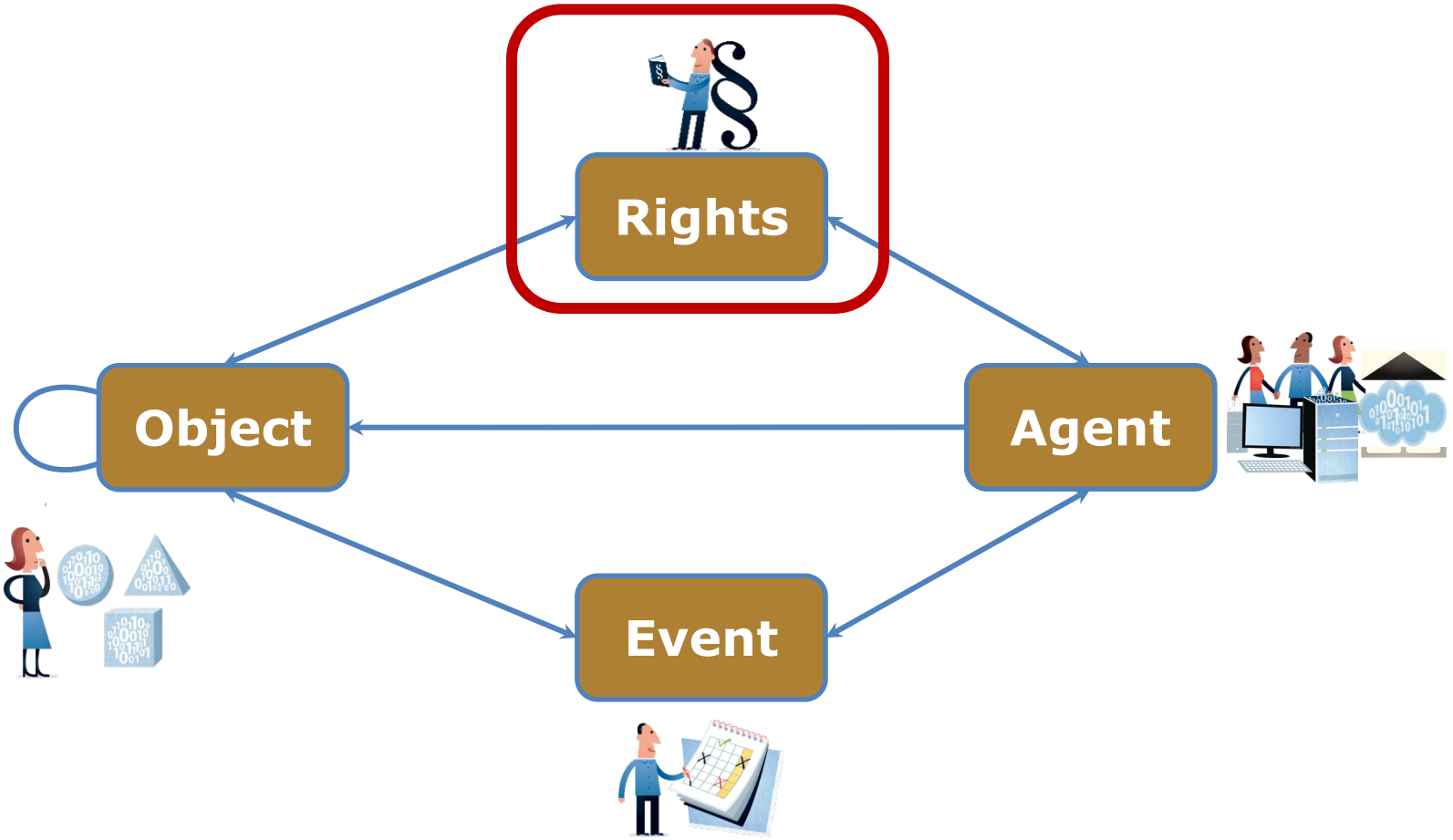


agent

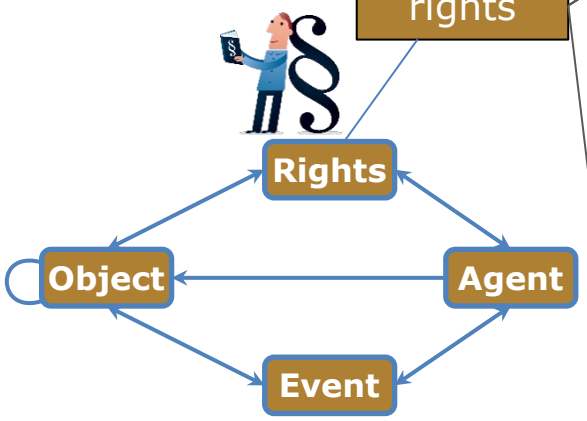
- agentIdentifier
- agentName
- agentType
- agentVersion
- agentNote
- agentExtension
- linkingEventIdentifier
- linkingRightsStatementIdentifier
- linkingEnvironmentIdentifier

Later entity linking environments

Properties of Entities - Semantic units



PREMIS Rights Entity – Semantic Units



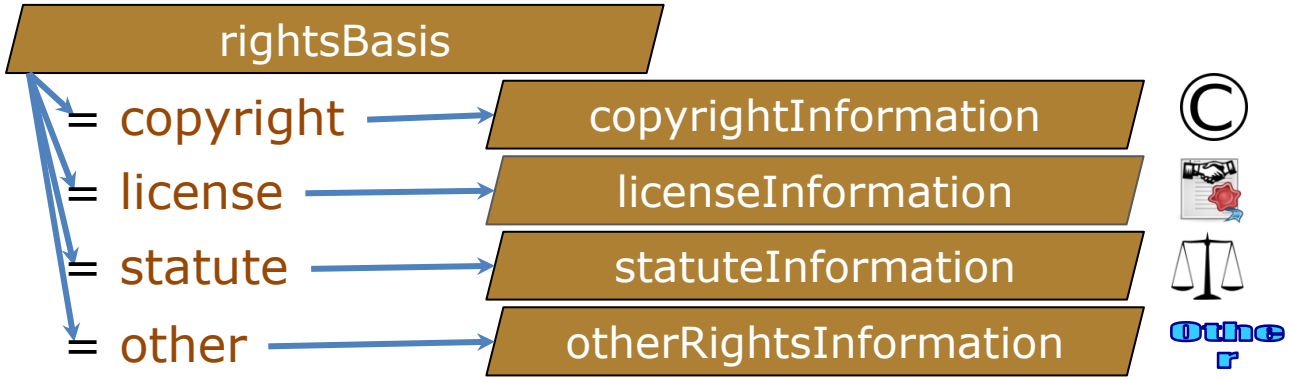
rights

- rightsStatement
- rightsStatementIdentifier
- rightsBasis*
- copyrightInformation
- licenseInformation
- statuteInformation
- otherRightsInformation
- rightsGranted*
- linkingObjectIdentifier
- linkingAgentIdentifier
- rightsExtension

One of these must be present!

Dependent units about rights

Specifying different types of rights



If more than one basis applies, the entire rights entity should be *repeated*.

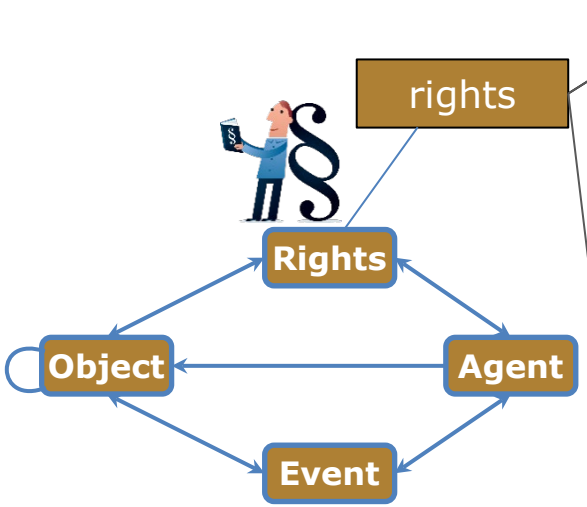
Example copyrightInformation

rightsBasis = copyright

copyrightInformation

- **copyrightStatus** = copyrighted
- **copyrightJurisdiction** = us
- **copyrightStatusDeterminationDate** = 2008-09-10
- **copyrightNote** = Copyright expiration expected in 2022
- **copyrightDocumentationIdentifier** = [link]

PREMIS Rights Entity – Semantic Units



- rights
  - rightsStatement
  - rightsStatementIdentifier
  - rightsBasis
  - copyrightInformation
  - licenseInformation
  - statuteInformation
  - otherRightsInformation
  - rightsGranted*
  - linkingObjectIdentifier
  - linkingAgentIdentifier
  - rightsExtension

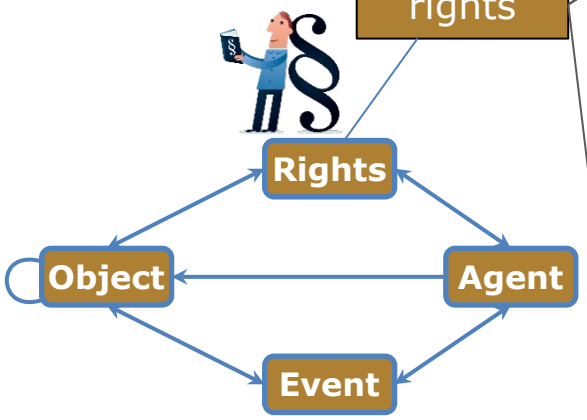
rightsGranted

- What action is allowed?
- Under what conditions?
- Are there time constraints?

Contains

- **Act** (e.g. migrate, modify)
- **Restriction** (description)
- **termOfGrant** (start and end date)
- **termOfRestriction** (start and end date)
- **rightsGrantedNote** (additional inf.)

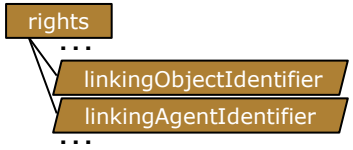
PREMIS Rights Entity – Semantic Units

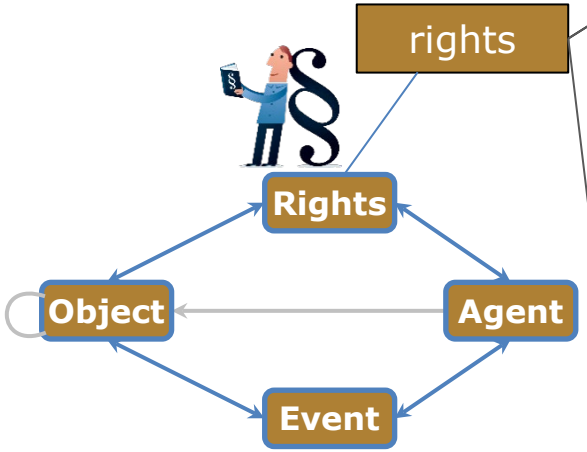


- rights
  - rightsStatement
  - rightsStatementIdentifier
  - rightsBasis
  - copyrightInformation
  - licenseInformation
  - statuteInformation
  - otherRightsInformation
  - rightsGranted
  - linkingObjectIdentifier
  - linkingAgentIdentifier
  - rightsExtension

Next entity linking

PREMIS Rights Entity – Semantic Units

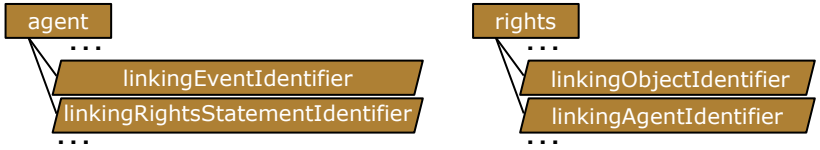


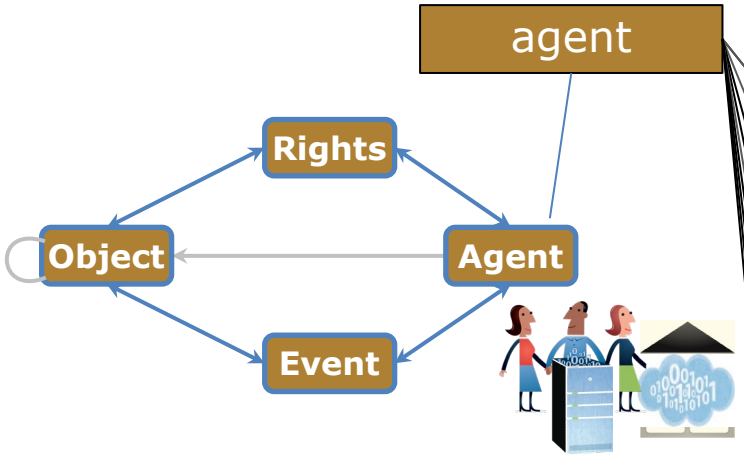


- rightsStatement
- rightsStatementIdentifier
- rightsBasis
- copyrightInformation
- licenseInformation
- statuteInformation
- otherRightsInformation
- rightsGranted
- **linkingObjectIdentifier**
- **linkingAgentIdentifier**
- rightsExtension

Next entity linking

PREMIS Agent Entity – Semantic Units



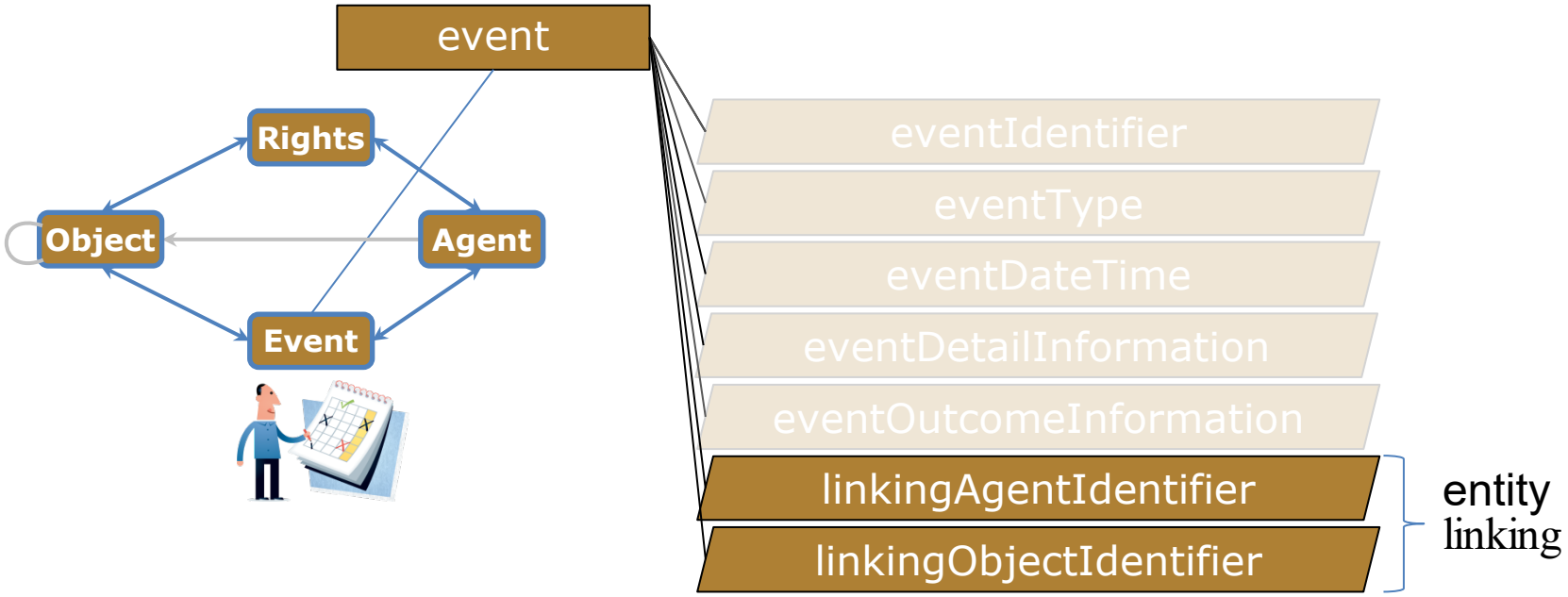


agent

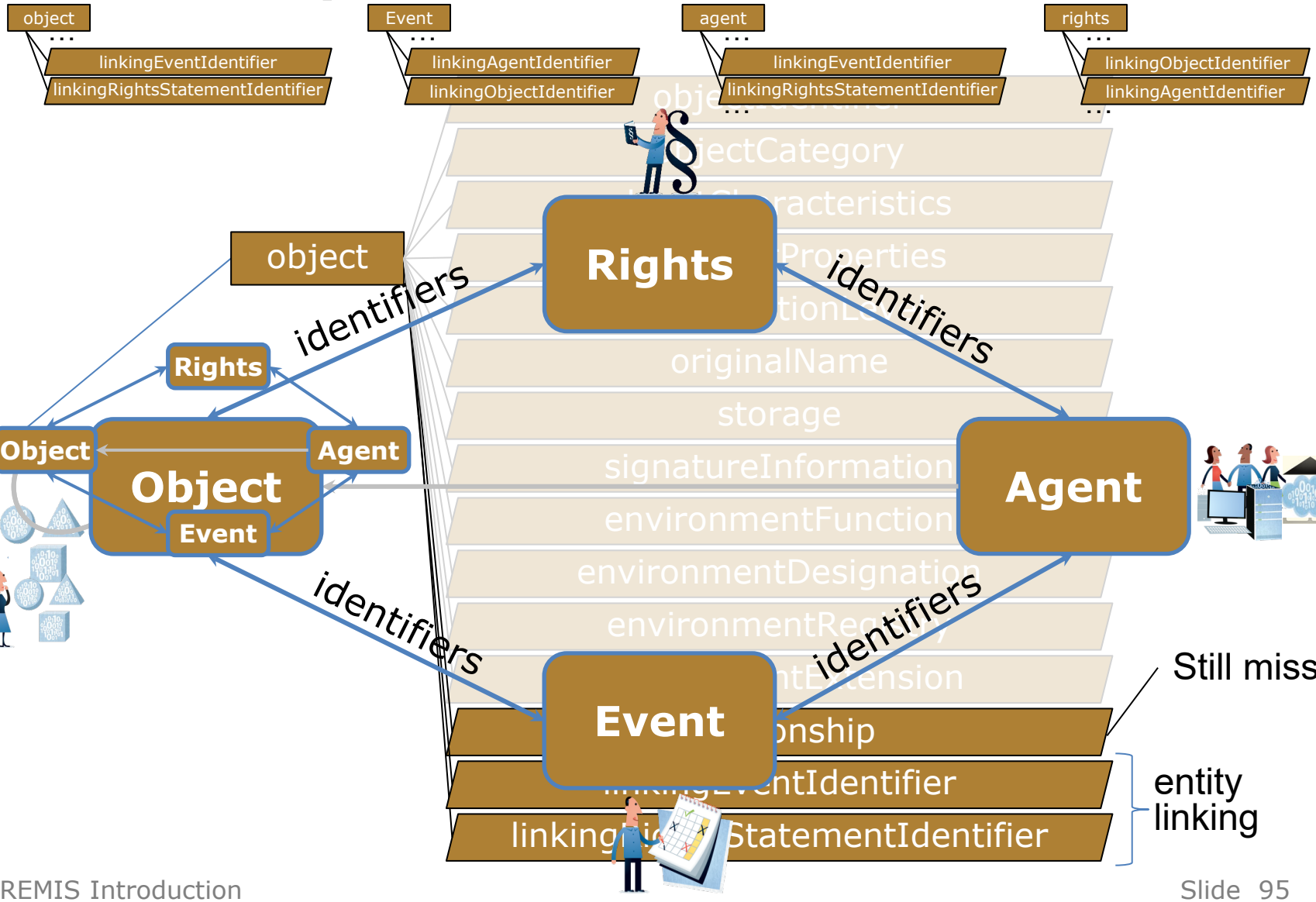
- agentIdentifier
- agentName
- agentType
- agentVersion
- agentNote
- agentExtension
- **linkingEventIdentifier**
- **linkingRightsStatementIdentifier**
- linkingEnvironmentIdentifier

entity linking

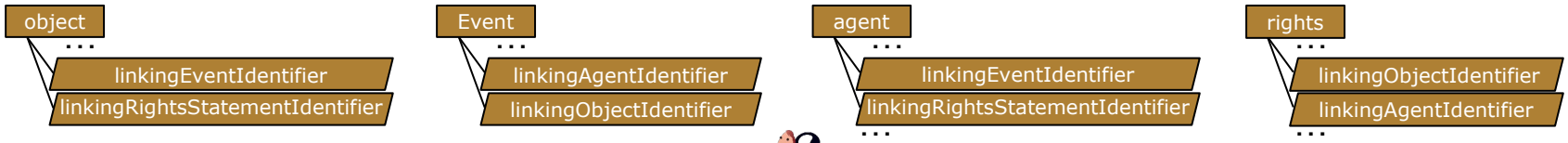
PREMIS Event Entity – Semantic Units



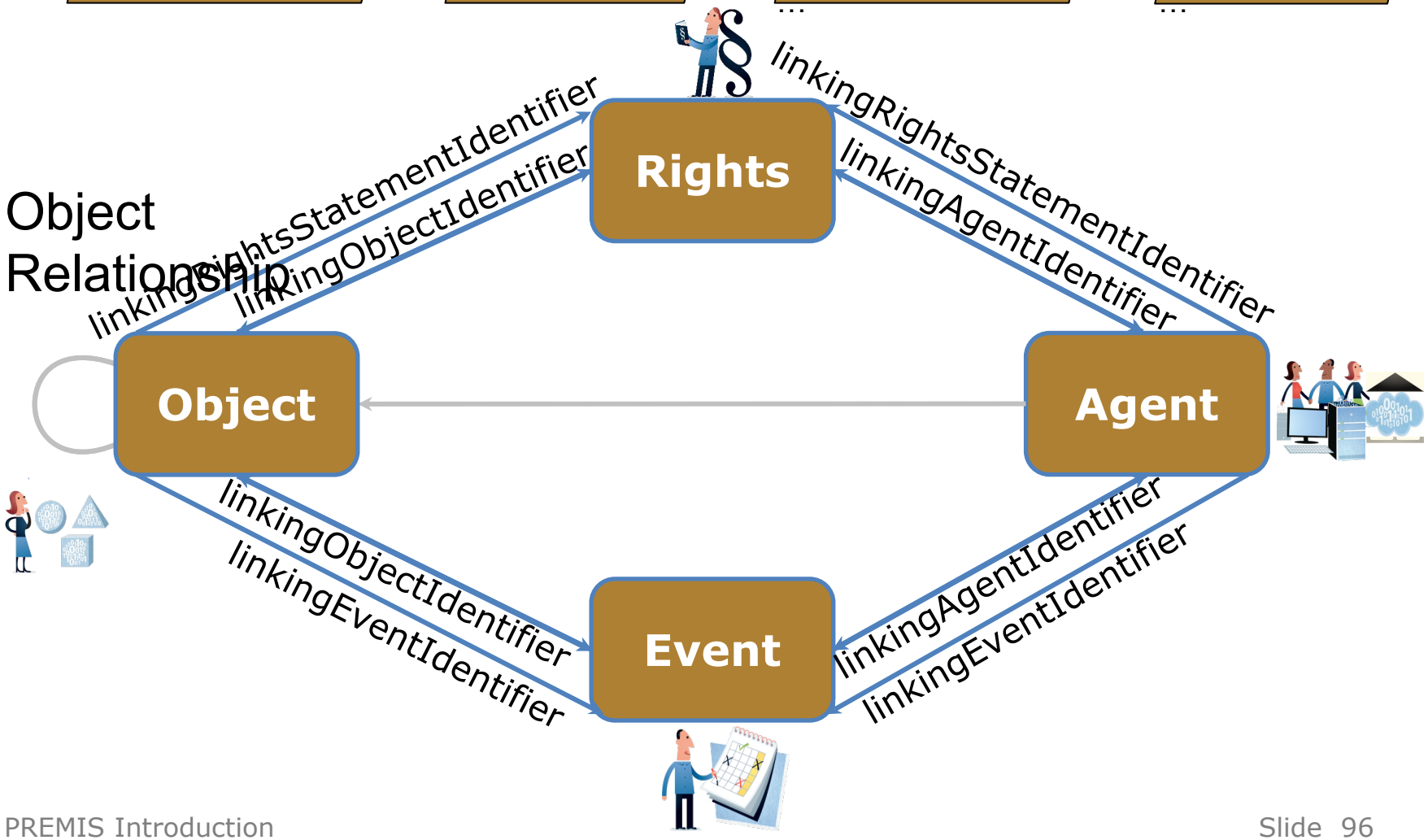
Relationships: Semantic Units



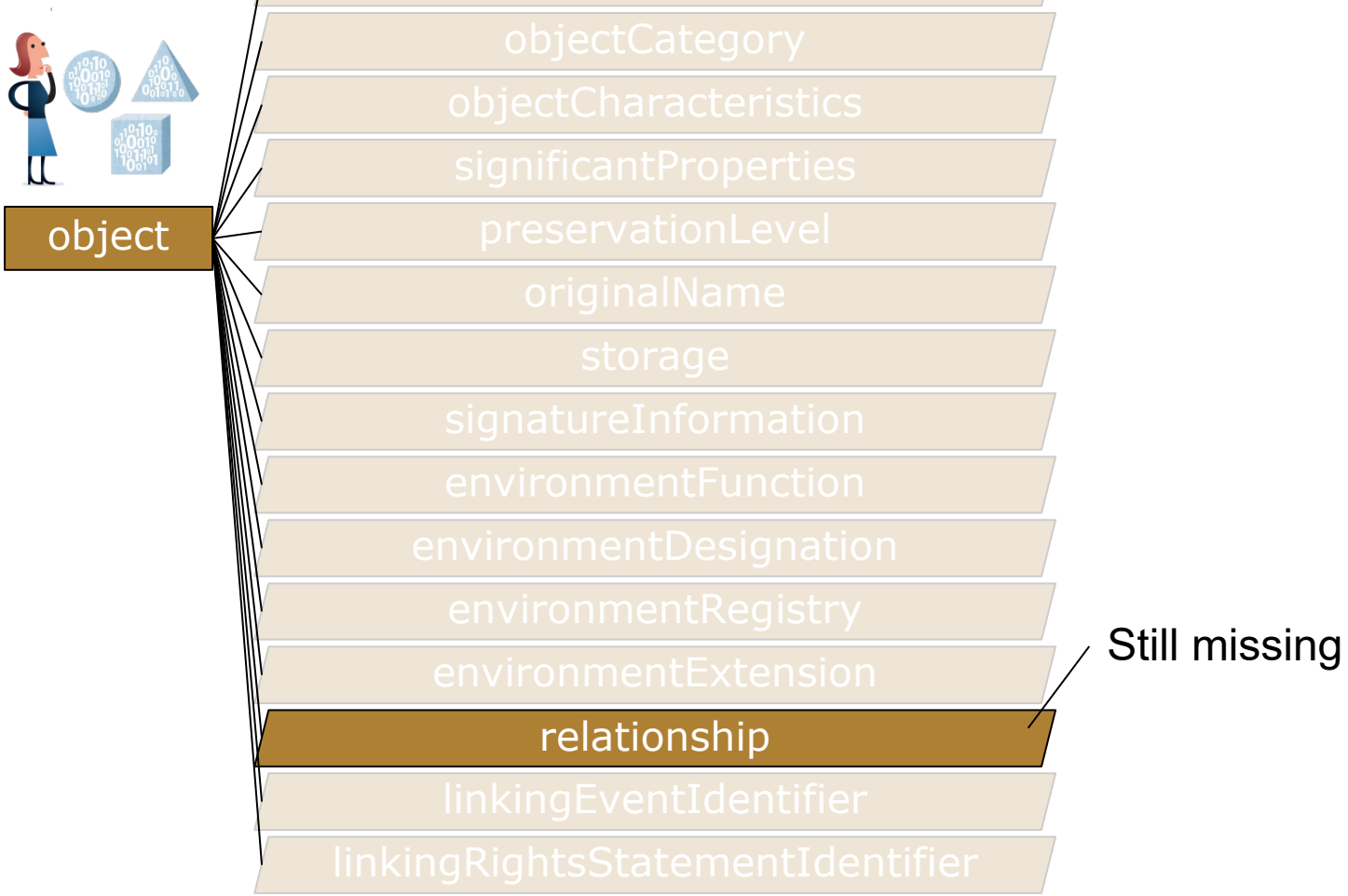
Relationships: Semantic Unit Identifiers

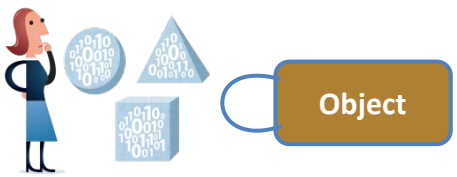


Object Relationship

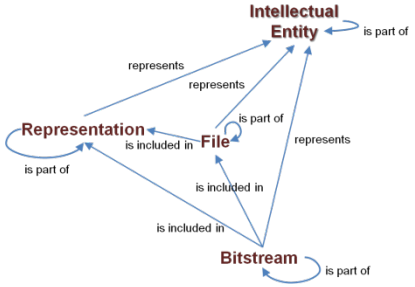


PREMIS Object Entity – Semantic Units





Objects and their interrelations



is part of Intellectual Entity

is part of Representation

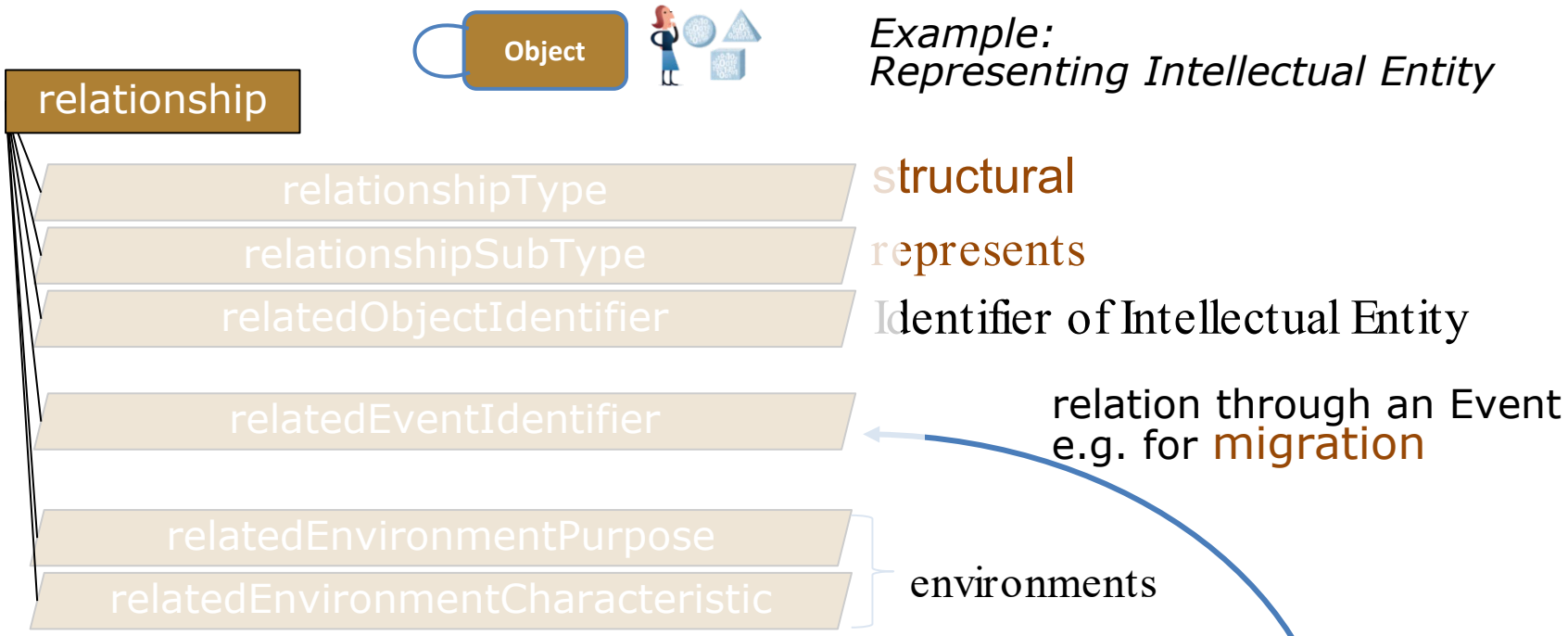
represents

is included in

is part of File

is part of Bitstream

PREMIS Object Entity – Semantic Units

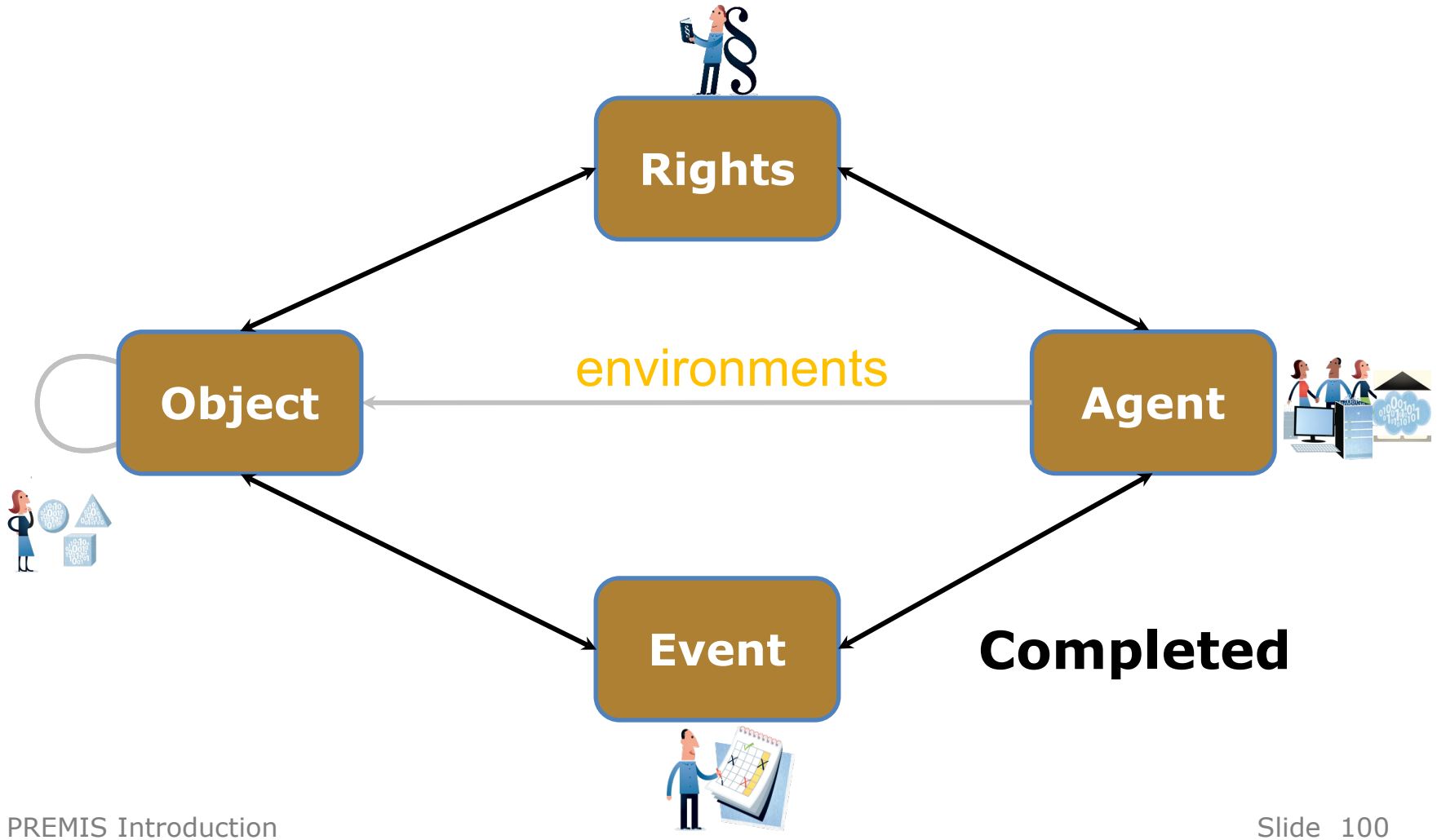


Example:
Representing Intellectual Entity

Objects can be associated with Events in two ways:

- If the Object has an associated Event with ***relationship***
- If the Object has an associated Event with ***no relationship*** to a second Object, e.g. ingest: use **linkingEventIdentifier**

Relationships Between Semantic Unit Identifiers



PREMIS PREservation Metadata Implementation Strategies

Micky Lindlar
TIB - German National Library of Science and Technology

HOW TO USE PREMIS?

The Data Dictionary in action: PREMIS Conformance and repository interoperability

PREMIS Conformance statement

- http://www.loc.gov/standards/premis/premis-conformance-20150429.pdf

Baseline requirements:

- For every implemented Entity (Objects, Events, Rights, Agents) mandatory semantic units must be captured
  - For those levels of Object that the repository supports (IE, representation, file, bitstream)
- Requirements for
- Shared name = Shared definition!
- Shared definition without shared name -> needs documentation

PREMIS Conformance Levels

	A – Object Entity Only	B – Object, Event & Agent
Level 1 – Mapping	Internal metadata is mapped to PREMIS & documented	
Level 2 – Export	Internal metadata can be exported (via a tool-/process-supported routine) to PREMIS	
Level 3 – Internal Implementation	PREMIS is implemented as internal metadata schema	

Example: What's conformant and what isn't?

1.1 objectIdentifier (M, R)
1.1.1 objectIdentifierType (M, NR)
1.1.2 objectIdentifierValue (M, NR)

✗ None

✓ DOI: 10.5281/zenodo.5569542

✓ Eindeutiger Bezeichner:10.5281/zenodo.5569542

✓
<dc:identifier>https://zenodo.org/record/5569578</dc:identifier>
<dc:identifier>10.5281/zenodo.5569578</dc:identifier>
<dc:identifier>oai:zenodo.org:5569578</dc:identifier>

✗ <objectIdentifier>fmt/18</objectIdentifer>

✓
premis:objectIdentifier
premis:objectIdentiferType=„doi“
premis:objectIdentifierValtue=„10.5281/zenodo.5569542“

Which Entities to implement?

- Object is the core Entity (level A);
- Event and Agent are closely related (level B); implementing Agents has strong implications: it means the repository is able to manage and follow the use of its Agents in the Object lifecycle.
- The Rights Entity (excluded from the conformance statement) helps a repository tracking the intellectual property rights governing the Object, or some institutional policy.

PREMIS can be used as a(n) …

- basis for other standards / locally defined metadata catalogues *(no conformance, but inspiration)*
  - e.g., Netherlands Institute for Sound and Vision https://publications.beeldengeluid.nl/pub/389/BIJLAGE-C_Metadatadictionary-English.pdf
- self-assessment tool *(Conformance Level 1)*
  - Am I able to provide information about my digital assets following the Data Dictionary structure and requirements?
- export format *(Conformance Level 2)*
  - Preferably in a PREMIS-endorsed expression (XML or RDF)
- native format of the repository Data Management module *(Conformance Level 3)*
  - Any technology, using a PREMIS-endorsed expression or not, can be used

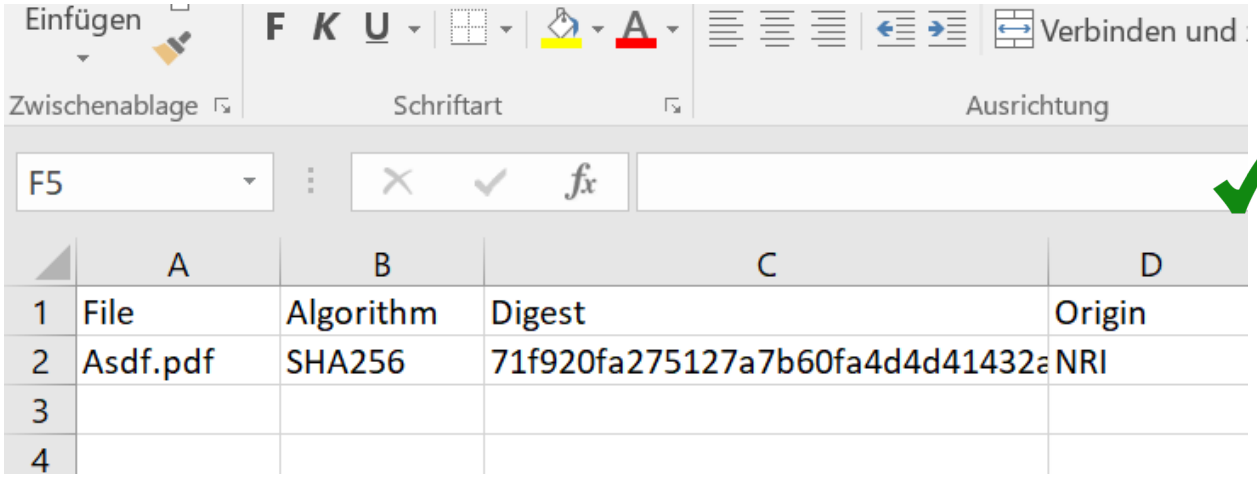
Examples for different implementations: RDF and XML

```
<http://nri.library.ca/5143-026.nrw> a premis:File ;
    premis:fixity <5143-026Fixity>
.

<5143-026Fixity> a crypHashFunc:sha256 ;
    rdf:value "71f920fa275127a7b60fa4d4d41432a3 ";
    dce:creator "hashlib.sha256"
.
```

```
<premis:fixity>
  <premis:messageDigestAlgorithm>SHA-256</premis:messageDigestAlgorithm>
  <premis:messageDigest>
d2bed92b73c7090bb30a0b30016882e7069c437488e1513e9deaacbe29d38d92
  </premis:messageDigest>
  <premis:messageDigestOriginator>NRI</premis:messageDigestOriginator>
</premis:fixity>
```

Examples for different implementations: CSV and spreadsheet

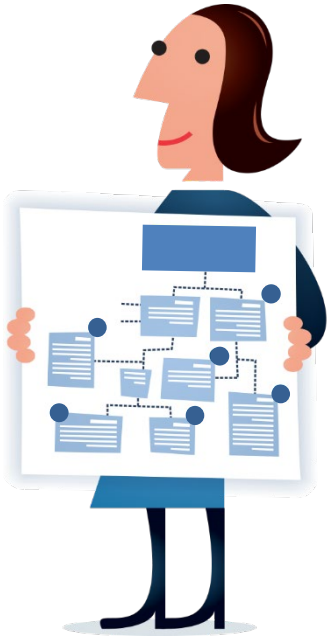
File, Algorithm, Digest, Origin;
Asdf.pdf; SHA-256, 71f920fa275127a7b60fa4d4d41432a3, NRI;



	A	B	C	D
1	File	Algorithm	Digest	Origin
2	Asdf.pdf	SHA256	71f920fa275127a7b60fa4d4d41432a	NRI
3				
4				

menti.com

Karin Bredenberg
Kommunalförbundet Sydarkivera



WRAP UP

- DD
- Where to find it and What it is for
- Current activity in PREMIS EC
- Book
- Exercises

But first

Go to Menti, link in the chat and respond to the question we wanted you to think about in the beginning

What kind of digital objects will you use PREMIS for?

Sum up – Data Dictionary

Lots of other information

Where? and What??

- Where
  - Resources: http://www.loc.gov/standards/premis/
  - PREMIS Implementors Group Forum: PIG@listserv.loc.gov
- What PREMIS is for today have given you a good start and you now need to explore it further

Images in this style is taken from digitalbevaring.dk

Resources

- Understanding PREMIS
- PREMIS-in-METS guidelines
- Conformance statement
- Examples of implementation

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

Understanding PREMIS – Entender PREMIS

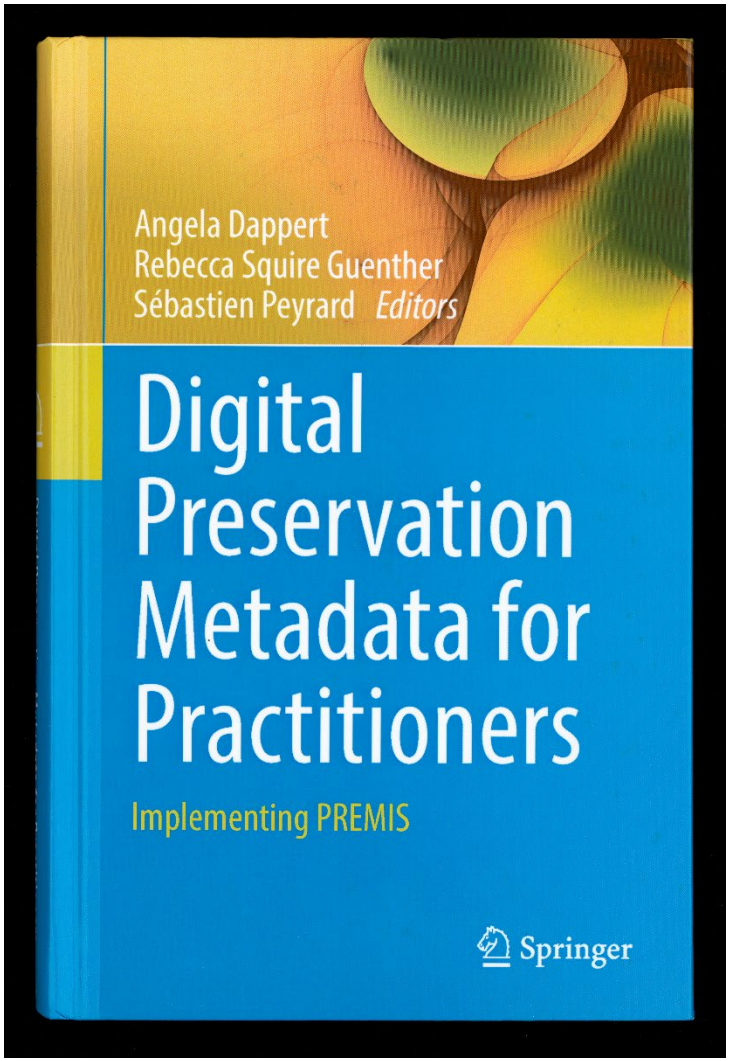
https://www.loc.gov/standards/premis/understandingPREMIS_spanish_2021.pdf

Current activity

- Move the DD to a TEI-format to simplify maintenance and transformations to publications
- DD updates following the ontology work
- Enhance our use of Zenodo
- Setup of new wiki
- Rights overhaul

Book

- ISBN E-book: 978-3-319-43763-7
- ISBN Hardcover: 978-3-319-43761-3
- http://www.springer.com/gp/book/9783319437613



Exercises

- Today have been really filled!
- Three exercises to start working with PREMIS metadata
  - Print them out!
- Solutions is also published!
- The aid is seen on next slide

Sample Data Dictionary table of contents

- Version 3 Hierarchical listing of semantics units.pdf
  https://doi.org/10.5281/zenodo.5569578

On your own!

Entity semantic units

NB: Semantic units are applicable for Intellectual Entities, Representations, Files and Bitstreams unless otherwise indicated.

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) [Intellectual Entity, Representation, File]
- 1.3.1 preservationLevelType (O, NR) [Intellectual Entity, Representation, File]
- 1.3.2 preservationLevelValue (M, NR) [Intellectual Entity, Representation, File]
- 1.3.3 preservationLevelRole (O, NR) [Intellectual Entity, Representation, File]
- 1.3.4 preservationLevelRationale (O, R) [Intellectual Entity, Representation, File]
- 1.3.5 preservationLevelDateAssigned (O, NR) [Intellectual Entity, 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 (O, NR) [File, Bitstream]
- 1.5.2 fixity (O, R) [File, Bitstream]
  - 1.5.2.1 messageDigestAlgorithm (M, NR) [File, Bitstream]
  - 1.5.2.2 messageDigest (M, NR) [File, Bitstream]

PREMIS Object Entity – Exercise

- Exercise and answers to get a feeling for the object!
  - https://doi.org/10.5281/zenodo.5569614
- Page 1
  - Find the different object types!
- The rest of the pages
  - With the data stated, fill your PREMIS semantic units.
  - Take help from the hand-out with all the semantic units!

PREMIS Events, Agents and Rights Entity – Exercise

- Exercise to get a feeling for the events, agents and rights!
  - https://doi.org/10.5281/zenodo.5569644
- For the pages
  - With the data stated, fill your PREMIS semantic units.
  - Take help from the hand-out with all the semantic units!

PREMIS Environments – Exercise

- Exercise to get a feeling for the environments!
  - https://doi.org/10.5281/zenodo.5569651
- For the pages
  - With the data stated, fill your PREMIS semantic units.
  - Take help from the hand-out with all the semantic units!

Extra example to look at on your own

- Main page: http://id.kb.dk/index_UK.html
- Showed example: http://id.kb.dk/metadata/structure.html
- Postcard: http://id.kb.dk/metadata/postcardExample.html
- Link to paper: https://mfr.de-1.osf.io/render?url=https://osf.io/kfetm/?direct%26mode=render%26action=download%26mode=render

Today

- You have had an introduction to PREMIS!
- Use the resources to learn more!
- Participate in the discussions!

menti.com

Finally...

PREMIS is a community standard.

- Send examples
- Ask questions
- Send suggestions
- Take part!

Thank you!

Karin, Eld and Micky