

UNDERSTANDING AND IMPLEMENTING PREMIS

A tutorial in two parts

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
- Answer questions:
 - We might need to put them in parking but all questions will be answered!

Who are we?

From the PREMIS Editorial Committee:

- Karin
- Eld
- Micky

Form the audience via show of hands or online reactions:

- Have you heard of PREMIS?
- Do you know the PREMIS data model?
- Have you used PREMIS in practice?
- Are you from
 - a library?
 - an archive?
 - a university?
 - Something else?



What do we want you to think about?

- During this first part 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

9:30-10:15

Introduction to PREMIS

Welcome

Background (brief history and rationale)

On-line presence

Benefits of implementing PREMIS

10:15-11:00

Implementation

Outline of the main entities and the data model

Your use cases

Break

11.00-11.30

11.30-12.45

Implementation and case studies

PREMIS Conformance & interoperability

Data Dictionary

12:45-13.00

Wrap Up

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

Answers to questions

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 (after main presentation in slide deck)

Background
(brief history
and rationale)

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



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 entities 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
 - 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



Background
(brief history
and rationale)

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

Off-the-shelf systems

- Predefined metadata profiles
- Out-of-the-box tools

Configured, extended, adapted

- Metadata profiles and tools

Custom-built systems

- Metadata profiles and tools

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_\(Preservation_Metadata_Implementation_Strategies\)](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>

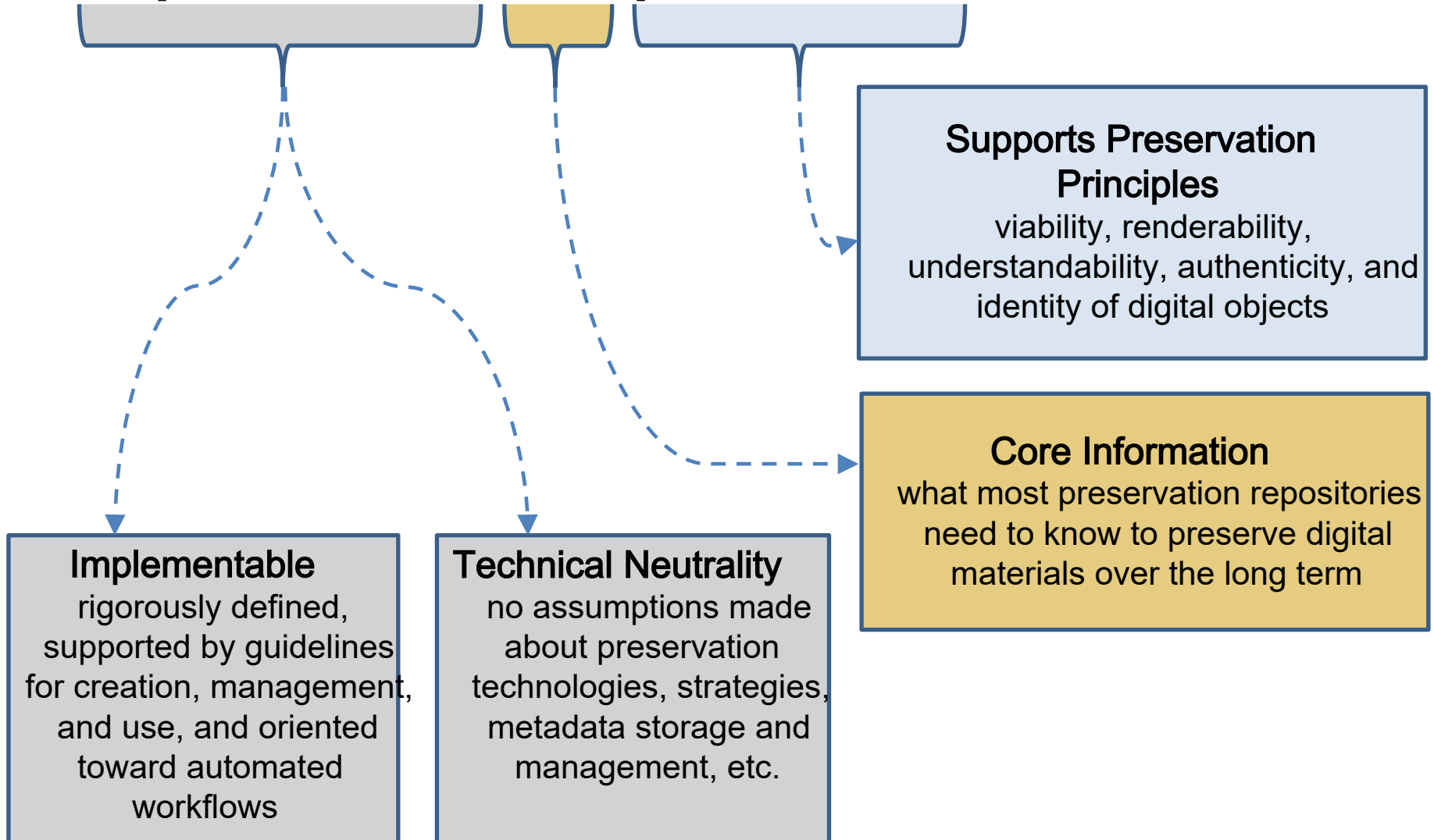
Micky Lindlar

TIB – Leibniz Information Centre for Science and
Technology

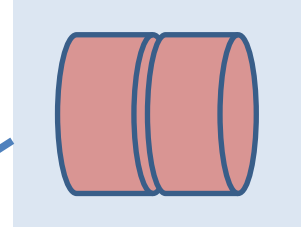


BENEFITS OF IMPLEMENTING PREMIS

Implementable, core preservation metadata



Benefits of implementing PREMIS (1)

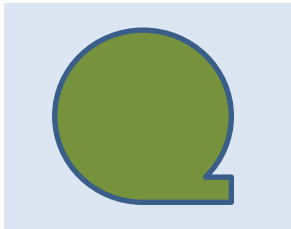


`IDobj="hdl.galapagos.6754.1"`

`hdl="galapagos.6754.1"`

`ID="6754.1"`
`GID="galapagos"`

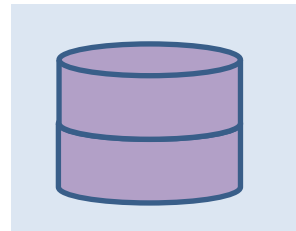
`ID="1"`
`TYPE="hdl"`
`GROUPID="galapagos.6754"`



Interoperability through implementable preservation metadata

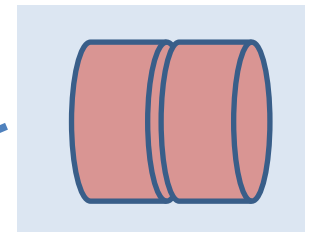
IDobj="hdl.galapagos.6754.1"

hdl="galapagos.6754.1"



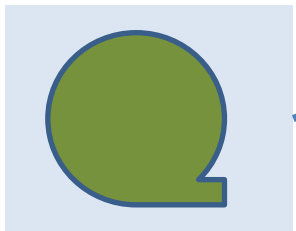
```

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



```

<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 .
  
```



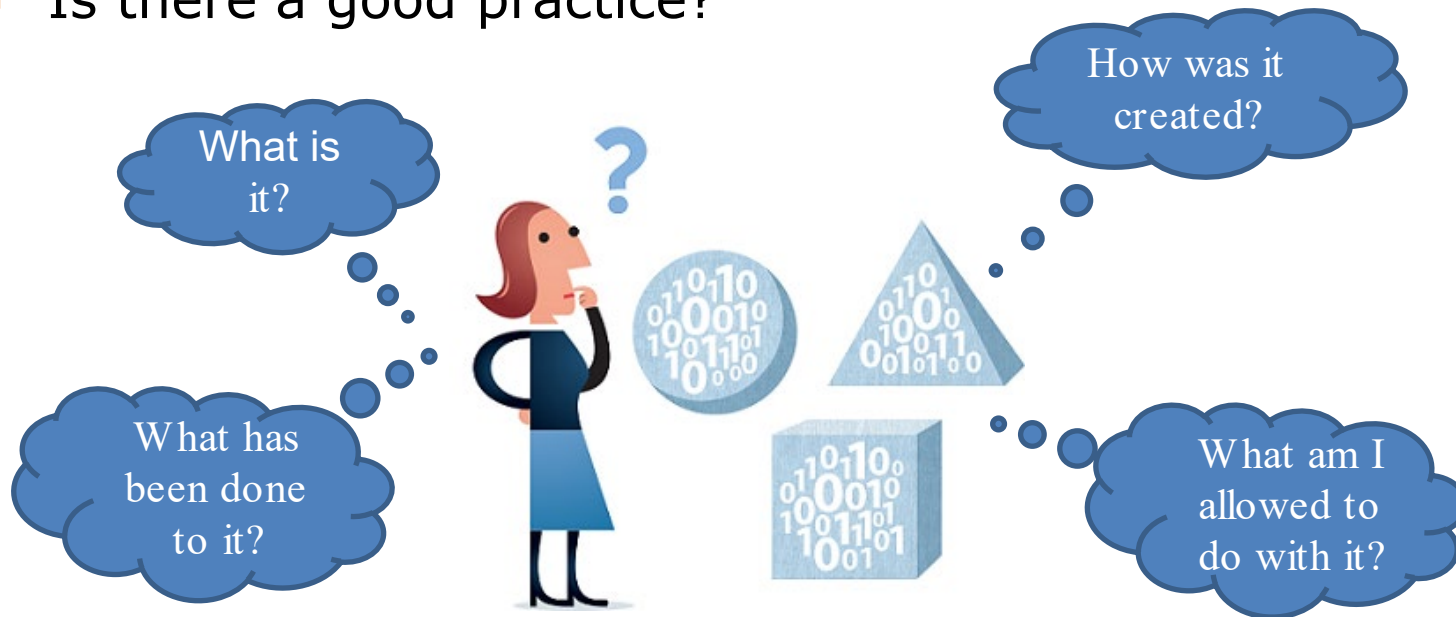
ID="6754.1"
GID="galapagos"



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

Benefits of implementing PREMIS (2)

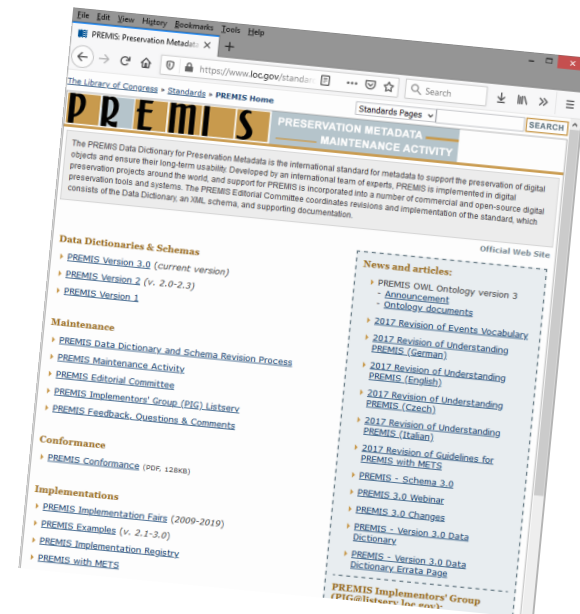
- What do I need to know about a digital object I am preserving?
 - ... today
 - ... in the future
- How can I best capture that information?
- Is there a good practice?



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



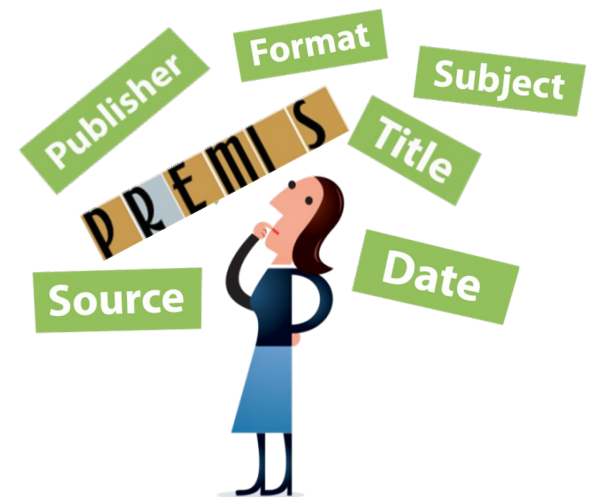
Eld Zierau

Royal Danish Library



OUTLINE OF MAIN ENTITIES

THE DATA MODEL & KEY CONCEPTS





The PREMIS Data Model



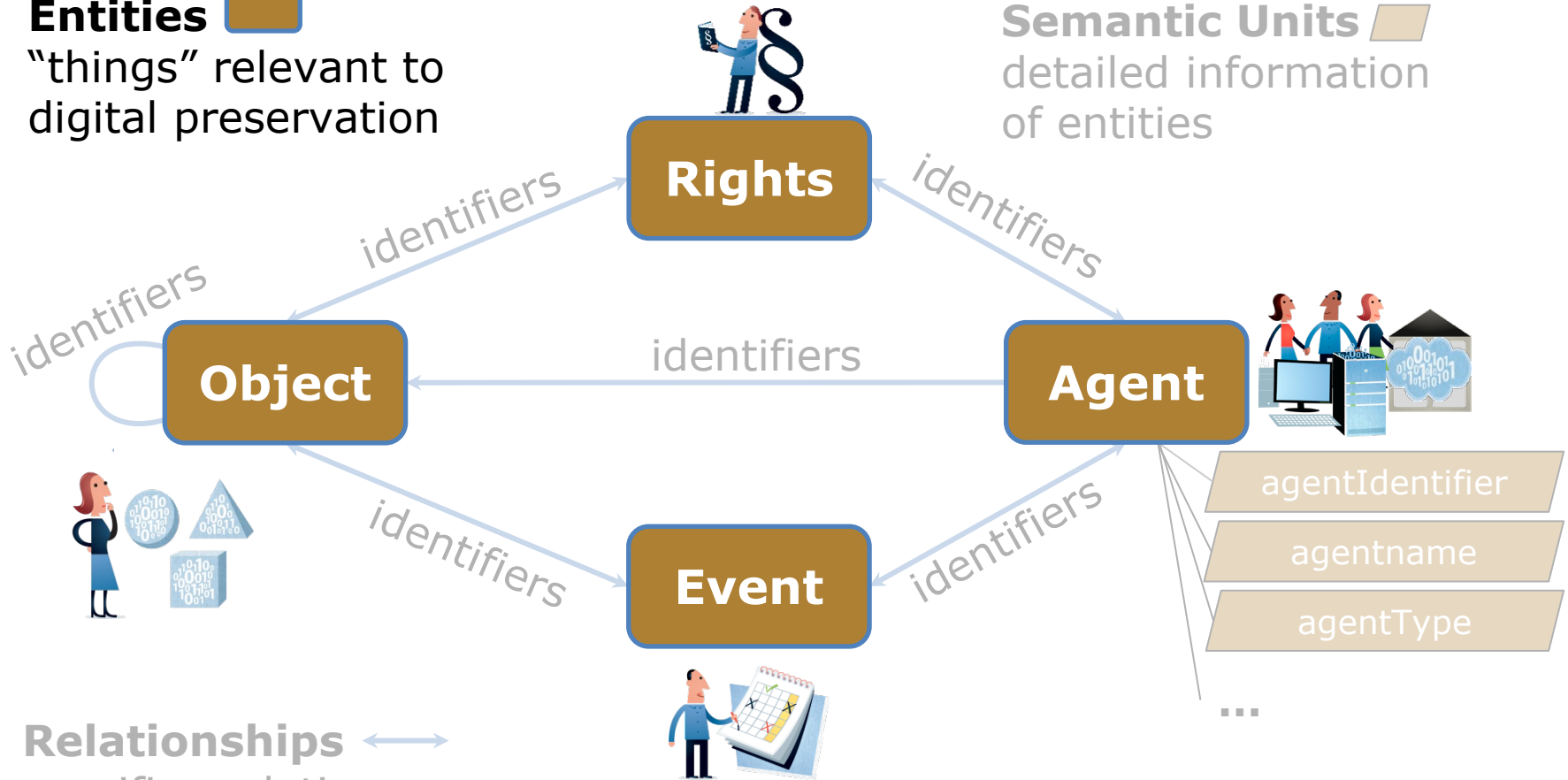
- Why have a data model?
 - Organizational convenience (for development and use)
 - But: not a formal entity-relationship model; not sufficient to design databases

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

PREMIS - Data model includes:

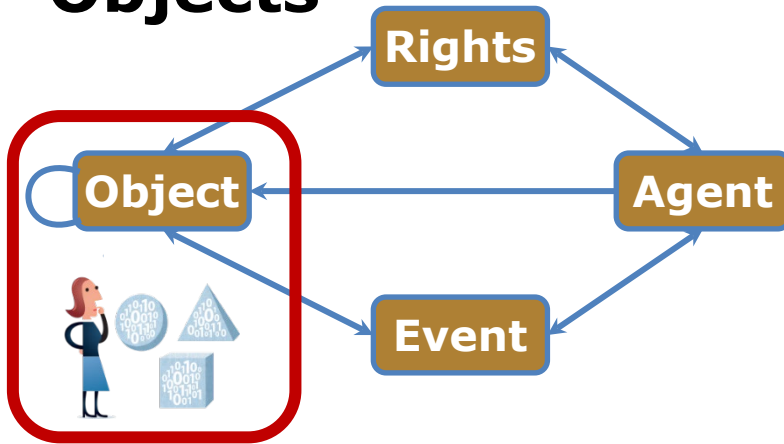
Entities 
 "things" relevant to digital preservation

Semantic Units 
 detailed information of entities



Relationships 
 specifies relations between entities

Objects



Examples:

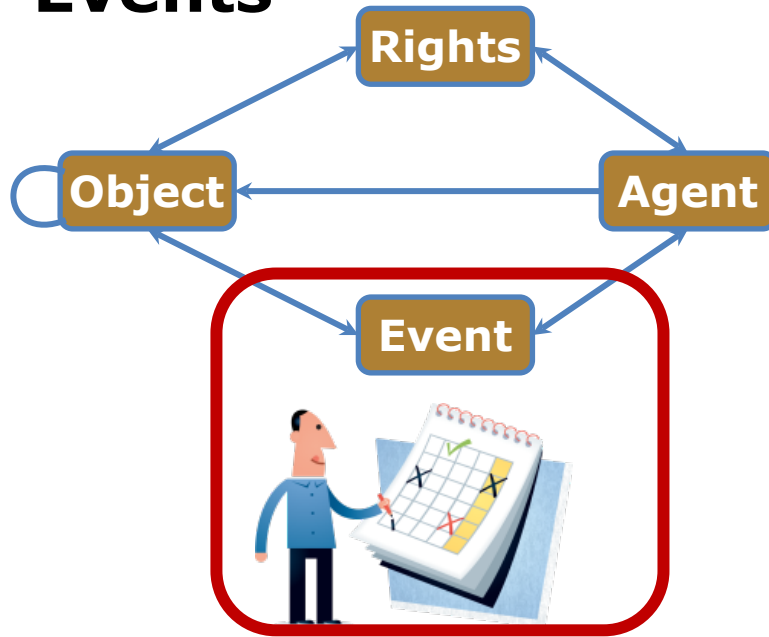
- A PDF file
- An image within a file
- A book
- A book representation

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

Implementation choices :

- A repository does NOT have to manage all types of Objects

Events



Examples:

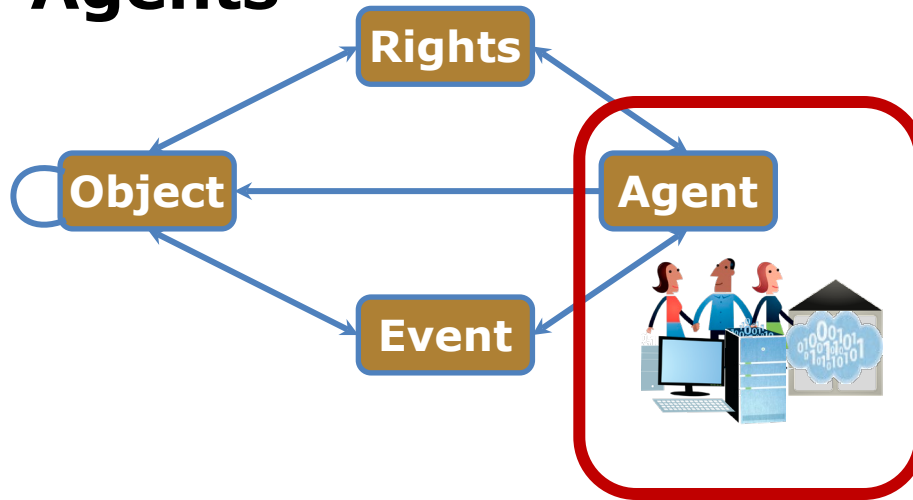
- Validation Event
e.g. using JHOVE
- Ingest Event

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

Implementation choices:

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

Agents



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

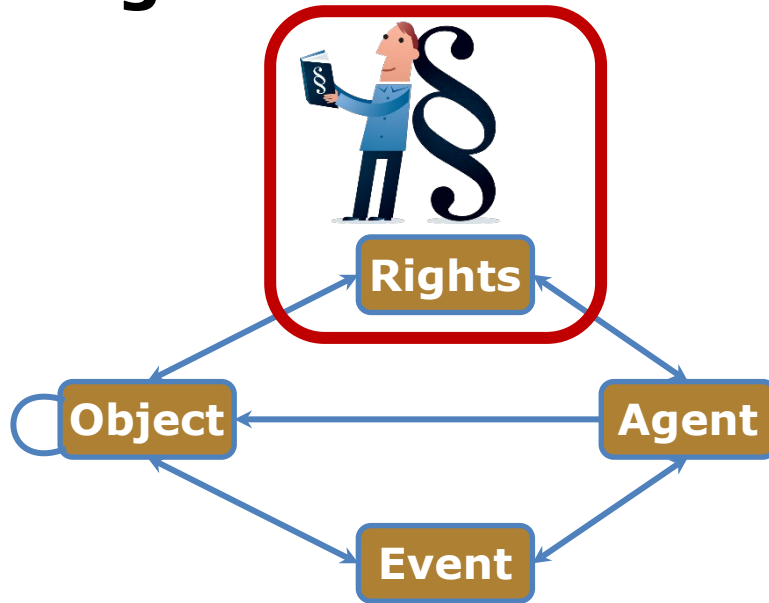
Examples:

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

Implementation choices:

- use richer scheme that may be appropriate.

Rights Statements



Example:

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

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

Implementation choices :

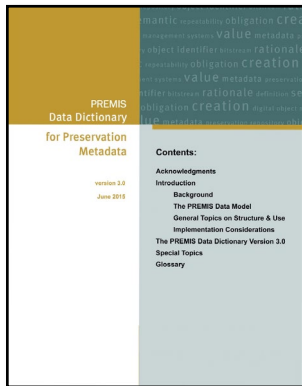
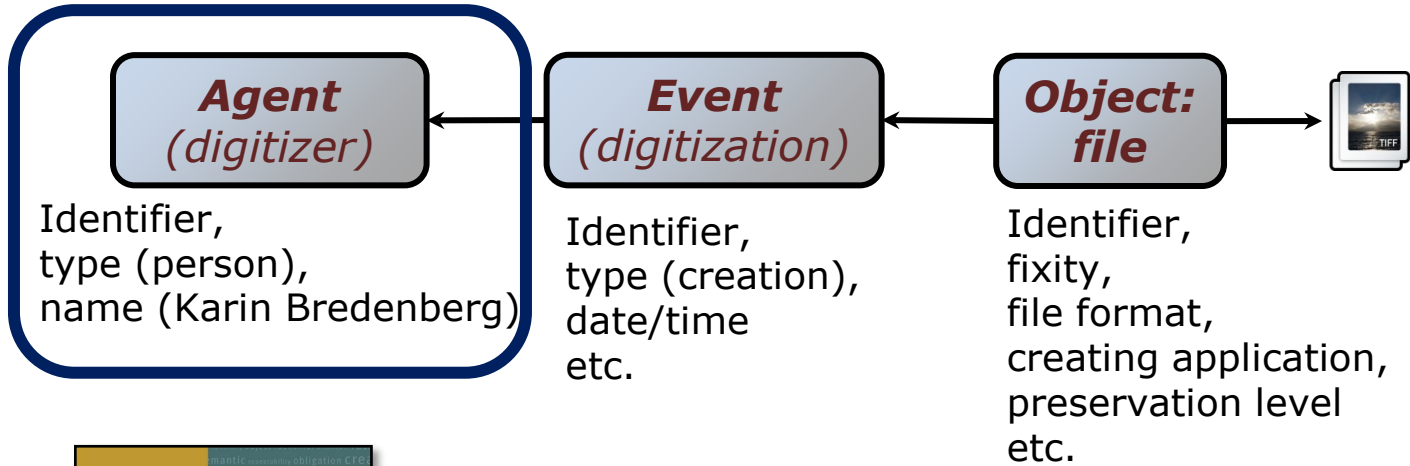
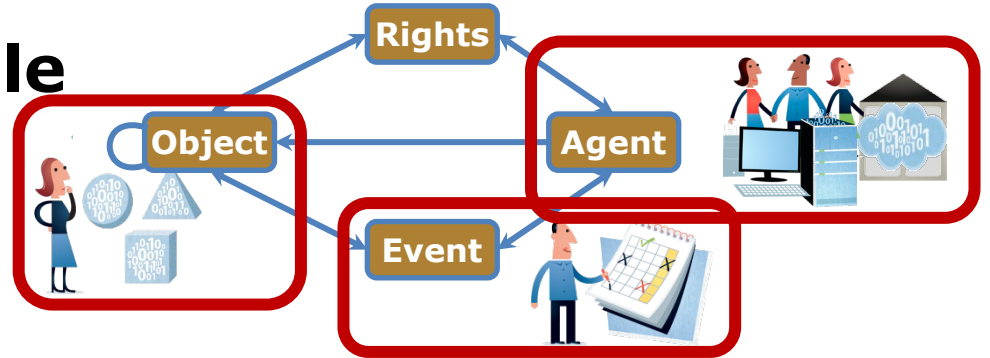
- Can use other schemes

PREMIS form:

Agent X

grants **Permission Y** to the repository in regard to Object **Z**.

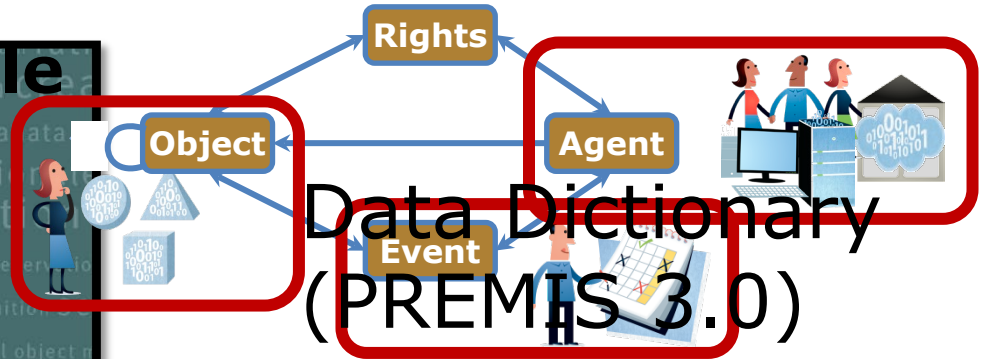
A digitization example



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 app
preservatio
etc.

The PREMIS Data Dictionary Version 3.0

Data Dictionary
(PREMIS 3.0)

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

The only mandatory semantic unit is *agentIdentifier*.



Entity properties

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.

Sample Data Dictionary Entry



Entity semantic units

3.1 agentIdentifier (M, R)

Semantic unit	3.1 agentIdentifier
Semantic components	3.1.1 agentIdentifierType 3.1.2 agentIdentifierValue

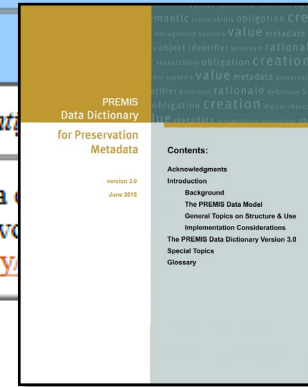
agentIdentifier

Semantic unit	3.1.1 agentIdentifierType
Semantic components	None
Definition	A designation of the domain of the <i>agentIdentifier</i> property.
Data constraint	Value should be taken from a controlled vocabulary.
Examples	LCNAF SAN DLC URI local
Repeatability	Not repeatable
Obligation	Mandatory

agentIdentifierType

agentIdentifierValue

Semantic unit	3.1.2 agentIdentifierValue
Semantic components	None
Definition	The value of the <i>agentIdentifier</i> property.
Data constraint	Value may be taken from a controlled vocabulary, such as http://id.loc.gov/vocabulary/agents .
Examples	92-79971 Owens, Erik C. 234-5676 MH-CS info:lccn/n78890351
Repeatability	Not repeatable
Obligation	Mandatory
Usage notes	May be a unique key or a controlled textual form of name.



agentType is

Data Dictionary for Preservation Metadata

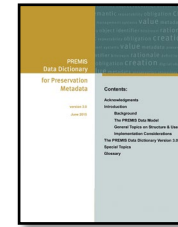
3.1.1 agentIdentifier

3.1.2 agentIdentifierValue

3.2 agentName (O, R)

XML example

AgentIdentifier
 AgentIdentifierType
 AgentIdentifierValue



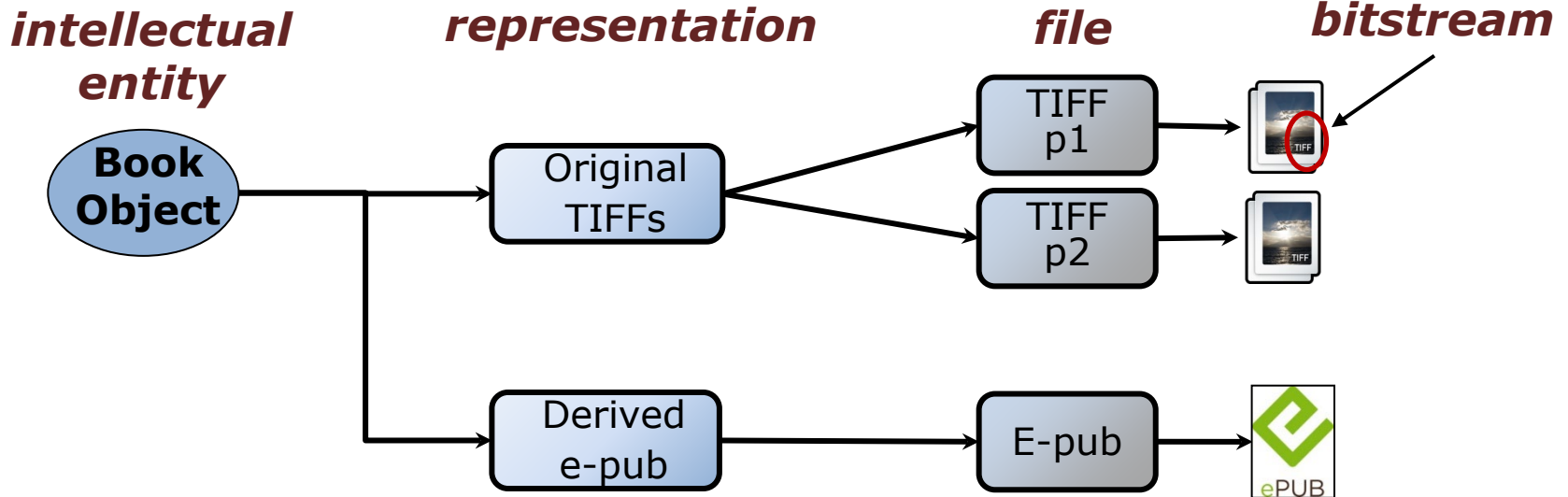
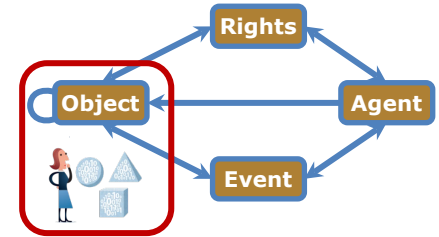
Exemplified by XML using XML schema v3.0:

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

```
<premis>
  <object xsi:type="file" ... </object> ←
  <event> ... </event>
  <agent> ... </agent>
    <agentIdentifier> ←
    <rights> </rights> ←
  </premis>
  <agentIdentifierType> ←
  <agentIdentifierValue> ←
    41d10-099-1e2-9
  </agentIdentifierValue>
</agentIdentifier>
  ...
</agent>
```



objectCategory (types of objects)



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

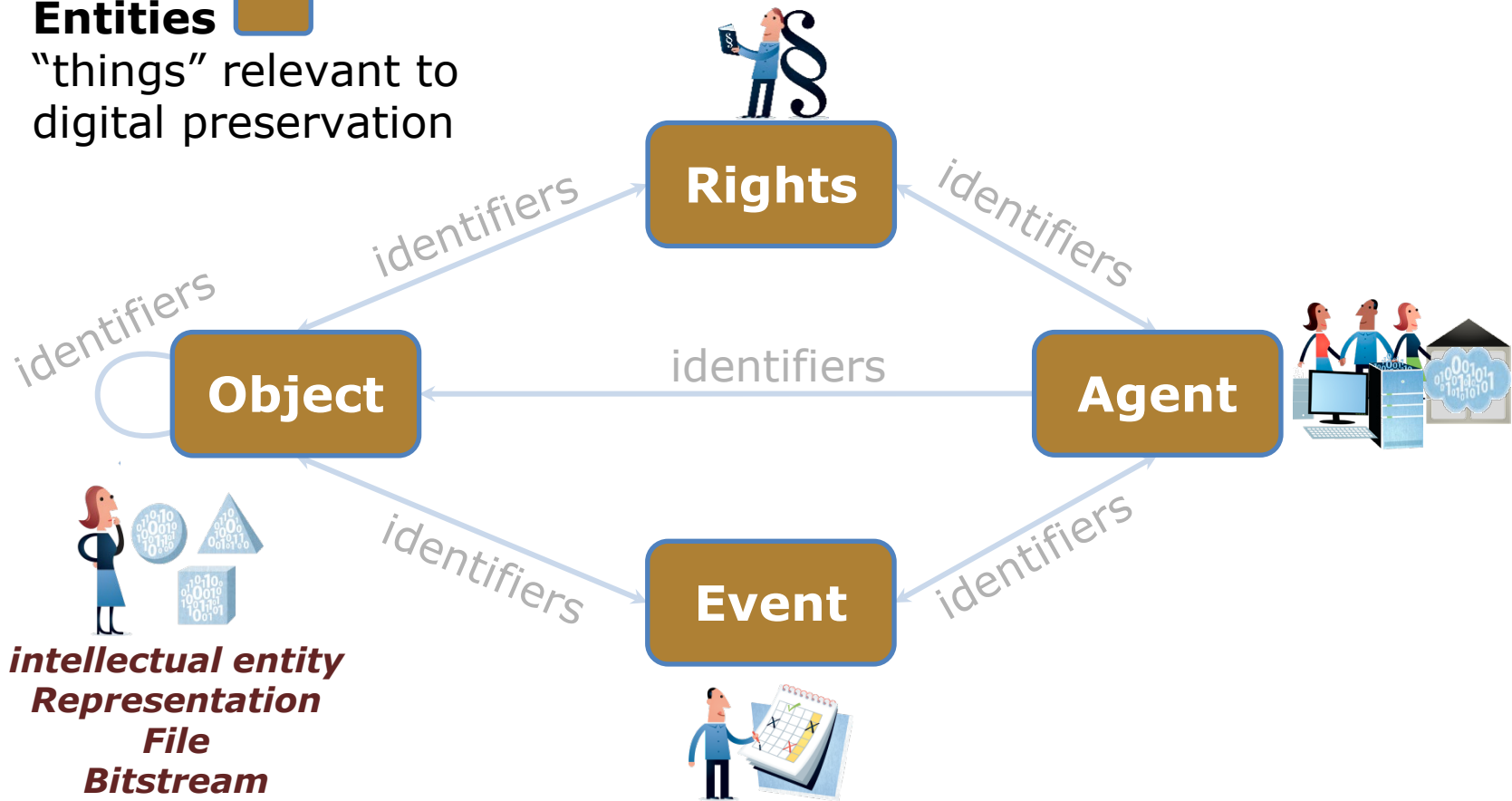
named and ordered sequence of bytes that is known by an operating system

data within a file with properties relevant for preservation purposes

PREMIS 3 - Data model includes:

Entities

"things" relevant to digital preservation



WHAT USE CASES DO YOU SEE?





LET'S HAVE A BREAK!
11.00-11.30

PREMIS
Conformance &
interoperability

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?

✗ 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</objectIdentifier>`

✓ `premis:objectIdentifier`
`premis:objectIdentifierType=„doi“`
`premis:objectIdentifierValue=„10.5281/zenodo.5569542“`

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

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.

Use case: For building other standards

- As a basis extended with locally-defined elements: Preservation Metadata Dictionary (Netherlands Institute for Sound and Vision).
https://publications.beeldengeluid.nl/pub/389/BIJLAG E-C_Metadatadictionary-English.pdf
- As a free source of inspiration: DEPIP (Data Exchange Protocol for Interoperability and Preservation), ISO 20614.
<https://www.iso.org/standard/68562.html>
- No conformance, inspiration!

Use case: As a self-assessment tool

- Am I able to provide information about my digital assets following the Data Dictionary structure and requirements?
 - I.e., documenting the mapping between my metadata structure and PREMIS semantic units.
- Conformance level 1 « through mapping » (see the Conformance Statement document, p. 5: <http://www.loc.gov/standards/premis/premis-conformance-20150429.pdf#page=5>)

Use Case: As an export format

- Preferably in a PREMIS-endorsed expression (XML or RDF)
- Conformance level 2 « through export » (see the Conformance Statement document, p. 5:
<http://www.loc.gov/standards/premis/premis-conformance-20150429.pdf#page=5>)

Use case: As the native format of the repository Data Management module

Where to store PREMIS data?

- Any technology, using a PREMIS-endorsed expression or not, can be used
 - Excel
 - CVS file
 - XML database
 - RDF triple store
 - (relational) database
 - etc.
- Conformance level 3 « through internal implementation » (see the Conformance Statement document, p. 6: <http://www.loc.gov/standards/premis/premis-conformance-20150429.pdf#page=6>)

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	71f920fa275127a7b60fa4d4d41432a3	NRI
3				
4				

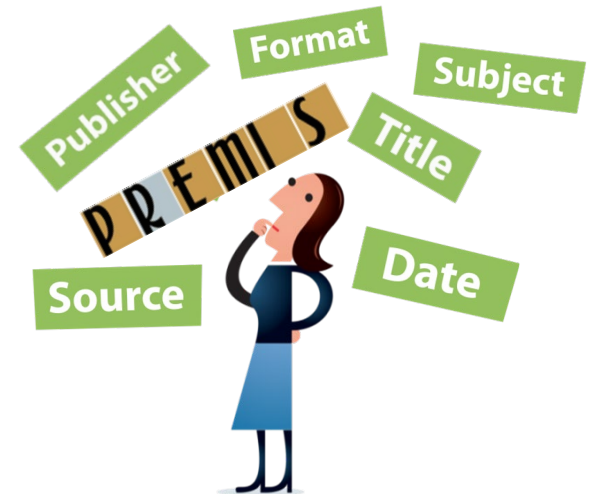
Eld Zierau

Royal Danish Library



DATA DICTIONARY

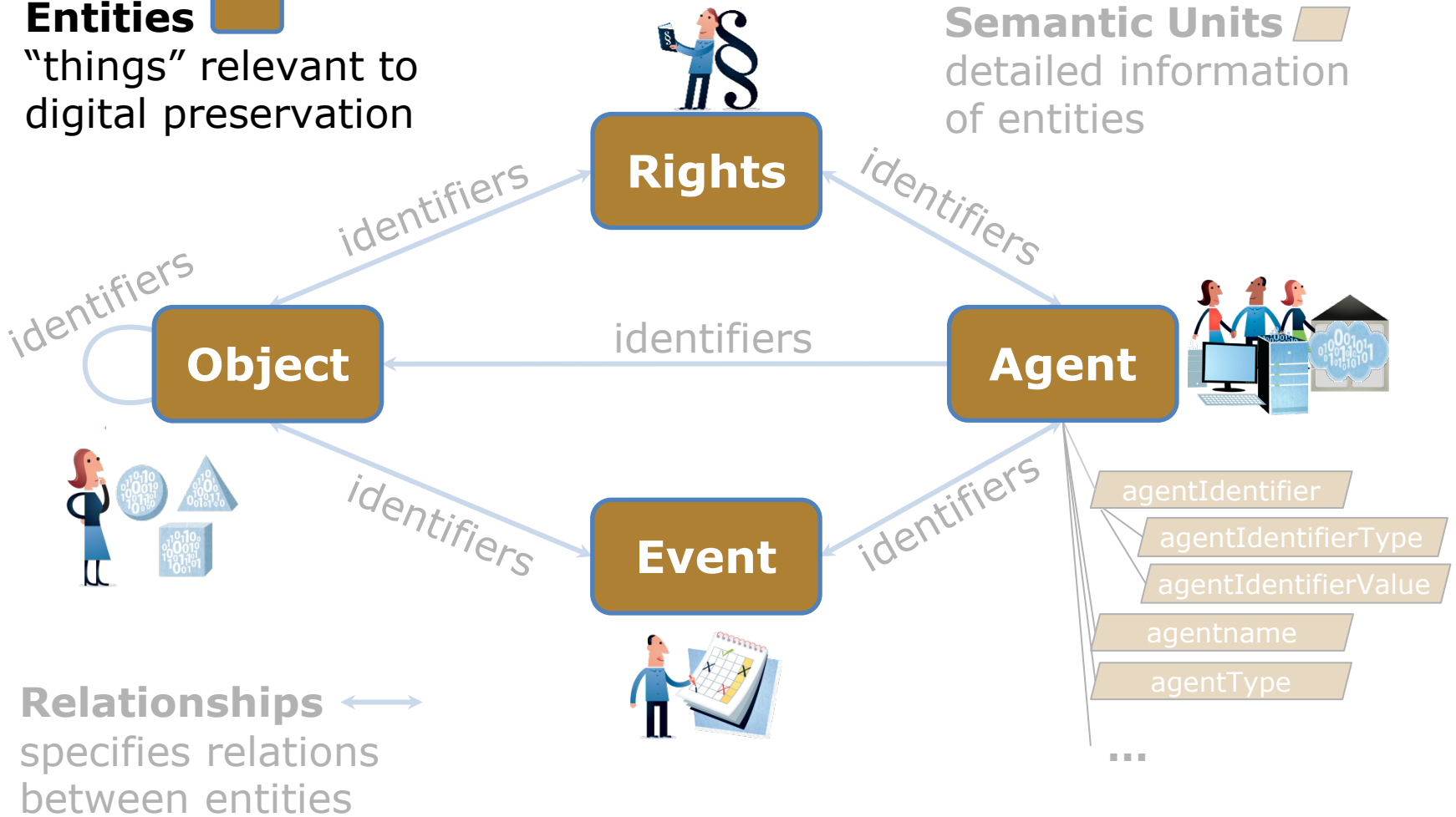
DESCRIPTION OF DATA MODEL



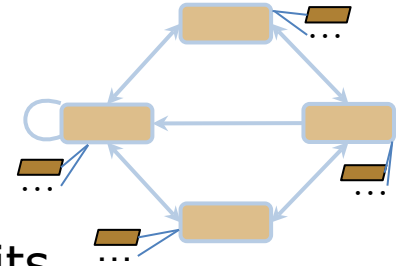
PREMIS - Data model includes:

Entities 
 "things" relevant to digital preservation

Semantic Units 
 detailed information of entities



Semantic Units



- Two kinds of semantic unit:
 - **Container:** groups together related semantic units
 - **Semantic components:** semantic units grouped under the same container

- **Example:**

AgentIdentifier [container]

AgentIdentifierType [semantic component]

AgentIdentifierValue [semantic component]

Semantic Units

A semantic unit is a property of an entity

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

Two kinds of semantic unit

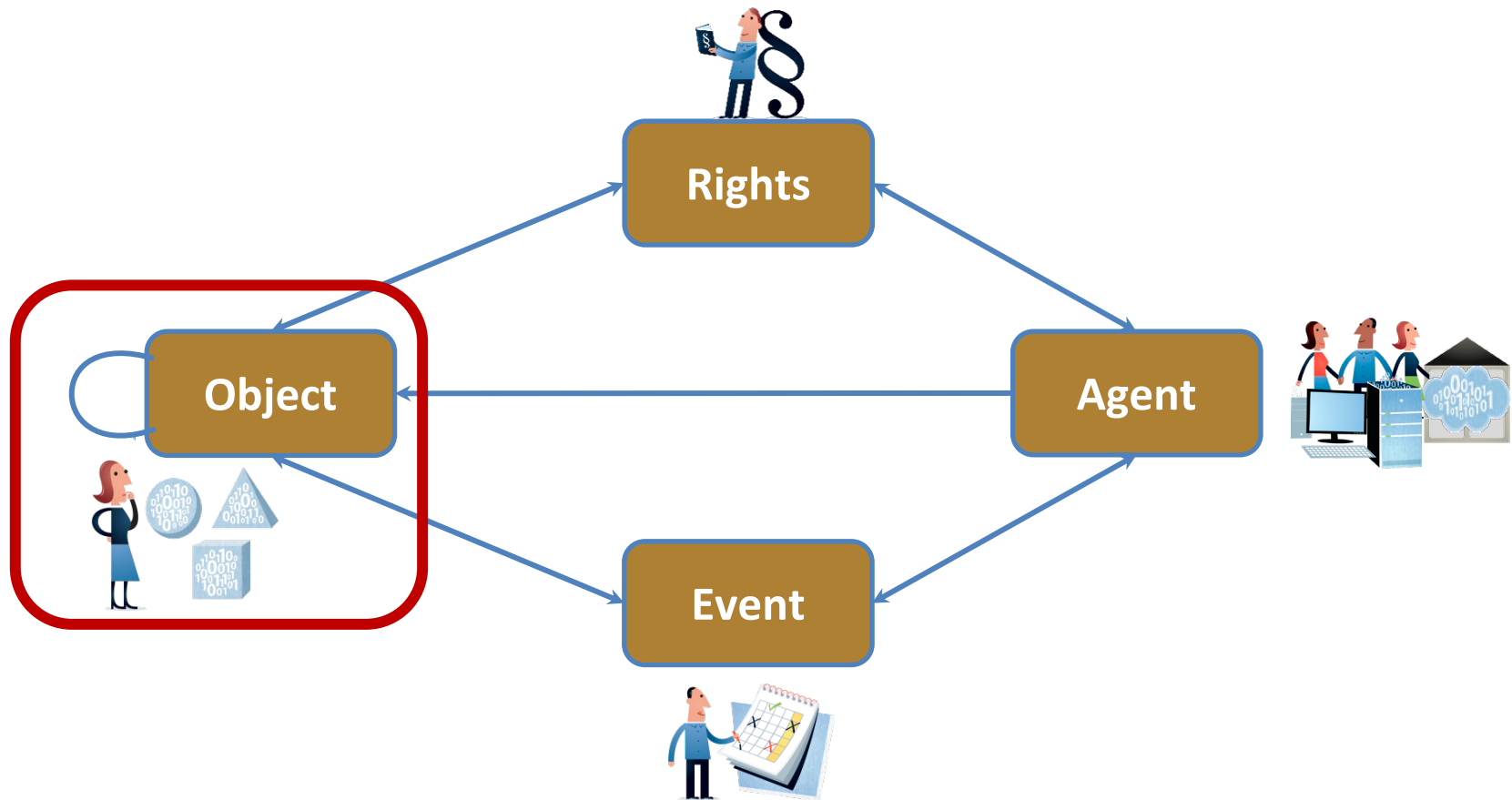
- Example:

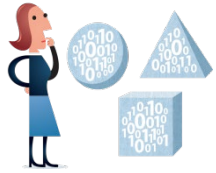
`agentIdentifier` [container] groups together related semantic units

`agentIdentifierType` [semantic component]

`agentIdentifierValue` [semantic component]

Semantic units - Properties of Entities





High level semantic units for Objects

what technical information on it?



which object is it?

`ark:/12148/btp6k102002g/f1`

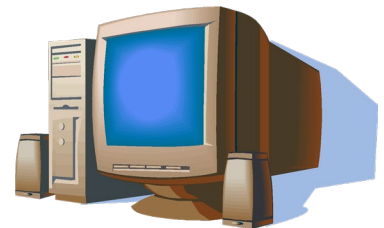
what kind of object?



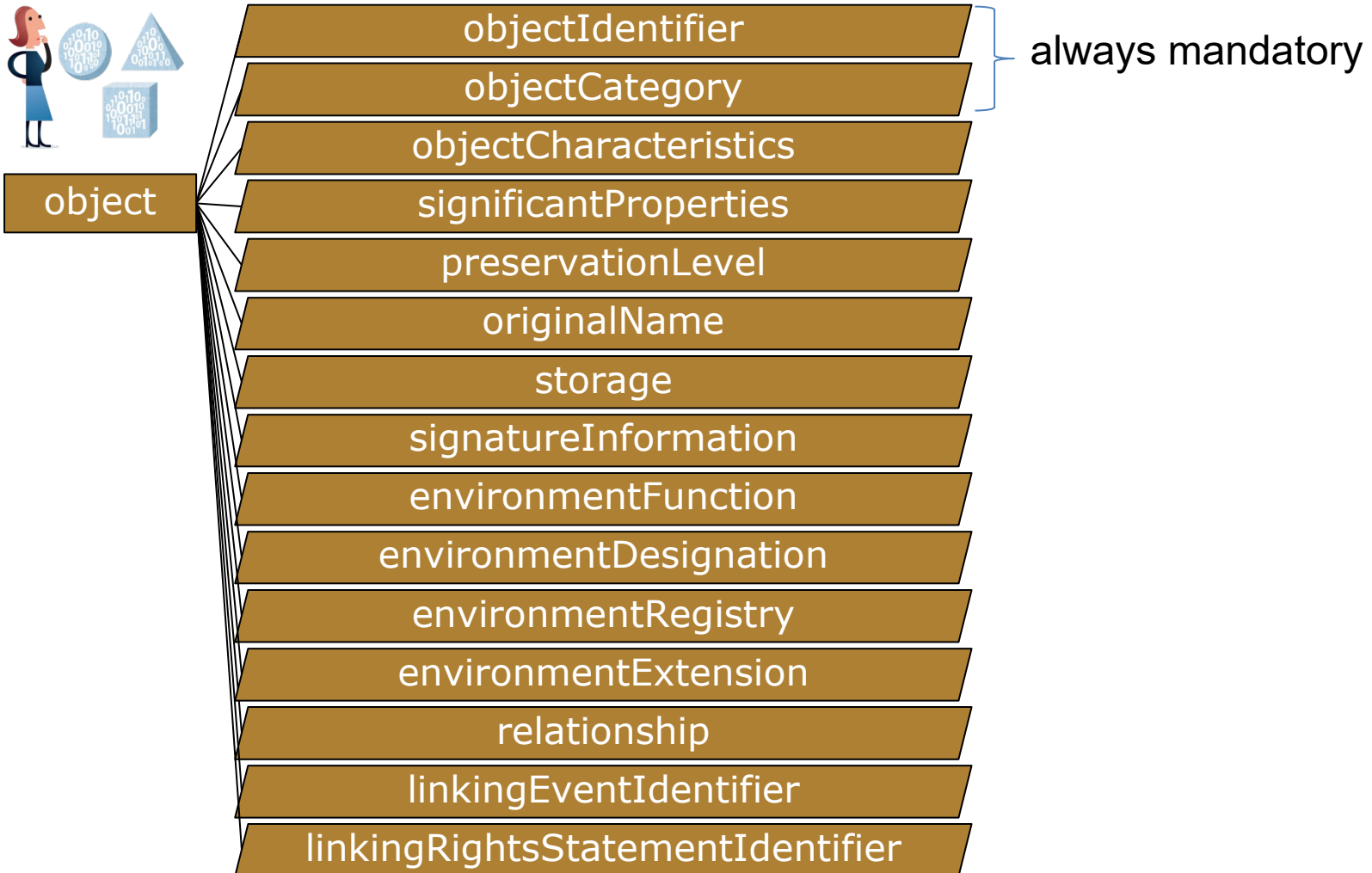
which of its characteristics do I want to preserve in it?

what is my preservation strategy for this object?

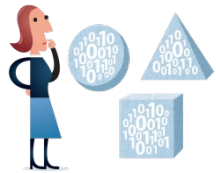
what software or hardware should be used to handle the object?



PREMIS Object Entity – Semantic Units



PREMIS Object Entity – Semantic Units



object

objectIdentifier

objectCategory

objectCharacteristics

significantProperties

preservationLevel

originalName

storage

signatureInformation

environmentFunction

environmentDesignation

environmentRegistry

environmentExtension

relationship

linkingEventIdentifier

linkingRightsStatementIdentifier

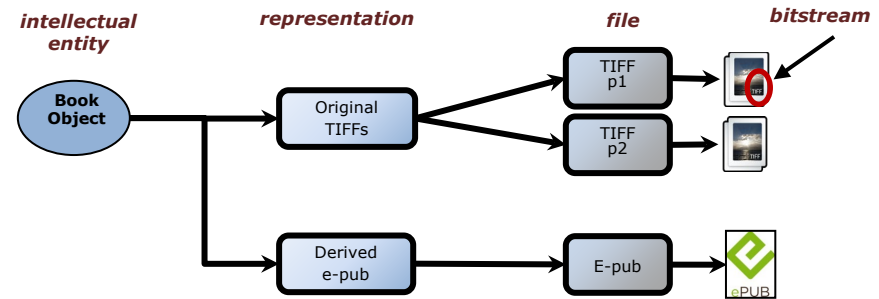
objectIdentifierType

objectIdentifierValue

objectCategory



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



- Implemented as an xsi:type in PREMIS XML-schema so not explicitly recorded

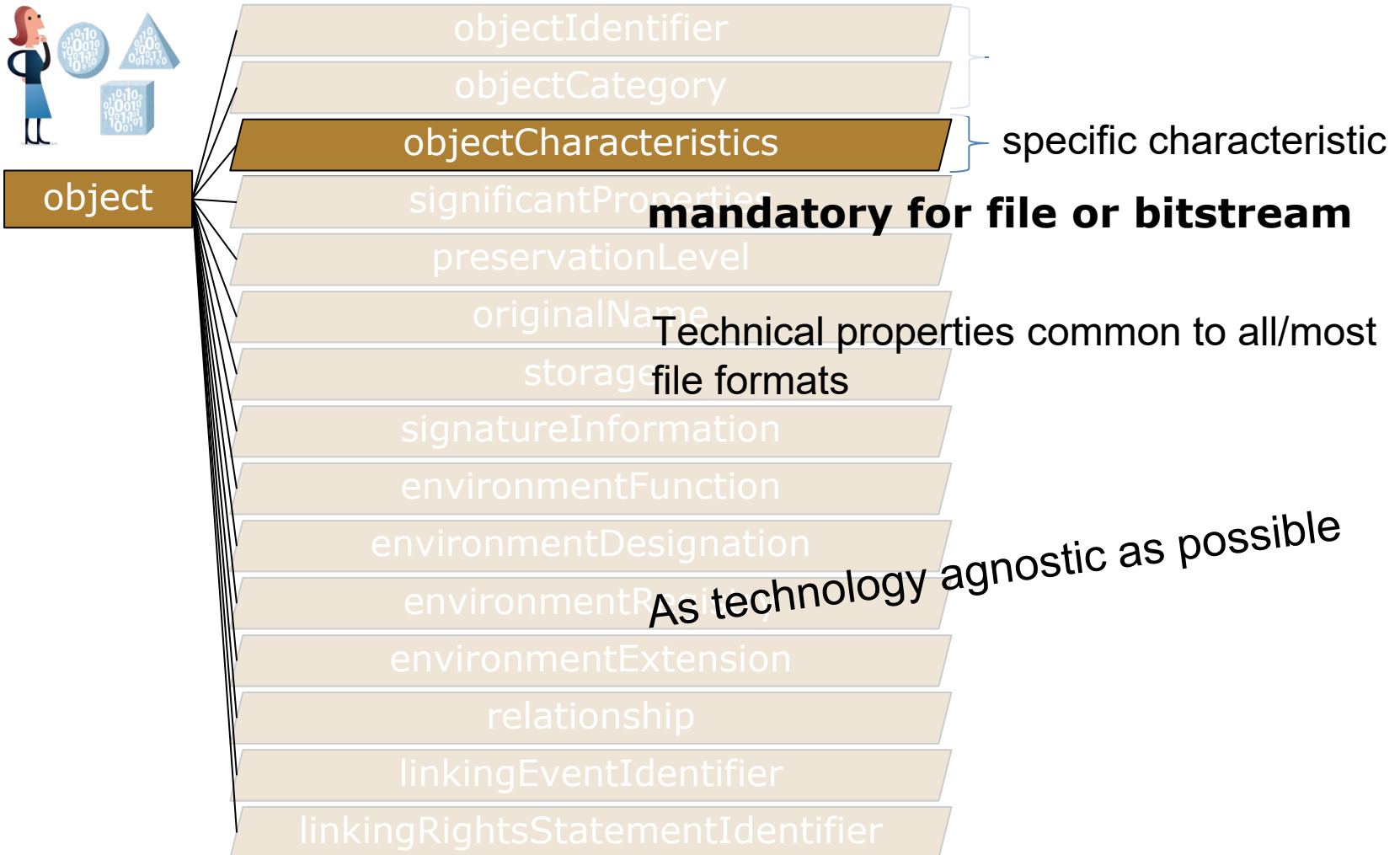
```
<premis>
```

```
<object xsi:type="file">
```

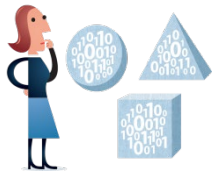
```
  ...
</object>
```

```
  ...
</premis>
```

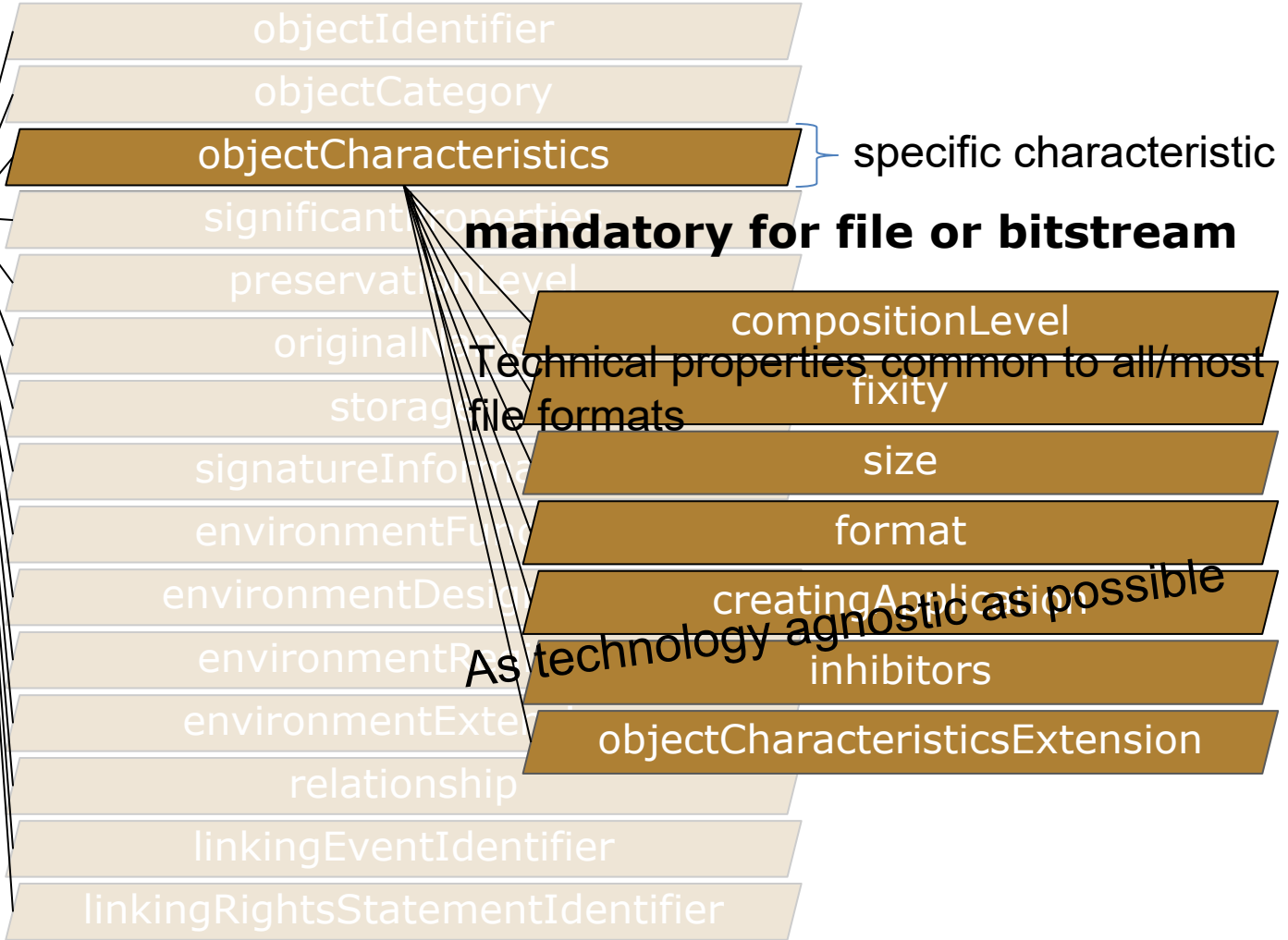

PREMIS Object Entity – Semantic Units



PREMIS Object Entity



object



Composition Level

sometimes there is more than one layer of characteristics



Or they be part of other files e.g.

- Mail attachments
- Images in PDF's etc

objectCharacteristicsExtension

Container to include externally defined semantic units – e.g. for more granularity

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

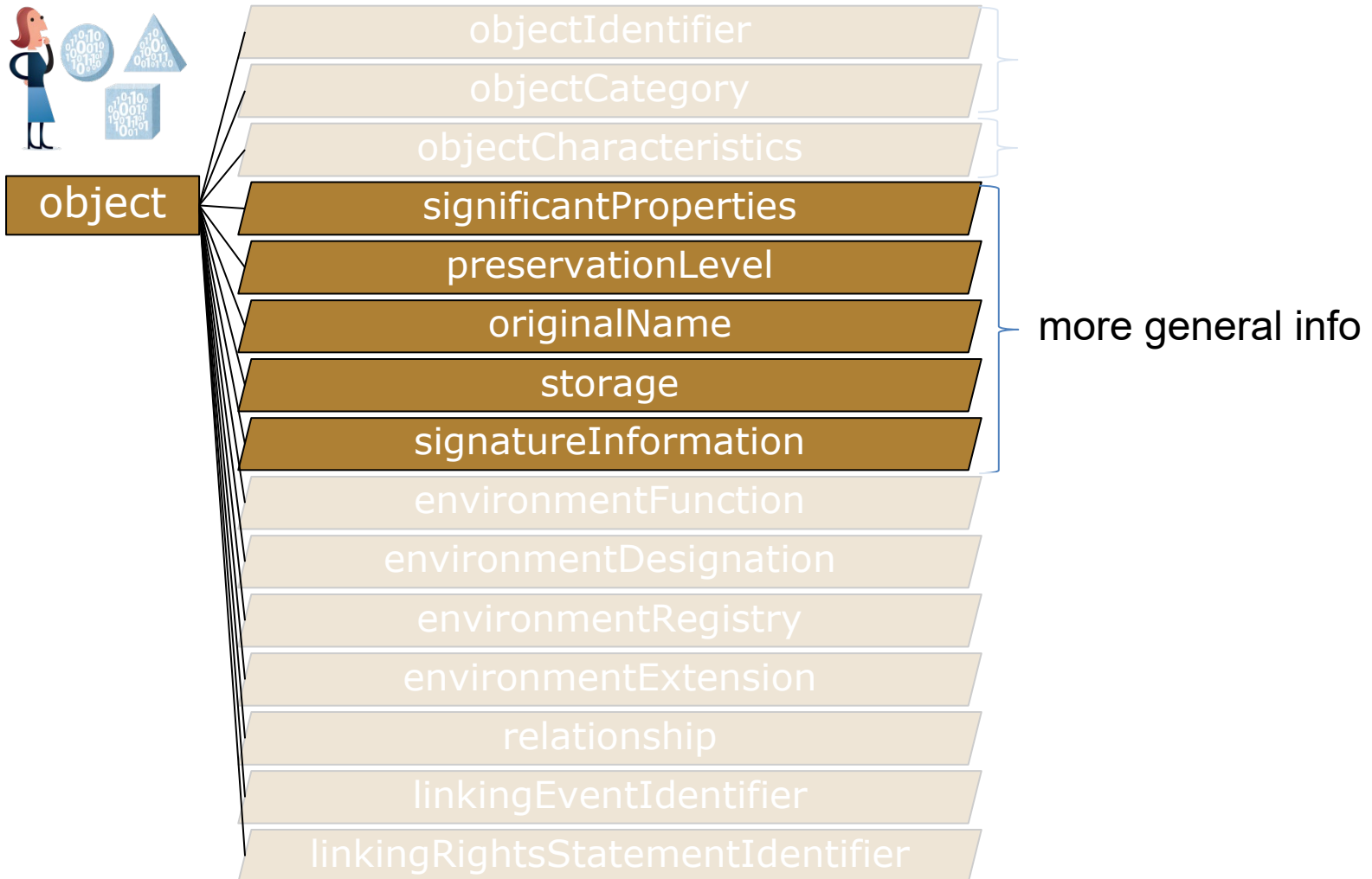


objectCharacteristicsExtension - example

```
<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



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

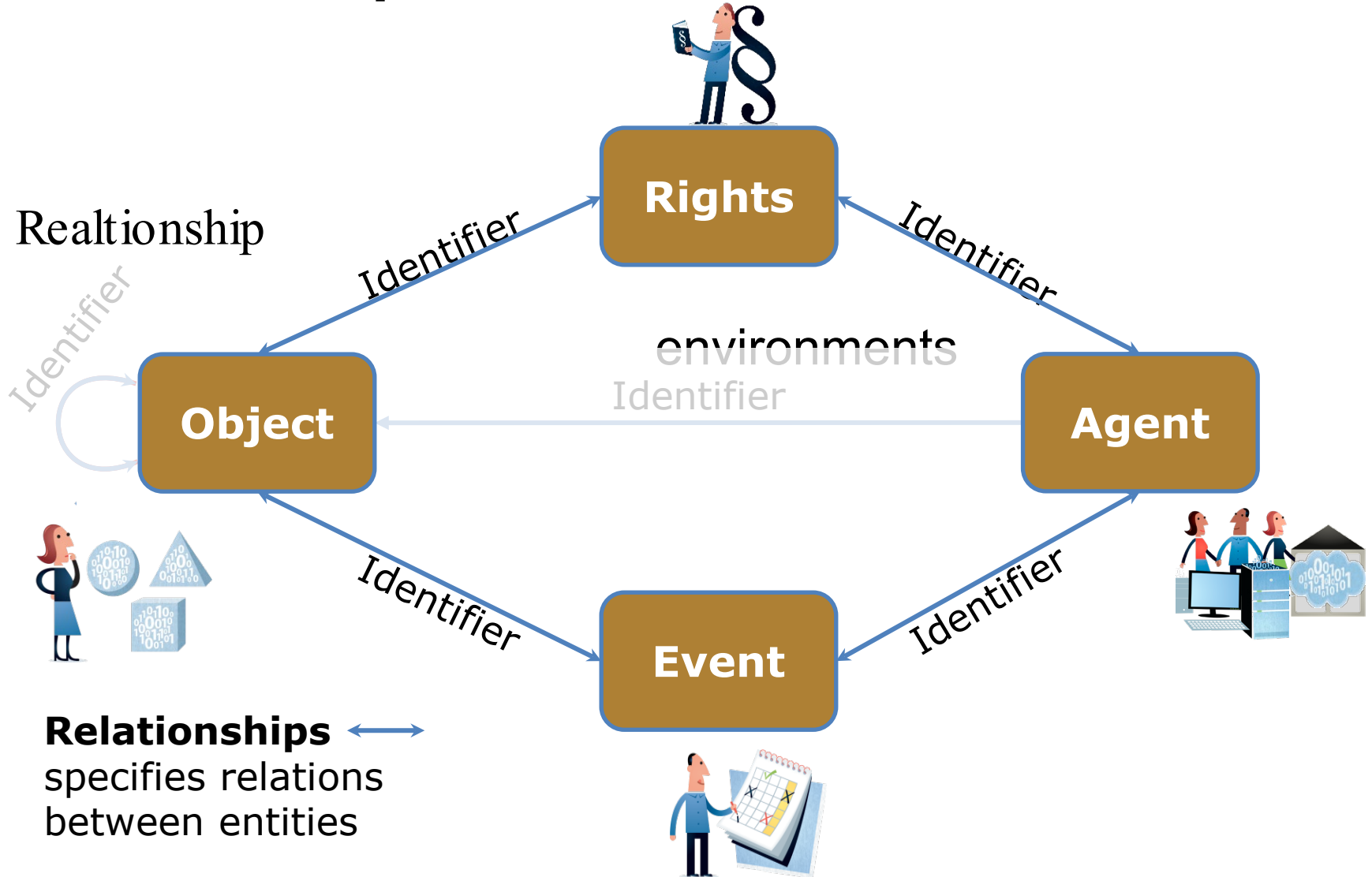
Contains

- **preservationLevelType**, e.g. *logicalStrategy* or *BitSafety*
- **preservationLevelValue**, e.g. *migration* or *High*
- **preservationLevelRole** (context), e.g. *intention* or *requirement*
- **preservationLevelRationale**, when differs from policy
- **preservationLevelDateAssigned** when Level was assigned

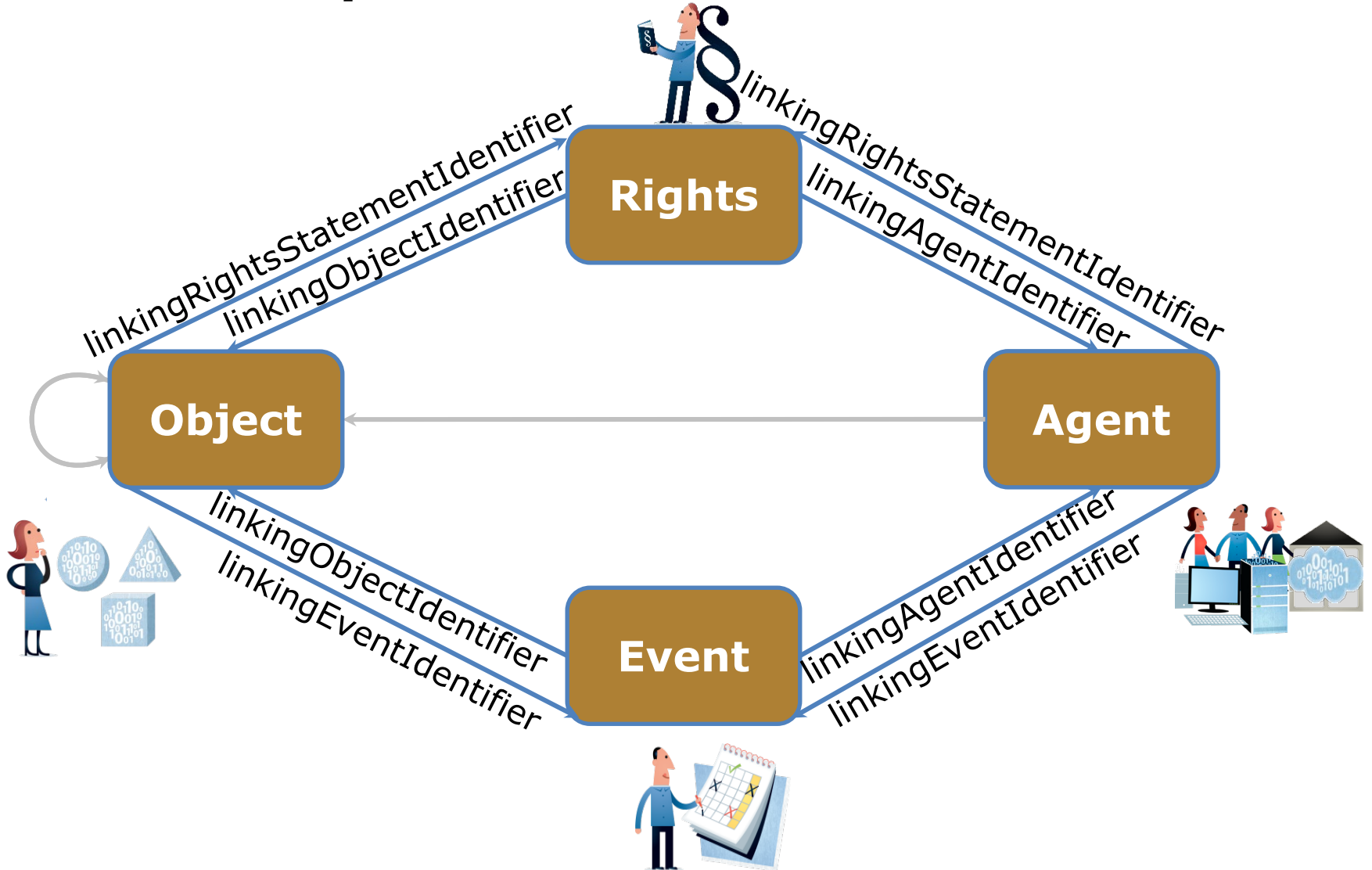
PREMIS Object Entity – Semantic Units



Relationships: Semantic Unit Identifiers



Relationships: Semantic Unit Identifiers



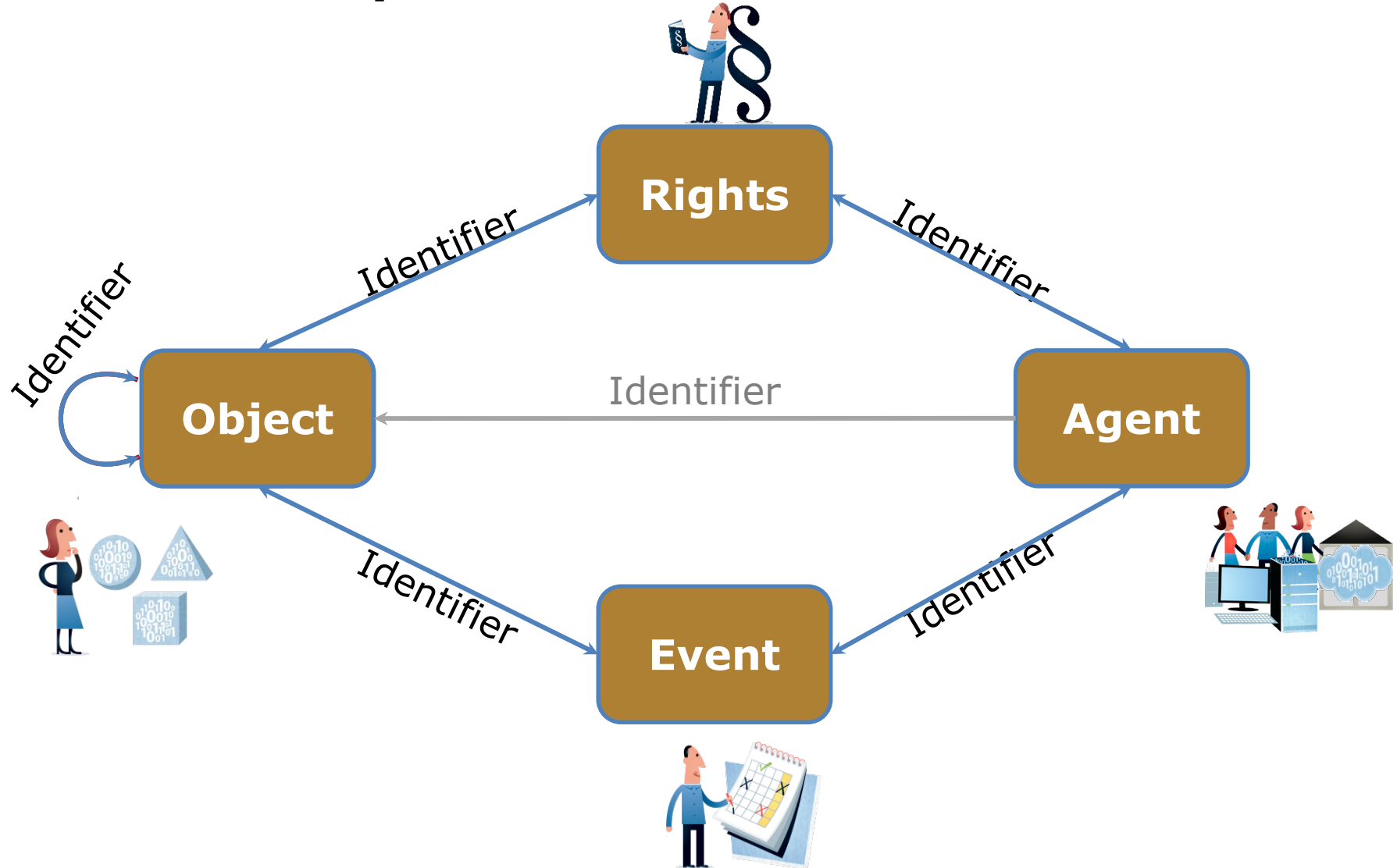
Linking Objects with Agents and Events

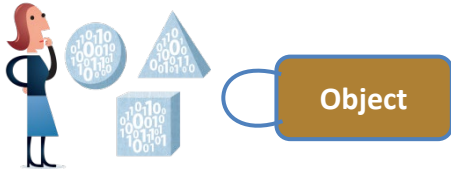
- **linkingRightsStatementIdentifier**
 - **linkingRightsStatementIdentifierType**
 - **linkingRightsStatementIdentifierValue**
- **linkingEventIdentifier**
 - **linkingEventIdentifierType**
 - **linkingEventIdentifierValue**

Likewise for the other entities

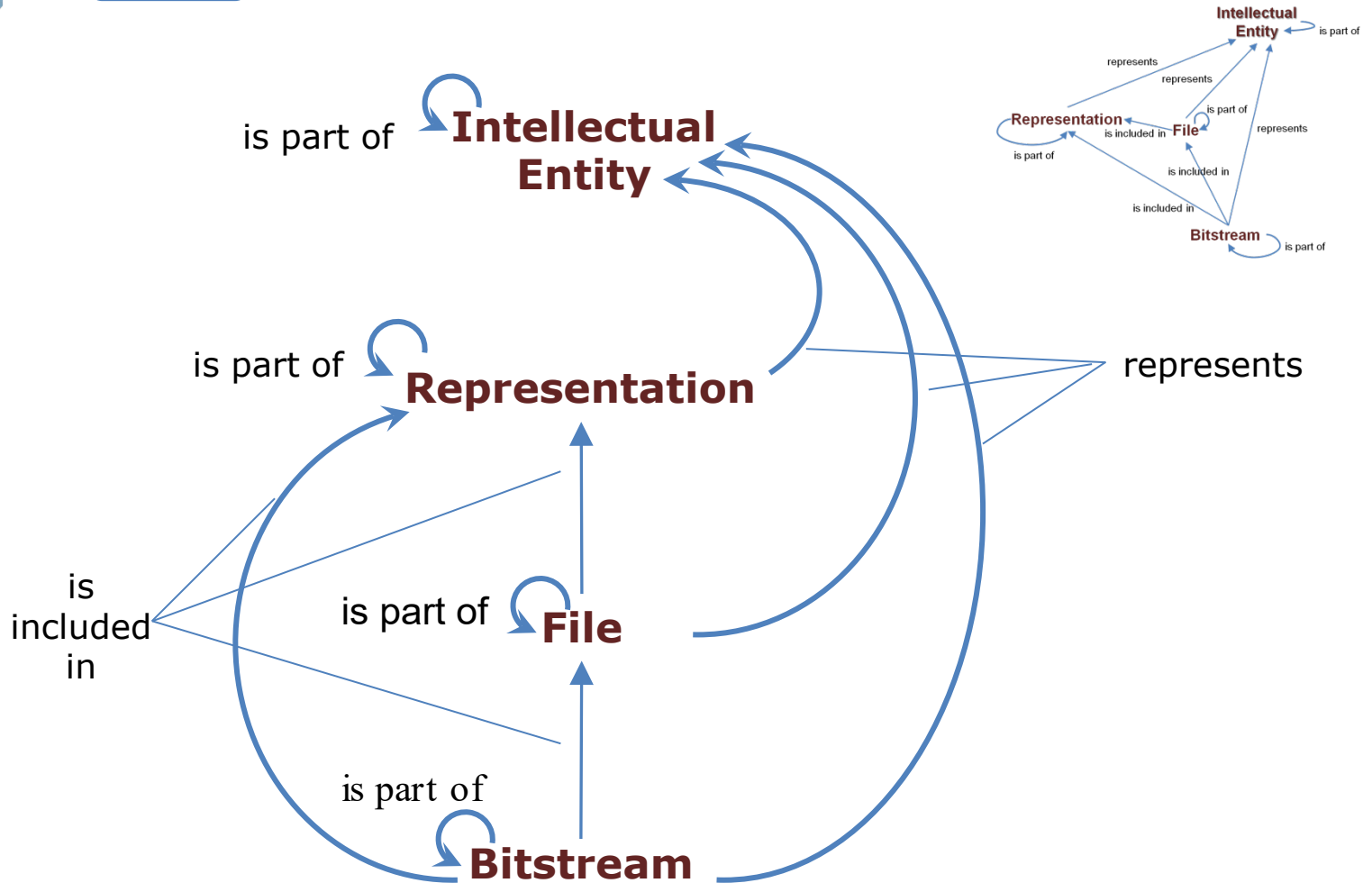
except there may be possibility of specifying roles, e.g. for Event because the same Agent may have a different Role in the digital Archive system

Relationships: Semantic Unit Identifiers





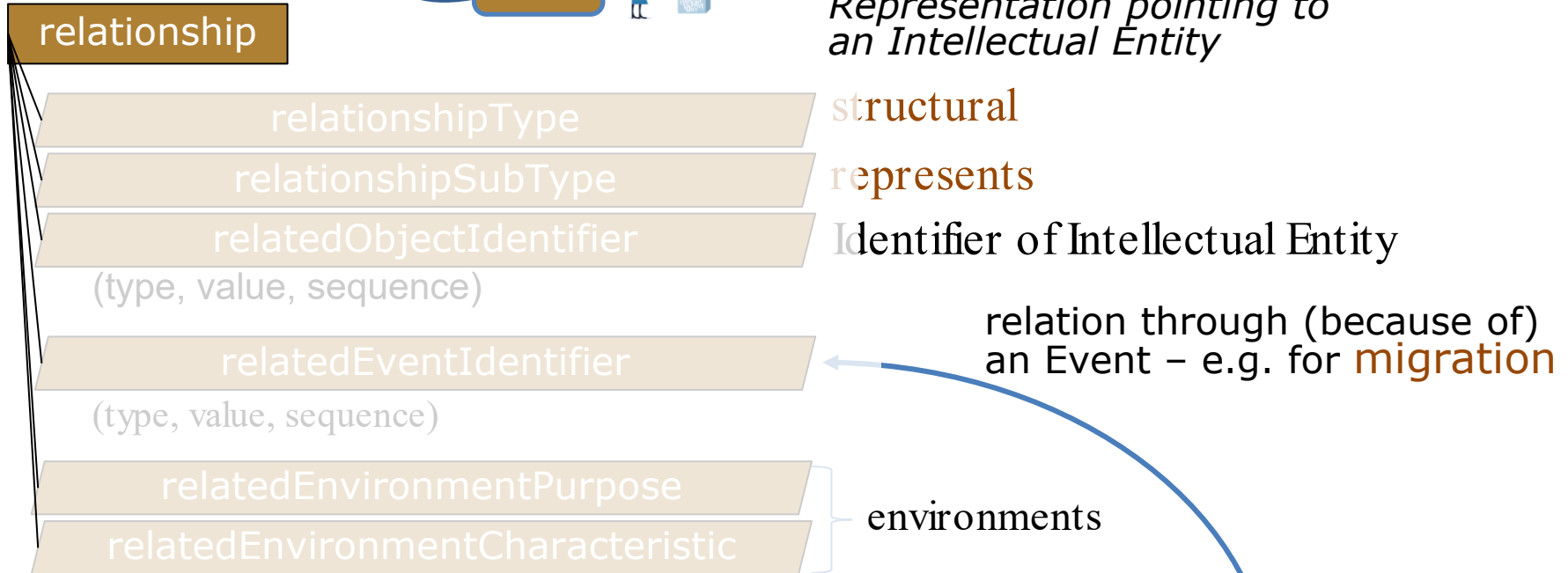
Objects and their interrelations



PREMIS Object Entity – Semantic Units



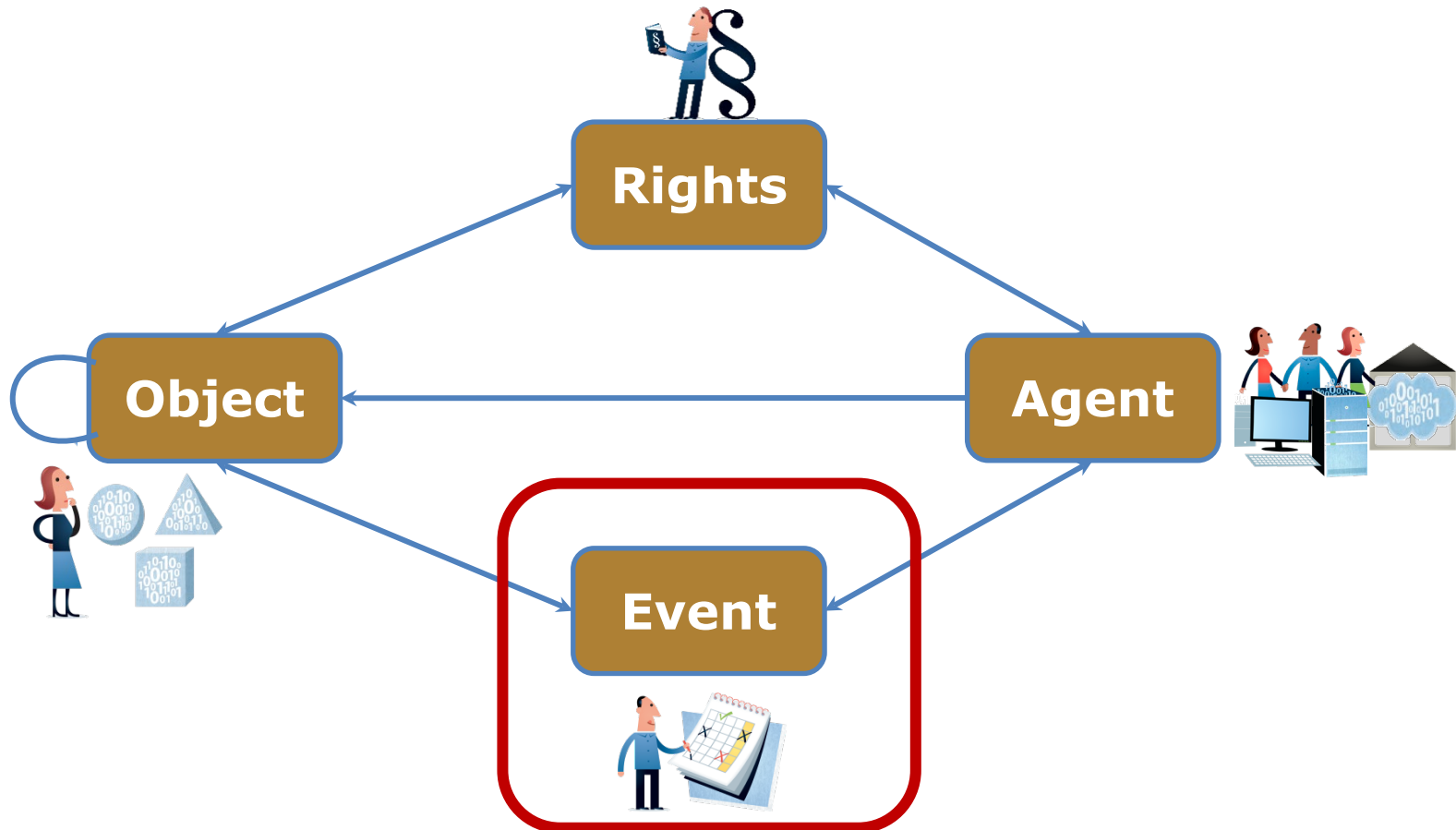
*Example:
Representation pointing to
an Intellectual Entity*



Objects can be associated with Events in two ways:

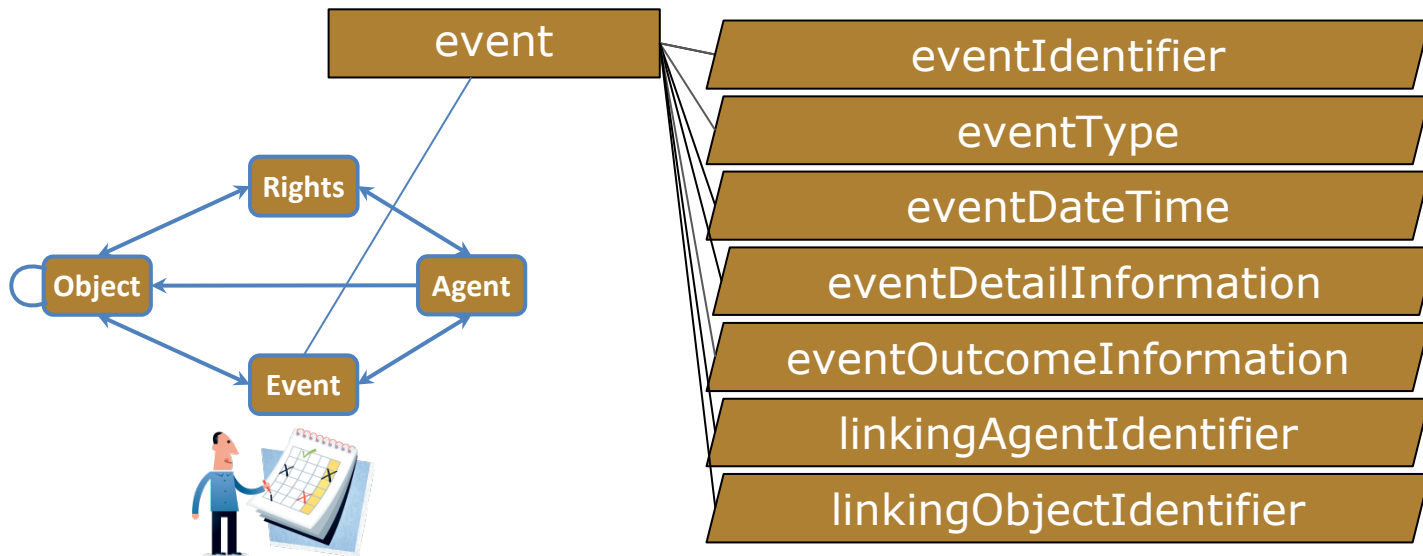
- If the Object has an associated Event with **relationship**
- If the Object has an associated Event with **no relationship** to a second Object, e.g. **ingest**: use **linkingEventIdentifier**

Properties of Entities - Semantic units



PREMIS Event Entity

- Mandatory semantic units are: *eventIdentifier*, *eventType*, and *eventDateTime*.
- Must be related to one or more Objects.
- Can be related to one or more Agents.



eventType

- Names the event

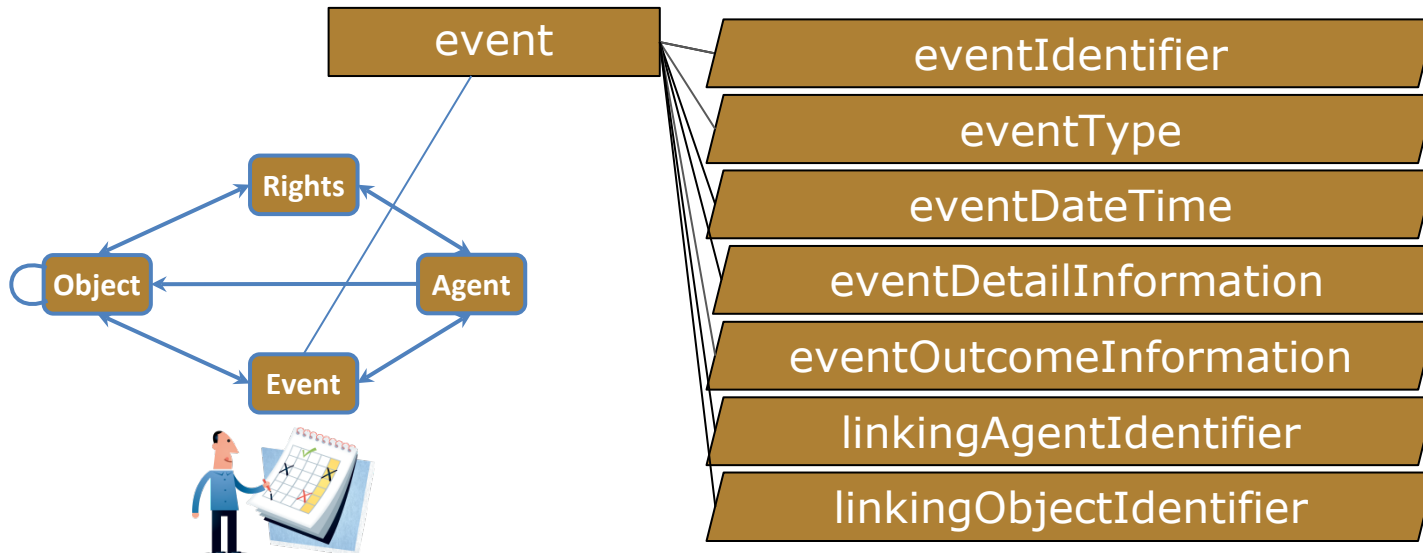
Ingestion **Virus check** **creation**
Validation
Message digest calculation **Compression**
migration **Fixity check** **Decompression** ...

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

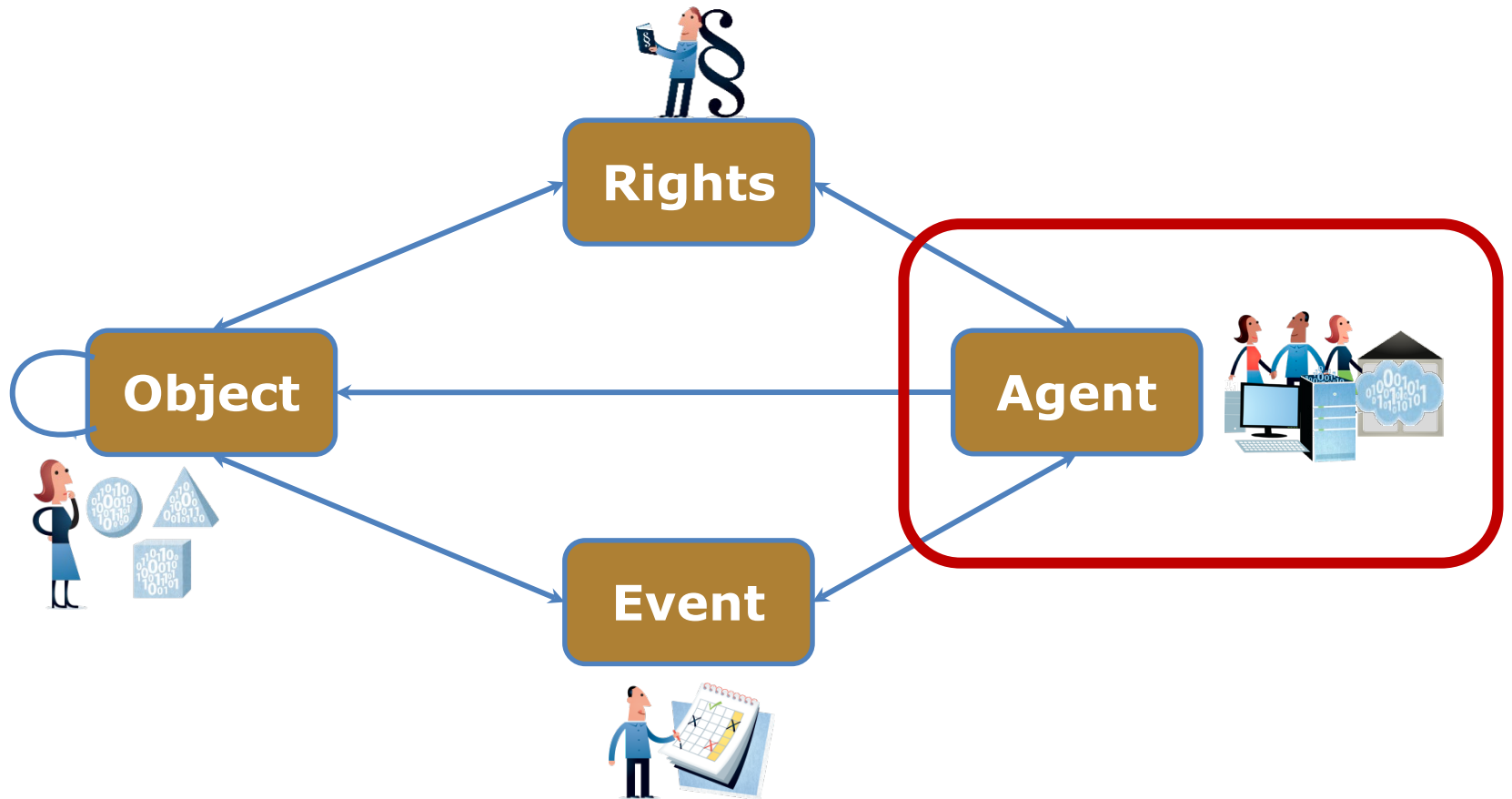
PREMIS Event Entity

eventDetailInformation and eventOutcomeInformation
Includes extension to allow use of other standards for the information

■



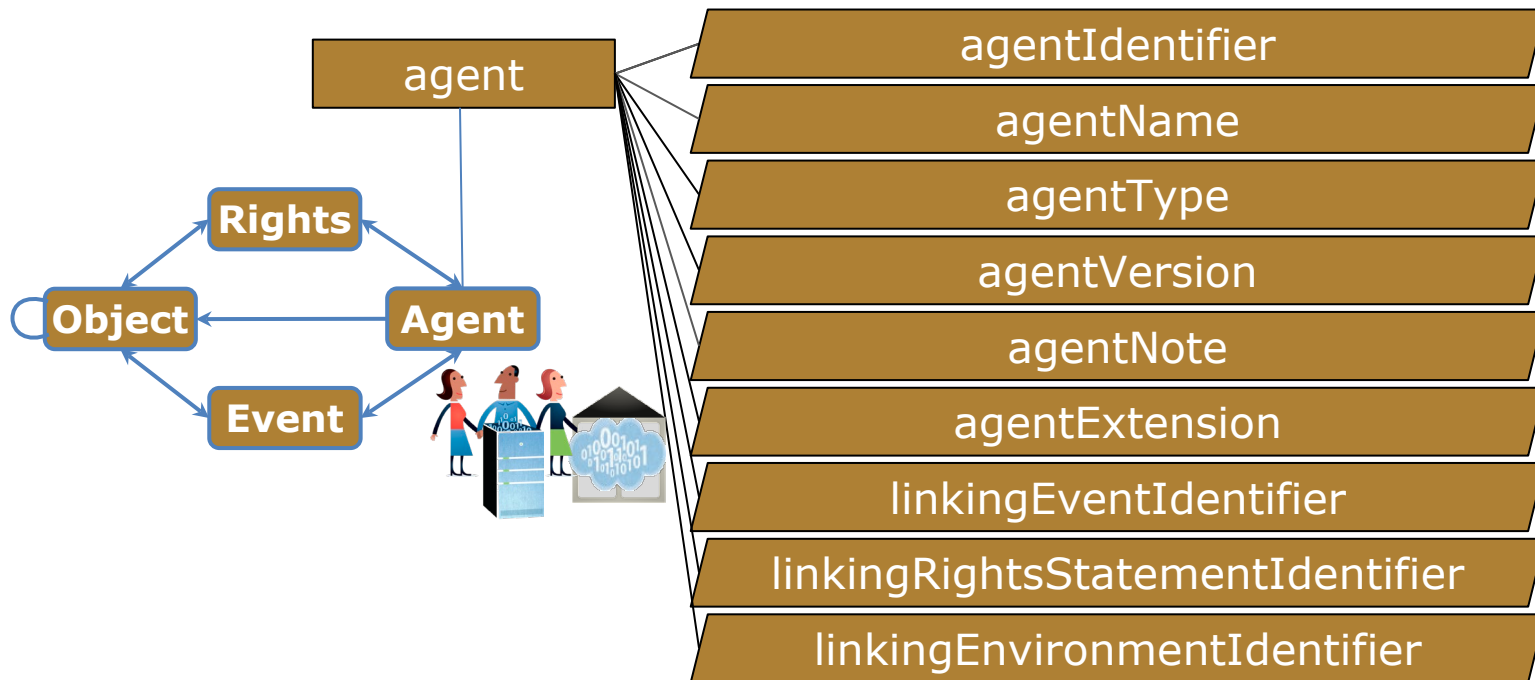
Properties of Entities - Semantic units



PREMIS Agent Entity

The only mandatory semantic unit is *agentIdentifier*

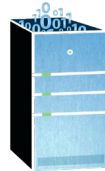
- May hold or grant one or more Rights.
- May carry out, authorize, or compel one or more Events.
- May create or act upon one or more Objects through an Event or with respect to a Rights statement.



agentType

- Can use controlled vocabulary, e.g.
<http://id.loc.gov/vocabulary/preservation/agentType.html>

- hardware



- organization



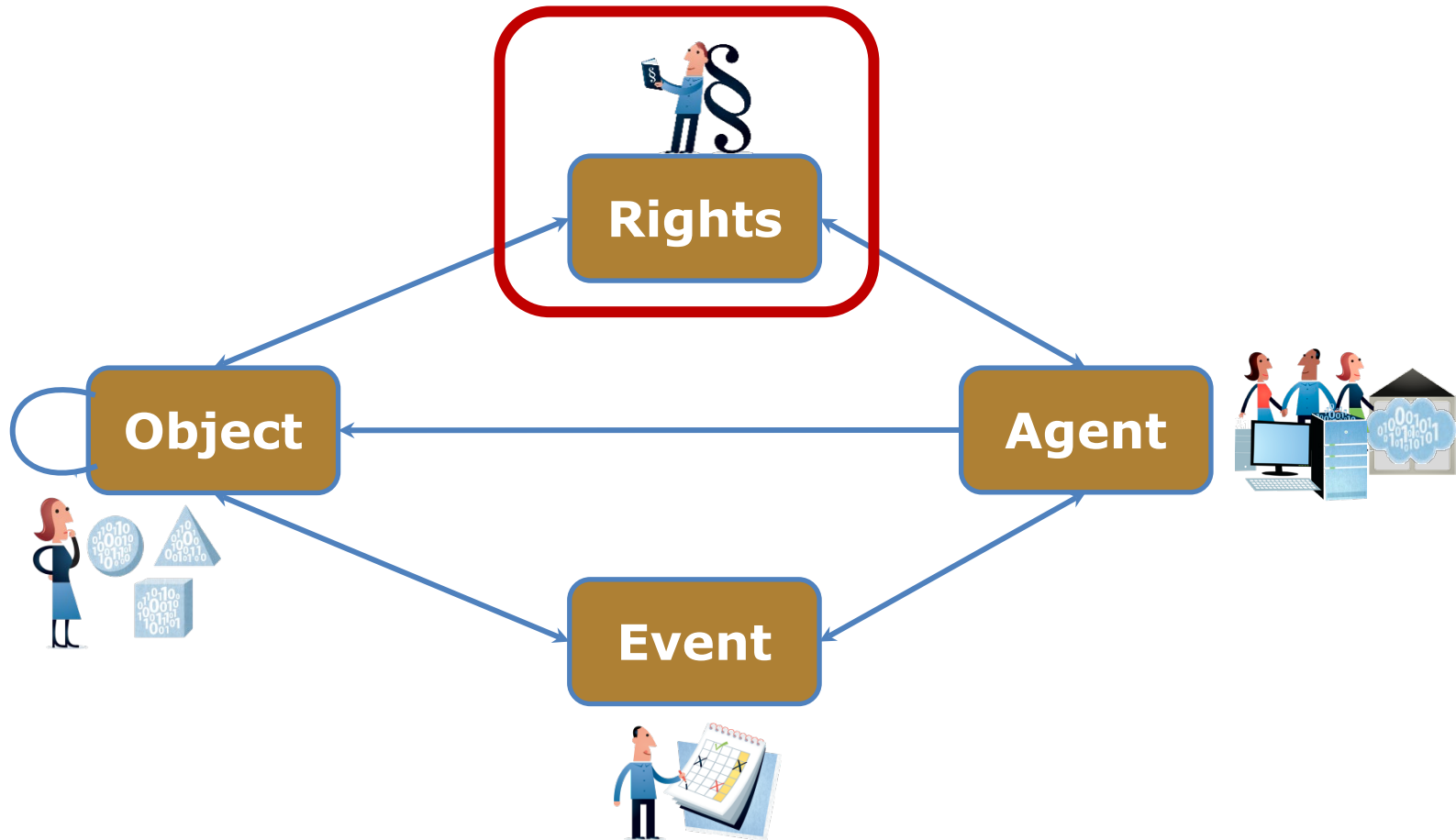
- person



- software



Properties of Entities - Semantic units



PREMIS Rights Entity

copyright



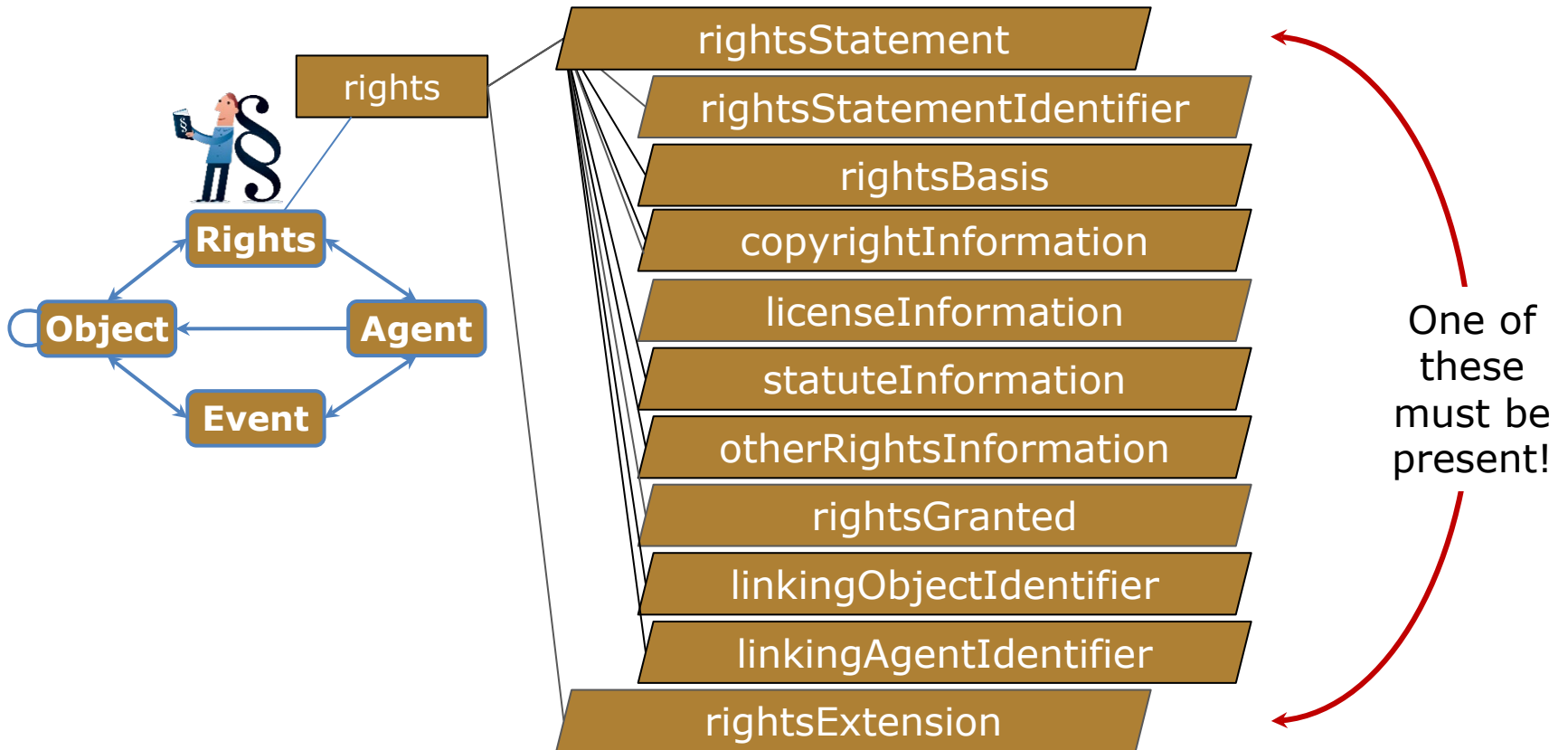
statute

license



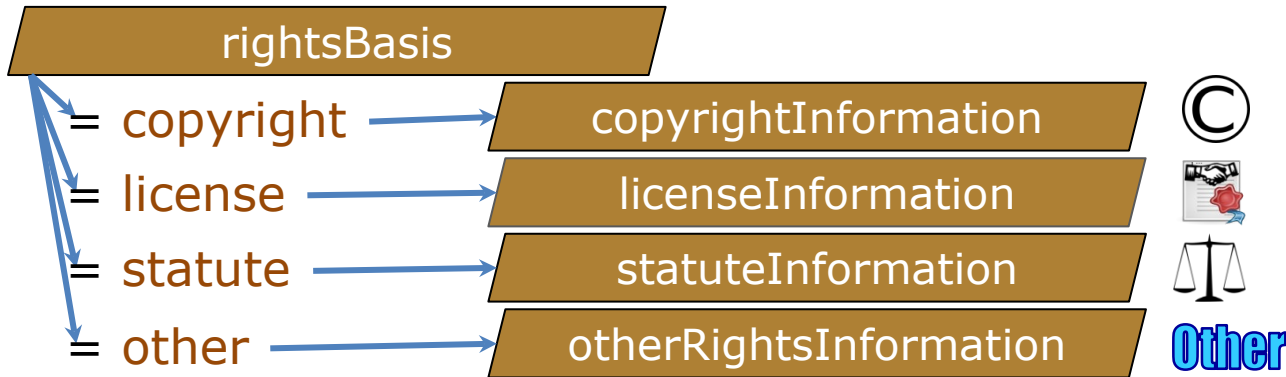
Other

other



Dependent units about rights

Specifying different types of rights



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

Example rightsBasis and copyrightInformation

rightsBasis = `copyright`

copyrightInformation

copyrightStatus = `copyrighted`

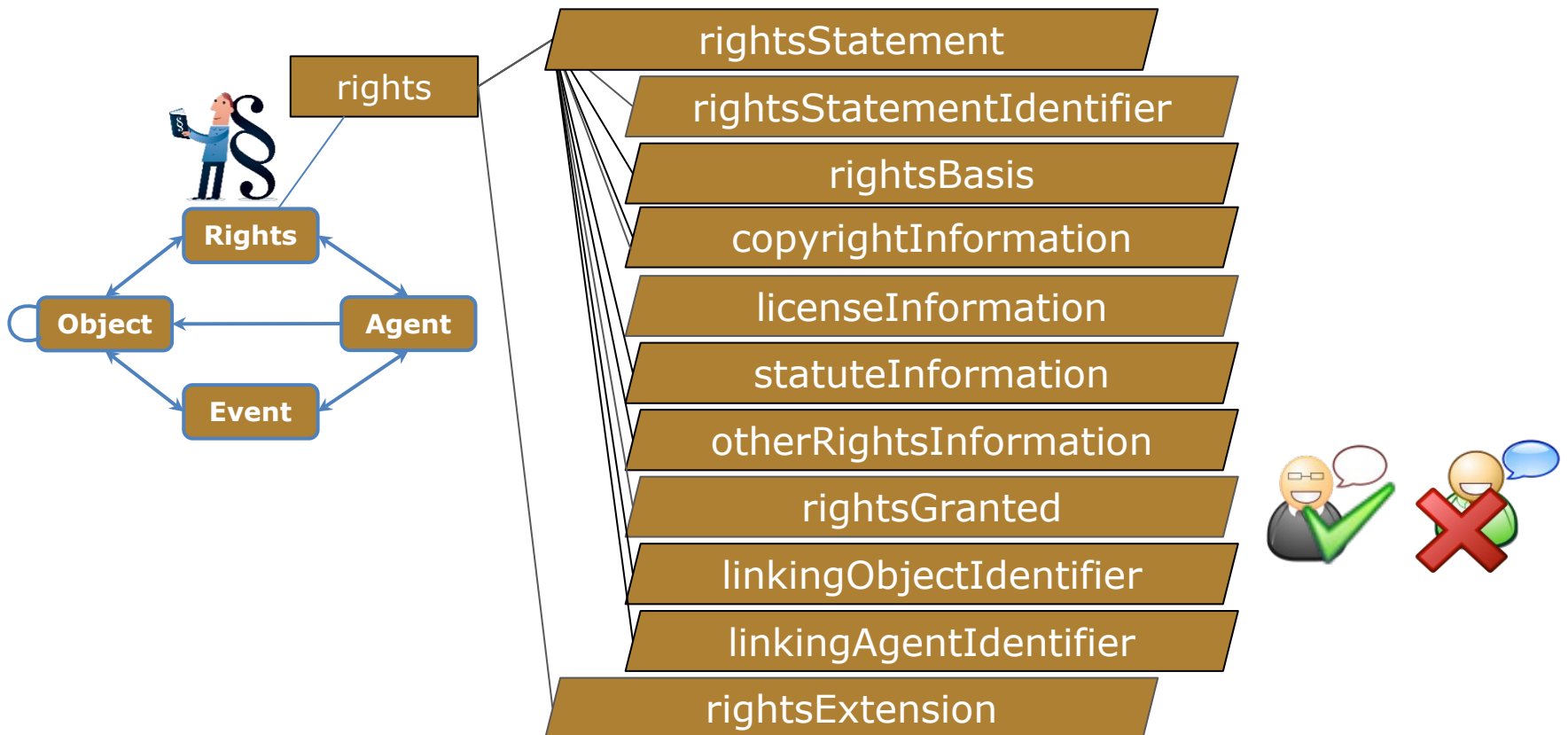
copyrightJurisdiction = `us`

copyrightStatusDeterminationDate = `2008-09-10`

copyrightNote = `Copyright expiration expected in 2022`

copyrightDocumentationIdentifier = `[link]`

PREMIS Rights Entity



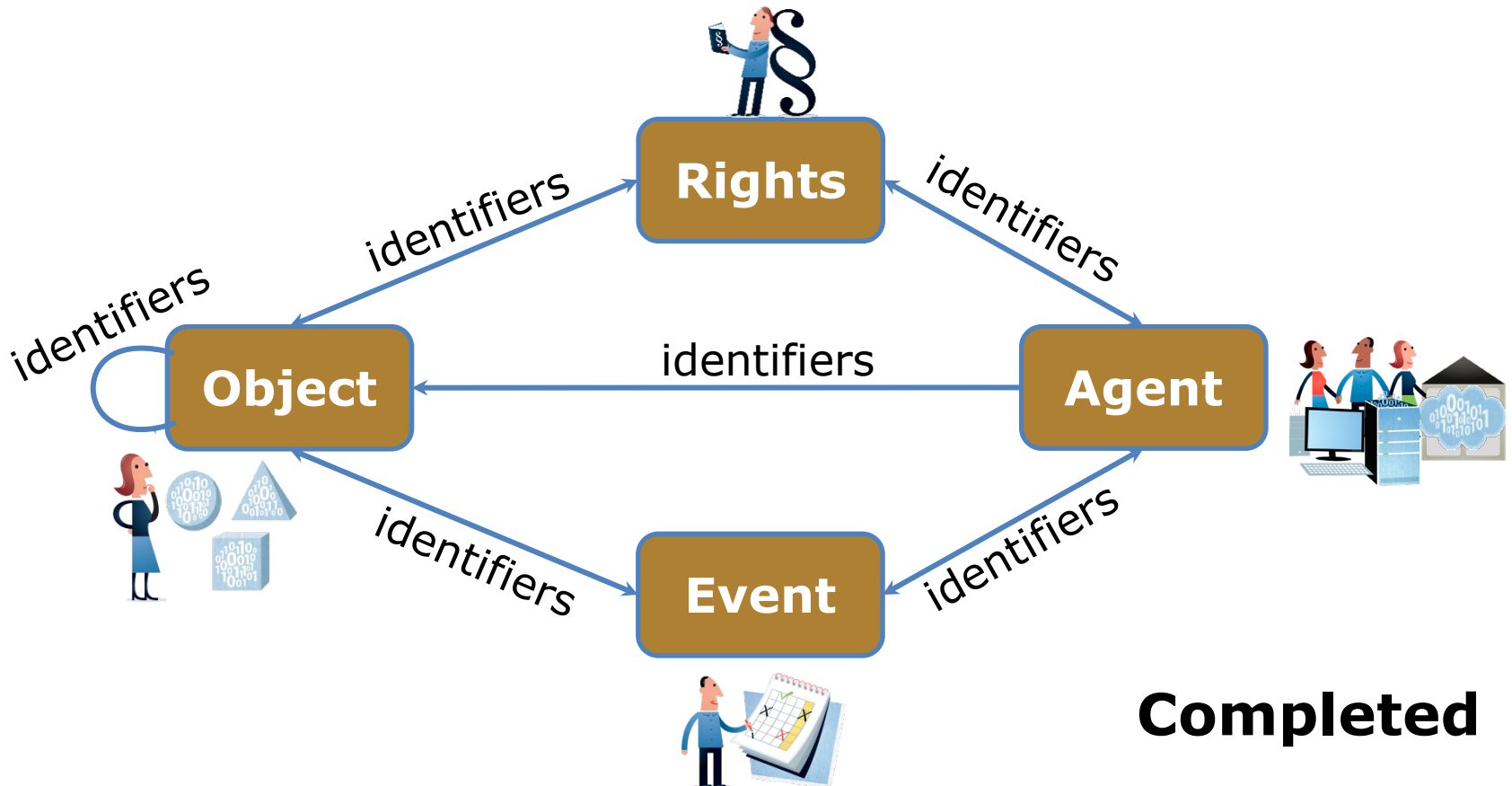
rightsGranted

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

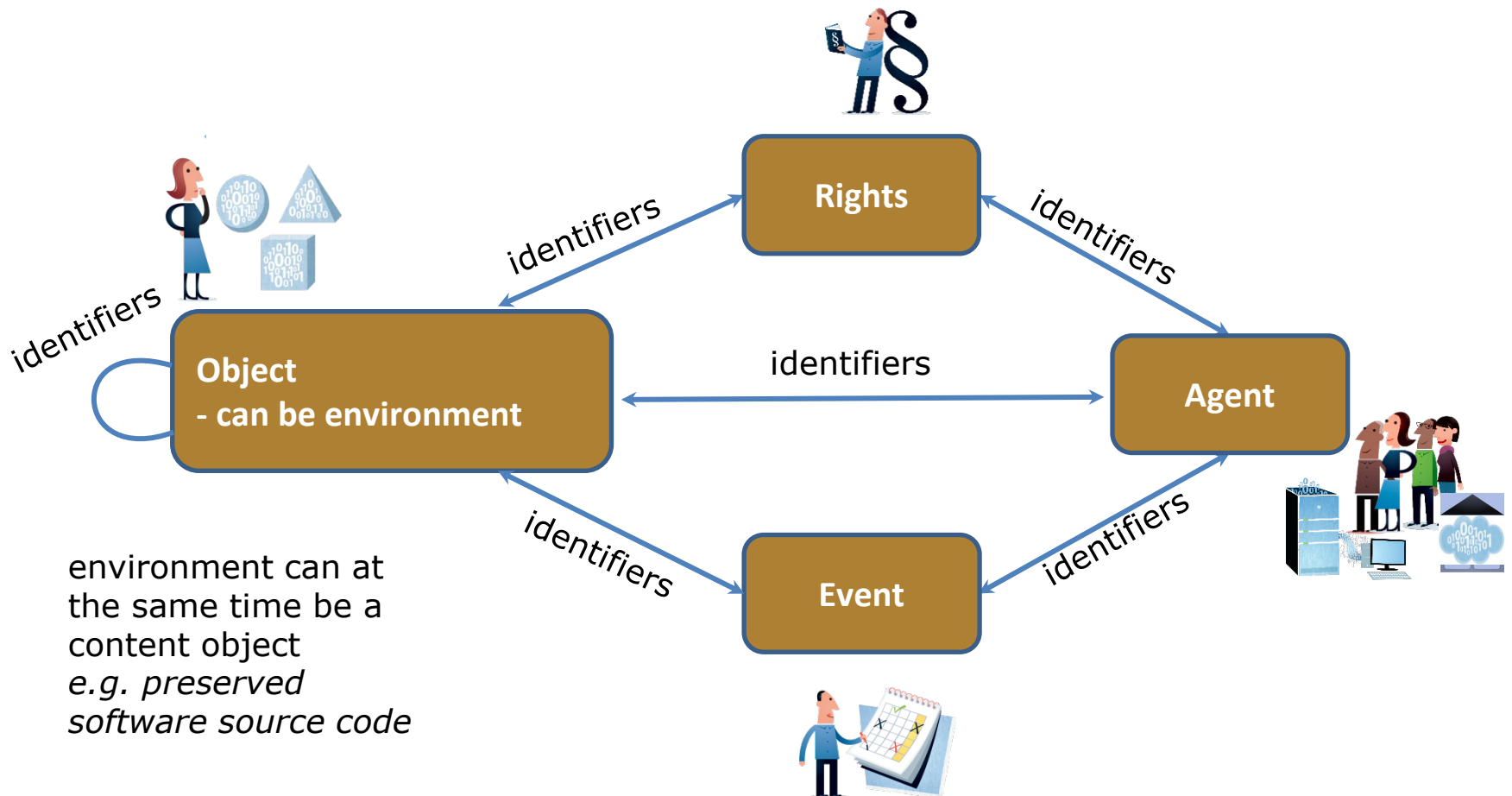
Contains

- **Act** (e.g. *migrate*, *modify*, ... could use *eventType values*)
- **Restriction** (description of condition or limitation on act)
- **termOfGrant** (*start and end dates of rights granted*)
 - **startDate** (e.g. *2005-01-01*)
 - **endDate** (e.g. *2005-01-01*)
- **termOfRestriction** (*start and end dates of restriction granted*)
 - **startDate** (e.g. *2005-01-01*)
 - **endDate** (e.g. *OPEN*)
- **rightsGrantedNote** (*additional inf. about the rights granted*)

PREMIS 3 - Entities



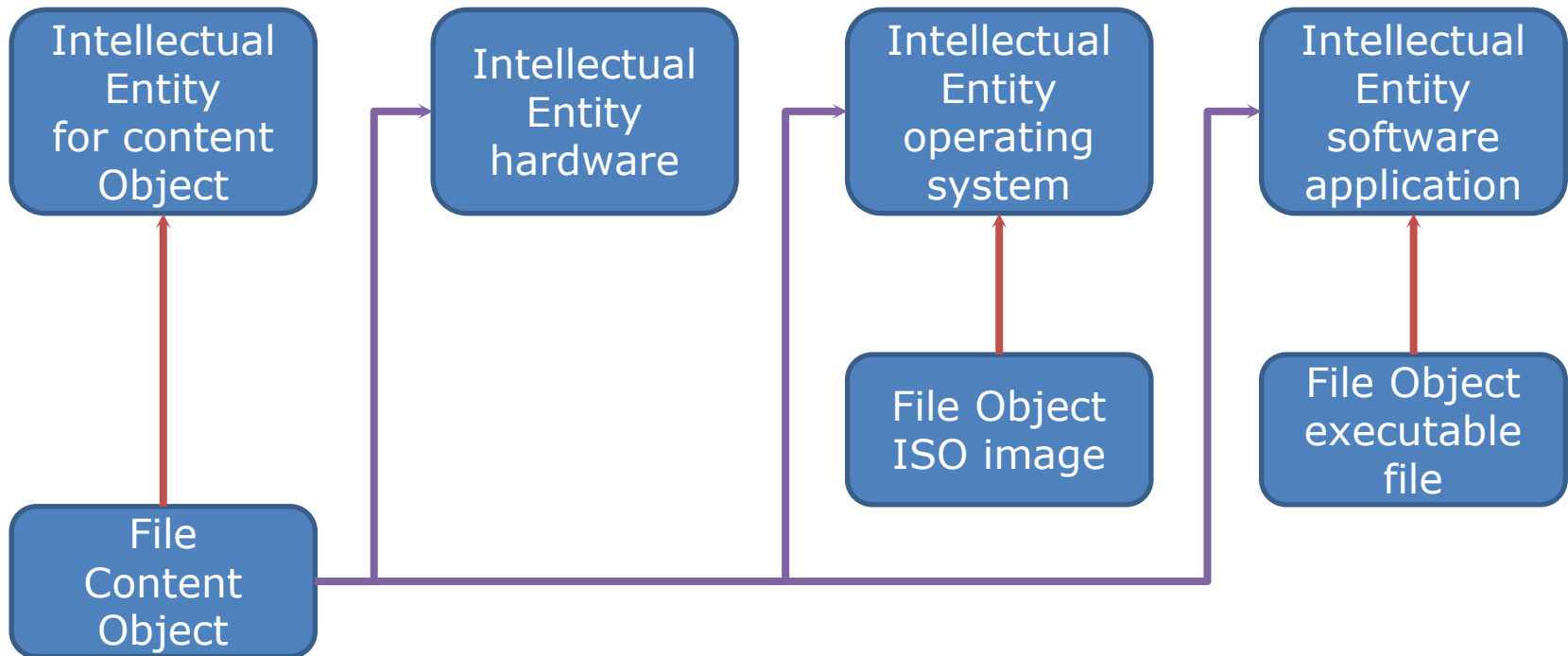
Data Model - environments



Environments as independent objects

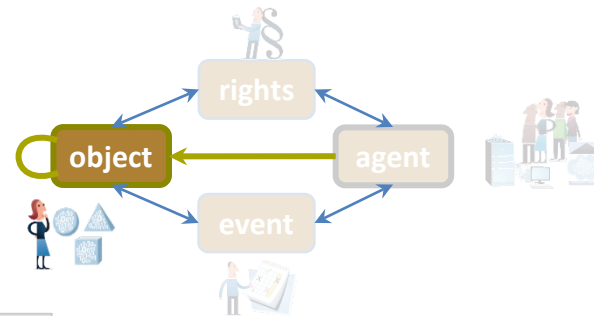
- What is needed to render or use a content object
 - Operating system
 - Application software
 - Hardware
 - Computing resources

Example: An object and its rendering environment



represents =
relationshipType: structural
relationshipSubType: represents

requires =
relationshipType: dependency
relationshipSubType: requires

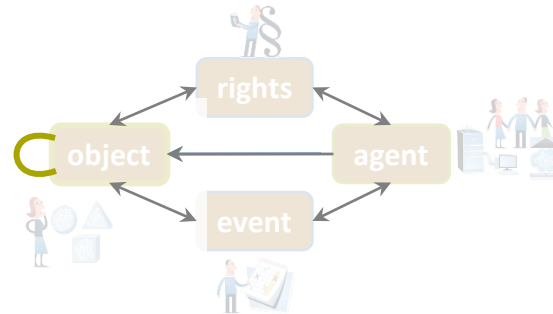


An agent is actually an environment acting as an agent

e.g. a format migration software agent involved in a preservation action

Relation points out the environment object acting as the agent

object



objectIdentifier

objectCategory

objectCharacteristics

significantProperties

preservationLevel

originalName

storage

signatureInformation

environmentFunction

environmentDesignation

environmentRegistry

environmentExtension

relationship

linkingEventIdentifier

linkingRightsStatementIdentifier

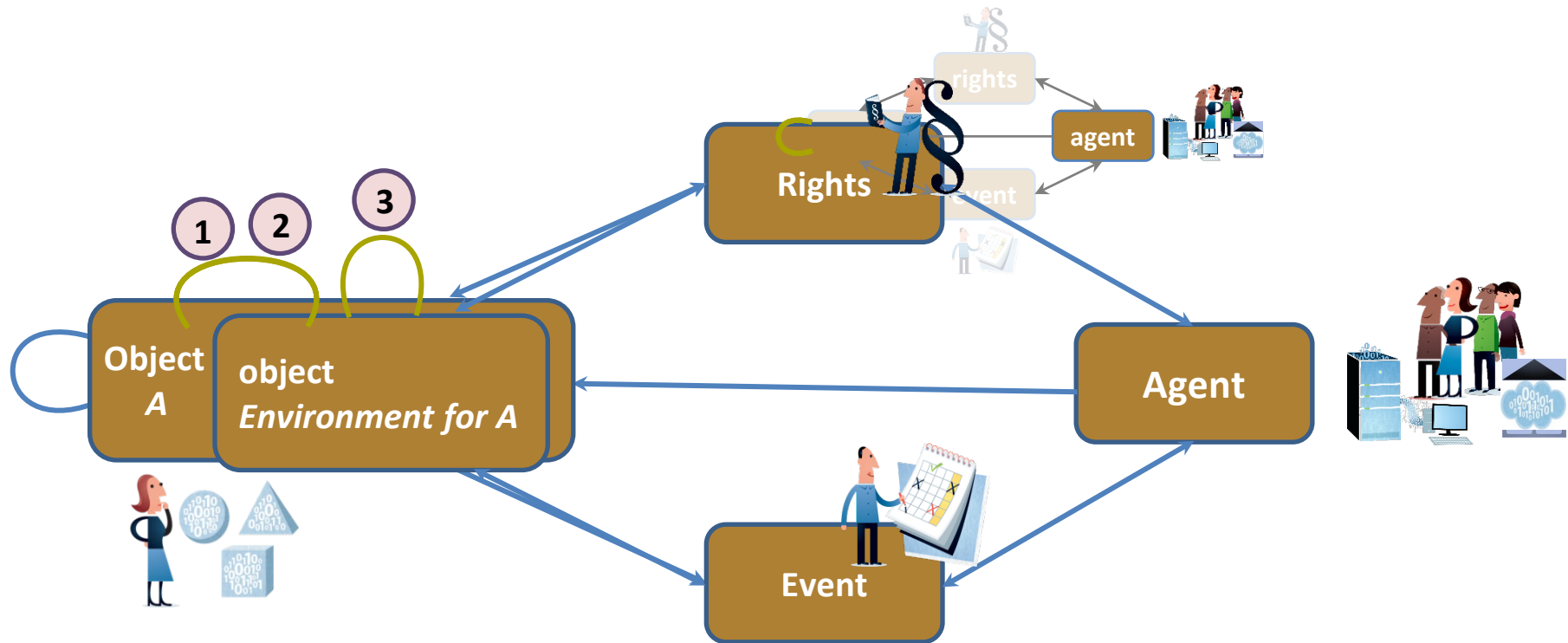
ntifier

e.g. function: SW, level: 1
function: OS, level: 2

name, version, origin, ...
e.g. for SW Windows XP Professional

name, key, role
-alternative: Link to an external registry

relationshipType: dependency
relationshipSubType: requires

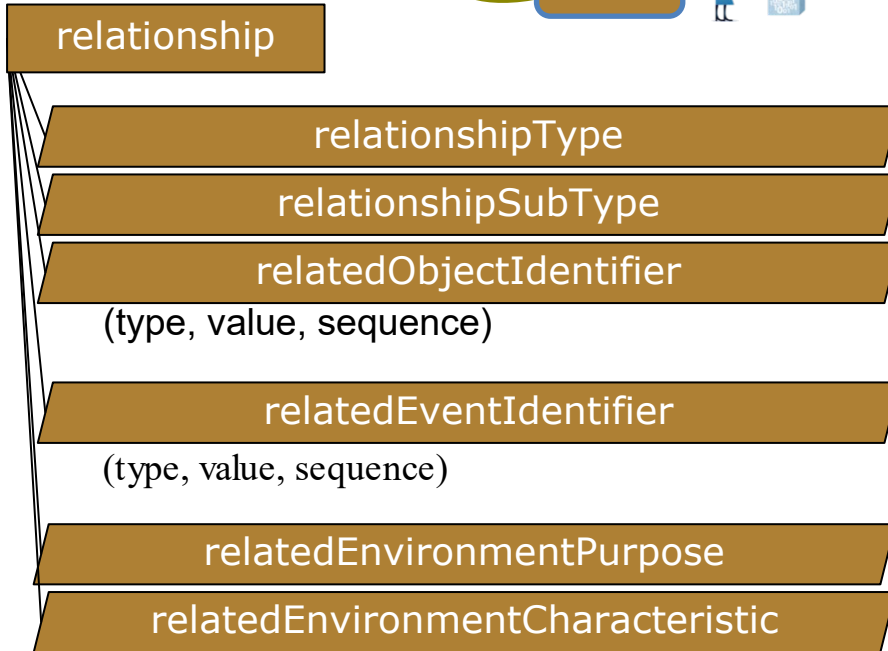


1. object to environment - specify computational context
2. environment to object - documentation, specifications, surrogates
3. environment to environment - inclusion, dependency, derivation, other

Additional environment information

- relationship
Different environments can support different uses/purposes of objects
create, edit, modify, render...
- relationship
Characteristics describing how the environment supports its purpose
unspecified, minimum, known to work, recommended ...

PREMIS Object Entity – Semantic Units



A controlled vocabulary is available at:

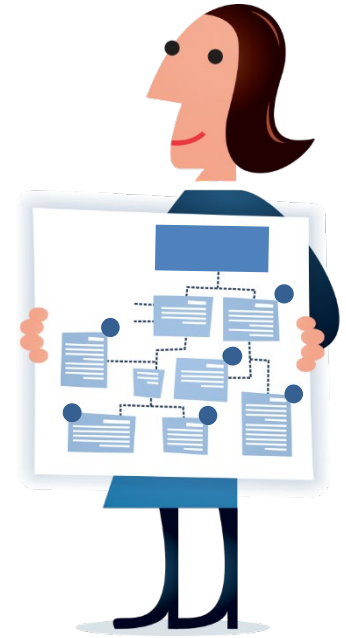
<http://id.loc.gov/vocabulary/preservation/environmentPurpose>

e.g. create, edit, modify, render

e.g. unspecified, minimum, recommended, known to work

Karin Bredenberg

Kommunalförbundet Sydarkivera



WRAP UP

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



Sum up – Data Dictionary



Lots of other information

Semantic	Entity sets	CONTENTS	iii
Semantic	<i>NB: Semantic Bitstreams</i>	Acknowledgments	v
Definition	1.1 object	PREMIS Editorial Committee members	v
Rational	1.2 object	Special thanks	v
Data con	1.3 preser	PREMIS Web Sites and E-Mail	viii
Object c	1.3.1	Introduction	1
Applica	1.3.2	Background	1
Repeata	1.3.3	Development of the original PREMIS Data Dictionary	1
Obligati	1.3.4	Implementable, core preservation metadata	2
Creation	1.3.5	PREMIS Maintenance Activity	3
Mainten	1.4 signi	Version History	4
Usage notes	1.4.1	PREMIS Awards and Recognition	5
	1.4.2	The PREMIS Data Model	6
	1.4.3	More on Objects	8
	1.4.4	More on Events	15
	1.4.5	More on Agents	16
	1.4.6	More on Rights	17
	1.5 object	General Topics on the Structure and Use of the Data Dictionary	17
	1.5.1	Identifiers	17
	1.5.2	Relationships between Objects	19
	1.5.3	Relationships between entities of different types	21
	1.5.4	The 1:1 principle	21
		Implementation Considerations	22
		PREMIS conformance	22
		Implementation of the data model	24
		Storing metadata	25
		Supplying metadata values	25
		Extensibility	27
		Date and time formats in PREMIS	29
		The PREMIS Data Dictionary Version 3.0	30
		Limits to the scope of the Data Dictionary	31
		Object Entity	33
		Entity types	33

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 explanation to 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

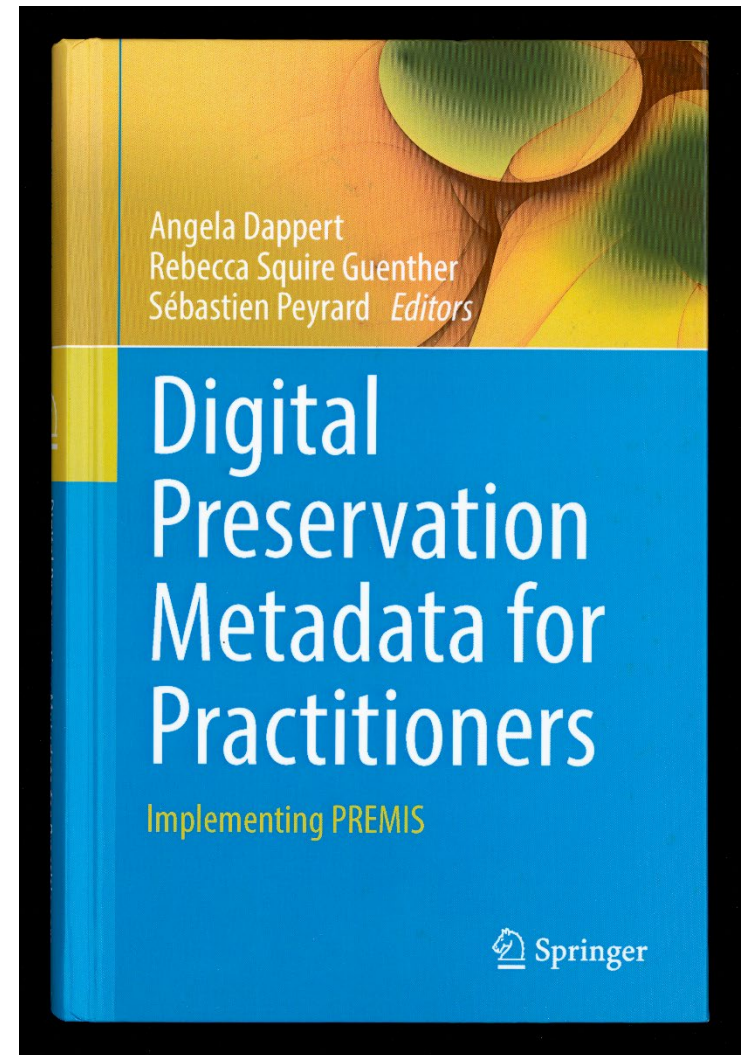
<http://www.loc.gov/standards/premis/>

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!

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



On your own!

Sample Data Dictionary table of contents

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

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]

On your own!

PREMIS Object Entity – Exercise

- Exercise to get a feeling for the object!
 - Objectexercise.pdf
 - <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!



On your own!

PREMIS Object Entity – Exercise

Objects in PREMIS

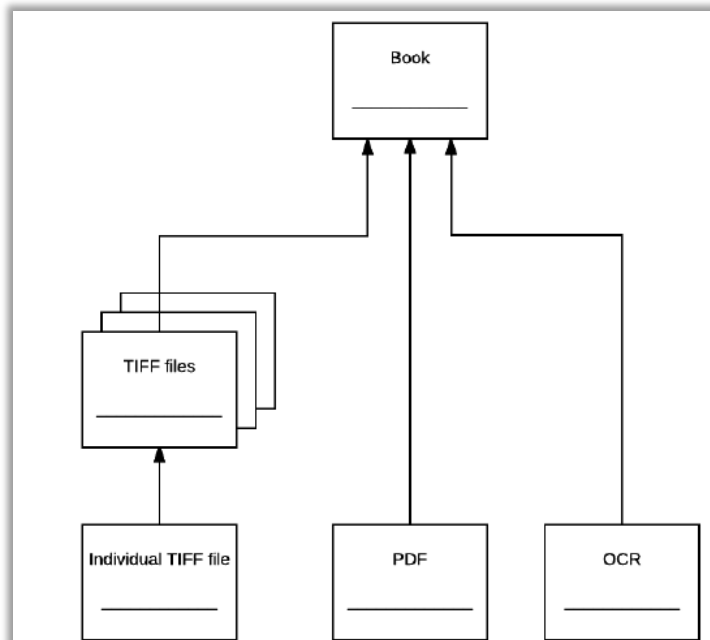
An institution is preserving a digitized book

(<https://open.library.ubc.ca/collections/bcbooks/items/1.0226324>). In their digital repository,

they preserve:

- A TIF file for each original scanned page of the book
- A PDF file of the entire book which is used in their access system.
- An OCR file of the text

Fill out the category [IE, REP, File, Bitstream] of object in this diagram:



For each of Book, PDF and individual TIF file, use the metadata provided to fill out the table.

On your own!

PREMIS Object Entity – Exercise

On your own!

Book:

Local identifier: II-378

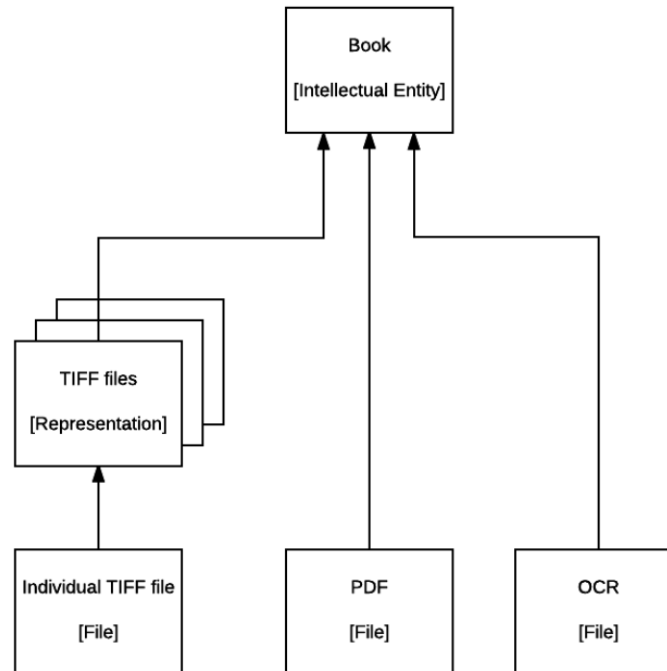
Local identifier for OCR file: II-378-ocr

Data Element (Field)	Data value (content)
objectIdentifier objectIdentifierType*	
objectIdentifier objectIdentifierValue	
objectCategory*	
relationship relationshipType*	
relationship relationshipSubType*	
relatedObjectIdentification relatedObjectIdentifierType*	
relatedObjectIdentification relatedObjectIdentifierValue	

Exercise answers

- Objectexercise-answers.pdf

On your own!



PREMIS Events, Agents and Rights Entity – Exercise

- Exercise to get a feeling for the events, agents and rights!
 - Eventsagentsrightsexercise.pdf
 - <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!



On your own!

PREMIS Events, Agents and Rights Entity – Exercise

On your own!

Events, agents and rights in PREMIS

The institution decides to make access derivatives of each page of the book because they are frequently requested as individual images. In the table below, describe this event for the individual tif image.

Also, the institution wishes to make it clear that the material is in the public domain and is allowed to be disseminated. Add the rights information in the table below.

PREMIS Events, Agents and Rights Entity – Exercise

Agent:

Assume that you are the individual who made the derivatives, and that the in practice of recording individuals as firstinitialLastName (e.g. Jamie Smith bec Hint: Skip ahead to the Event information for the linkingEvent fields.

On your own!

Data Element (Field)	Data Value (Content)
agentIdentifier agentIdentifierType*	
agentIdentifier agentIdentifierValue	
agentName	
agentType*	
agentNote	
linkingEventIdentifier linkingEventIdentifierType*	
linkingEventIdentifier linkingEventIdentifierValue	

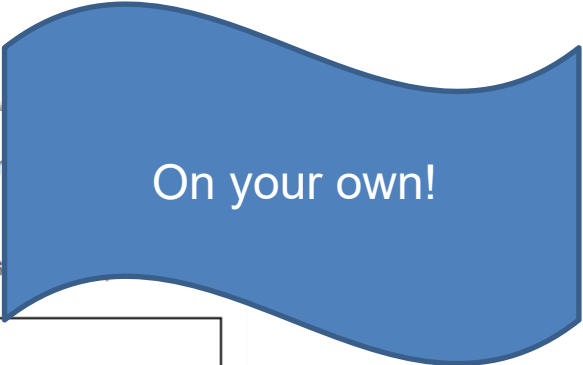
* indicates a controlled vocabulary should be used

PREMIS Events, Agents and Rights Entity – Exercise

Event:

Assume that the system you use to record events automatically assigns a Unique Identifier to each event record.

Assume that it's the local practice to base the identifier for the access object for the preservation object with "_a" appended (hint: refer back to the objects

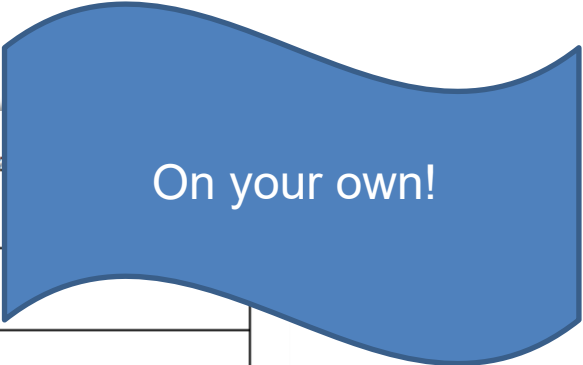


Data Element (Field)	Data Content (Value)
eventIdentifier eventIdentifierType*	
eventIdentifier eventIdentifierValue	
eventType*	
eventDateTime	
eventDetailInformation eventDetail	
linkingAgentIdentifier linkingAgentIdentifierType*	
linkingAgentIdentifier linkingAgentIdentifierValue	
linkingAgentIdentifier linkingAgentRole*	
linkingObjectIdentifier linkingObjectIdentifierType*	
linkingObjectIdentifier linkingObjectIdentifierValue	

PREMIS Events, Agents and Rights Entity – Exercise

Rights:

Assume that the system you use to enter rights statements automatically assigns a Universal Identifier to each rights statement created



Data element (Field)	Data content (Value)
rightsStatementIdentifier rightsStatementIdentifierType*	
rightsStatementIdentifier rightsStatementIdentifierValue	
rightsBasis*	
copyrightInformation copyrightStatus*	
copyrightInformation copyrightJurisdiction*	
copyrightInformation copyrightStatusDeterminationDate	
copyrightInformation copyrightNote	
copyrightApplicableDates startDate	
copyrightApplicableDates endDate	
rightsGranted act*	
rightsGranted restriction	

Exercise answers

- [Eventsagentsrightsexercise-Answers.pdf](#)



On your own!

PREMIS Environments – Exercise

- Exercise to get a feeling for the environments!
 - Environmentsexercise.pdf
 - <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!



On your own!

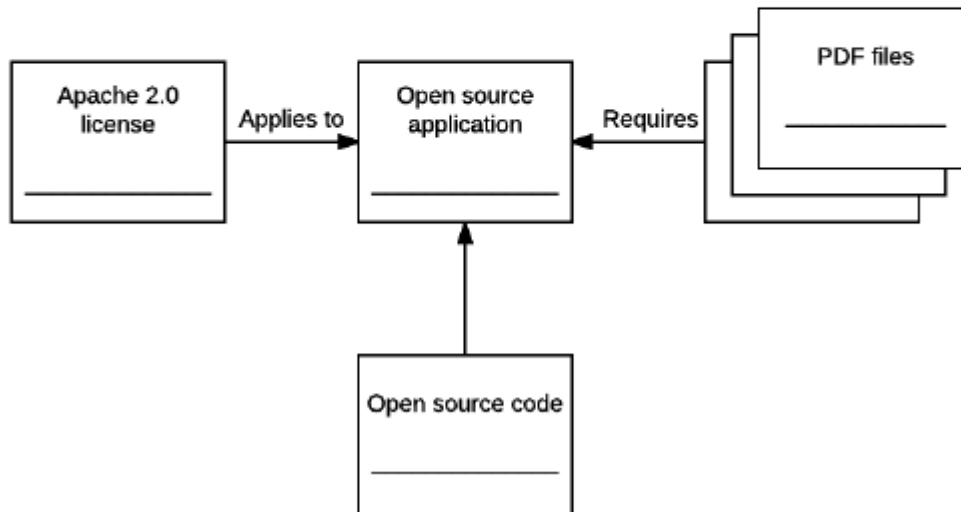
PREMIS Environments – Exercise

On your own!

Environments in PREMIS

The institution develops an open-source reader for the content in the repository more easily disseminate it to the public. They wish to preserve the source code of the newly developed application as well. They've made the application available under an Apache 2.0 software license

Fill out the types of entities on this diagram:



PREMIS Environments – Exercise

On your own!

Fill out the following fields for the application:

Application:

Local identifier: app-1234

Local identifier for the source code: code-1234

Name of application: Open Viewer version 1.0

Assume that the rights statement will be generated automatically by the system as a UUID.

License URI: <http://www.apache.org/licenses/LICENSE-2.0>

Data Element (Field)	Data value (content)
objectIdentifier objectIdentifierType*	
objectIdentifier objectIdentifierValue	
objectCategory*	
environmentFunction environmentFunctionType*	

Exercise answers

- [Environmentsexercise-Answers.pdf](#)



On your own!

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>

Questions in parking?!

Today

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

Finally...

PREMIS is a community standard.

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

Thank you!

Karin, Eld and Micky