UNDERSTANDING AND IMPLEMENTING PREMIS
A 90 minute introduction

ENTENDER E IMPLEMENTAR PREMIS
Una introducción de 90 minutos

Sarah Romkey
Artefactual

Leslie Johnston
NARA – National Archives and
Records Administration

Purpose of the Tutorial

Objetivo del taller

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

- Presentar PREMIS
- Dar una primera base de entendimiento común
- Presentarle a la comunidad PREMIS

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

Copyright © 2009 The Library of Congress, excepto en EE. UU.

Se debe mencionar esta publicación como procedencia de cualquier extracto de ella que se cite.

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

Who are we?

From the PREMIS Editorial Committee:

- Sarah and Leslie

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

¿Quiénes somos?

Del Comité Editorial de PREMIS:

- Sarah y Leslie

Sabemos que:

- Acaba de empezar a conocer PREMIS
- El modelo de datos de PREMIS da un poco de miedo
- Puede que haya estado utilizando PREMIS en la práctica
- Usted trabaja en:
  ◦ una biblioteca
  ◦ un archivo
  ◦ una universidad
  ◦ o algo diferente

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.

¿En qué queremos que piense?

- Piense en su caso de uso
  - Quiero utilizar PREMIS porque ...

- Ejemplo: Estoy digitalizando una serie de publicaciones realizadas en mi institución y necesito preservarlas para el futuro y asegurarme de que sean accesibles dentro de 50 años.

PREMIS
PREservation Metadata Implementation Strategies

Leslie Johnston

United States National Archives and Records Administration (NARA)

TODAY IS JUST A FIRST INTRODUCTION

PREMIS

PREservation Metadata Implementation Strategies

Leslie Johnston

United States National Archives and Records Administration (NARA)

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

What is digital preservation metadata?

- Digital preservation metadata =
  Metadata to ensure <u>long-term accessibility</u> of <u>digital resources</u>

- 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

¿Qué son los metadatos de preservación digital?

- Metadatos para la preservación digital =
  Metadatos para garantizar la <u>accesibilidad</u> a <u>largo plazo</u> de los <u>recursos digitales</u>

- Los objetos digitales deben ser autodescriptivos

- Deben poder describirse, gestionarse y descubrirse independientemente de los sistemas utilizados para crearlos.
  XML (legible por máquinas y personas)

- A menudo se incluye con los archivos de contenido en un paquete de información

Domain / Dominio

Born digital

Nacido digital

News

Digitized

Digitalizado

DP metadata supports preservation goals

Authentication

Format strategies

Media management

Secure storage

Documentation

Description

Capture
Selection

Means

Authenticity

Renderability

Viability

Fixity

Understandability

Identity

Availability

Preservation Goals

Preservation Pyramid
(from Priscilla Caplan)

PREMIS
PREservation Metadata Implementation Strategies

Leslie Johnston

United States National Archives and Records Administration (NARA)

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

Leslie Johnston

United States National Archives and Records Administration (NARA)

ON-LINE RESOURCES

- Website
  - 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

Leslie Johnston

United States National Archives and Records Administration (NARA)

BENEFITS OF IMPLEMENTING PREMIS

Implementable, core preservation metadata

Supports Preservation Principles
viability, renderability, understandability, authenticity, and identity of digital objects

Core Information
what most preservation repositories need to know to preserve digital materials over the long term

Implementable
rigorously defined, supported by guidelines for creation, management, and use, and oriented toward automated workflows

Technical Neutrality
no assumptions made about preservation technologies, strategies, metadata storage and management, etc.

Benefits of implementing PREMIS (1)

hdl="galapagos.6754.1"

IDobj="hdl.galapagos.6754.1"

exchange consistency

?

ID="6754.1"
GID="galapagos"

ID="1"
TYPE="hdl"
GROUPID="galapagos.6754"
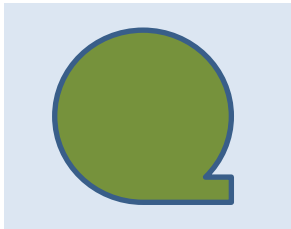
Interoperability through implementable preservation metadata

IDobj="hdl.galapagos.6754.1"

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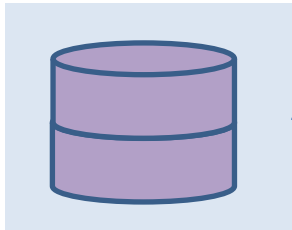
```
premis:objectIdentifier
premis:objectIdentifierType="hdl"
premis:objectIdentifierValue="galapagos.6754.1"
```
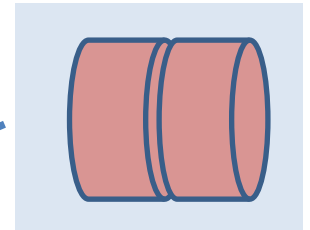
PREMIS

```
<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?
- Are there community best practices?

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*
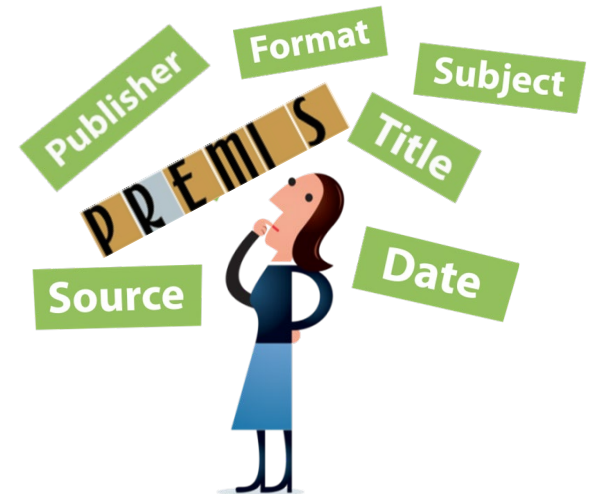
- Valuable resource to know what you need to capture

- 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
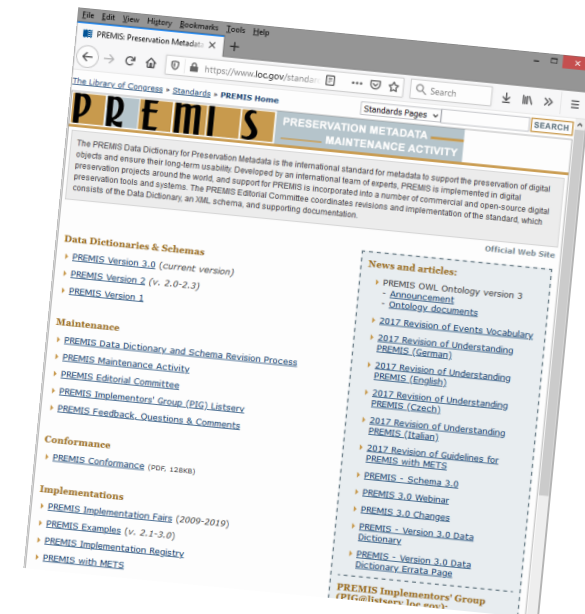
Sarah Romkey

Artefactual

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:

Entities ▭
"Things" relevant to digital preservation

Semantic Units ▱
Detailed information on entities
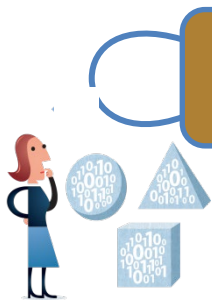
Rights

Object

Agent

Rights

Object

Agent

Event

Event

Relationships
Specifies relations between entities

agentIdentifier

agentIdentifierType

agentIdentifierValue

agentname

agentType

Objects
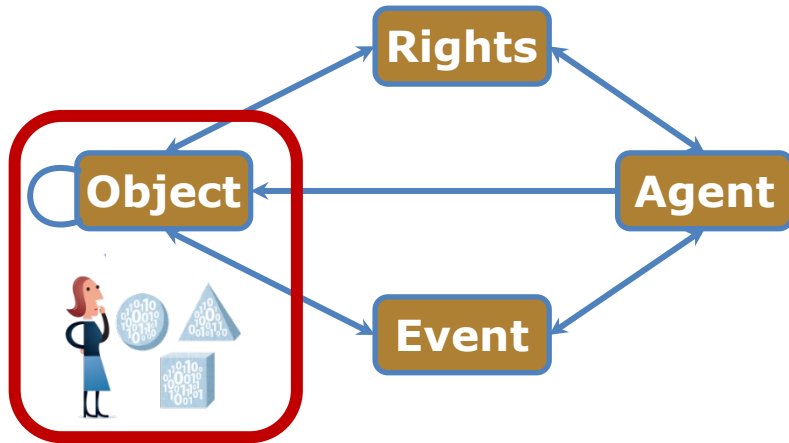
- These are the objects that you preserve

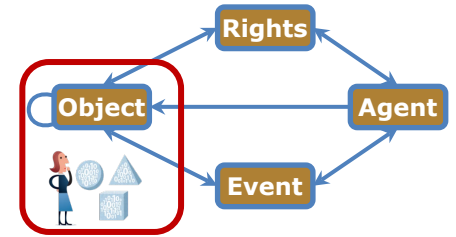- Objects can be intellectual entities, representations or bitstreams (more later)

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)

intellectual entity — *representation* — *file* — *bitstream*

Book Object → Original TIFFs → TIFF p1 / TIFF p2 → Image in page

Derived e-pub → E-pub
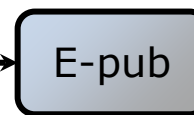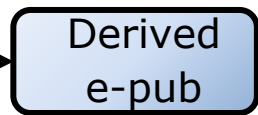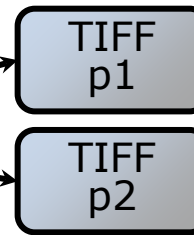
A distinct intellectual object that is considered relevant in the context of digital preservation

A 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
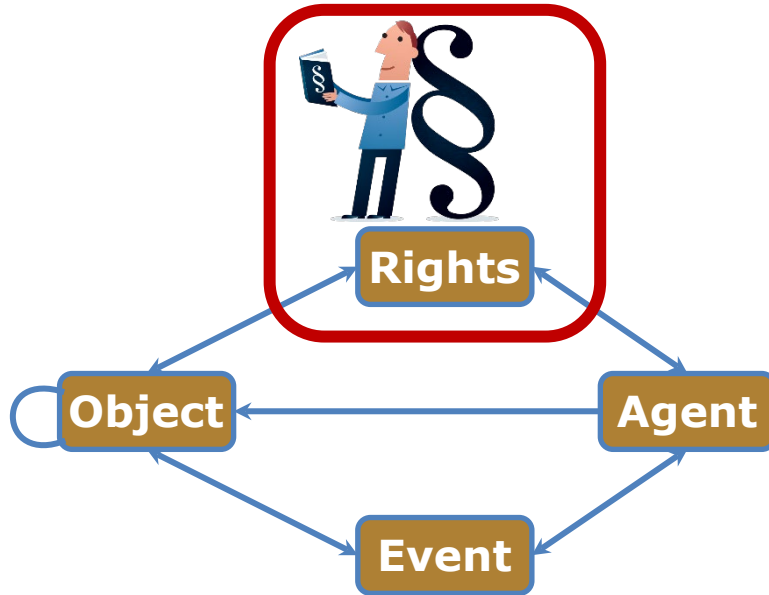
Events
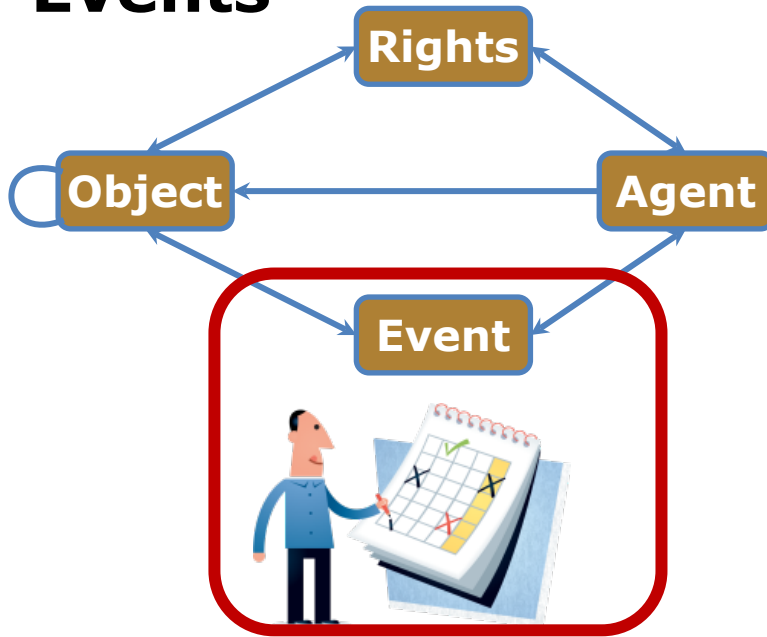
Examples:
- Validation Event e.g. using JHOVE
- Ingest Event

- An action that involves or impacts at least one Object or Agent

- Can document digital provenance, which is needed to track history of Object

Implementation choices:

- Determining which Events are in scope

- Determining which Events should be recorded, as well as the level of granularity

Rights Statements

Rights

Object

Agent

Event

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

Implementation choices :

- Can use other schemes

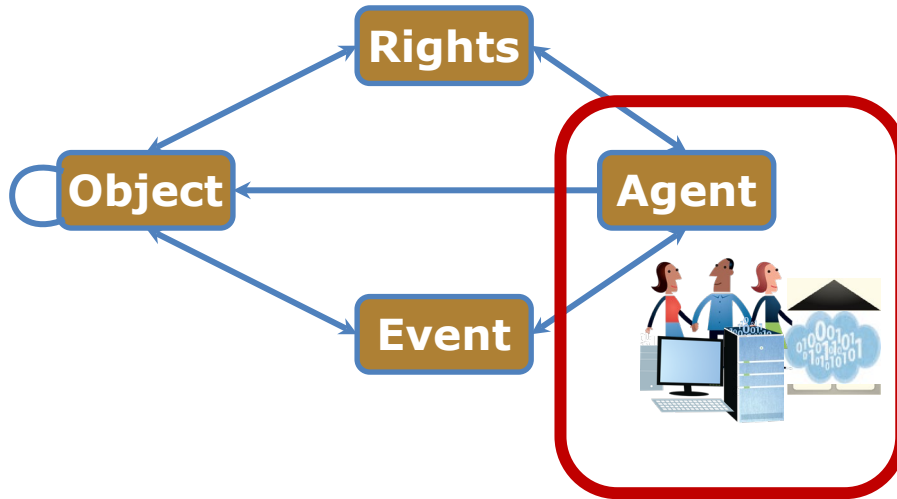 PREMIS form:
  Agent X
  grants **Permission Y**
  to the repository in regard to Object **Z.**

Example:

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

Agents

- Person, organization, or software program/system

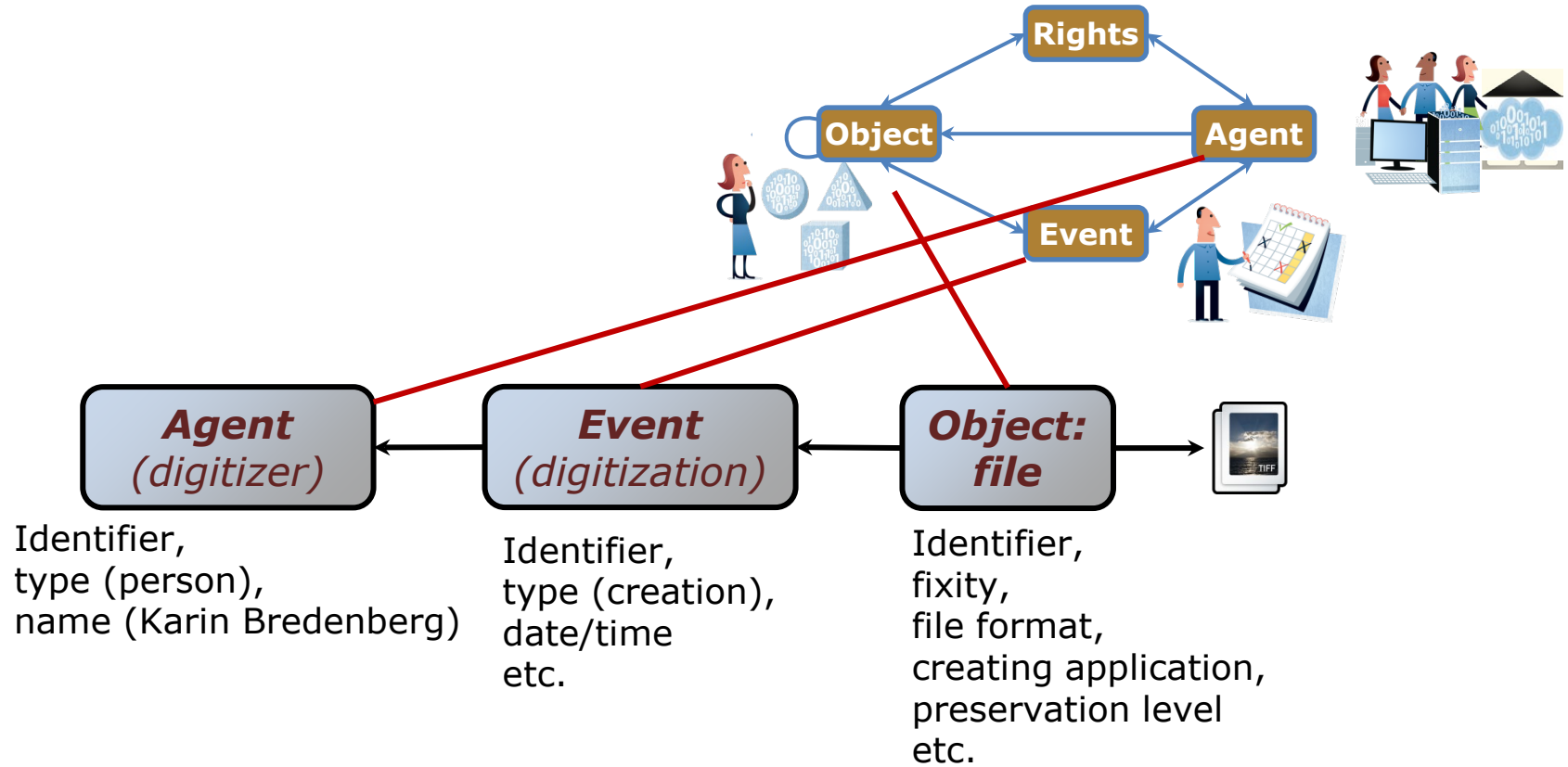- Intended only to identify the agent, and to allow linking from other entity types.

Implementation choices:
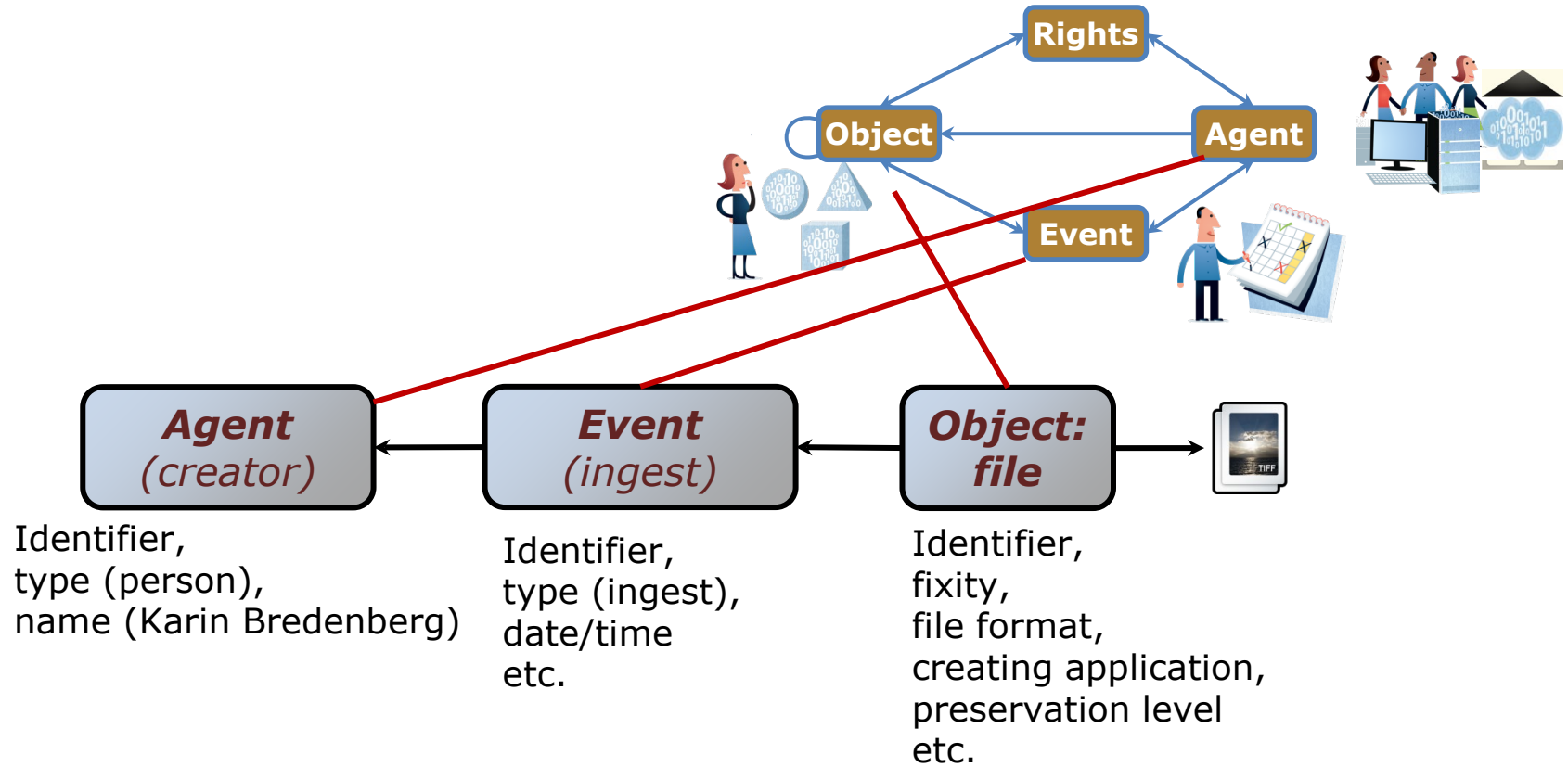
- Can use a richer scheme that may be appropriate.

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

A Digitization Example

Agent
(digitizer)

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

Event
(digitization)

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

**Object:
file**

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

A Born-Digital Example

Agent
(creator)

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

Event
(ingest)

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

**Object:
file**

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

PREservation Metadata Implementation Strategies

menti.com

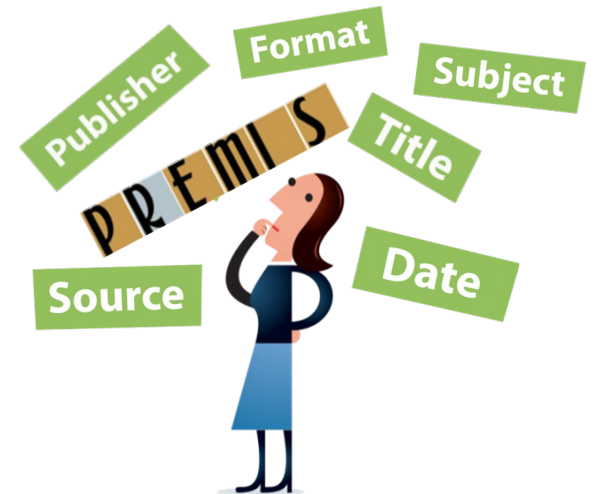PREservation Metadata Implementation Strategies
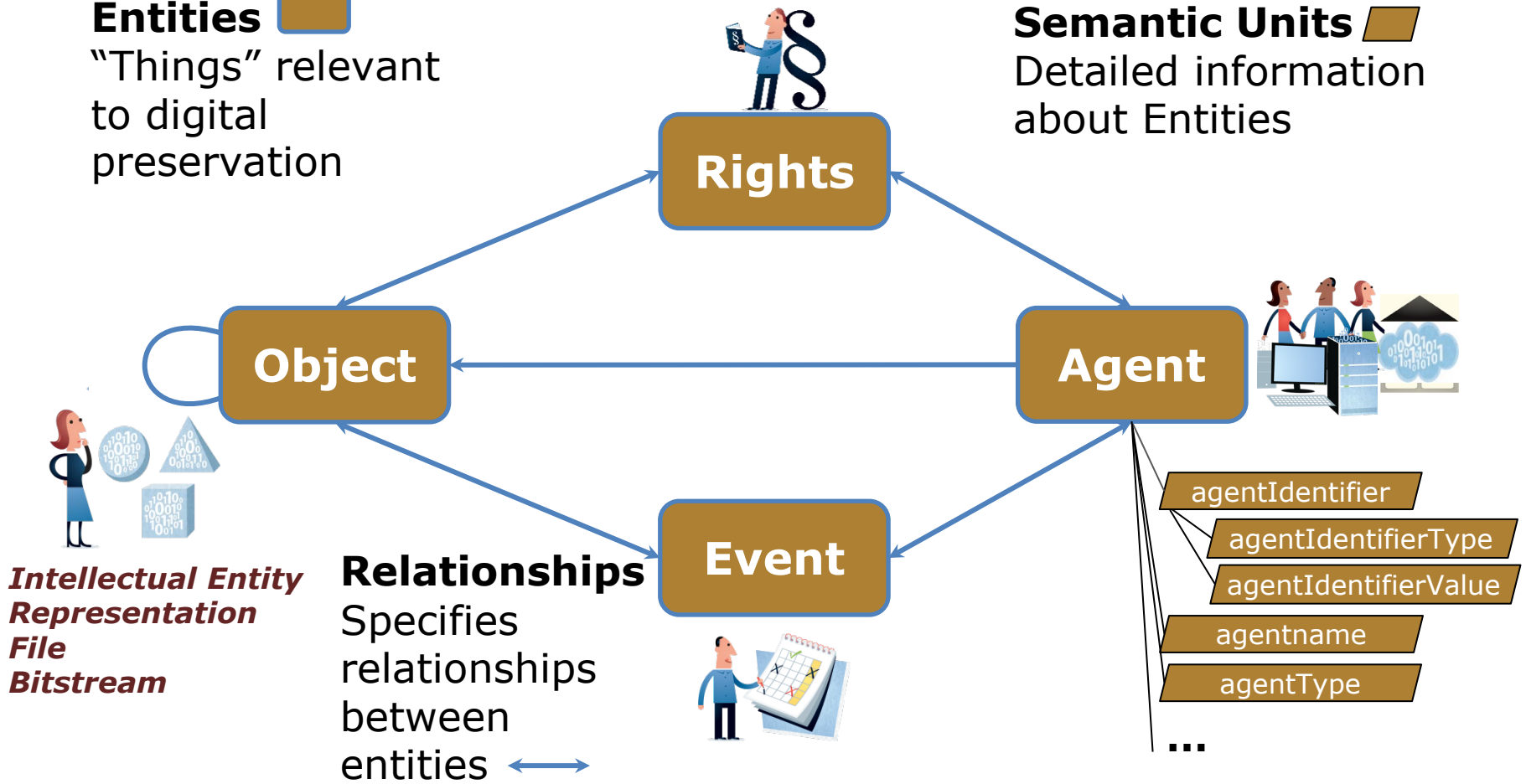
Sarah Romkey

Artefactual

DATA DICTIONARY
DESCRIPTION OF DATA MODEL

PREMIS Data model

Entities
"Things" relevant to digital preservation

Semantic Units
Detailed information about Entities

Rights

Object

Agent

Event

Intellectual Entity
Representation
File
Bitstream

Relationships
Specifies relationships between entities ⟷

agentIdentifier
agentIdentifierType
agentIdentifierValue
agentname
agentType

...

Semantic Units

Properties of Entities

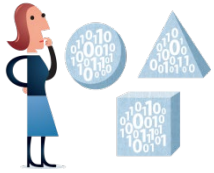A piece of information most repositories need to know in order to carry out their digital preservation functions
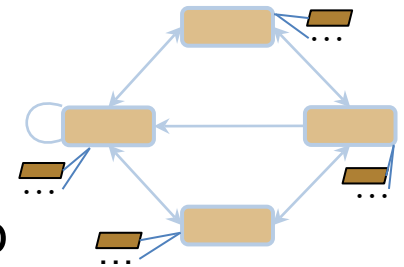
● Example: Two kinds of semantic unit

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

objectIdentifierType [**semantic component**]

objectIdentifierValue [**semantic component**]

High level semantic units for Objects

What technical information is documented?

What is my preservation strategy for this object?

Which object is it?

`ark:/12148/btp6k102002g/f1`

What kind of object is it?

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

Which of its characteristics do I have to preserve??

A digitization example

Rights — Object — Agent — Event
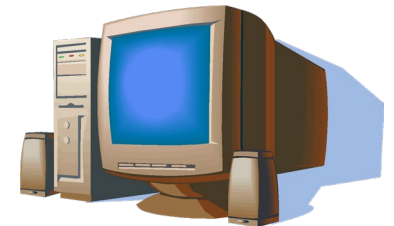
Agent *(digitizer)* ← **Event** *(digitization)* ← **Object: file**
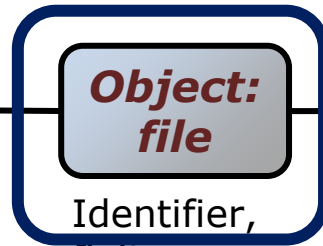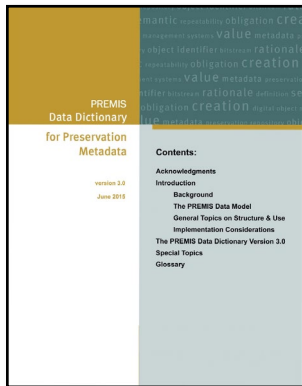
Identifier,
type (person),
name (Karin Bredenberg)

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

Strategies

DATA DICTIONARY

Entity semantic units

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

Object

objectIdentifier

objectIdentifier

1.1   objectIdentifier (M, R)

objectIdentifierType

objectCategory

    1.1.1   objectIdentifierType (M, NR)

    1.1.2   objectIdentifierValue (M, NR)

objectIdentifierValue

PreservationLevel

1.2   objectCategory (M, NR)

Mandatory (M)   ...vationLevel (O, R) [Intellectual Entity, Representation, File]

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

Optional (O)   preservationLevelValue (M, NR) [Intellectual Entity, Representation, File]

    1.3.3   ...servationLevelRole (O, NR) [Intellectual Entity, Representation, File]

Repeatable (R)   ...ationLevelRationale (O, R) [Intellectual Entity, Representation, File]

Not Repeatable (NR)   ...ationLevelDateAssigned (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]

XML example

Using XML schema v3.0:

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

```
<premis>
  <object xsi:type="file">    ⬅
    <objectIdentifier>
      <objectIdentifierType>
        UUID
      </objectIdentifierType>
      <objectIdentifierValue>
        41d10-099-1e2-9
      </objectIdentifierValue>
    </objectIdentifier>
    ...
  </object>
      <rights> ... </rights>    ⬅
      <event> ... </event>    ⬅
      <agent> ... </agent>    ⬅
</premis>
```
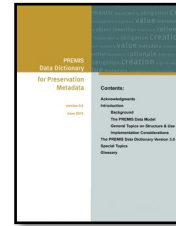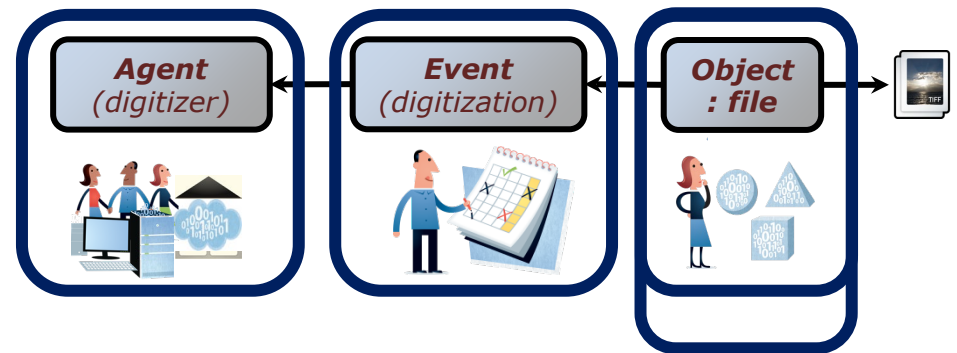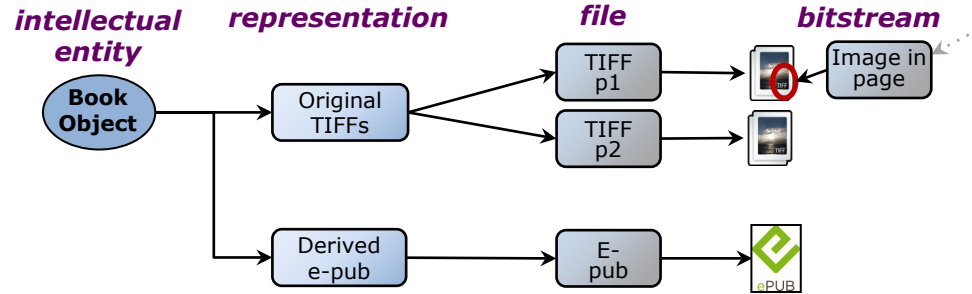
Agent (digitizer) | *Event* (digitization) | *Object : file*

objectCategory

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

- Implemented as a ~~~~
  so not explicitly re~~~~

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

intellectual entity · representation · file · bitstream

Book Object → Original TIFFs → TIFF p1 → Image in page
                              → TIFF p2
          → Derived e-pub → E-pub

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

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 PDFs, etc.

objectCharacteristicsExtension

NEEDS Change

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)

textMD

mix

MPEG7

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

eventType

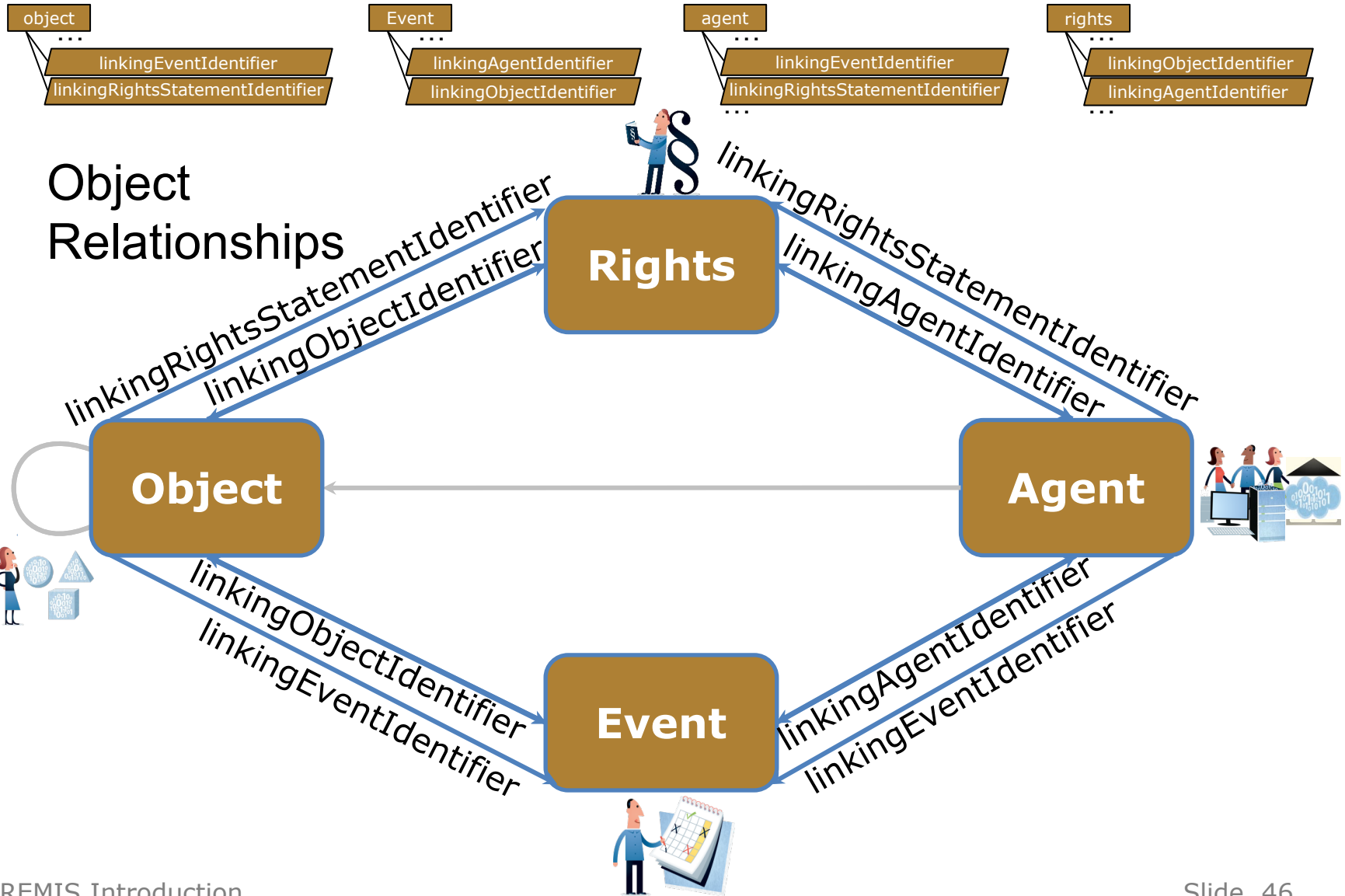> NEEDS Change so its about vocabularies instead

- Names the event

  Ingestion **Virus check** **creation**

  Validation **Compression**

  Message digest calculation

  Decompression ...

  migration **Fixity check**

- This is an element where it is strongly recommended to use a controlled vocabulary, e.g.

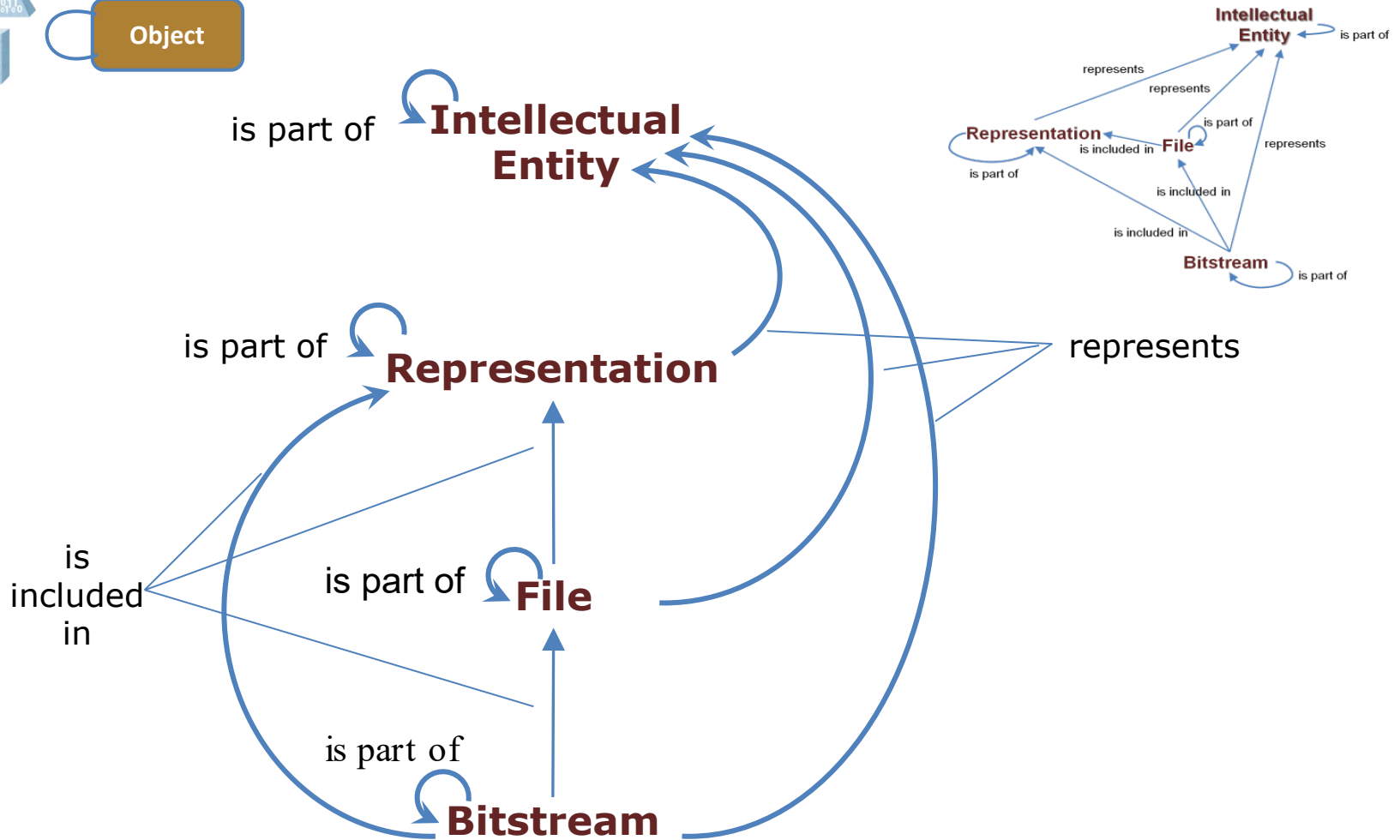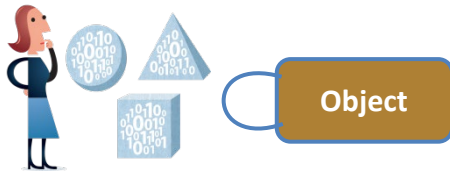 http://id.loc.gov/vocabulary/preservation/eventType.html

  - Can use coded values

  - Granularity is implementation-specific, not mandated by PREMIS

Relationships: Semantic Unit Identifiers

object	Event	agent	rights
...	...	...	...
linkingEventIdentifier	linkingAgentIdentifier	linkingEventIdentifier	linkingObjectIdentifier
linkingRightsStatementIdentifier	linkingObjectIdentifier	linkingRightsStatementIdentifier	linkingAgentIdentifier
		...	...

Object
Relationships

Objects and their interrelations

Object

is part of **Intellectual Entity**

is part of **Representation**

is included in

is part of **File**

is part of **Bitstream**

represents

PREMIS Object Entity – Semantic Units

Object

Example:
Representing Intellectual Entity

relationship

relationshipType — structural

relationshipSubType — represents

relatedObjectIdentifier — Identifier of Intellectual Entity

relatedEventIdentifier — relation through an Event
e.g. for migration

relatedEnvironmentPurpose
relatedEnvironmentCharacteristic — environments
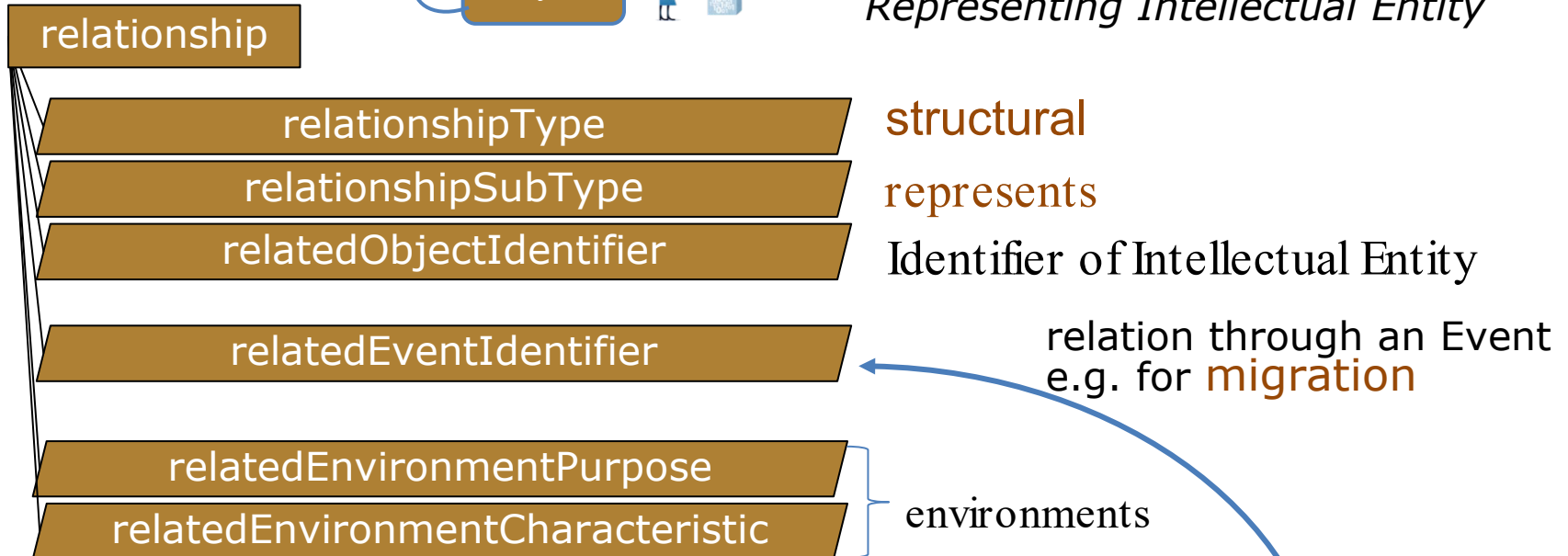
Objects can be associated with Events in two ways:
- If the Object has an associated Event with a ***relationship***
- If the Object has an associated Event with ***no relationship*** to a second Object, e.g. ingest: use **linkingEventIdentifier**

PREMIS — PREservation Metadata Implementation Strategies

Sarah Romkey
Artefactual

USING 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

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

✔ 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/an …

- 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 versus 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 versus a spreadsheet

File, Algorithm, Digest, Origin;

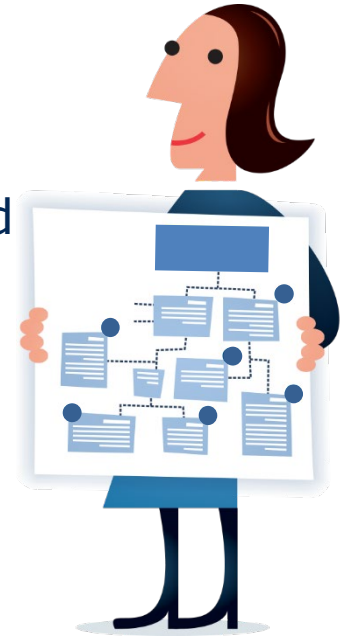Asdf.pdf; SHA-256, 71f920fa275127a7b60fa4d4d41432a3, NRI; ✔

menti.com

Leslie Johnston

United States National Archives and Administration (NARA)

WRAP UP

- Resources
- Current PREMIS EC activities
- 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?

Resources

What you learned about PREMIS today is a good start, but there is much more to explore

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

- Understanding PREMIS

- PREMIS-in-METS guidelines

- Conformance statement

- Examples of implementation

PREMIS Implementors Group Forum:

PIG@listserv.loc.gov

Images in this style are taken from digitalbevaring.dk

Data Dictionary

Lots of other information

Semanti...

Semanti...

Entity ser...

NB: Sema...
Bitstreams...

Definiti...

1.1   objec...
    1.1.1
    1.1.2

Rationa...

1.2   objec...
1.3   prese...
    1.3.1
    1.3.2

Data co...

    1.3.3

Object ...

    1.3.4
    1.3.5

1.4   signi...

Applica...

    1.4.1
    1.4.2

Repeata...

    1.4.3

1.5   objec...

Obligati...

    1.5.1
    1.5.2

Creatio...
Mainten...

    1.5.3
    1.5.4

Usage no...

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

Copyright © 2009 The Library of Congress, excepto en EE. UU.

Se debe mencionar esta publicación como procedencia de cualquier extracto de ella que se cite.

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
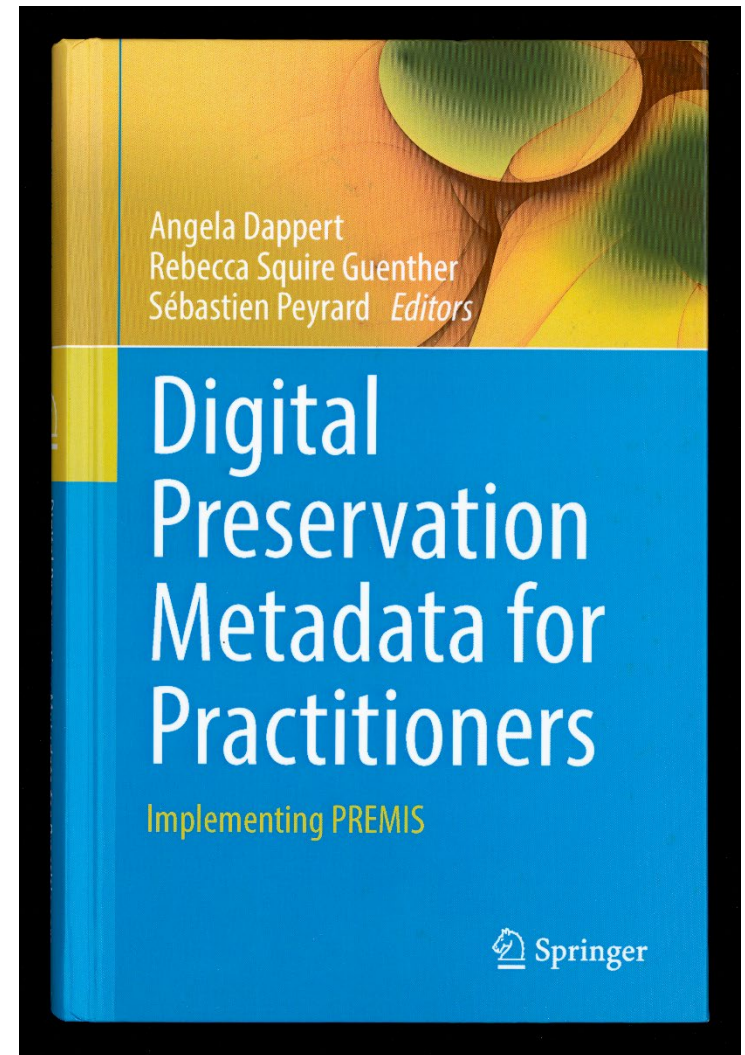
Exercises

- There are 3 exercises you can use to start working with PREMIS metadata

  - https://zenodo.org/communities/premis/search?page=1&size=20&q=exercise

  - The solutions are available

On your own!

Today

- You have had an introduction to PREMIS! Today has covered a lot of information.

- Use the resources and exercises to learn more.

- Participate in the discussions on the email list!

- Watch the recording in English with more detail that will be available ????

menti.com

Finally...

PREMIS is a **<u>community</u>** standard.

- Participate!

    - Send examples

    - Ask questions

    - Send suggestions

Thank you!

Sarah and Leslie