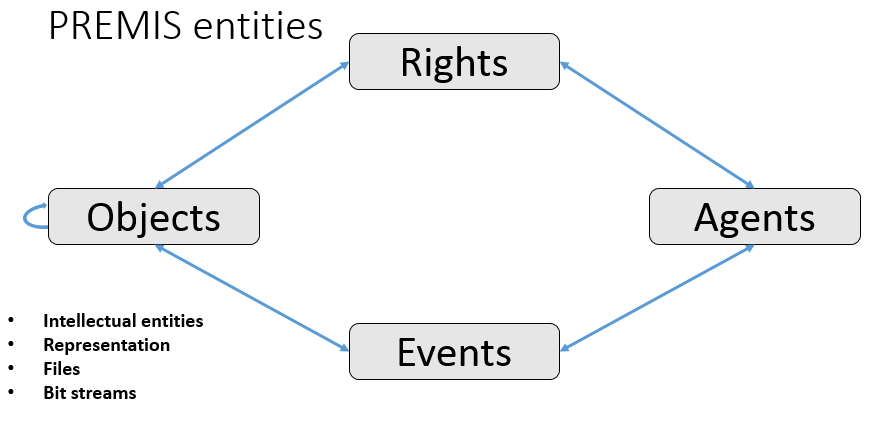
**Introduction to PREMIS**

**Exercise booklet**

**Bodleian Libraries**

**07/06/2018**

**Version 1.1**



**Object modelling**

**PREMIS has four categories of objects:**

**Intellectual entity:** *A conceptual way to describe a work*

**Representation:** *A particular representation of an intellectual entity*

**File:** *Binary information that is available to a computer program*

**(Bit stream**: *A stream of data in binary form)*

*Below is an example of the work Alice in Wonderland modelled in PREMIS*

**Intellectual Entity**

Alice in Wonderland

**Representation**

Alice’s Adventures in Wonderland, audiobook recording by Kara Shallenberg, 2010, LibriVox publishing

**Representation**

Alice’s Adventures in Wonderland, digitized version of 1865 print edition held by the British Library

**Representation**

Alice’s Adventures in Wonderland, Lewis Carol  
  
1865, MacMillian and Co London

**File**   
  
TIFF file

**File**

QuickTime audio file

TIFF file

TIFF file

TIFF file

TIFF file

**Exercise A.1:**

**Scenario:** The book Brave New World by Aldous Huxley was published by Chatto and Windus in 1932. A Kindle e-book version (MOBI file format) was published by Vintage Digital in 2008.

How would you model the following items to the PREMIS object model?

**Intellectual Entity**

……………………………..

**Representation**

………………………………..

**Representation**

…………………………………

**File (s)**  
………………

……………….

……………….

**Exercise A.2:**

**Scenario:** The song *Hallelujah* was written by Leonard Cohen in 1984. In 2009 Susan Boyle received standing ovations after she performed the song on the TV show “Britain’s Got Talent”.

How would you model the following items to the PREMIS object model?

1. A 1985 radio recording (MP2) of Leonard Coen singing Hallelujah
2. A YouTube video recording (MP4) of Susan Boyle performing Hallelujah on Britain’s Got Talent
3. Fan photographs (JPEG) from a mobile phone taken by an audience member during the above performance

**Intellectual Entity**

……………………………..

**Representation**

………………………………..

**Representation**

…………………………………

**File (s)**  
………………

……………….

……………….

**File (s)**  
………………

……………….

……………….

**Exercise A.3:**

**Scenario:** Harry Potter and the Philosopher’s Stone by J.K. Rowling was first published in 1997. The Harry Potter books are now part of a huge international franchise.

Describe the following items in the PREMIS object model, using the blank space below.

* J.K. Rowling, 1997. Harry Potter and the Philosophers’ Stone. Bloomsbury. Hardback.
* J.K. Rowling, 2014 (ed). Harry Potter and the Philosophers’ Stone. Bloomsbury. Paperback.
* Harry Potter and the Philosopher’s Stone, book 1. Pottermore audiobook. Unabridged. 20 November 2015. Narrated by Stephen Fry. (AA – Audible Audio file format)
* Harry Potter and the Philosopher’s Stones. EA GAMES. PC game. 16 November 2001. (Executable file)

**Data modelling**

**Exercise B.1:**

**Scenario:** Steve, a digital archivist, migrates a TIFF image file to a JPEG using ImageMagick software. The TIFF file was digitized by Steve’s own organisation and made available under a CC-BY NC license.

In this scenario – who/what is:

* The agent(s)?…………………………………………………………………………………….
* The object(s)?………………………………………………………………………………..…
* The event(s)?.......................................................................................
* And what are the legal rights?..............................................................

**Exercise B.2:**

**Scenario**: A National Archive runs the file format identification tool DROID over a new deposit from a governmental department.

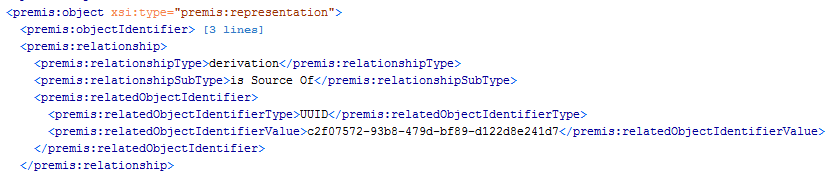
In this scenario – who/what is:

* The agent(s)?…………………………………………………………………………………….
* The object(s)?………………………………………………………………………………..…
* The event(s)?.......................................................................................

**Exercise B.3:** PREMIS Object

What type of object is this?

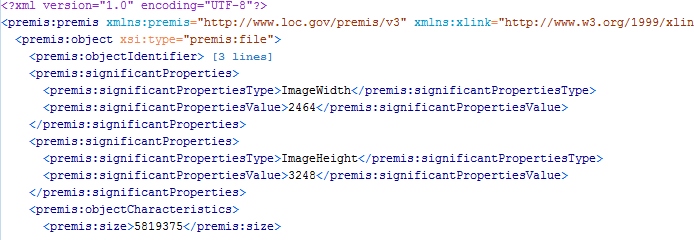
1. **Intellectual entity**
2. **Representation**
3. **Digital file**
4. **Bit stream**



**Exercise B.4:** PREMIS Object

What type of object is this?

1. **Intellectual entity**
2. **Representation**
3. **Digital file**
4. **Bit stream**



PREMIS lists the size of files and bit streams in bytes. How many bytes is this object?...........................

What else can you tell about the properties of this object?....................................................................

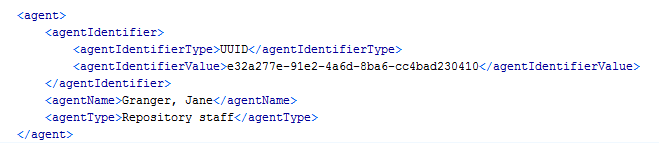
……………………………………………………………………………………………………………………………………………………………

**Exercise B.5: Agent**

What is the name of the agent?...................................................................................................

What is the agent’s role in their organisations?.............................................................................

………………………………………………………………………………………………………………………………………………….

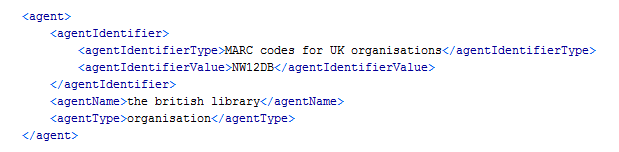


**Exercise B.6: Agent**

What is the name of the agent?.......................................................................................................

What type of agent is this?..............................................................................................................

What is the agents identification code?...........................................................................................

****

**Exercise B.7: Events**

What type of event is this (eventType)? ………………………………………….………………………………………………..

What information can you tell from the event (eventOutcomeInformation)?......................................

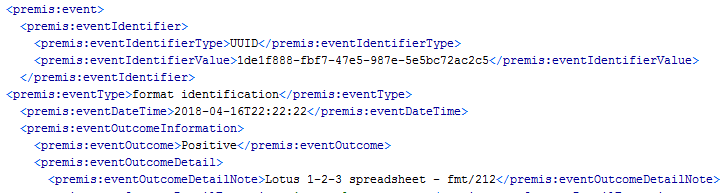
................................................................................................................................................................

................................................................................................................................................................

How do you think this information is useful to an organisation?.........................................................

................................................................................................................................................................

................................................................................................................................................................

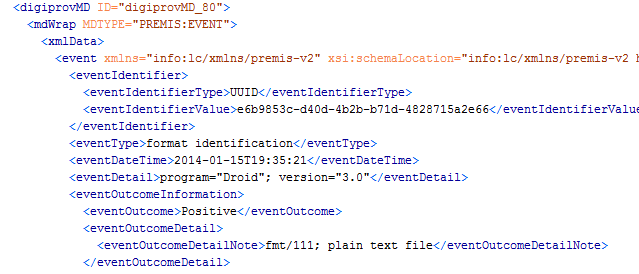


**Exercise B.8: Event (METS file)**

What type of event is this?..................................................................................................................

What software program was used? ………………..………………………………………………………………………….…..

When was the tool ran?...........................................................................................................................



**Exercise B.9: Events and agents**

What kind of event is described below?................................................................................................

What was the outcome of the event?...................................................................................................

What is the name of the software which executed the event?............................................................

What version of the software was it?...................................................................................................

What is the (internal) identifier number of the software/agent?.........................................................

How do you think this information is useful to an organisation?.........................................................

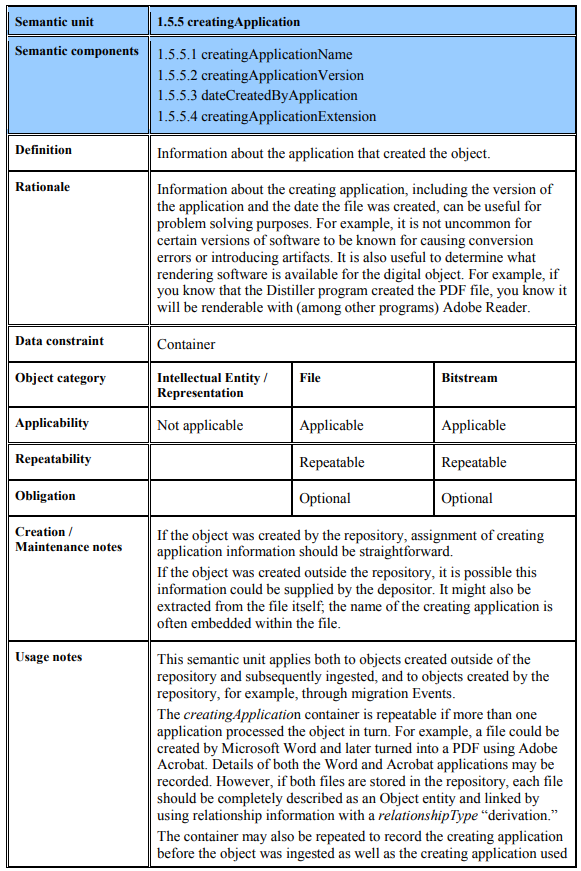
................................................................................................................................................................



**Exercise C.1: Using the data dictionary**

This is an extract from the PREMIS Data Dictionary. The full dictionary is available at: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

Look at the metadata element/semantic unit called “creatingApplication” (below) from page 75 of the PREMIS Data Dictionary (version 3.0).



Why might a repository want to record this information?

…………………………………………………………………………………………………………………………………………………

…………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………….

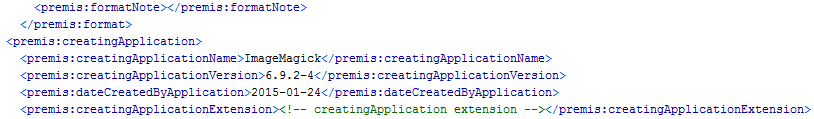
How many “semantic components” (or sub-categories) does the unit have?

1. **2**
2. **7**
3. **4**

At what object level can this information be recorded? Circle all that apply.

1. **Intellectual entity**
2. **Representation**
3. **Digital file**
4. **Bitstream**

*Look at the PREMIS XML extract below and circle the semantic components from page 75.*



**C.2 - Exercise: using the data dictionary**

Go onto the library of congresses website and open the PREMIS 3.0 Data Dictionary. <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

**Read about the following metadata elements:**

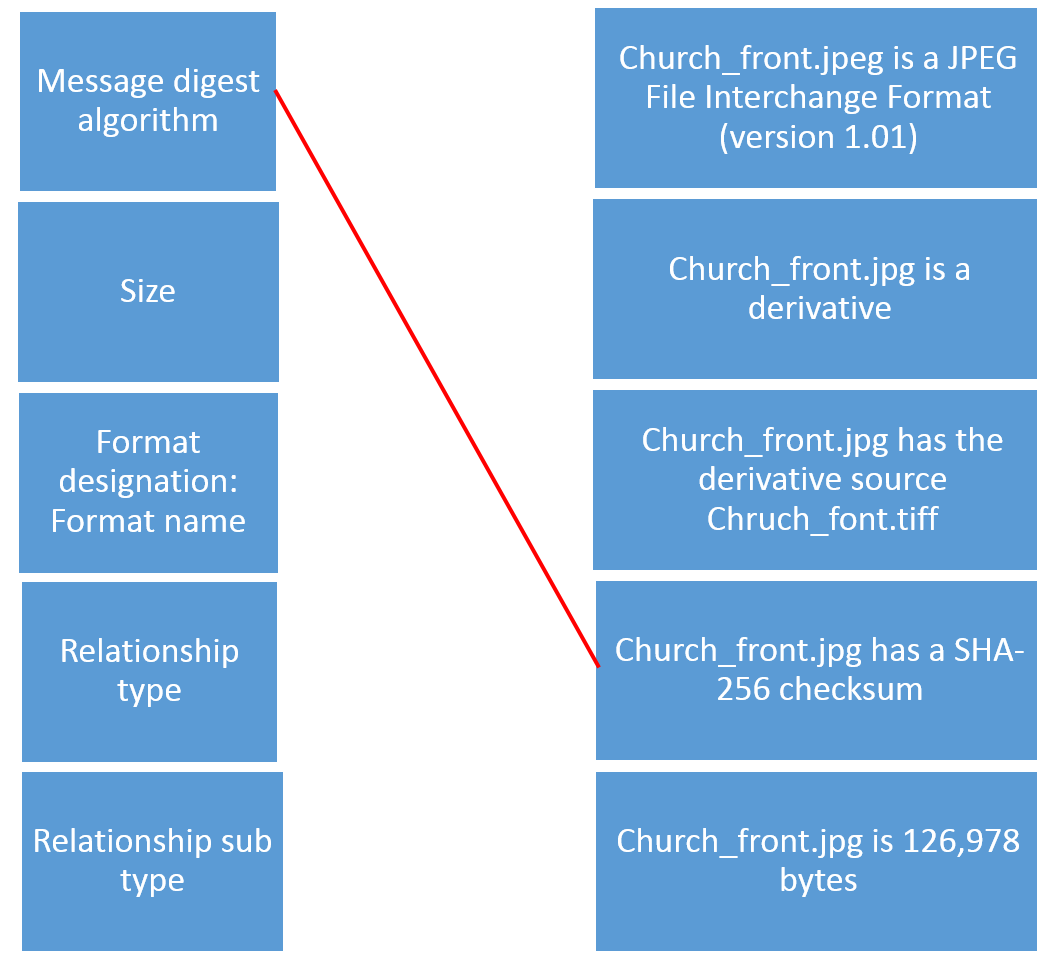
**1.5.2.1 -** Message Digest Algorithm *(page 61)*  
**1.5.3 –** Size *(page 64)*

**1.5.4.1.1 -** Format Designation: Format name *(page 68)*

**1.13.1 -** Relationship type *(page 119)*

**1.13.3 -** Relationship sub type *(page 120)*

On the right is a list of information about the image “Church\_front.jpeg”. Draw a line between the information on the right and the metadata elements (on the left) which could be used to record it. (An answer has already been provided for the metadata element “message digest algorithm” to get you started)

****

**C.3: Fill in the blanks**

Read about the following metadata elements in the Data Dictionary: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

**EventType (page 141)  
EventDataTime (page 142)**

**EventOutcome (p.147)**

**AgentName (p.163)**

**AgentType (p.164)**

**AgentVersion (p.165)**

Using what you have learnt, fill in the following information in the missing sections of the PREMIS XML document ***on the next page.***  
**Scenario:** A repository runs a routine virus check over its collections using CLAM AV (v.0.100.0) on the 12th January 2018. It finds a virus in a file which has the unique identifier e32a277e-91e2-4a6d-8ba6-cc4bad230410.

<event>  
 <eventIdentifier>  
 <eventIdentifierType>UUID</eventIdentifierType>  
 <eventIdentifierValue>00d60462-23a0-48e7-ae56-4e325d3df4d9 </eventIdentifierValue>  
 </eventIdentifier>  
 <eventType>…………………………………..</eventType>  
 <eventDateTime> ………………………………….</eventDateTime>  
 <eventOutcomeInformation>  
 <eventOutcome>………………………………….</eventOutcome>  
 </eventOutcomeInformation>   
 <linkingAgentIdentifier>  
 <linkingAgentIdentifierType>Agent ID </linkingAgentIdentifierType>  
 <linkingAgentIdentifierValue>Agent-123-123 </linkingAgentIdentifierValue>  
 </linkingAgentIdentifier>  
 <linkingObjectIdentifier>  
 <linkingObjectIdentifierType>UUID</linkingObjectIdentifierType>  
 <linkingObjectIdentifierValue>e32a277e-91e2-4a6d-8ba6-cc4bad230410 </linkingObjectIdentifierValue>  
 </linkingObjectIdentifier>   
 </event>  
 <agent>  
 <agentIdentifier>  
 <agentIdentifierType>Agent ID</agentIdentifierType>  
 <agentIdentifierValue>Agent-123-123</agentIdentifierValue>  
 </agentIdentifier>  
 <agentName>………………………………….</agentName>   
 <agentType>………………………………….</agentType>   
 <agentVersion>………………………………….</agentVersion>  
 </agent>

**C.4: Fill in the blanks**

Fill in the following information in the missing sections of the PREMIS XML document.

Use the Data Dictionary to look up how the different metadata elements should be used. (i.e. size, formatName etc.)  
**Scenario:** The object **“**e32a277e-91e2-4a6d-8ba6-cc4bad230410”, which contained the virus, is a PDF file. By extracting technical metadata from the file, the archive finds out that it is a PDF version 1.7 which is 10819375 bytes. It was originally created on the 23ed of May 2014 using the software Adobe Acrobat (version 11.0).

<premis:object xsi:type="premis:file">  
 <premis:objectIdentifier>  
 <premis:objectIdentifierType>UUID</premis:objectIdentifierType>  
 <premis:objectIdentifierValue> e32a277e-91e2-4a6d-8ba6-cc4bad230410 </premis:objectIdentifierValue>

[…..]

[…..]

<premis:size>………………………………………..</premis:size>  
 <premis:format>  
 <premis:formatDesignation> <premis:formatName>……………………………………….</premis:formatName>  
 <premis:formatVersion>………..</premis:formatVersion>  
 </premis:formatDesignation>  
 <premis:formatRegistry>  
 <premis:formatRegistryName>PRONOM</premis:formatRegistryName>  
 <premis:formatRegistryKey>fmt/276</premis:formatRegistryKey>  
 <premis:formatRegistryRole>identification</premis:formatRegistryRole>  
 </premis:formatRegistry>  
 </premis:format>  
 <premis:creatingApplication> <premis:creatingApplicationName>…………....…………</premis:creatingApplicationName>  
 <premis:creatingApplicationVersion>………………….…</premis:creatingApplicationVersion> <premis:dateCreatedByApplication>…………………………………</premis:dateCreatedByApplication>  
 </premis:creatingApplication>

[…..]

[…..]

</object>

**C.5: Fill in the blanks**

**Scenario:** An archive receives the papers of a famous poet. These includes a number of floppy disks with files from the 1980’s. The archive extract technical information from the files which they include in their catalogue.   
  
They retrieve the following information from a WordPerfect (v.5.1) file containing a draft poem which was later published.

*Character count: 397*

*Word count: 79*

*Size = 2859385*

**Task:**

Fill in the extracted information in the missing sections of the PREMIS XML document **on the next page.**

Use the Data Dictionary to look up how the different metadata elements should be used.

**Additional question:**

What was the name of the poem?..........................................................................................................

<premis:object xsi:type="premis:file">  
 <premis:objectIdentifier>  
 <premis:objectIdentifierType>UUID</premis:objectIdentifierType>  
 <premis:objectIdentifierValue> 90f5a61f-5d51-4282-b1da-8480457cc43b</premis:objectIdentifierValue>  
 </premis:objectIdentifier>  
 <premis:significantProperties>  
 <premis:significantPropertiesType>character count</premis:significantPropertiesType>  
 <premis:significantPropertiesValue>…….</premis:significantPropertiesValue>  
 </premis:significantProperties>  
 <premis:significantProperties>  
 <premis:significantPropertiesType>word count</premis:significantPropertiesType>  
 <premis:significantPropertiesValue>…….</premis:significantPropertiesValue>  
 </premis:significantProperties>  
 <objectCharacteristics>  
 <premis:size>…………..….</premis:size>  
 <premis:format>  
 <premis:formatDesignation>  
 <premis:formatName>WordPerfect for DOS</premis:formatName>  
 <premis:formatVersion>..….</premis:formatVersion>  
 </premis:formatDesignation>  
 </premis:format>  
 </objectCharacteristics>  
<premis:originalName>%transferDirectory%poetry/LoveLettersToNina.wpd</premis:originalName>

[…..]

[…..]

</object>

**Exercise D.1: Rights**

What is the name of the person who holds copyright in the object?.....................................................

…………………………………………………………………………………………………………………….…………………………………….

What country (jurisdiction) is copyright held in? ……………………………………………………………………………..

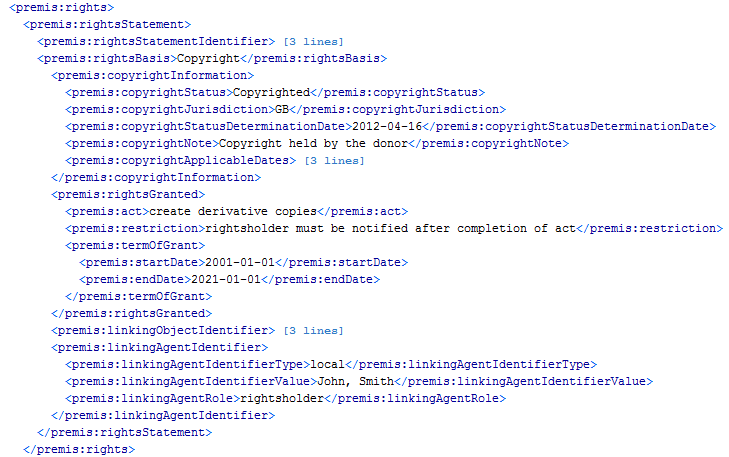
The copyright holder has allowed the repository to create “derivatives” (new versions) of the files which they deposited. However, under what conditions can new derivatives be created?...................

…………………………………………………………………………………………………………………………………………………………..

…………………………………………………………………………………………………………………………………………………………..

The copyright holder has only granted the right to create “derivatives” for a limited period. What date does this right stop?

……………………………………………………………………………………………………………………………………………………..…

……………………………………………………………………………………………………………………………………………………..…

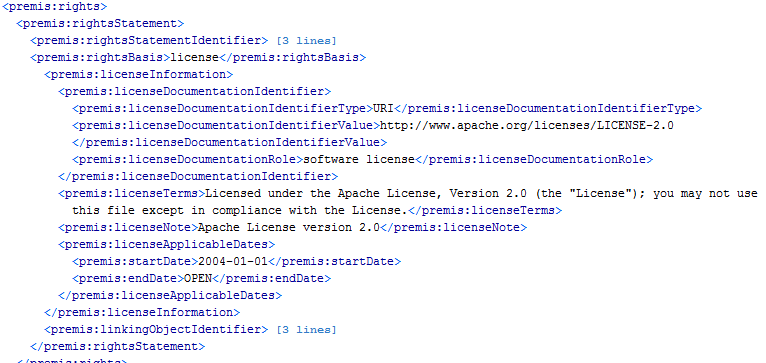
**Exercise D.2: Rights**

What type of license does this refer to?..................................................................................................

When did the license start?.....................................................................................................................

How do you think this information is useful to an organisation?.........................................................

................................................................................................................................................................

****

**Exercise E.1: Finished the exercises?**

Log onto the Library of Congress’ main PREMIS page (<https://www.loc.gov/standards/premis/>)

1. Click on ***“PREMIS Implementation Registry***” from the main menu and read about how different organisations have chosen to implement PREMIS in their local institutions.
2. Click on “Preservation Events Controlled Vocabulary at id.loc.gov” and look at the controlled vocabularies for: *AgentType, EventType, and ObjectCategory*