

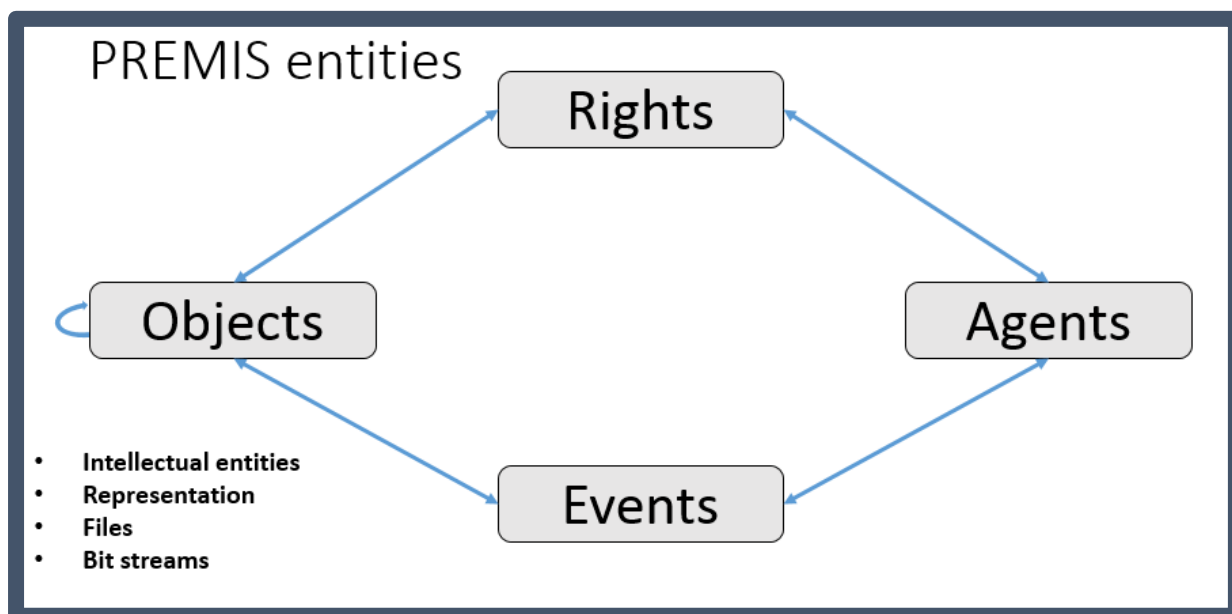
# Introduction to PREMIS

## Exercise booklet

**Bodleian Libraries**

**07/06/2018**

**Version 1.1**



# Object modelling

PREMIS has four categories of objects:

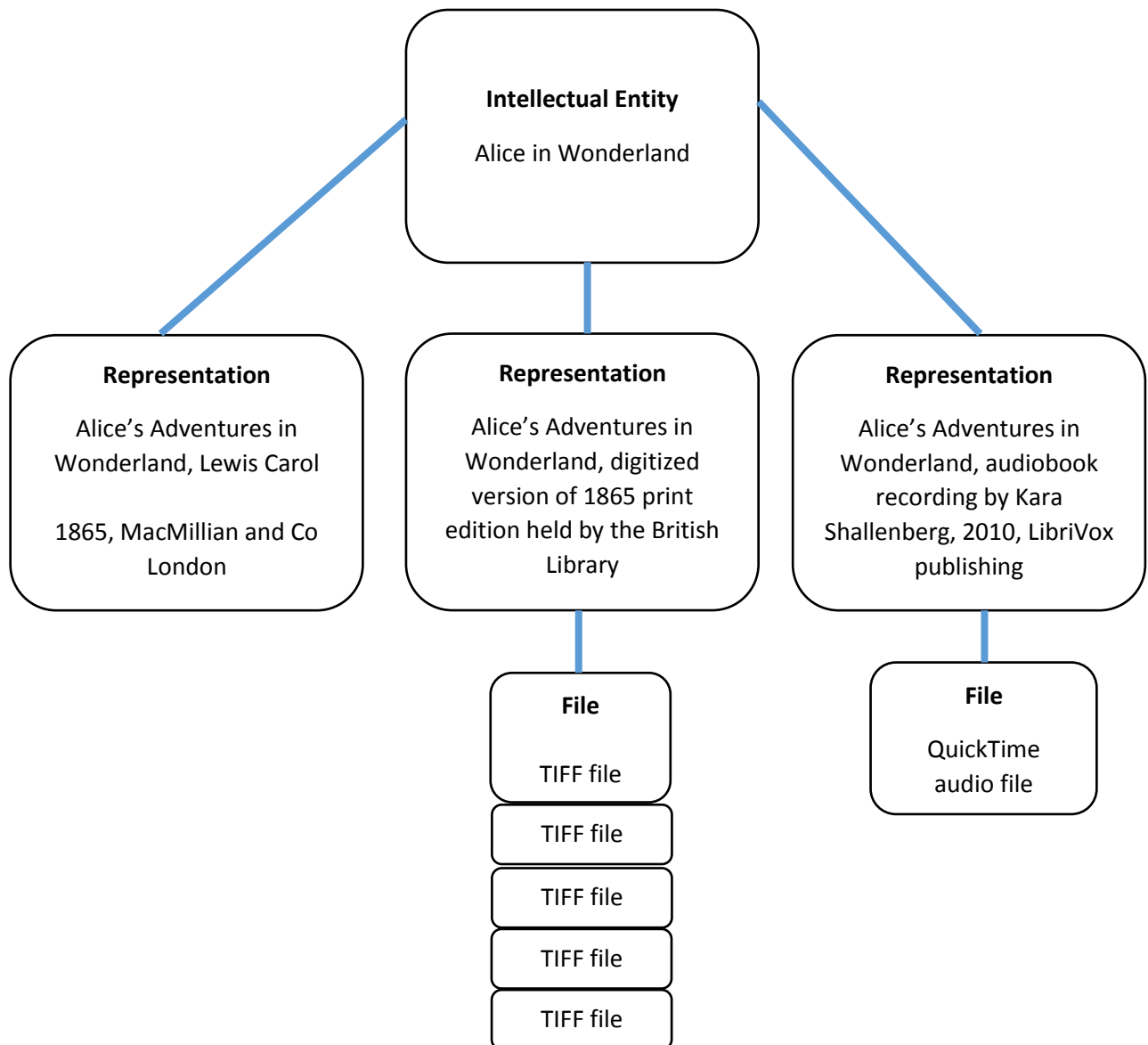
**Intellectual entity:** *A conceptual way to describe a work*

**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)*

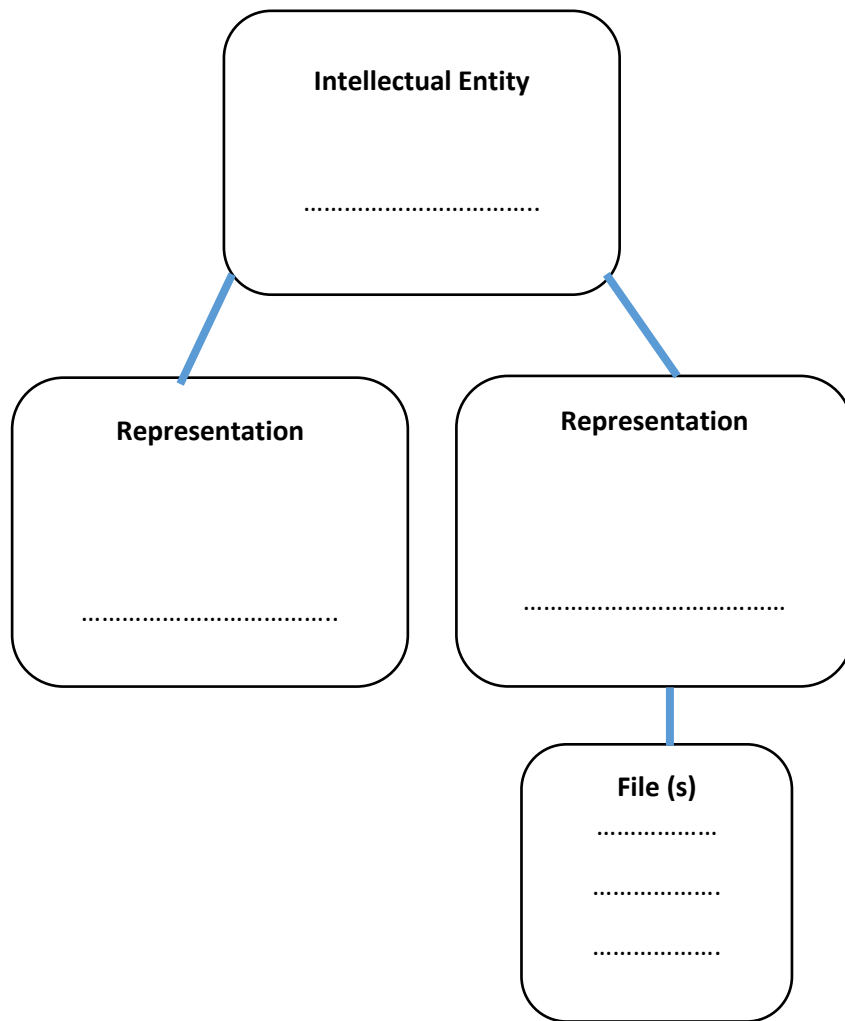
*Below is an example of the work Alice in Wonderland modelled in PREMIS*



## 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?

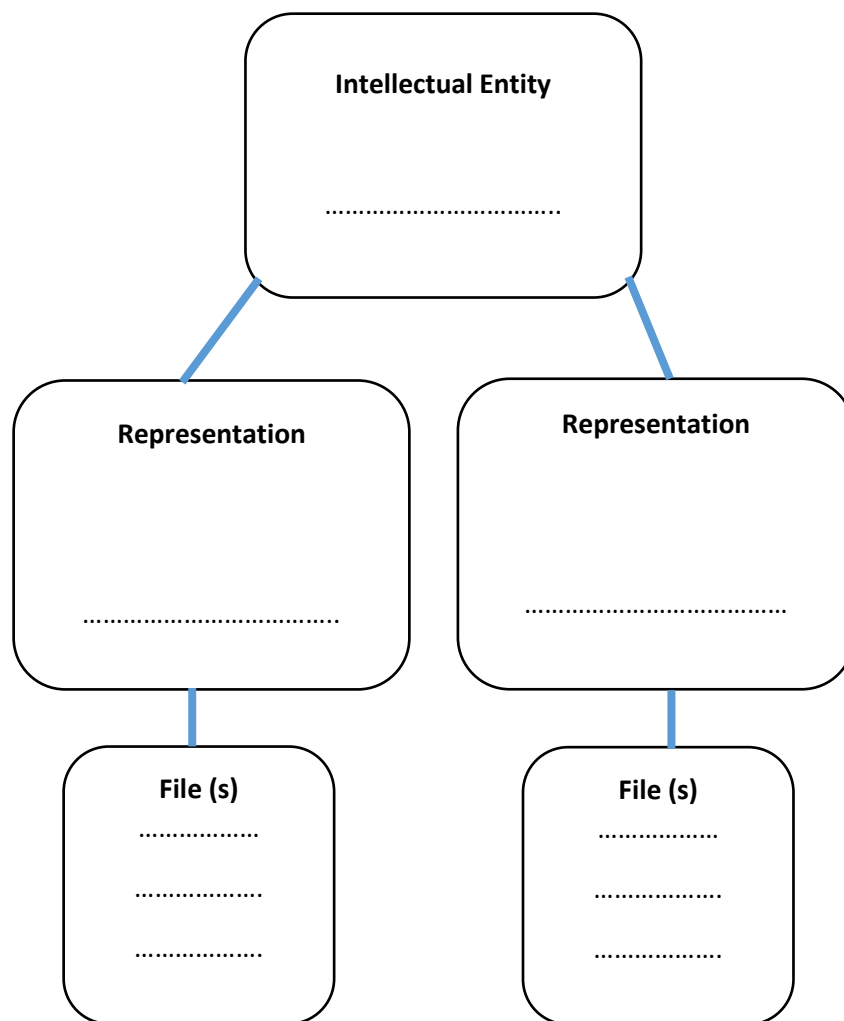


## Exercise A.2:

**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



### 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?

- A. Intellectual entity
- B. Representation
- C. Digital file
- D. Bit stream

```
<premis:object xsi:type="premis:representation">
  <premis:objectIdentifier> [3 lines]
  <premis:relationship>
    <premis:relationshipType>derivation</premis:relationshipType>
    <premis:relationshipSubType>is Source Of</premis:relationshipSubType>
    <premis:relatedObjectIdentifier>
      <premis:relatedObjectIdentifierType>UUID</premis:relatedObjectIdentifierType>
      <premis:relatedObjectIdentifierValue>c2f07572-93b8-479d-bf89-d122d8e241d7</premis:relatedObjectIdentifierValue>
    </premis:relatedObjectIdentifier>
  </premis:relationship>
```

---

## Exercise B.4: PREMIS Object

What type of object is this?

- A. Intellectual entity
- B. Representation
- C. Digital file
- D. Bit stream

```
<premis:object xsi:type="premis:file">
  <premis:objectIdentifier> [3 lines]
  <premis:significantProperties>
    <premis:significantPropertiesType>ImageWidth</premis:significantPropertiesType>
    <premis:significantPropertiesValue>2464</premis:significantPropertiesValue>
  </premis:significantProperties>
  <premis:significantProperties>
    <premis:significantPropertiesType>ImageHeight</premis:significantPropertiesType>
    <premis:significantPropertiesValue>3248</premis:significantPropertiesValue>
  </premis:significantProperties>
  <premis:objectCharacteristics>
    <premis:size>5819375</premis:size>
```

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?.....

.....

```
<agent>
  <agentIdentifier>
    <agentIdentifierType>UUID</agentIdentifierType>
    <agentIdentifierValue>e32a277e-91e2-4a6d-8ba6-cc4bad230410</agentIdentifierValue>
  </agentIdentifier>
  <agentName>Granger, Jane</agentName>
  <agentType>Repository staff</agentType>
</agent>
```

## Exercise B.6: Agent

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

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

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

```
<agent>
  <agentIdentifier>
    <agentIdentifierType>MARC codes for UK organisations</agentIdentifierType>
    <agentIdentifierValue>NW12DB</agentIdentifierValue>
  </agentIdentifier>
  <agentName>the british library</agentName>
  <agentType>organisation</agentType>
</agent>
```

## 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?.....

.....

.....

```
<premis:event>
  <premis:eventIdentifier>
    <premis:eventIdentifierType>UUID</premis:eventIdentifierType>
    <premis:eventIdentifierValue>1de1f888-fbf7-47e5-987e-5e5bc72ac2c5</premis:eventIdentifierValue>
  </premis:eventIdentifier>
  <premis:eventType>format identification</premis:eventType>
  <premis:eventDateTime>2018-04-16T22:22:22</premis:eventDateTime>
  <premis:eventOutcomeInformation>
    <premis:eventOutcome>Positive</premis:eventOutcome>
    <premis:eventOutcomeDetail>
      <premis:eventOutcomeDetailNote>Lotus 1-2-3 spreadsheet - fmt/212</premis:eventOutcomeDetailNote>
    </premis:eventOutcomeDetail>
  </premis:eventOutcomeInformation>
</premis:event>
```

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

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

What software program was used? .....

When was the tool ran?.....

```
<digiprovMD ID="digiprovMD_80">
  <mdWrap MDTYPE="PREMIS:EVENT">
    <xmlData>
      <event xmlns="info:lc/xmlns/premis-v2" xsi:schemaLocation="info:lc/xmlns/premis-v2 http://www.loc.gov/standards/premis/v2/premis-v2.xsd">
        <eventIdentifier>
          <eventIdentifierType>UUID</eventIdentifierType>
          <eventIdentifierValue>e6b9853c-d40d-4b2b-b71d-4828715a2e66</eventIdentifierValue>
        </eventIdentifier>
        <eventType>format identification</eventType>
        <eventDateTime>2014-01-15T19:35:21</eventDateTime>
        <eventDetail>program="Droid"; version="3.0"</eventDetail>
        <eventOutcomeInformation>
          <eventOutcome>Positive</eventOutcome>
          <eventOutcomeDetail>
            <eventOutcomeDetailNote>fmt/111; plain text file</eventOutcomeDetailNote>
          </eventOutcomeDetail>
        </eventOutcomeInformation>
      </event>
    </xmlData>
  </mdWrap>
</digiprovMD>
```

## 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?.....

.....

```
<event>
  <eventIdentifier>
    <eventIdentifierType>Local ID</eventIdentifierType>
    <eventIdentifierValue>123456789</eventIdentifierValue>
  </eventIdentifier>
  <eventType>virus scan</eventType>
  <eventDateTime>2014-02-02T12:09:07Z</eventDateTime>
  <eventOutcomeInformation>
    <eventOutcome>Failed</eventOutcome>
    <eventOutcomeDetail>
      <eventOutcomeDetailNote>Virus log.txt</eventOutcomeDetailNote>
    </eventOutcomeDetail>
  </eventOutcomeInformation>
  <linkingAgentIdentifier>
    <linkingAgentIdentifierType>Agent ID</linkingAgentIdentifierType>
    <linkingAgentIdentifierValue>A00004</linkingAgentIdentifierValue>
  </linkingAgentIdentifier>
  <linkingObjectIdentifier>
    <linkingObjectIdentifierType>UUID</linkingObjectIdentifierType>
    <linkingObjectIdentifierValue>e32a277e-91e2-4a6d-8ba6-cc4bad230410</linkingObjectIdentifierValue>
  </linkingObjectIdentifier>
</event>
<agent>
  <agentIdentifier>
    <agentIdentifierType>Agent ID</agentIdentifierType>
    <agentIdentifierValue>A00004</agentIdentifierValue>
  </agentIdentifier>
  <agentName>CLAM AV</agentName>
  <agentVersion>0.95.2</agentVersion>
</agent>
```

## 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).

<b>Semantic unit</b>	<b>1.5.5 creatingApplication</b>		
<b>Semantic components</b>	1.5.5.1 creatingApplicationName 1.5.5.2 creatingApplicationVersion 1.5.5.3 dateCreatedByApplication 1.5.5.4 creatingApplicationExtension		
<b>Definition</b>	Information about the application that created the object.		
<b>Rationale</b>	Information about the creating application, including the version of the application and the date the file was created, can be useful for problem solving purposes. For example, it is not uncommon for certain versions of software to be known for causing conversion errors or introducing artifacts. It is also useful to determine what rendering software is available for the digital object. For example, if you know that the Distiller program created the PDF file, you know it will be renderable with (among other programs) Adobe Reader.		
<b>Data constraint</b>	Container		
<b>Object category</b>	<b>Intellectual Entity / Representation</b>	<b>File</b>	<b>Bitstream</b>
<b>Applicability</b>	Not applicable	Applicable	Applicable
<b>Repeatability</b>		Repeatable	Repeatable
<b>Obligation</b>		Optional	Optional
<b>Creation / Maintenance notes</b>	If the object was created by the repository, assignment of creating application information should be straightforward. If the object was created outside the repository, it is possible this information could be supplied by the depositor. It might also be extracted from the file itself; the name of the creating application is often embedded within the file.		
<b>Usage notes</b>	This semantic unit applies both to objects created outside of the repository and subsequently ingested, and to objects created by the repository, for example, through migration Events. The <i>creatingApplication</i> container is repeatable if more than one application processed the object in turn. For example, a file could be created by Microsoft Word and later turned into a PDF using Adobe Acrobat. Details of both the Word and Acrobat applications may be recorded. However, if both files are stored in the repository, each file should be completely described as an Object entity and linked by using relationship information with a <i>relationshipType</i> “derivation.” The container may also be repeated to record the creating application before the object was ingested as well as the creating application used		

Why might a repository want to record this information?

.....

.....

.....

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

- A. 2
- B. 7
- C. 4

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

- A. Intellectual entity
- B. Representation
- C. Digital file
- D. Bitstream

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

```
<premis:formatNote></premis:formatNote>
</premis:format>
<premis:creatingApplication>
  <premis:creatingApplicationName>ImageMagick</premis:creatingApplicationName>
  <premis:creatingApplicationVersion>6.9.2-4</premis:creatingApplicationVersion>
  <premis:dateCreatedByApplication>2015-01-24</premis:dateCreatedByApplication>
  <premis:creatingApplicationExtension><!-- creatingApplication extension --></premis:creatingApplicat
```

## 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)



Message digest algorithm

Size

Format designation:  
Format name

Relationship type

Relationship sub type

Church\_front.jpeg is a JPEG  
File Interchange Format  
(version 1.01)

Church\_front.jpg is a  
derivative

Church\_front.jpg has the  
derivative source  
Chruch\_font.tiff

Church\_front.jpg has a SHA-  
256 checksum

Church\_front.jpg is 126,978  
bytes

### 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 12<sup>th</sup> 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 23rd 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>
  [redacted]
  <premis:size>.....</premis:size>
  <premis:format>
    <premis:formatDesignation>
      <premis:formatName>.....</premis:formatName>
    </premis:formatName>
    <premis:formatVersion>.....</premis:formatVersion>
    </premis:formatDesignation>
    <premis:formatRegistry>
      <premis:formatRegistryName>PRONOM</premis:formatRegistryName>
      <premis:formatRegistryKey>fmt/276</premis:formatRegistryKey>
    </premis:formatRegistry>
    <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>
    [redacted]
  </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:significantPropert
iesValue>
  </premis:significantProperties>
  <premis:significantProperties>
    <premis:significantPropertiesType>word
count</premis:significantPropertiesType>

<premis:significantPropertiesValue>.....</premis:significantPropert
iesValue>
  </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/LoveLettersToNi
na.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?

.....

.....

```
<premis:rights>
  <premis:rightsStatement>
    <premis:rightsStatementIdentifier> [3 lines]
    <premis:rightsBasis>Copyright</premis:rightsBasis>
    <premis:copyrightInformation>
      <premis:copyrightStatus>Copyrighted</premis:copyrightStatus>
      <premis:copyrightJurisdiction>GB</premis:copyrightJurisdiction>
      <premis:copyrightStatusDeterminationDate>2012-04-16</premis:copyrightStatusDeterminationDate>
      <premis:copyrightNote>Copyright held by the donor</premis:copyrightNote>
      <premis:copyrightApplicableDates> [3 lines]
    </premis:copyrightInformation>
    <premis:rightsGranted>
      <premis:act>create derivative copies</premis:act>
      <premis:restriction>rightsholder must be notified after completion of act</premis:restriction>
      <premis:termOfGrant>
        <premis:startDate>2001-01-01</premis:startDate>
        <premis:endDate>2021-01-01</premis:endDate>
      </premis:termOfGrant>
    </premis:rightsGranted>
    <premis:linkingObjectIdentifier> [3 lines]
    <premis:linkingAgentIdentifier>
      <premis:linkingAgentIdentifierType>local</premis:linkingAgentIdentifierType>
      <premis:linkingAgentIdentifierValue>John, Smith</premis:linkingAgentIdentifierValue>
      <premis:linkingAgentRole>rightsholder</premis:linkingAgentRole>
    </premis:linkingAgentIdentifier>
  </premis:rightsStatement>
</premis:rights>
```

## 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?.....

.....

```
<premis:rights>
  <premis:rightsStatement>
    <premis:rightsStatementIdentifier> [3 lines]
    <premis:rightsBasis>license</premis:rightsBasis>
    <premis:licenseInformation>
      <premis:licenseDocumentationIdentifier>
        <premis:licenseDocumentationIdentifierType>URI</premis:licenseDocumentationIdentifierType>
        <premis:licenseDocumentationIdentifierValue>http://www.apache.org/licenses/LICENSE-2.0
        </premis:licenseDocumentationIdentifierValue>
        <premis:licenseDocumentationRole>software license</premis:licenseDocumentationRole>
      </premis:licenseDocumentationIdentifier>
      <premis:licenseTerms>Licensed under the Apache License, Version 2.0 (the "License"); you may not use
        this file except in compliance with the License.</premis:licenseTerms>
      <premis:licenseNote>Apache License version 2.0</premis:licenseNote>
      <premis:licenseApplicableDates>
        <premis:startDate>2004-01-01</premis:startDate>
        <premis:endDate>OPEN</premis:endDate>
      </premis:licenseApplicableDates>
    </premis:licenseInformation>
    <premis:linkingObjectIdentifier> [3 lines]
  </premis:rightsStatement>
</premis:rights>
```

## 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*