Archives Ready To the AIPs Transmission (ARTAT) – developments in 2010

Angela Di Iorio
Metadata specialist at Università di Roma ‘La Sapienza’

PREMIS Implementation Fair
Austrian National Library
September 22nd, 2010

Reminding the iPRES2010 Presentation

The exported repositories’ AIPs including a PML will be received by selected repositories and ingested into their archival systems.

Hopefully, because of the common PREMIS knowledge base, the receiving repositories will be able…

… to locate information objects and data objects contained in the AIPs transmitted by the originating repositories.
Reminding the iPRES2010 Presentation

Ridley Scott, *Alien*, 1979

If it can be scary to preserve AIPs coming from different repositories.

Even more scary can be to preserve AIPs encoded in different Metadata Standards and... ...

... maybe receiving repositories' technologists can sweat blood in interpreting AIPs' semantics and structures.


Hopefully this project will contribute to avoid this “bloody process”, to understand better alien AIPs coming from other repositories and differently characterized

.. and hopefully will help to look at them as cheering hosts.

How a metadata specialist can become a metadata spatialist
Inquiry phase results and considerations

The ICCU’s aggregator repository named MAGTECA, the grounding archive of the Italian national Digital Library Portal and Culture-Tourist Network. More than 2.400.000 of digitalized images for 29.000 documents.

MD is a project undertaken by Fondazione Raiti and National Library of Florence. The repository contains and preserves the doctoral theses harvested from the Italian universitaries institutional repositories.

BSR is the digital repository of the Library & Archive of the British School at Rome contains digitazed collections of historical photographs, prints and maps. It comprehends around 40.000 images with more than 13.900 documents.

<table>
<thead>
<tr>
<th>Institution / Project</th>
<th>Metadata type</th>
<th>XML Scheme name</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICCU</td>
<td>Container</td>
<td>MAG</td>
<td>1.0.2.01</td>
</tr>
<tr>
<td></td>
<td>Descriptive</td>
<td>DC</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>Technical</td>
<td>MIX</td>
<td>6.1.1</td>
</tr>
<tr>
<td>MD</td>
<td>Container</td>
<td>MPEG21-DDDL</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Descriptive</td>
<td>DC</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>Technical</td>
<td>Jhove</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Technical</td>
<td>MIX</td>
<td>0.2</td>
</tr>
<tr>
<td>BSR</td>
<td>Container</td>
<td>METS</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Descriptive</td>
<td>MODS</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Descriptive</td>
<td>DC</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>Technical</td>
<td>MIX</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Legend

ICCU = Union Catalogue of Italian Libraries and Bibliographic Information [www.internetculturale.it]
MD = Magazzini Digitali [www.rinascimento-digitale.it/index.php?SEZ=28)
BSR = British School at Rome Digital Collections [digitalcollections.bsrome.it]

The digital repository of the Library & Archive of the British School at Rome contains digitazed collections of historical photographs, prints and maps. It comprehends around 40.000 images with more than 13.900 documents.

Legend Institution / Project
ICCU = Union Catalogue of Italian Libraries and Bibliographic Information [www.internetculturale.it]
MD = Magazzini Digitali [www.rinascimento-digitale.it/index.php?SEZ=28)
BSR = British School at Rome Digital Collections [digitalcollections.bsrome.it]

Legend XML Schema name
MAG = Metadati amministrativi e gestionali [www.iccu.sbn.it/genera.jsp?id=267]
METS = Metadata Enclosing Transmission Standard [www.loc.gov/standards/mets]
DC simple = Dublin Core Metadata Element Set [dublincore.org/documents/dces/]
Jhove = JSTOR/Harvard Object Validation Environment [hul.harvard.edu/jhove/]
MIX = NISO Technical Metadata for Digital Still Images [www.loc.gov/standards/mix]
MODS = Metadata Object Description Schema [www.loc.gov/standards/mods]

IPRES 2010 Austrian National Library Sept. 22nd, 2010 Archives Ready To the AIPs Transmission (ARTAT) - developments in 2010 [ipress2010.austria.org]

Metadata Containers >> structure and semantics

<table>
<thead>
<tr>
<th>Structure</th>
<th>MAG</th>
<th>METS</th>
<th>MPEG21_DIDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>root</td>
<td>mag:metadata</td>
<td>mets:mets</td>
<td>did:DIDL</td>
</tr>
<tr>
<td>descriptive wrapper</td>
<td>mag:bib</td>
<td>mets.dm/desc/mets.mxf/mets.xmlData</td>
<td>did:Item/did:Descriptor</td>
</tr>
<tr>
<td>descriptive reference</td>
<td>mets.dm/desc/mets.mxf/mets.xmlData</td>
<td>did:Item/did:Component/did:Descriptor/did:Statement</td>
<td></td>
</tr>
<tr>
<td>descriptive sec prefix</td>
<td>mets:ref</td>
<td>did:Item/did:Component/did:Descriptor/did:Statement</td>
<td></td>
</tr>
<tr>
<td>technical container</td>
<td>mag:group[video</td>
<td>audio</td>
<td>doc</td>
</tr>
<tr>
<td>technical sec prefix</td>
<td>mets:techMD</td>
<td>mets:rel</td>
<td>did:Item/did:Component/did:Descriptor</td>
</tr>
<tr>
<td>objects locations</td>
<td>mag:group[video</td>
<td>audio</td>
<td>doc</td>
</tr>
<tr>
<td>structural section</td>
<td>mag:xml</td>
<td>mets:structMap</td>
<td>did:Item/did:Component</td>
</tr>
<tr>
<td>agents involved</td>
<td>mets:agent</td>
<td>mets:agent</td>
<td>-</td>
</tr>
<tr>
<td>events</td>
<td>mets:agent</td>
<td>mets:agent</td>
<td>-</td>
</tr>
<tr>
<td>minimal obligation</td>
<td>mag:gen,mag:bib</td>
<td>mets:structMap</td>
<td>did:Item</td>
</tr>
</tbody>
</table>
The **PML** is composed of two parts:

The **PML core** is the part which essentially translates the container’s relevant metadata into PREMIS semantic units. The translation consists of a mapping from the original administrative, technical, provenance, rights and structural information into the PREMIS framework.

The **PML redundant** part simply describes the content objects in PREMIS terms replicating information like objectIdentifier, compositionLevel, fixity, size, format, originalName, and storage from the object’s related metadata.

---

**ARTAT - Preservation Metadata Layer >> Structure**

**Transmission Package**

---

**ARTAT - Preservation Metadata Layer >> Transmission scenario**

**Originating repository**

Original AIP objects

Metadata container

Metadata objects

Content objects

**Receiving repository**

Received AIP objects

Metadata container

Metadata objects

Content objects
creatingApplication

ARTAT and TIPR >> LESSONS LEARNED

TIPR: Towards Interoperable Preservation Repositories
(http://wiki.fcla.edu/;B00/TIPR)
INTERCHANGE YOU CAN BELIEVE IN!

ART

Archives Ready to the AIPs Transmission – ARTAT
(http://www.riuscimento-digitale.it/artat.phtml)

TIPR found information pertaining to the exchange package (history, description, and high level rights) must at this time be recorded at the intellectual entity level, because the highest level of object describable in PREMIS is a representation object.

The PML core gathers events and rights at the exchange package level.

both TIPR and ARTAT found problems with the unambiguous identification of entities

details about RXP composition by the source repository
how a packages will be transferred from source to target repository
actions to be performed
rights and permissions
archiving and preservation treatment
financial and legal aspects of agreement

relationships’ information of PML core
devolving partnership’s agreement and transmission conditions applicable to the massive transmission of AIPs;
providing a common controlled vocabulary about actions that must be selected at PML production time and associated with agents;
rights framework system
partnership’s agreement level
should be provided in ARTAT partnership agreement

Archives Ready To the AIPs Transmission
(ARTAT) – developments in 2010 (http://www.rinascimento-digitale.it/artat.phtml)
Identification system

**PMLCIT MD** [Local object identifier]

Core or Redundant Agent ID value

ISO 3166 code

Examples
PMLCIT-MD-cb8e12ad-5591-4220-a779-6b5bdf871d2e.xml
PMLCIT-itri-CB0007_MSM_0000024.xml
PMLCIT-itrobs-0000076.xml

Agent’s names system

<table>
<thead>
<tr>
<th>agentIDType</th>
<th>agentID/Value</th>
<th>agentName</th>
<th>agentType</th>
<th>agentNote</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARTAT narrows</td>
<td>IT-ARTAT-FRD</td>
<td></td>
<td></td>
<td>organization</td>
</tr>
<tr>
<td>ARTAT narrows</td>
<td>IT-ARTAT-</td>
<td>Preservation Metadata</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PMLCORE-CR</td>
<td>Core Creator</td>
<td>software</td>
<td></td>
</tr>
<tr>
<td>ARTAT narrows</td>
<td>IT-ARTAT-</td>
<td>Preservation Metadata</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PMLREDUCR-CR</td>
<td>Redundant Creator</td>
<td>software</td>
<td></td>
</tr>
<tr>
<td>ARTAT narrows</td>
<td>FRD</td>
<td>Fondazione Rinascimento Digitale</td>
<td>organization</td>
<td><a href="http://www.rinascimento-digitale.it/cont_1236">http://www.rinascimento-digitale.it/cont_1236</a>...</td>
</tr>
<tr>
<td>MARC organization codes</td>
<td>itn</td>
<td>Istituto centrale per il catalogo unico della bibl.</td>
<td>organization</td>
<td></td>
</tr>
<tr>
<td>SBN</td>
<td>BNCF</td>
<td>Biblioteca Nazionale Centrale di Firenze</td>
<td>organization</td>
<td><a href="http://www.bncf.firenze.sbn.it">http://www.bncf.firenze.sbn.it</a></td>
</tr>
<tr>
<td>MARC organization codes</td>
<td>itrobs</td>
<td>British School at Rome</td>
<td>organization</td>
<td></td>
</tr>
<tr>
<td>ARTAT narrows</td>
<td>MD</td>
<td>Magazzini Digitali</td>
<td>project</td>
<td><a href="http://www.depositoriologi.it">http://www.depositoriologi.it</a></td>
</tr>
</tbody>
</table>
What are the significant properties and how do we convey them?

What are relationships between content objects and metadata objects?

Are the significant properties, relationships?

Are relationships already expressed in PML by the linking identifiers?

---

Applying the INSPECT workflow to Metadata Container Objects

http://www.significantproperties.org.uk

defined by the INSPECT project significant properties are
“The characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what they purport to record”

Dissecting the MCO

Identifier: MCO identifier [value and type]
Title: MODS
Description: Descriptive section for the intellectual entity
Function/class: metadata content
Function/subclass: descriptor
PreservationLevel:...
Identify purpose of technical properties of Metadata Container Objects

**Content:** is XML text;

**Context:** is the environment, where the participants manage metadata and its exchange;

**Rendering:** is considered the recreation of an AIP in a recipient repository by means of a translated MCO, where metadata values and relationships among metadata objects and content objects are replicated in a new container;

**Structure:** is metadata which contains information about intra-relationships and inter-relationships;

**Behaviour:** is how the information object is connected to other metadata or content objects (i.e. the mdRef for external metadata files used in METS).

---

Limiting the analysis to the transmission context, where a source and a recipient have to exchange AIPs between their heterogeneous archival systems, the stakeholders involved in transmission of AIPs are repositories’ systems that have to be able to make an interpretation of the alien AIPs and to ingest them as their own AIPs.

This particular “user” with a well defined objective may wish to perform the following main activities:
- selecting information relevant to preservation,
- interpreting technically the selected information, and
- understanding the relational structure conveyed.
Metadata Containers >> Significant Properties

<table>
<thead>
<tr>
<th>Behaviours</th>
<th>Functions</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>understand information architecture (technical, structural)</td>
<td>determine semantics used</td>
<td>descriptive section</td>
</tr>
<tr>
<td>understand relational structure of content objects</td>
<td>identify intellectual characteristics</td>
<td>technical section</td>
</tr>
<tr>
<td>understand relational structure of content objects</td>
<td>determine representations information</td>
<td>structural section</td>
</tr>
<tr>
<td>identify components (metadata and content objects)</td>
<td>identify intellectual information</td>
<td>external reference metadata</td>
</tr>
<tr>
<td>identify intellectual content</td>
<td>understand content's structural information</td>
<td>embedded content</td>
</tr>
<tr>
<td>identify rights to perform actions</td>
<td>determine provenance of content objects</td>
<td>provenance section</td>
</tr>
<tr>
<td>identify events chain</td>
<td>determine objects' location</td>
<td>rights and access conditions</td>
</tr>
</tbody>
</table>

Functions may be used as a basis for tailoring future manifestations of the Information Object to the need of the stakeholder.

---

IPA 2010 Austrian National Library

Archives Ready To the AIPs Transmission (ARTAT) – developments in 2010
**Metadata Containers >> Drafting Relationships’ model**

- relationSubType
  - external metadata/content
  - internal metadata/content
  - metadata wrapper

---

**INSPECT Significant Properties Data Dictionary**

Figure 1, based on the PREMIS Data Model\(^2\), indicates the relationship between the four entities.

![Diagram of relationships between object, agent, component, and property](insert_diagram)

- object
- agent
- component
- property
- relationships

Figure 1. INSPECT data model

The four entities may be linked to other entities using a defined set of rules:

1. An Object may be associated with one or more Components
2. A Component may be associated with one or more Properties or Components.
Drafting relationships’ model: metadata embedded

Agent=sender  
Agent=recipient  
Event=AIP transport package building  
Object=MCO

Agent=PML builder software  
Event=PML redundant building  
Object=OBJ

**relationship**

<table>
<thead>
<tr>
<th>relationshipType</th>
<th>relationshipSubType</th>
<th>direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>is referred by</td>
<td>xlink</td>
<td>from OBJ to MCO</td>
</tr>
<tr>
<td>is technically described by</td>
<td>xlink</td>
<td>from OBJ to OBJ</td>
</tr>
<tr>
<td>technically describes</td>
<td>xlink</td>
<td>from MCO to OBJ</td>
</tr>
<tr>
<td>is technically described by</td>
<td>embedded MIX</td>
<td>from OBJ to MCO</td>
</tr>
</tbody>
</table>

Drafting relationships’ model: metadata referred

Agent=sender  
Agent=recipient  
Event=AIP transport package building  
Object=MCO

Agent=PML builder software  
Event=PML redundant building  
Object=OBJ

**relationship**

<table>
<thead>
<tr>
<th>relationshipType</th>
<th>relationshipSubType</th>
<th>direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>is referred by</td>
<td>xlink</td>
<td>from OBJ to MCO</td>
</tr>
<tr>
<td>is technically described by</td>
<td>external MIX</td>
<td>from OBJ to OBJ</td>
</tr>
<tr>
<td>technically describes</td>
<td>xlink</td>
<td>from MCO to OBJ</td>
</tr>
<tr>
<td>is technically described by</td>
<td>xlink</td>
<td>from OBJ to MCO</td>
</tr>
</tbody>
</table>

The relation is the same but the object referred is different: the first is a content object and the second is a metadata object.
First prototype of the Preservation Metadata Layer

Two examples of METS files encoded in PREMIS semantics as PreservationMetadataLayer you can see:
- the original METS files as AIP
- the XML PreservationMetadataLayer
- Human readable version of the XML Preservation Metadata Layer

PML application builder website
www.demokrito.org/artat

Install
Encoding
Show first two prototypes:
original AIP – 1
original AIP – 2
PREMIS – 1
PREMIS – 2
XSLT of PREMIS – 1
XSLT of PREMIS – 2
Partners Info space

Thanking

Thanks for your kind attention

and Questions Time.

contacts information

Angela Di Iorio
Fondazione Rinascimento Digitale
Metadata specialist
angela_diiorio(at)gmail(dot)com

Maurizio Lunghi
Fondazione Rinascimento Digitale
Scientific Director
lunghi(at)rinascimento-digitale(dot)it