FONDAZIONE RINASCIMENTO DIGITALE

iPres 02010

7th International Conference on
Preservation of Digital Objects (iPRES2010)
September 19 - 24, 2010, Vienna, Austria

Foundation promoted by Ente Cassa di Risparmio of Florence

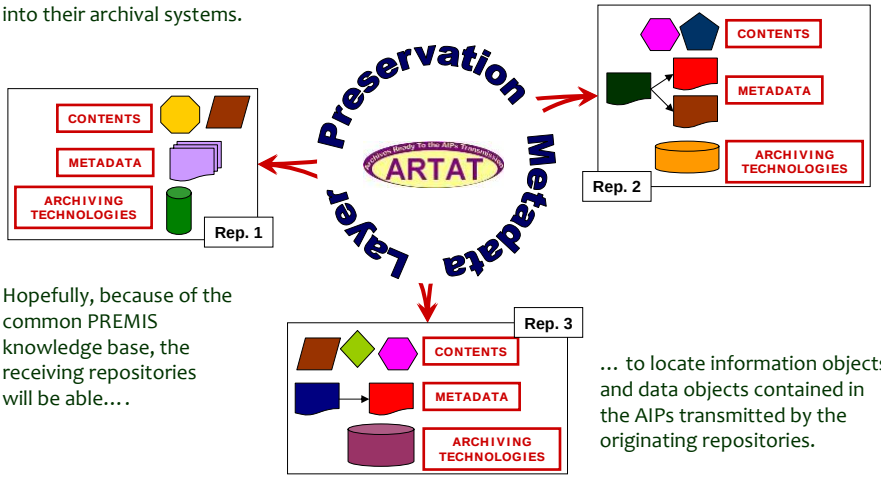
Archives Ready To the AIPs Transmission (ARTAT) – developments in 2010

Angela Di Iorio
Metadata specialist at Università di Roma "La Sapienza"

PREMIS Implementation Fair
Austian 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.

Reminding the iPRES2010 Presentation

Steven Spielberg, *E.T.: The Extra-Terrestrial*, 1982

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 Cultural -Tourist Network. More than 2.400.000 of digitalized images for 29.000 documents

MD is a project undertaken by Fondazione Rinascimento Digitale and National Library of Florence.
The repository contains and preserve the doctoral thesis harvested from the italian universities institutional repositories.

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 then 13.900 documents.

Institution/ Project	Metadata type	XML Schema name	Version
ICCU	Container	MAG	1.0-2.01
	Descriptive	DC simple	1.1
	Technical	MIX	0.1 draft
MD	Container	MPEG21-DIDL	-
	Descriptive	DC simple	1.1
	Technical	Jhove	1.5
	Technical	MIX	0.2
BSR	Container	METS	1.9
	Descriptive	MODS	3.3
	Descriptive	DC simple	1.1
	Technical	MIX	2.0

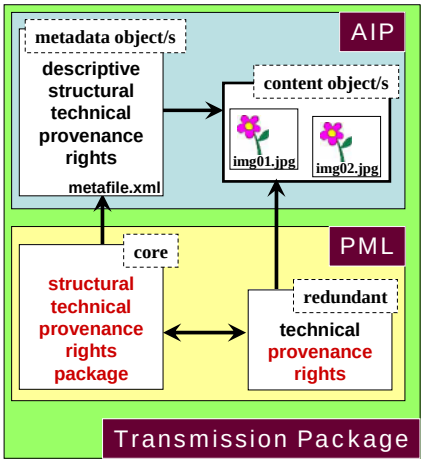
Legend Institution / Project
ICCU = Union Catalogue of Italian Libraries and Bibliographic Information www.internetculturale.it
MD = Magazzini Digitali www.rinascimento-digitale.it/index.php?SEZ=28)
BSR = British School at Rome Digital Collections digitalcollections.bsrome.it

Legend XML Schema name
MAG = Metadati amministrativi e gestionali **www.iccu.sbn.it/genera.jsp?id=267**
METS = Metadata Encosing Transmission Standard **www.loc.gov/standards/mets**
MPEG21-DIDL= MPEG21 Digital Item Declaration Language **www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm**
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/standard/mods**

Metadata Containers >> structure and semantics

Structure	MAG	METS	MPEG21_DIDL
root	mag:metadigit	mets:mets	didl:DIDL
descriptive wrapper	mag:bib	mets:dmdSec/ mets:mdWrap/ mets:xmlData	didl:Item/ didl:Descriptor
			didl:Item/ didl:Component/ didl:Descriptor/ didl:Statement
descriptive reference		mets:dmdSec/ mets:mdRef	didl:Item/ didl:Component/ didl:Descriptor/ didl:Statement
descriptive sec prefix	dc	mods	dc
technical container	mag:[img|altimg|audio| video|doc|ocr|dis] mag:[img_group| audio_group| video_group]	mets:techMD	didl:Item/ didl:Component/ didl:Descriptor
technical sec prefix	niso	mix	jhove; mix
objects locations	mag:[img|altimg|audio| video|doc|ocr|dis]/mag:file	mets:fileSec/ mets:fileGrp/mets:file/ mets:FLocat	didl:Item/ didl:Component/ didl:Resource
structural section	mag_stru	mets:structMap/ mets:div	didl:Item/ didl:Component/
agents involved	agency	mets:agent	-
events	stprog		
minimal obligation	mag:gen; mag:bib	mets:structMap	didl:Item

ARTAT - Preservation Metadata Layer >> Structure

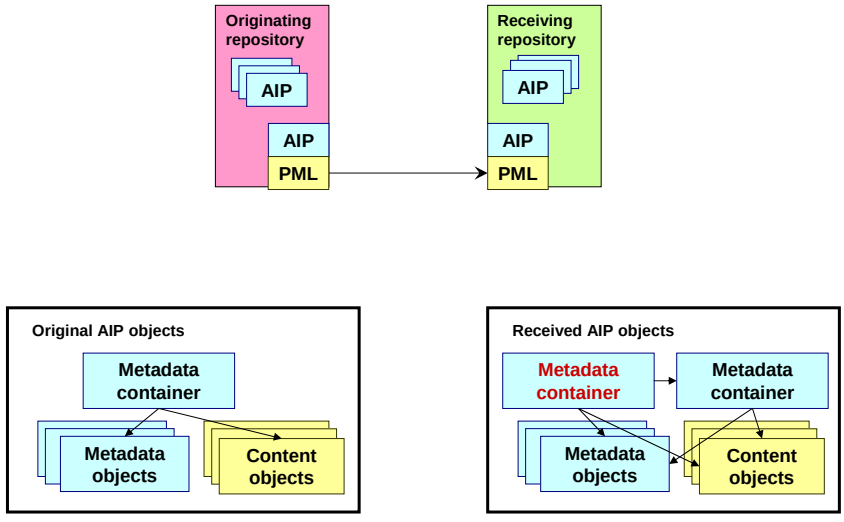


The **PML** is composed of two parts:

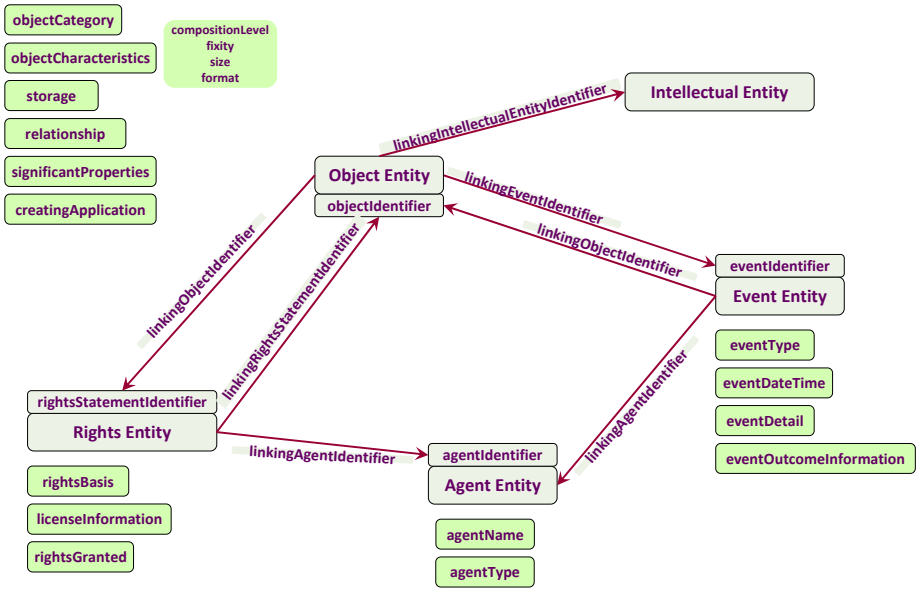
The **PML core** is the part which essentially translates the container's relevant metadata into PREMIS semantic units. The translation consists of a mapping from the original administrative, technical, provenance, rights and structural information into the PREMIS framework.

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 >> Transmission scenario



PML >> DATA MODEL



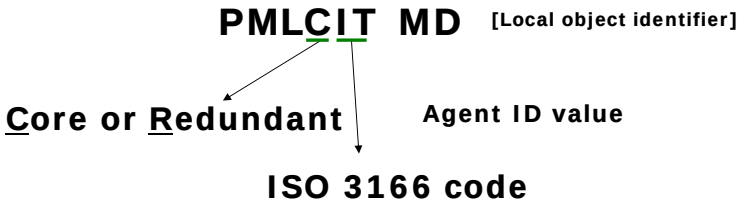
ARTAT and TIPR >> LESSONS LEARNED

TIPR: Towards Interoperable Preservation Repositories
(http://wiki.fcla.edu:8000/TIPR)
INTERCHANGE YOU CAN BELIEVE IN!

Archives Ready To the AIPs Transmission - ARTAT
(http://www.rinascimento-digitale.it/artat.phtml)

TIPR Towards Interoperable Preservation Repositories	Archives Ready To the AIPs Transmission ARTAT
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	relationships' information of PML core
how a packages will be transferred from source to target repository	devising partnership's agreement and transmission conditions applicable to the massive transmission of AIPs;
actions to be performed	providing a common controlled vocabulary about actions that must be selected at PML production time and associated with agents;
rights and permissions	rights framework system
archiving and preservation treatment	partnership's agreement level
financial and legal aspects of agreement	should be provided in ARTAT partnership agreement

Identification system



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

Agent's names system

agentIDType	agentIDValue	agentName	agentType	agentNote
ARTAT namesys	IT-ARTAT-FRD		organization	
ARTAT namesys	IT-ARTAT-PMLCORE-CR	Preservation Metadata Layer Core Creator	software	
ARTAT namesys	IT-ARTAT-PMLREDU-CR	Preservation Metadata Layer Redundant Creator	software	
ARTAT namesys	FRD	Fondazione Rinascimento Digitale	organization	http://www.rinascimento-digitale.it/cont__856_1226...
MARC organization codes	itri	Istituto centrale per il catalogo unico delle bibl...	organization	
SBN	BNCF	Biblioteca Nazionale Centrale di Firenze	organization	http://www.bncf.firenze.sbn.it
MARC organization codes	itrobs	British School at Rome	organization	
ARTAT namesys	MD	Magazzini Digitali	project	http://www.depositolegale.it

Metadata Containers >> Significant Properties

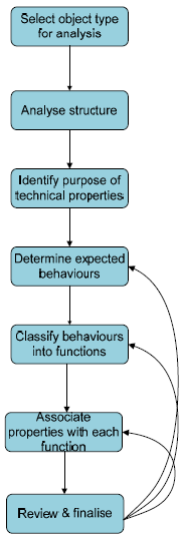
What are the significant properties and how do we convey them?

What are relationships between content objects and metadata objects?

Are the significant properties, relationships?

Are relationships already expressed in PML by the linkingidentifiers?

Metadata Containers >> Significant Properties



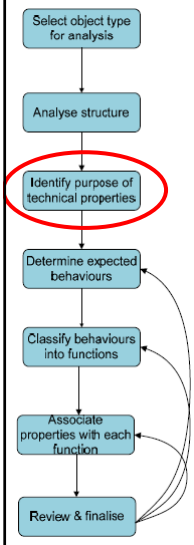
Applying the INSPECT workflow to Metadata Container Objects
http://www.significantproperties.org.uk

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 identiffier [value and type]
Title: MODS
Description: Descriptive section for the intellectual entity
Function/class: metadata content
Function/subclass: descriptor
PreservationLevel:…

Metadata Containers >> Significant Properties



Identify purpose of technical properties of Metadata Container Objects

Content: is XML text;

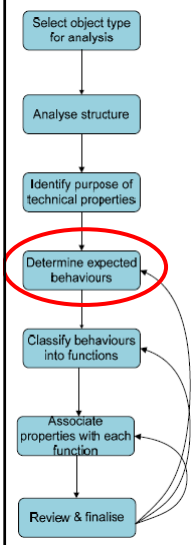
Context: is the environment, where the participants manage metadata and its exchange;

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).

Metadata Containers >> Significant Properties



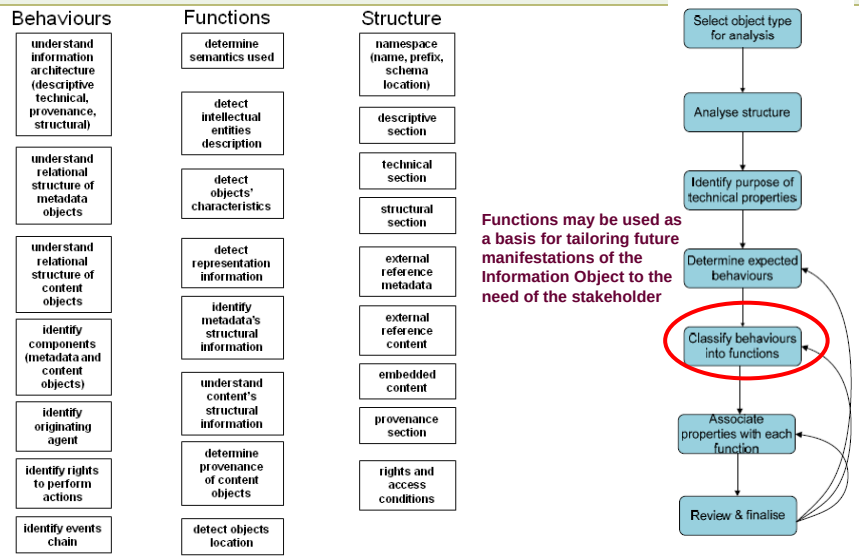
Determine expected behaviours

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.

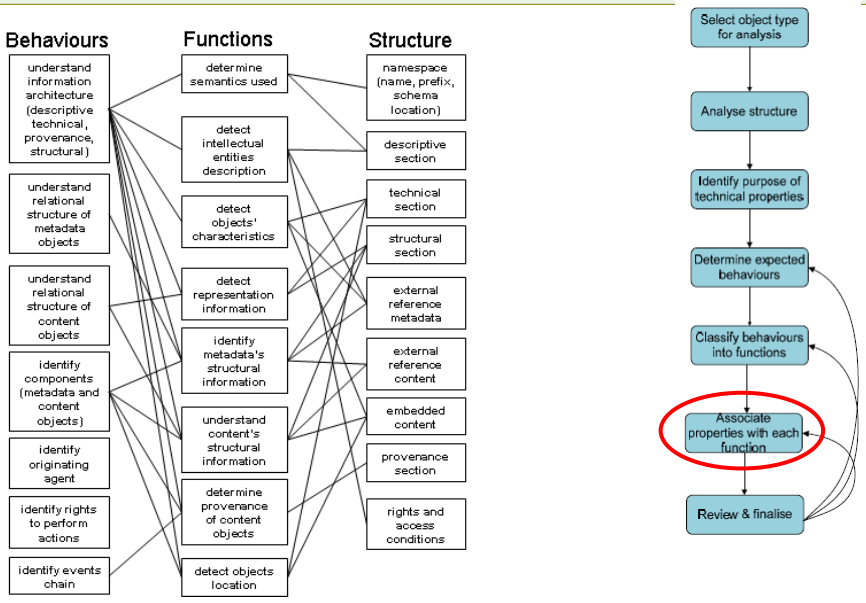
This particular "user" with a well defined objective may wish to perform the following main activities:
- selecting information relevant to preservation,
- interpreting technically the selected information, and
- understanding the relational structure conveyed.

iPRES 2010 | Austrian National Library Sept. 22nd, 2010 | Archives Ready To the AIPs Transmission (ARTAT) – developments in 2010 Angela Di Iorio

Metadata Containers >> Significant Properties



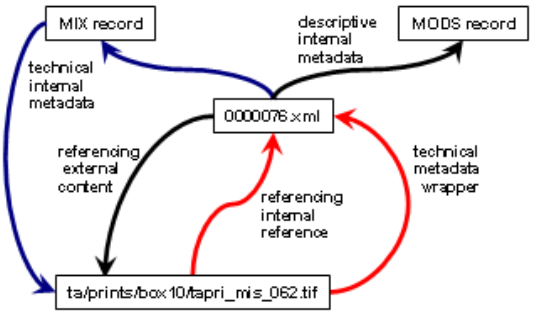
Metadata Containers >> Significant Properties



iPRES 2010 Austrian National Library Sept. 22nd, 2010 Archives Ready To the AIPs Transmission (ARTAT) – developments in 2010 Angela Di Iorio

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[20] indicates the relationship between the four entities.

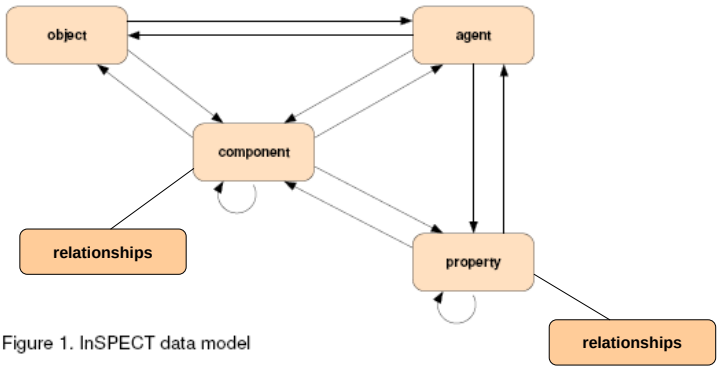


Figure 1. InSPECT data model

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

type=is referred by; subtype=xlink
relatedObjectIndentification=[MCO identifier]

relationship

type=is technically described by; subtype=embedded MIX
relatedObjectIndentification=[MCO identifier]

relationships among objects

relationshipType	relationshipSubType	direction
is referred by	xlink	from OBJ to MCO
is technically described by	xlink	from OBJ to MOBJ
technically describes	xlink	from MOBJ to OBJ
is technically described by	embedded MIX	from OBJ to MCOD

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

type=is referred by; subtype=xlink
relatedObjectIndentification=[MCO identifier]

relationship

type=is technically described by; subtype=external MIX
relatedObjectIndentification=[OBJ identifier]

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

relationship

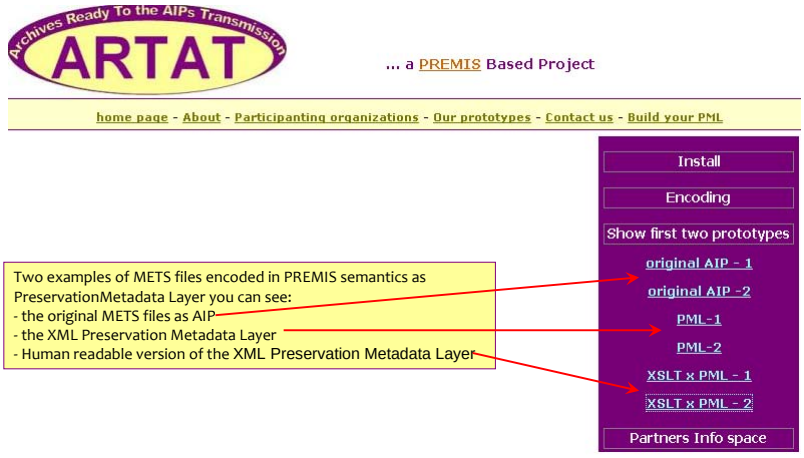
type=technically describes; subtype=xlink
relatedObjectIndentification=[OBJ identifier]

relationship

type=is referred by; subtype=xlink
relatedObjectIndentification=[MCO identifier]

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



PML application builder website

www.demokrito.org/artat

username: ipres2010
password: 22premisfair

Thanking

Thanks for your kind attention

....

and Questions Time....

contacts information

Angela Di Iorio
Fondazione Rinascimento Digitale
Metadata specialist
angeladiiorio[at]gmail[dot]com

Maurizio Lunghi
Fondazione Rinascimento Digitale
Scientific Director
lunghi[at]rinascimento-digitale[dot]it