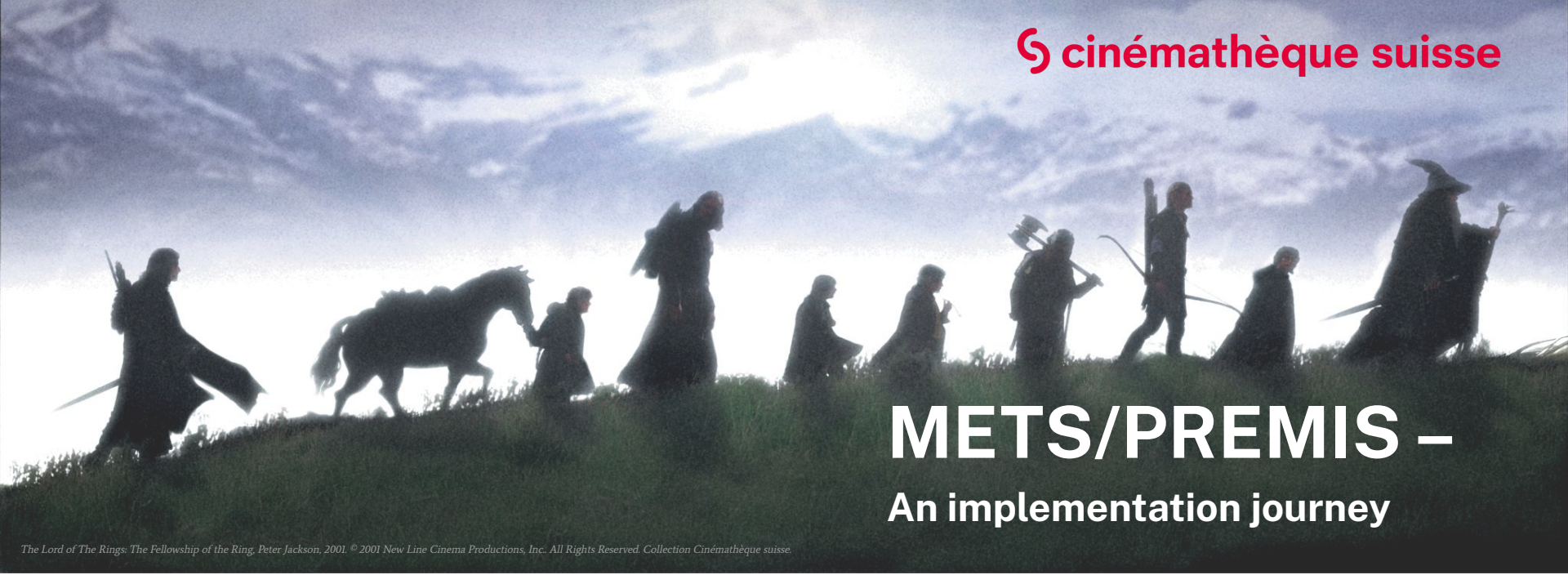


METS/PREMIS –

An implementation journey

The Lord of The Rings: The Fellowship of the Ring, Peter Jackson, 2001. © 2001 New Line Cinema Productions, Inc.. All Rights Reserved. Collection Cinémathèque suisse.

Rebecca ROCHAT, Software platform unit (speaker)
Robin FRANÇOIS, Software platform unit (speaker)
Maryline MONNERAT, Film department (contributor)

FIAF Commissions'
Thursday Online Workshops
19th October 2023

Agenda

1. In the beginning...
2. Why PREMIS ?
3. Everything is metadata
4. Ways to proceed
5. Workflows examples
6. Lasco, our digital archiving solution
7. Conclusions


MISSION CONTROL CENTER
RETURN TO ASSEMBLY
STS-115 / 12A
STS-115 ORBIT 2
FLIGHT CONTROL TEAM


MILESTONES IN
MOTION PICTURE HISTORY
1895 First Motion Picture
1927 The Introduction of Sound
1935 Color Film
1955 Cinemascope Wide Screen
1990 Digital Audio
1999 DLP Cinema™ Premieres
Digital projection replaces traditional
film projection for the first time.
DLP
CINEMA
STAR WARS
EPISODE I
THE PHANTOM MENACE

Why PREMIS ?

“

Context is everything - John Sheridan, National Archives UK

The paradox of data digital archiving

- Provide contextualised **time capsules**
- Provide **data access**

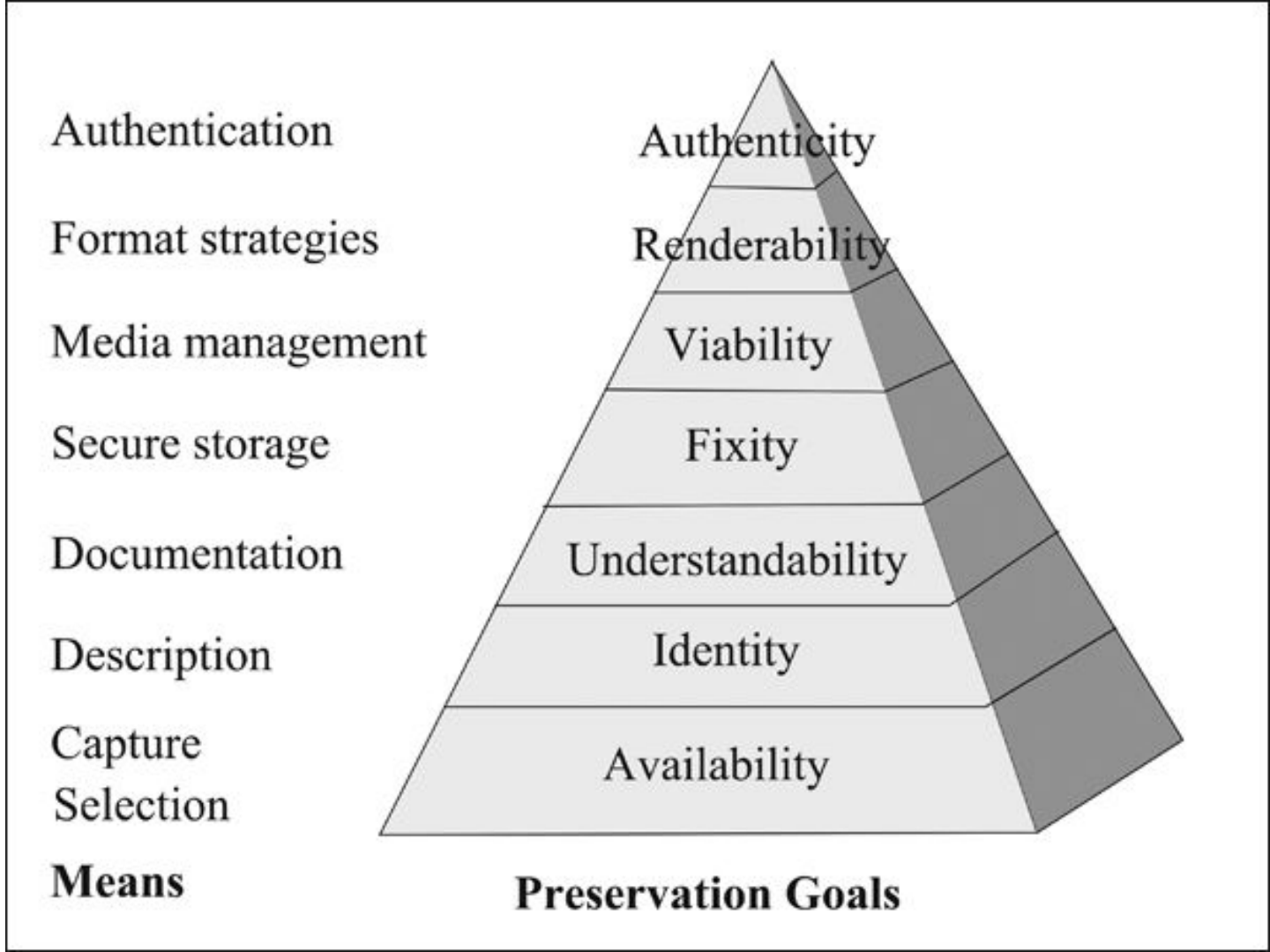
PREMIS → aims to structure preservation metadata

- Concept, agnostic of collections and formats
- Widely use, first introduced in 2005
- CERN, Library of Congress…

Reynold Brown/MGM

Lucasfilm Ltd./Paramount

Preservation Pyramid



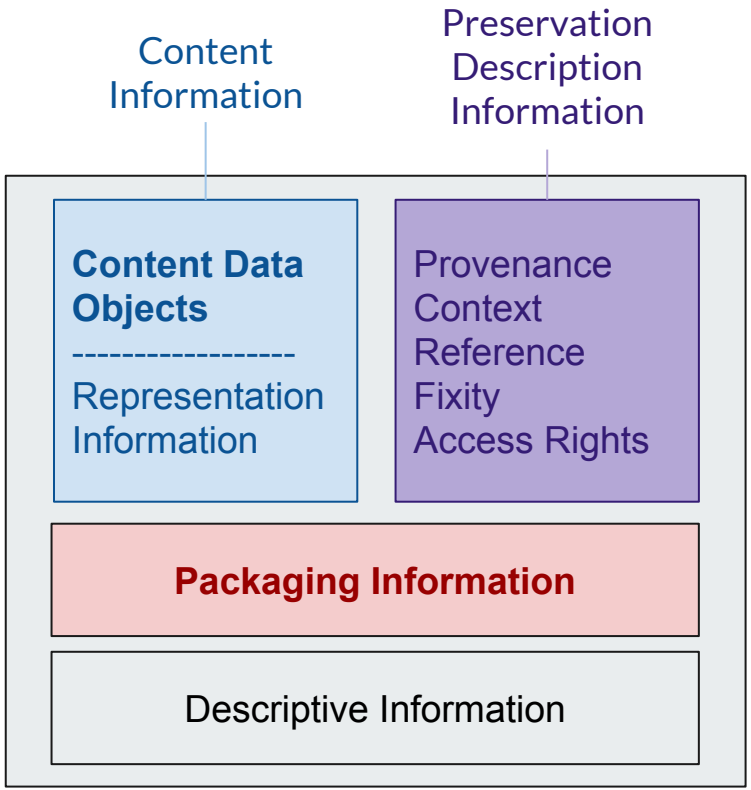
Source: Caplan, 2005

Information package (OAIS)

Free-standing

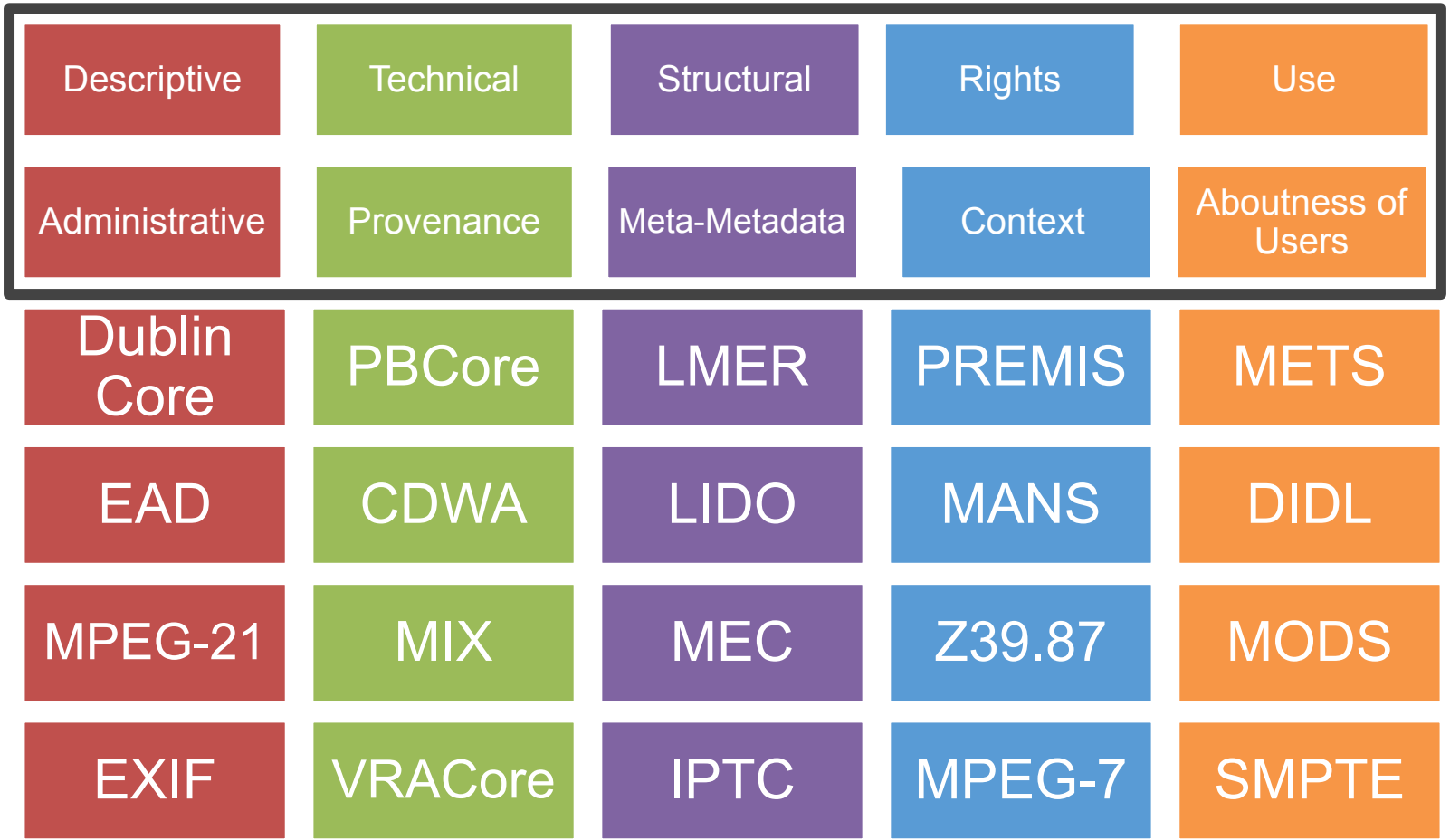
Self-contained

Self-sufficient




WHAT IF I TOLD YOU
EVERYTHING IS METADATA

Metadata of all horizons



Source: AMIA Preservation Metadata PREMIS, 14 July 2021

Want to preserve a digital object (now and in future)?

What questions do I need to ask?

What is it?

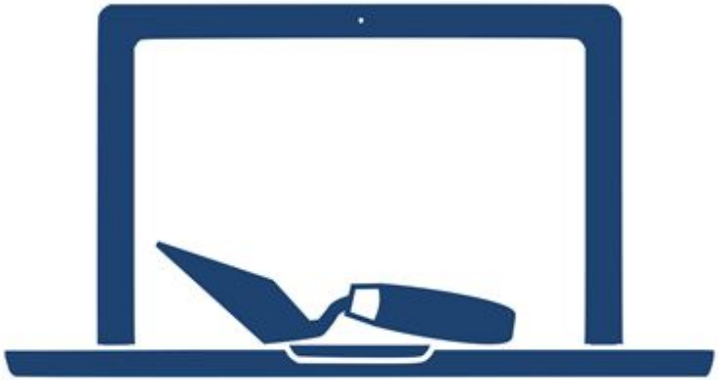
What has been done to it?

What can I do with it?

How was it created?

Package must be self sufficient

It is archaeology in reverse



R

Semantic units

for the object

Which object is it?

TIF

LZW

What are the technical informations?

ark:/21891/mwm0tz0jvg45wsj

What is the preservation strategy for this object?

TIF XML

ALTO OCR

What kind of object?

Which characteristics do I want to preserve?

What software or hardware are needed to handle the object?

What rights conditions?

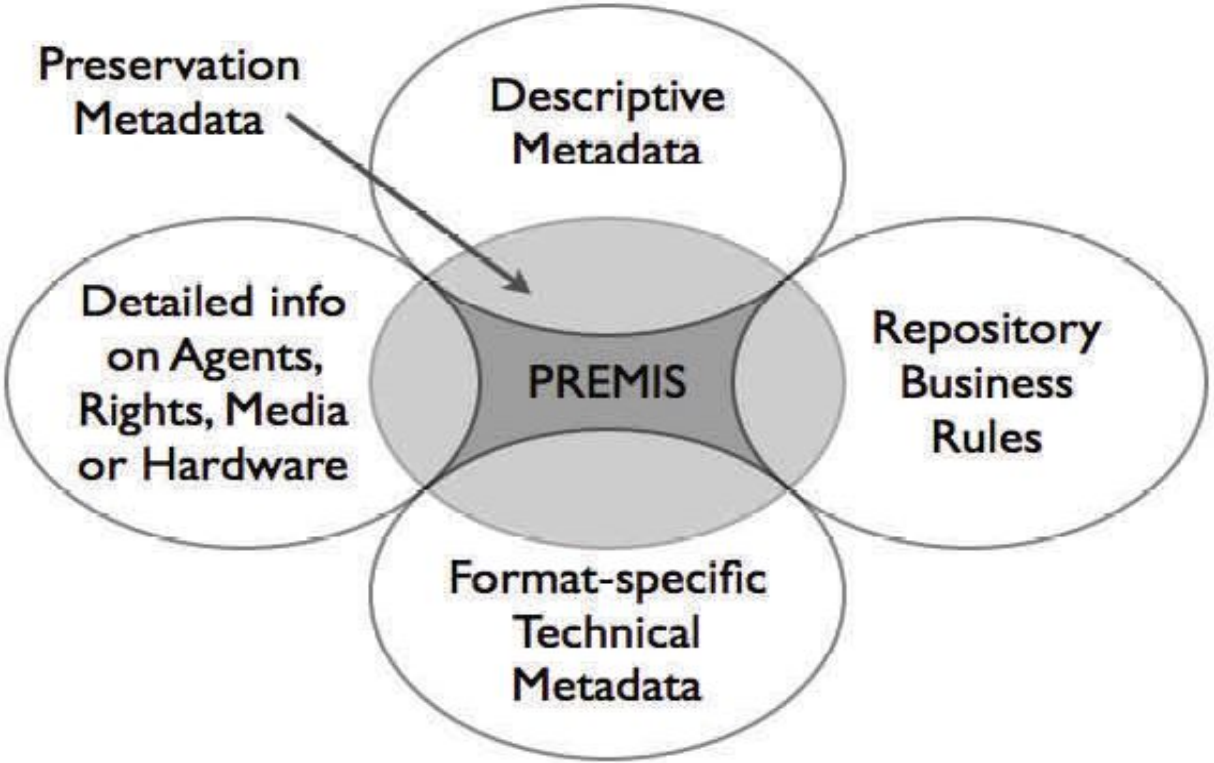
PREMIS IS

- A data dictionary
- Technically neutral
- Preservation metadata
- Metadata structuration
- Metadata aggregation

PREMIS IS NOT

- All-needed metadata
- Repository management
- Ready-made solution
- Only for archives/libraries
- Only in XML form

PREMIS Data Model



Source: Caplan, 2009

PREMIS Data Model

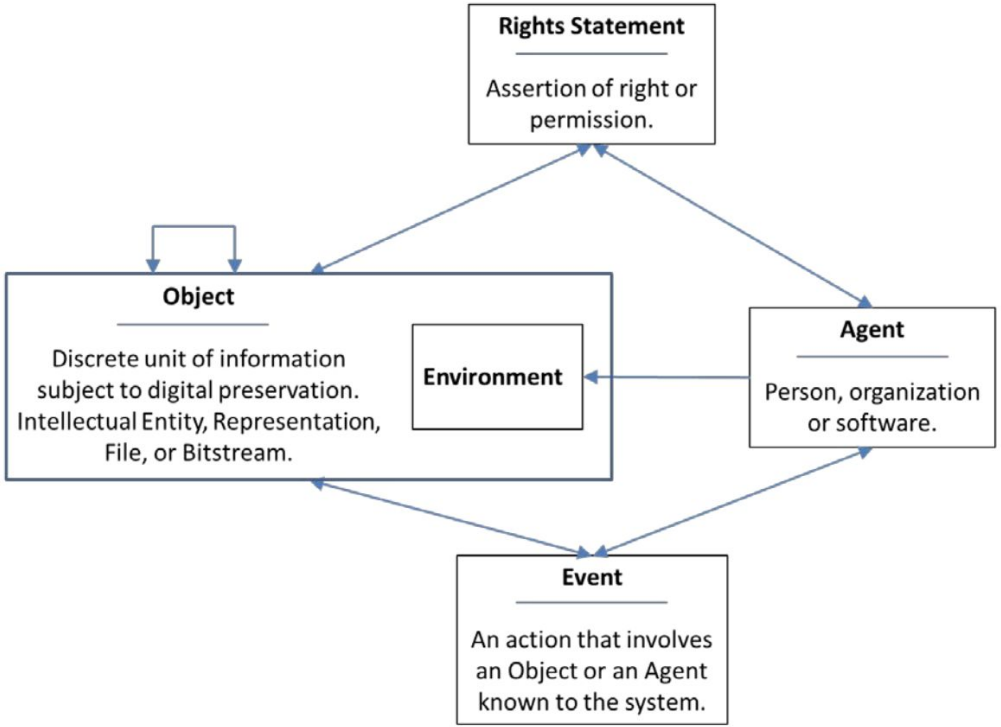
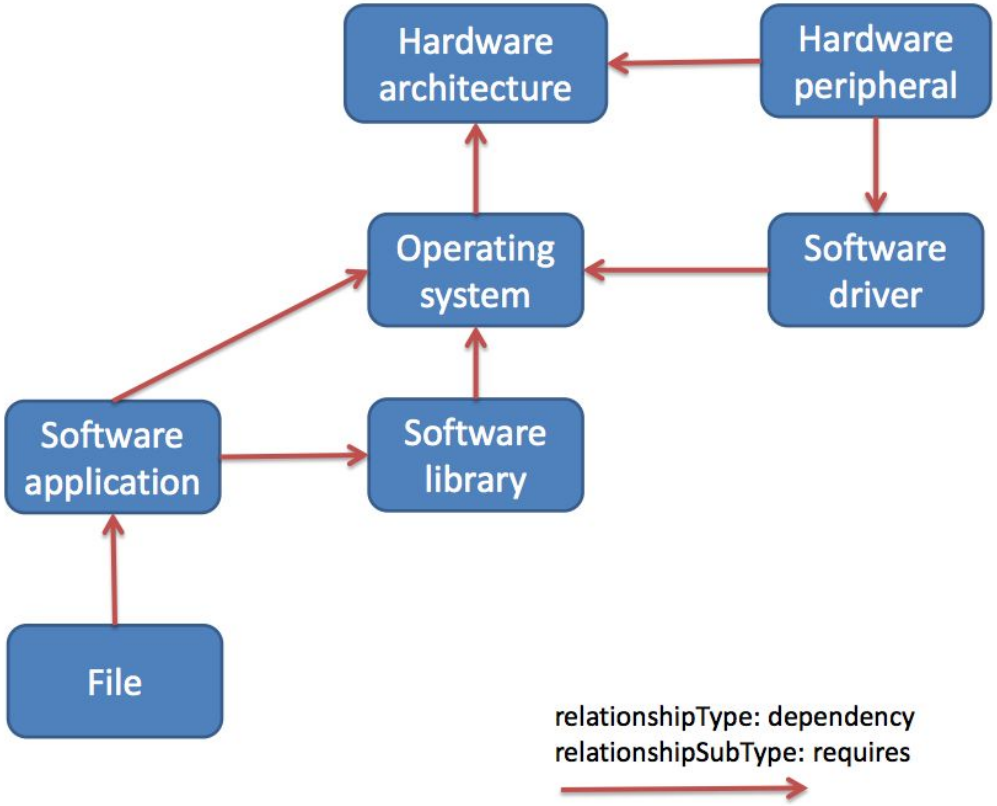


Figure 1: The PREMIS Data Model

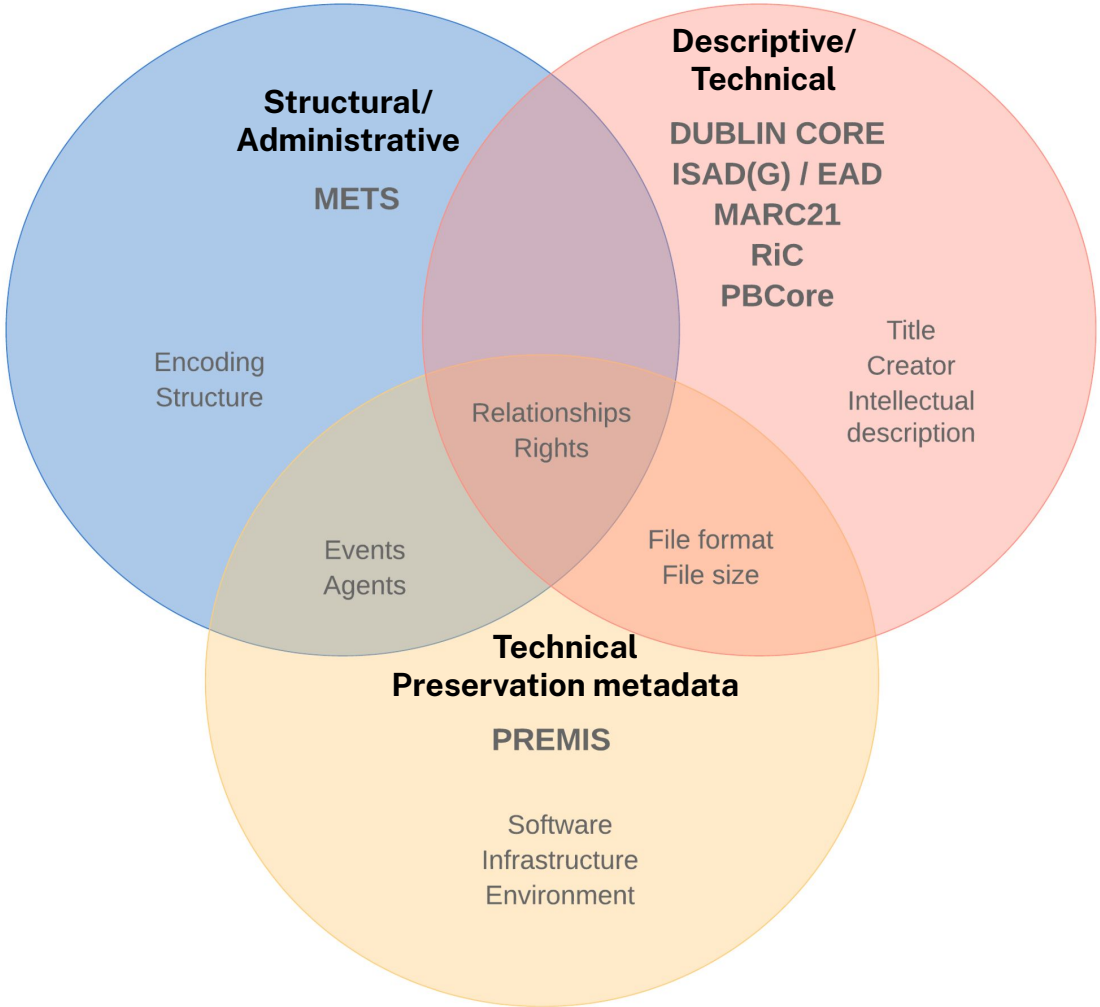
Source: Data Dictionary for Preservation Metadata, version 3

PREMIS - Environnement



Source: Data Dictionary for Preservation Metadata, version 3

Metadata aggregators



EN 17650

A framework for digital preservation of cinematographic works - The Cinema Preservation Package

Tableau 8 — Élément obligatoire de l'élément <digiprovMD>

Rubrique	Élément METS	Arité de l'élément	Attribut ou sous-élément	Sous-élément obligatoire	valeur
digiprovMD	mdRef	1	@LOCTYPE	x	"URL"
			@xlink:href	x	*relative path to file*
			@MDTYPE	x	"PREMIS"
			@MIMETYPE	x	"text/xml"
			@SIZE	x	*file size*
			@CHECKSUM	x	*file checksum*
			@CHECKSUMTYPE	x	*checksum algorithm*

Two ways to proceed ...

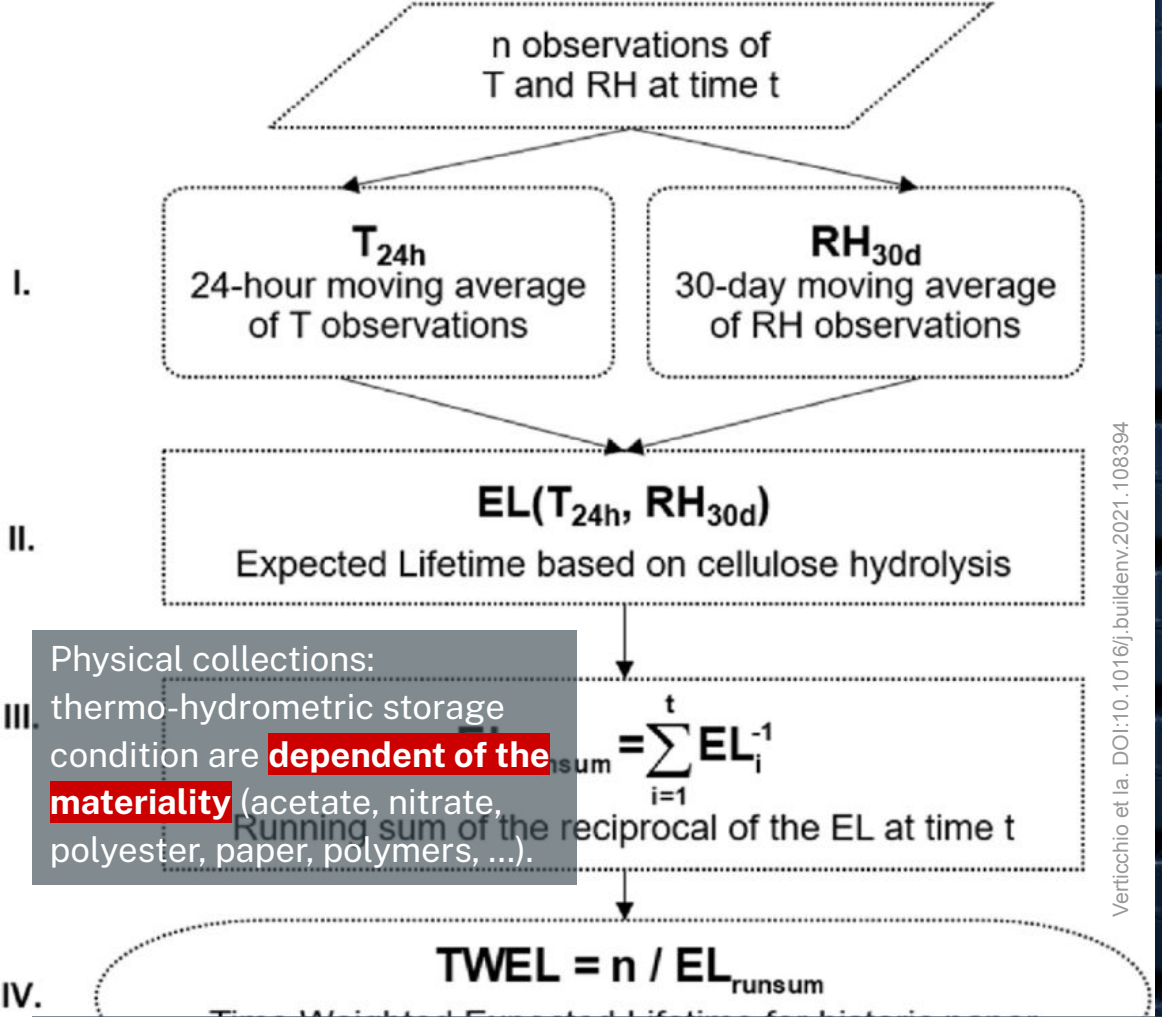
Two ways to proceed ...

cinémathèque suisse

Digitisation

Black Jack-Adobe Stock

Born digital



Verticchio et la. DOI:10.1016/j.buildenv.2021.108394

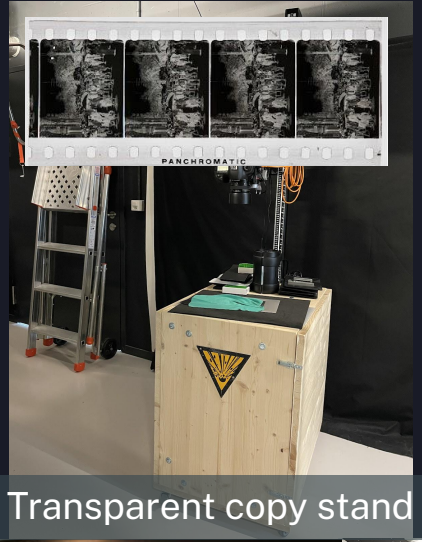
Physical collections: thermo-hydrometric storage condition are **dependent of the materiality** (acetate, nitrate, polyester, paper, polymers, ...).

Physical storage

Good news: digital files are preserved in the same vault, whatever it is Film or Film-related collections. Hence, methodologies are similar. It is, in the end, **binary**.

Digital storage: LTO libraries

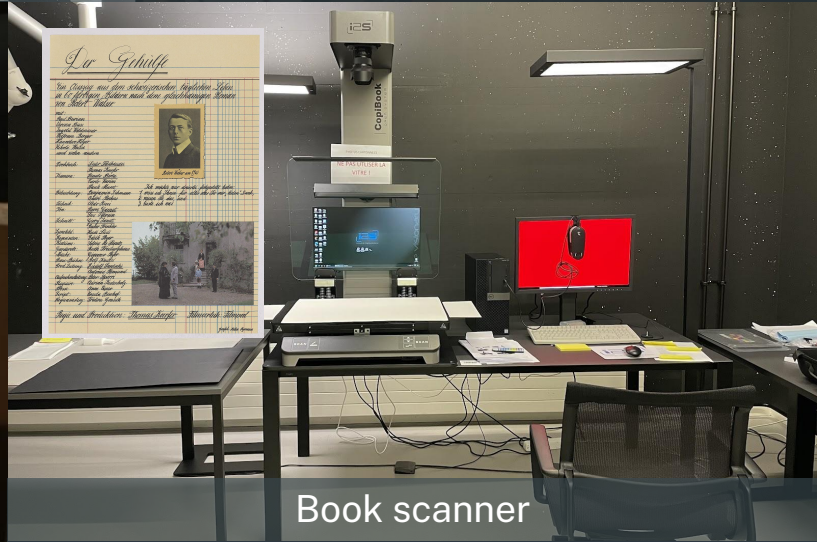
Copy stand



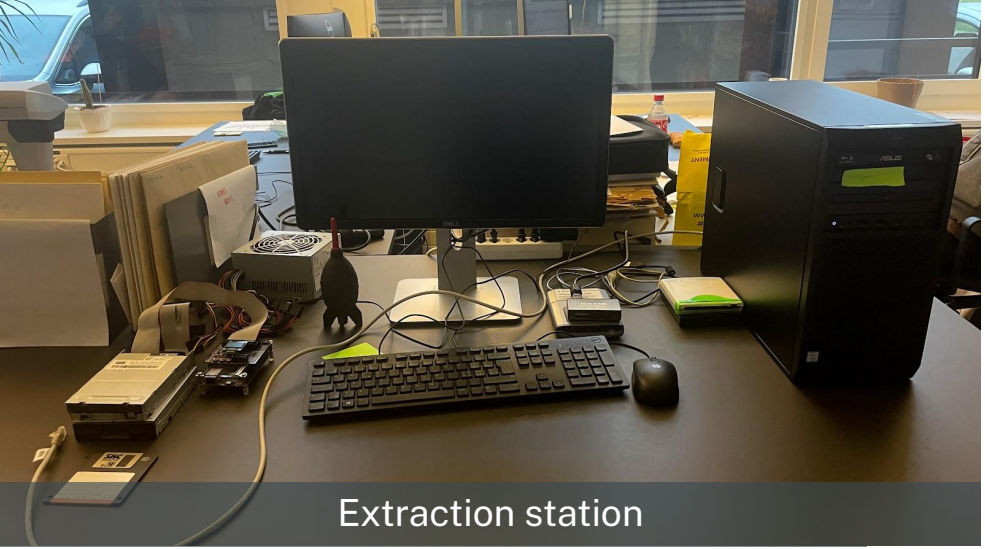
Transparent copy stand

Photogrammetry (3D digitisation)

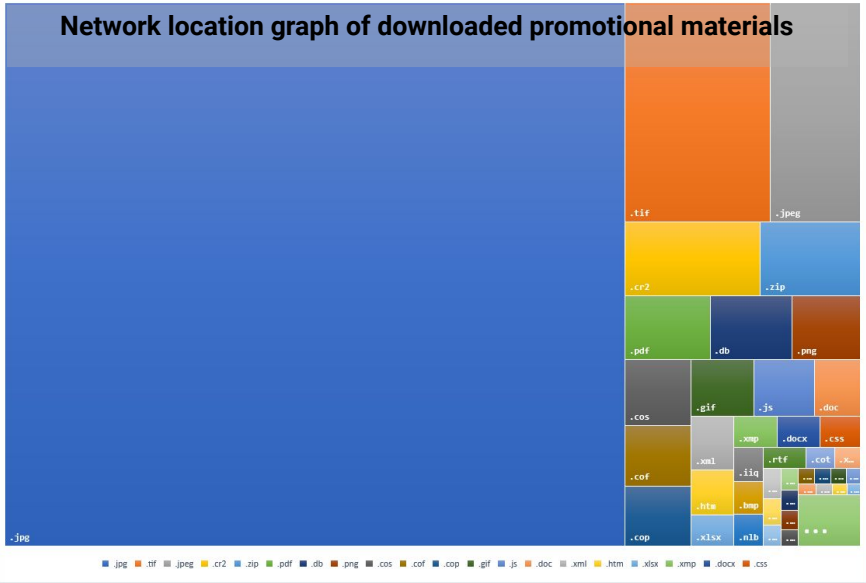
Digital Camera Back

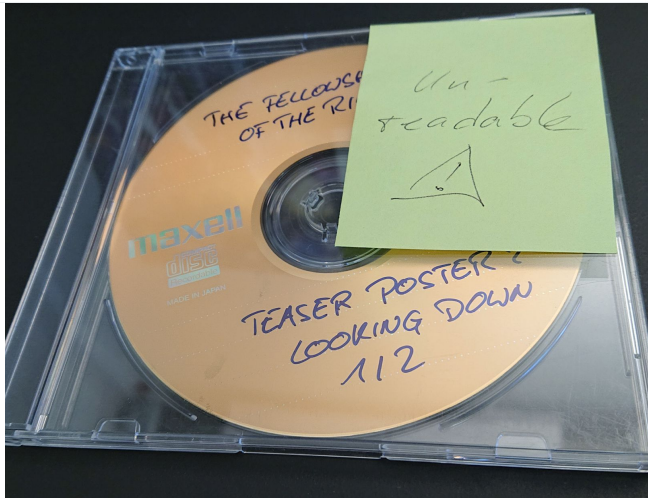


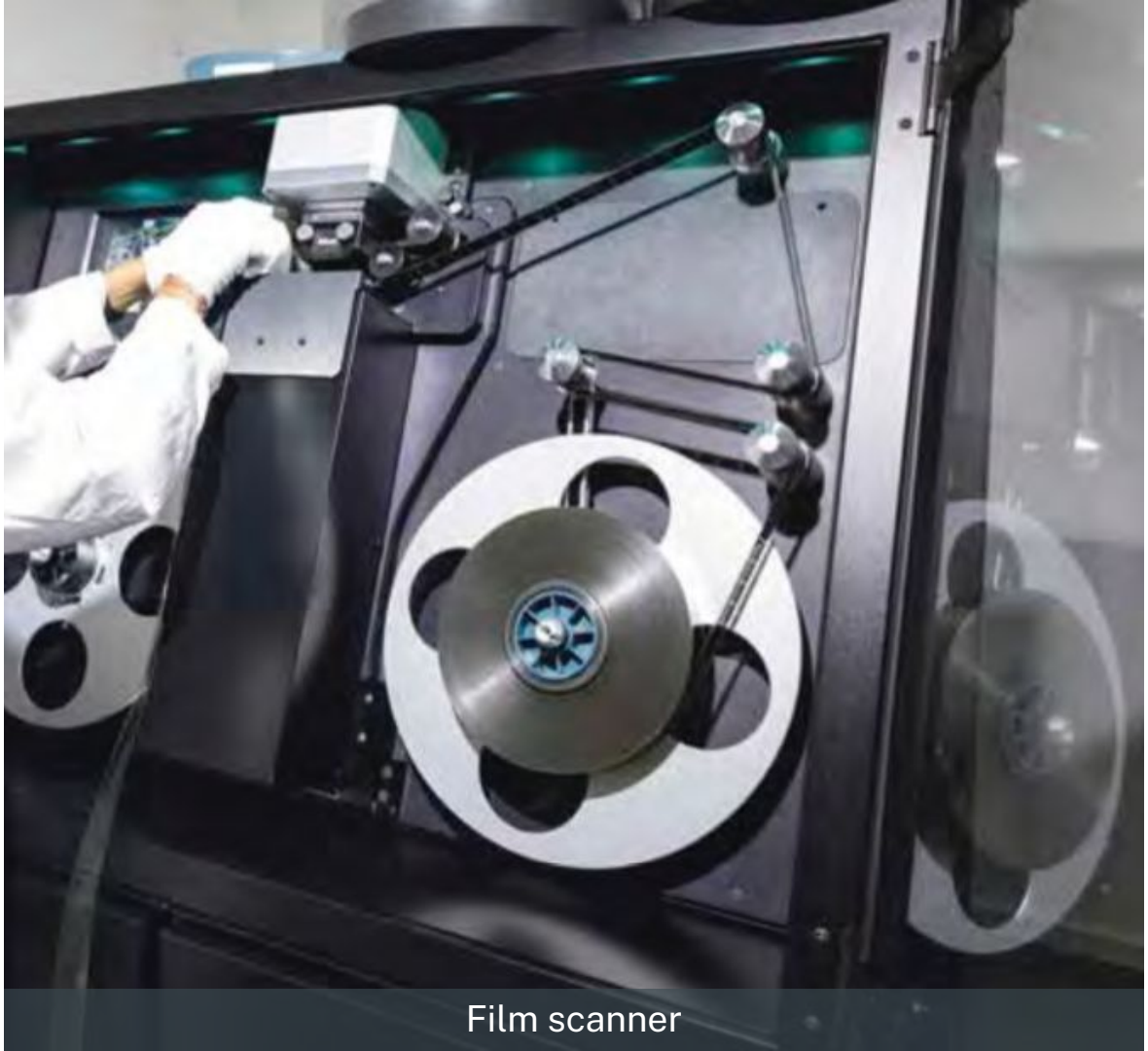
Book scanner



Extraction station

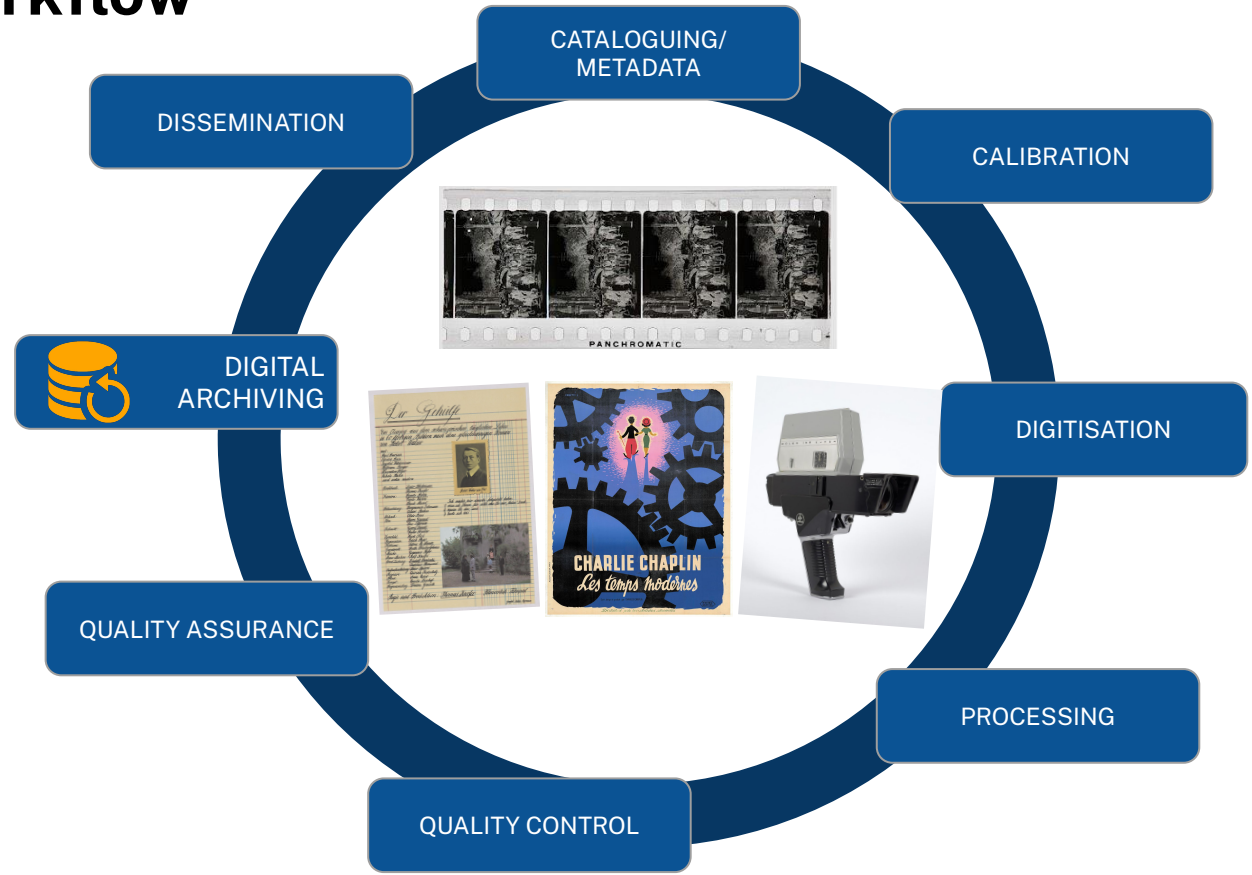






Film scanner

Core workflow



The journey

1. Get familiar with the **OAIS model**
2. Get familiar with **METS/PREMIS**
    a. Talk to your **neighbouring memory institutions** (not only film archives)
    b. Attend **conferences specialised on digital preservation** (iPres, No Time to Wait…), where you can discuss and meet other institutions (not only film archives).
3. **Analyse your workflows** and identify the main preservation metadata
4. Try to **ingest some sets of data** with existing tools that generate METS/PREMIS : Archivematica, DocuTeam Packer…
5. Have a look at the resulting METS/PREMIS from the tests
6. Go through the **PREMIS dictionary**

Going through the data dictionary

Evaluate and **choose** what you need

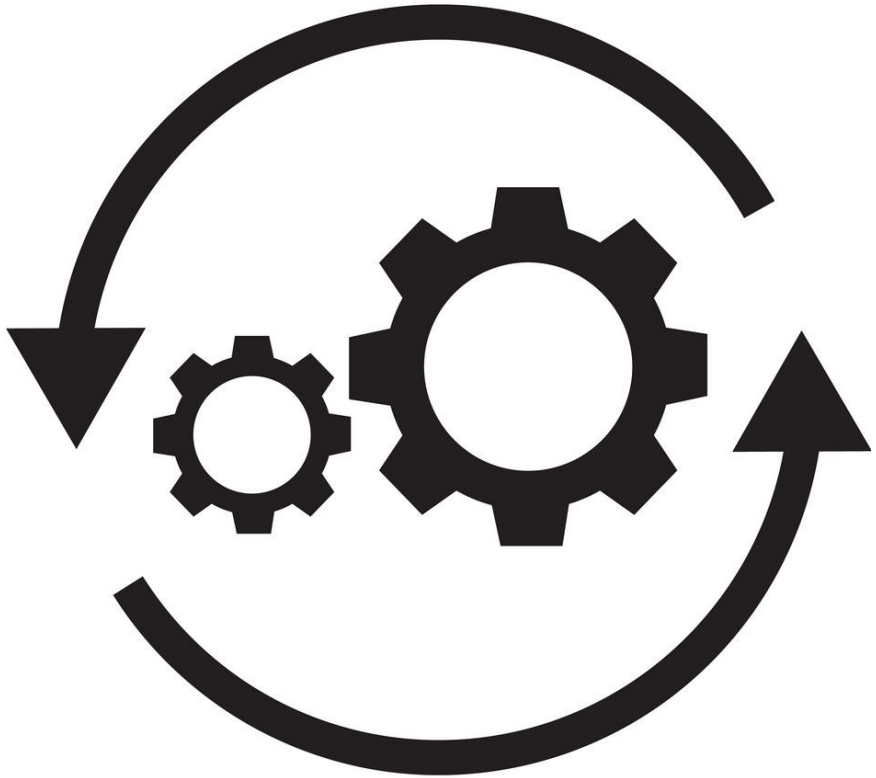
Not all the 180+ data dictionary entries

PREMIS - Implémentation du dictionnaire pour la CS

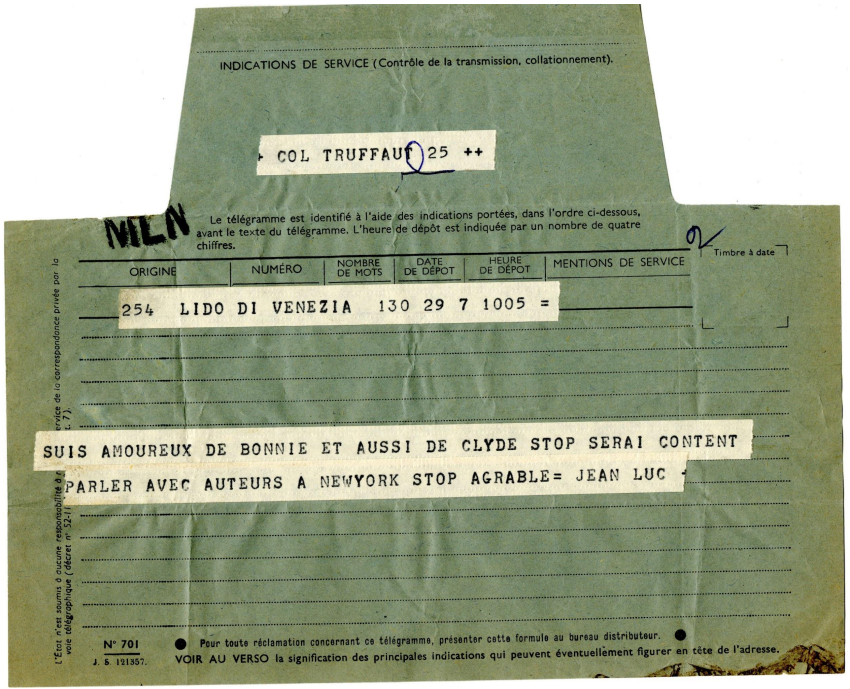
Unités sémantiques : https://www.loc.gov/standards/premis/v3/premis-3-0-datadictionary-only.pdf	Obligatoire Dictionnaire	Modèle CS Non-Film	Modèle CS Film	Exemple	METS	BnF/LoC
Exemple implémentation par *National digital preservation services of Finland*						
https://github.com/orgs/Digital-Preservation-Finland/repositories?type=all						
Unités sémantiques (semantic units)						
1. Object						
1.1 objectIdentifier (M, R)	oui				techMD	oui
1.1.1 objectIdentifierType (M, NR)	oui			ARK	techMD	oui
1.1.2 objectIdentifierValue (M, NR)	oui			No ARK	techMD	oui
1.2 objectCategory (M, NR)	oui			intellectual entity representation	techMD	
1.3 preservationLevel (O, R) [Intellectual Entity, Representation, File]		oui			techMD	oui
1.3.1 preservationLevelType (O, NR) [Intellectual Entity, Representation, File]		oui			techMD	oui
1.3.2 preservationLevelValue (M, NR) [Intellectual Entity, Representation, File]	oui			haute/moyen/basse	techMD	oui
1.3.3 preservationLevelRole (O, NR) [Intellectual Entity, Representation, File]		oui			techMD	oui
1.3.4 preservationLevelRationale (O, R) [Intellectual Entity, Representation, File]		oui			techMD	oui
1.3.5 preservationLevelDateAssigned (O, NR) [Intellectual Entity, Representation, File]		oui			techMD	oui
1.4 significantProperties (O, R)		non	non		techMD	oui
1.4.1 significantPropertiesType (O, NR)			non		techMD	oui
1.4.2 significantPropertiesValue (O, NR)			non		techMD	oui
1.4.3 significantPropertiesExtension (O, R)			non		techMD	oui
1.5 objectCharacteristics (M, R) [File, Bitstream]	oui				techMD	oui
1.5.1 compositionLevel (O, NR) [File, Bitstream]			non		techMD	oui
1.5.2 fixity (O, R) [File, Bitstream]		oui			techMD	oui
1.5.2.1 messageDigestAlgorithm (M, NR) [File, Bitstream]	oui			algorithme MD5	techMD	oui
1.5.2.2 messageDigest (M, NR) [File, Bitstream]	oui			somme de contrôle MD5	techMD	oui
1.5.2.3 messageDigestOriginator (O, NR) [File, Bitstream]		oui			techMD	oui
1.5.3 size (O, NR) [File, Bitstream]		oui			techMD	oui
1.5.4 format (M, R) [File, Bitstream]	oui				techMD	oui
1.5.4.1 formatDesignation (O, NR) [File, Bitstream]		oui			techMD	oui
1.5.4.1.1 formatName (M, NR) [File, Bitstream]	oui			TIFF / PDF	techMD	oui
1.5.4.1.2 formatVersion (O, NR) [File, Bitstream]		oui			techMD	oui
1.5.4.2 formatRegistry (O, NR) [File, Bitstream]		oui			techMD	oui
1.5.4.2.1 formatRegistryName (M, NR) [File, Bitstream]	oui			PRONOM	techMD	oui
1.5.4.2.2 formatRegistryKey (M, NR) [File, Bitstream]	oui			fmt/155	techMD	oui
1.5.4.2.3 formatRegistryRole (O, NR) [File, Bitstream]			non		techMD	oui
1.5.4.3 formatNote (O, R) [File, Bitstream]			non		techMD	
1.5.5 creatingApplication (O, R) [File, Bitstream]		oui			techMD	

Workflows

examples



Textual/Still documents



INDICATIONS DE SERVICE (Contrôle de la transmission, collationnement).

COL TRUFFAUT 25 ++

NILN

Le télégramme est identifié à l'aide des indications portées, dans l'ordre ci-dessous, avant le texte du télégramme. L'heure de dépôt est indiquée par un nombre de quatre chiffres.

Timbre à date

ORIGINE	NUMÉRO	NOMBRE DE MOTS	DATE DE DÉPOT	HEURE DE DÉPOT	MENTIONS DE SERVICE
254	LIDO DI VENEZIA	130	29	7 1005	=

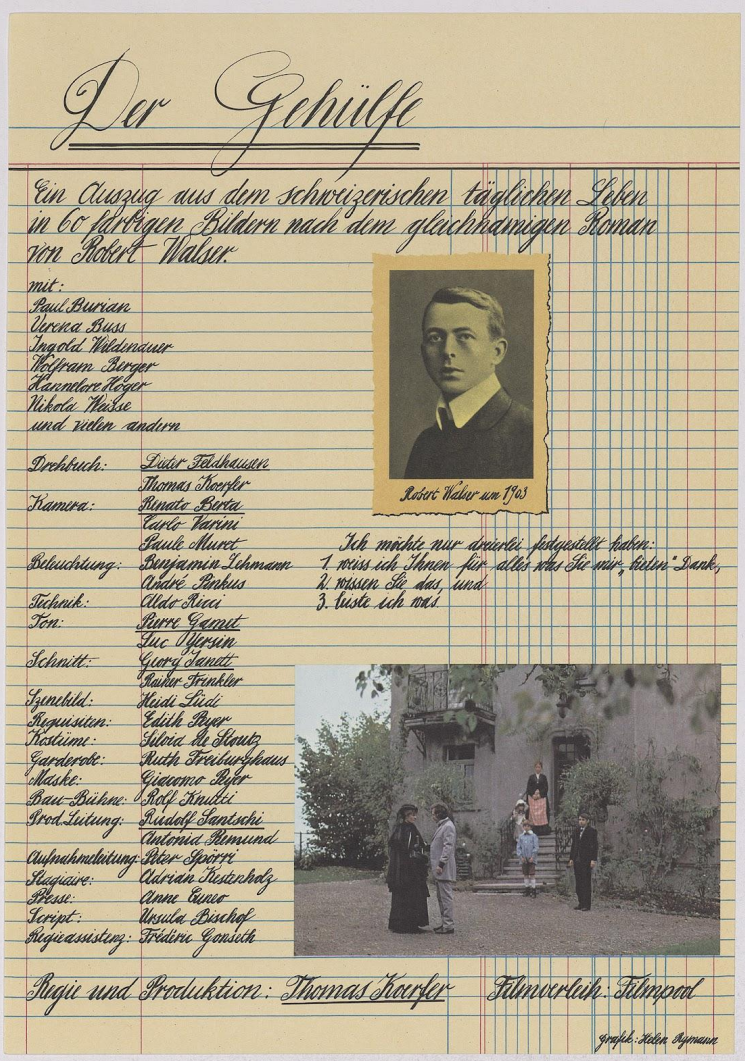
SUIS AMOUREUX DE BONNIE ET AUSSI DE CLYDE STOP SERAI CONTENT

PARLER AVEC AUTEURS A NEWYORK STOP AGRABLE= JEAN LUC

L'Etat n'est soumis à aucune responsabilité à raison du service de la correspondance privée par la voie télégraphique (décret n° 52-1 [illegible] c. 7).

N° 701
J. 5 121357.

● Pour toute réclamation concernant ce télégramme, présenter cette formule au bureau distributeur. ●
VOIR AU VERSO la signification des principales indications qui peuvent éventuellement figurer en tête de l'adresse.



Der Gehülfe

Ein Auszug aus dem schweizerischen täglichen Leben in 60 farbigen Bildern nach dem gleichnamigen Roman von Robert Walser.

mit:
Paul Burian
Verena Buss
Ingold Wildenauer
Wolfram Berger
Laurentore Roger
Nikola Weise
und vielen andern

Robert Walser um 1903

Drehbuch: Lutz Feldhausen, Thomas Koerfer
Kamera: Renato Berta, Carlo Varini, Paule Muret
Beleuchtung: Benjamin Lehmann, André Pinkus
Technik: Aldo Ricci
Ton: Pierre Gamet, Luc Yersin
Schnitt: Georg Janett, Walter Frinkler
Szenebild: Heidi Lüdi
Requisiten: Edith Meyer
Kostüme: Silvia de Stoutz
Garderobe: Ruth Freiburghaus
Maske: Giacomo Tyler
Bau-Bühne: Rolf Knubi
Prod.Leitung: Rudolf Santschi, Antonia Rimmund
Aufnahmeleitung: Peter Spörri
Standfoto: Adrian Riestenholz
Presse: Anne Cuneo
Script: Ursula Bischof
Regieassistenz: Frédéric Gonseth

Ich möchte nur draurlei festgestellt haben:
1. weiss ich Ihnen für alles was Sie mir "Lieben" Dank,
2. wissen Sie das, und
3. Liebe ich mich

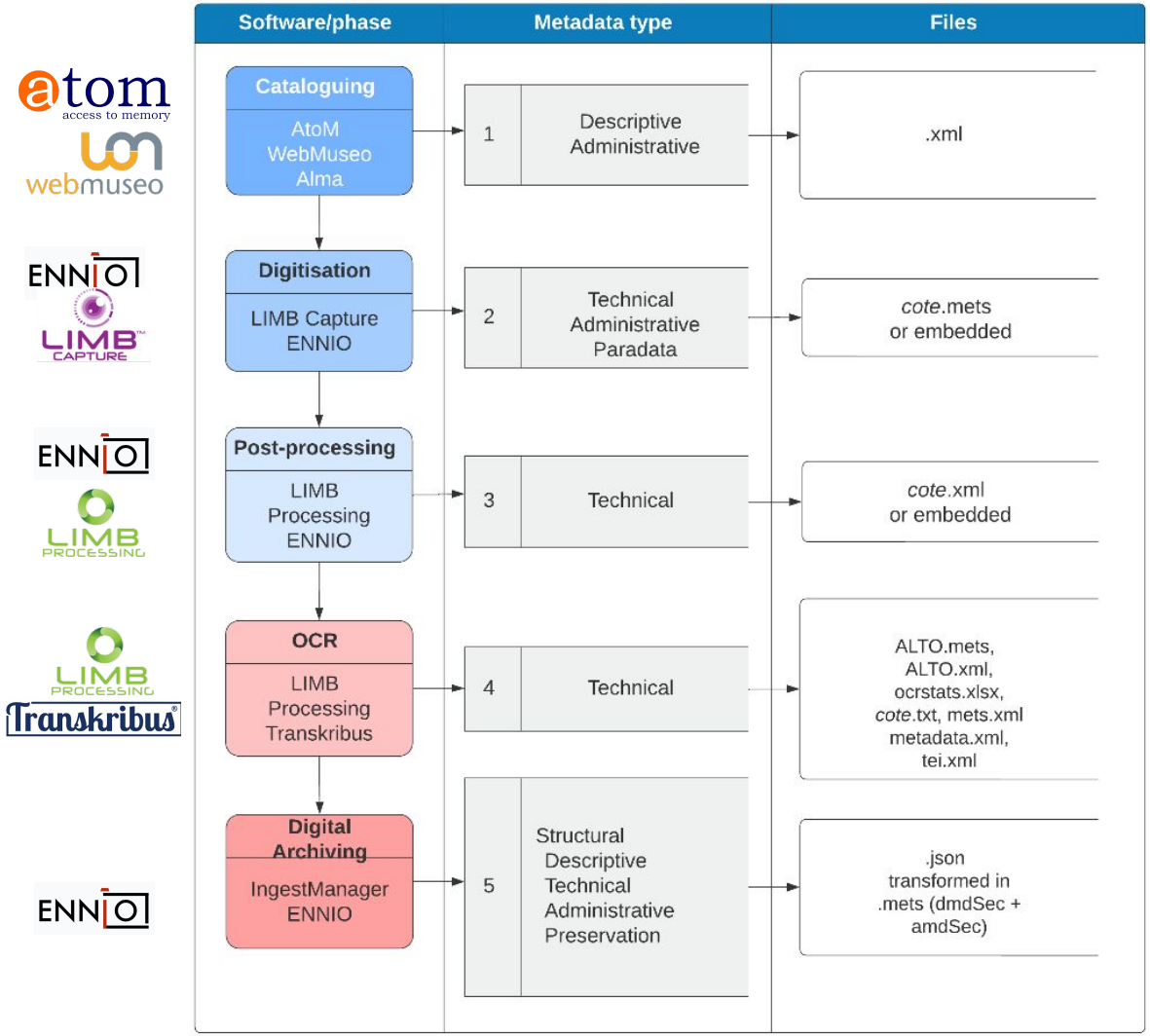
Regie und Produktion: Thomas Koerfer Filmverleih: Filmcoop

Graph. Helen Dymann

Metadata workflow

Book Scanner example,

with OCR (LIMB) and HTR (Transkribus)



atom access to memory

webmuseo

Dublin Core
EAD
MARC21
EN15907
RiC

ENNO

Digitsation validation date
Digitisation agent name
Errors logs
Ingest validation date
Post-processing validation date

LIMB CAPTURE

Type of scanner
Scanner brand and serial number
Software name and version
Name of files generated
Fixity check (MD5)

LIMB PROCESSING

Quality report
processingSoftware
softwareCreator
softwareName
softwareVersion
applicationDescription
pageLayout

Transkribus

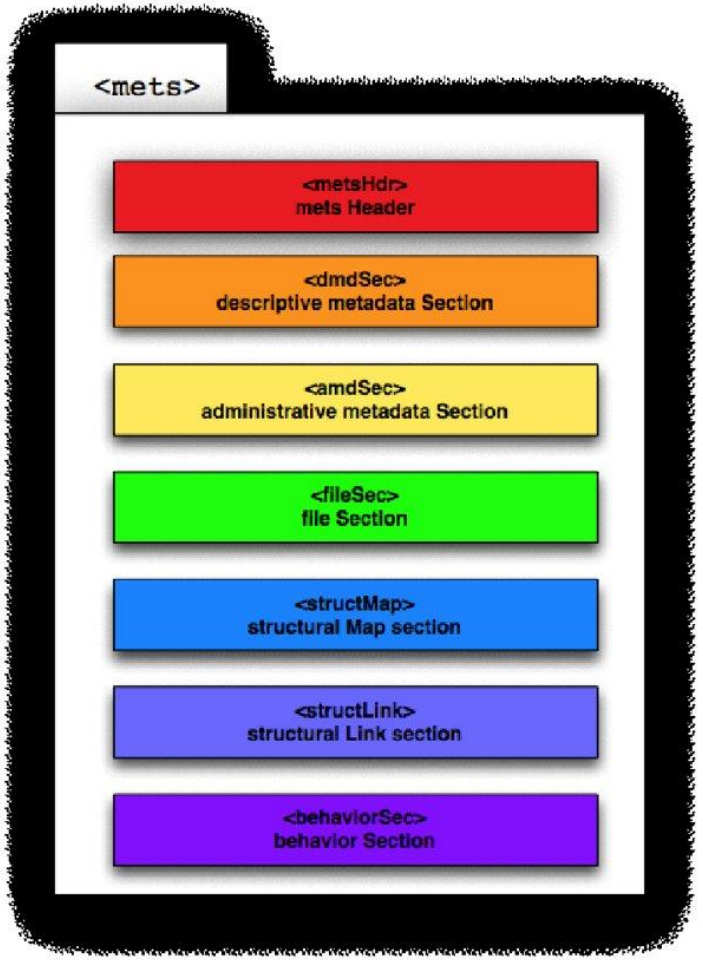
zonepoints
facs
uploadTimestamp
uploader
uploaderId

METS structure

7 sections

- metsHdr
- dmdSec
- amdSec
- fileSec
- structMap
- structLink
- behaviorSec

METS v2/v3, METS/PREMIS overlap



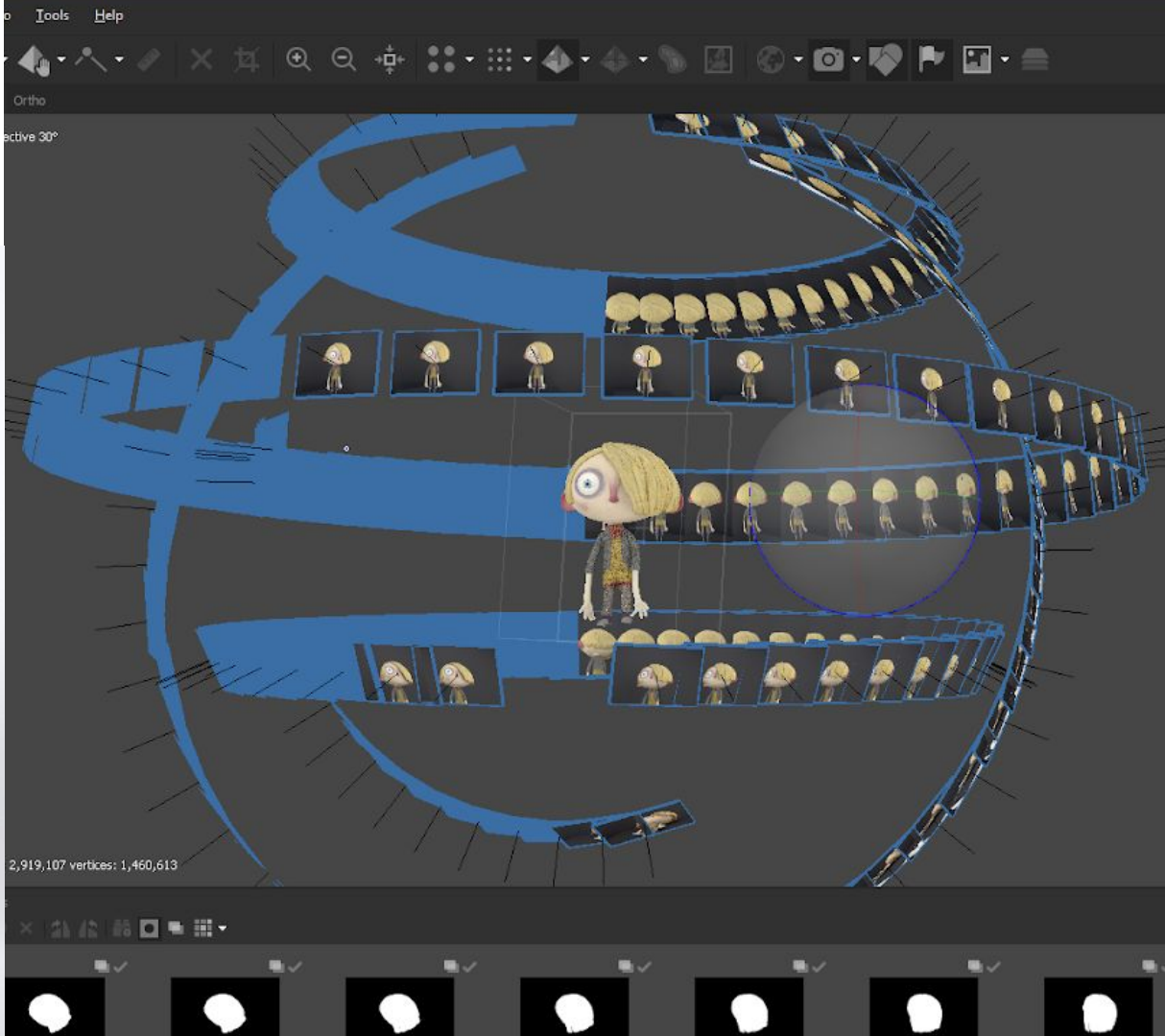
METS/PREMIS example with DVD ISO

Generated by Archivematica on a ISO file made from a DVD-Video


KEEP
CALM
and
Coffee
Break

3D objects


BOLEX 155 SUPER
PAILLARD-BOLEX
LENS AND CAMERA
MADE IN SWITZERLAND


Tools
Help
Ortho
2,919,107 vertices: 1,460,613

3D digitisation workflow: example

Data acquisition

Standard photogrammetry

Calibration
ICC profile

RAW image collection
(Canon, 70-200 images,
3 rings)

RAW image
converted into TIFF

HeliconSoft
RAW image
collection in
Helicon remote

HeliconSoft
Photos stacking
Batch process

Macro-photogrammetry

Data processing

Masks created in Ps

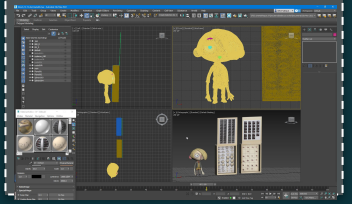
Import mask
Estimate image quality
Align photos
Metashape

Gradual selection
Dense clouds mild
Dense clouds cleaned
Mesh
Texture (8092x2)
Scaling
Metashape

Data assembling

blender
AUTODESK 3DS MAX

Import .obj+texture, mtl
Assemble elements
Animation
Export as .obj



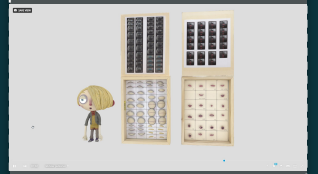
Deliverable output

3D models -
Preservation

Metadata
Paradata
Quality report

3D models
Dissemination

Decimated/animation



Paradata

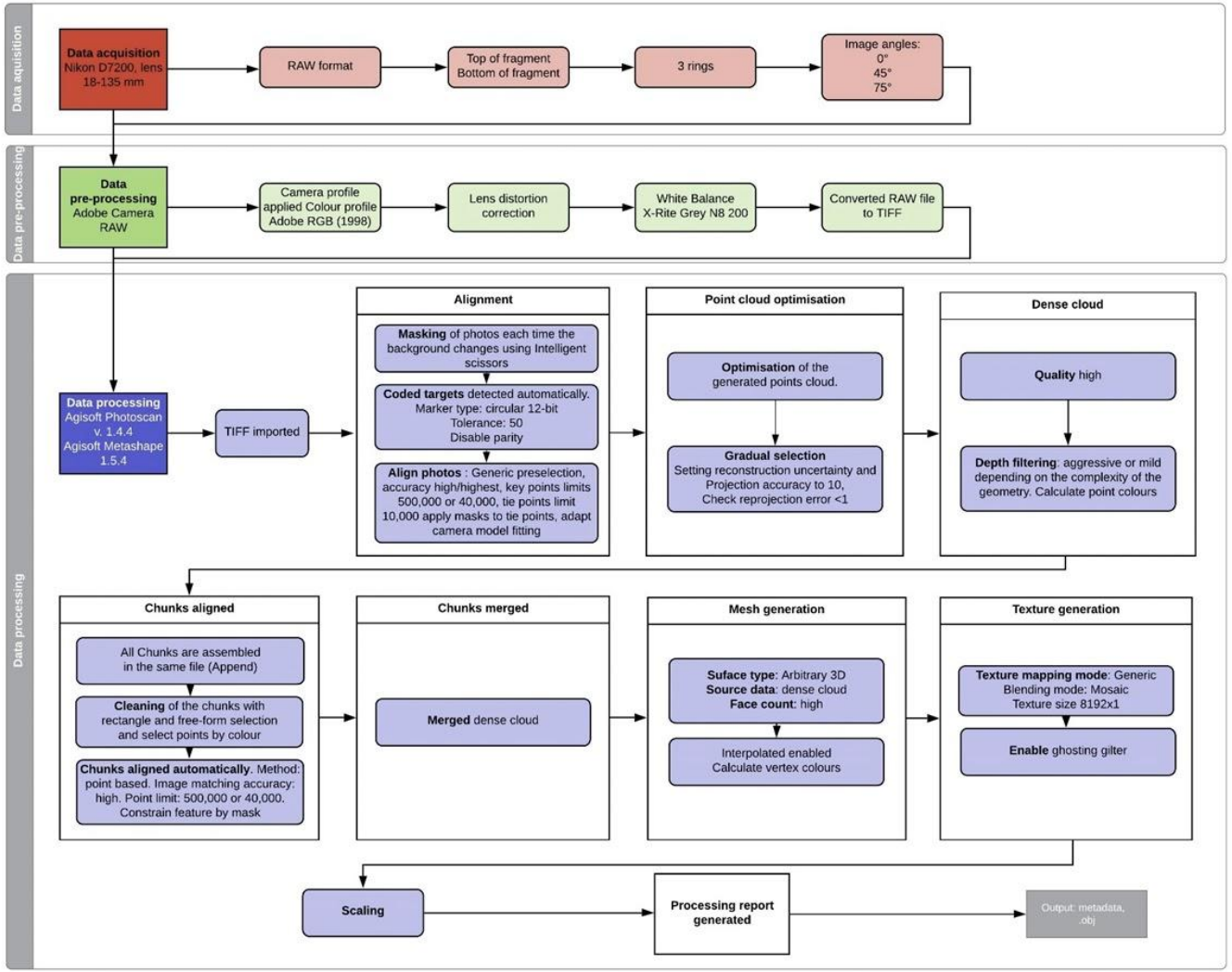
“

Digitisation is already an interpretation

- Joffres et al. 2017

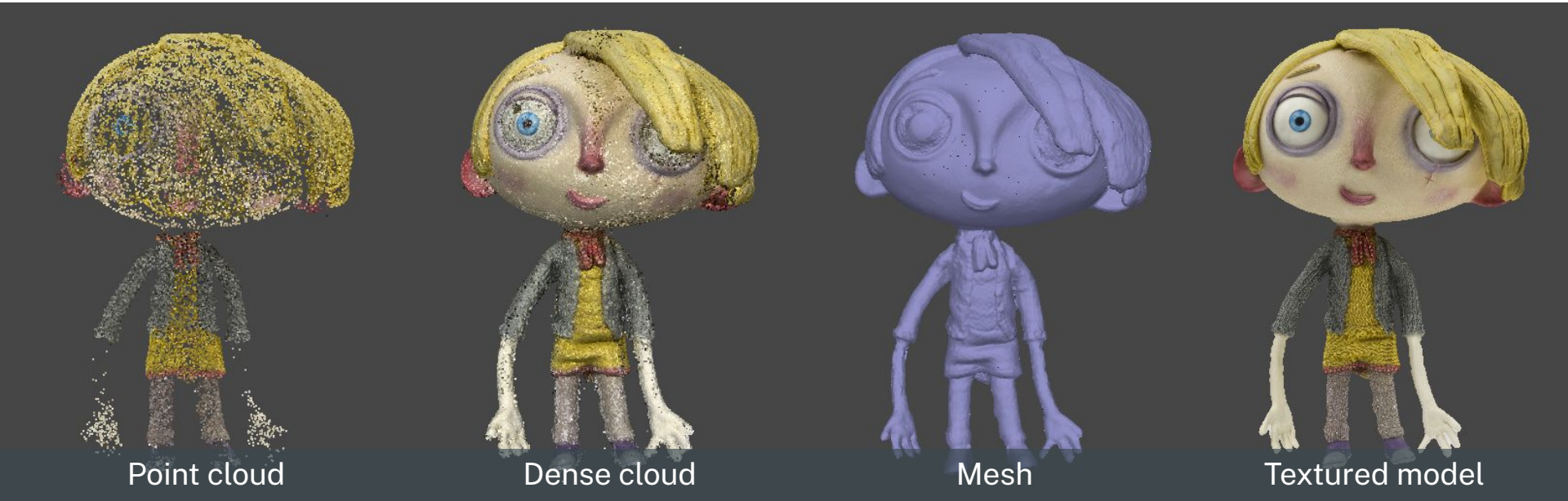
PARTHENOS, Horizon 2020

Explain like a **scientific experimentation**



3D complexity

- **Multiple source files created by proprietary software and hardware**
- **Preservation complication → "need to preserve a lot of additional metadata related to devices, scanners, tools and equipment used to capture the data."** Ahmad and Cassidy, DPC Preserving 3D, 2021



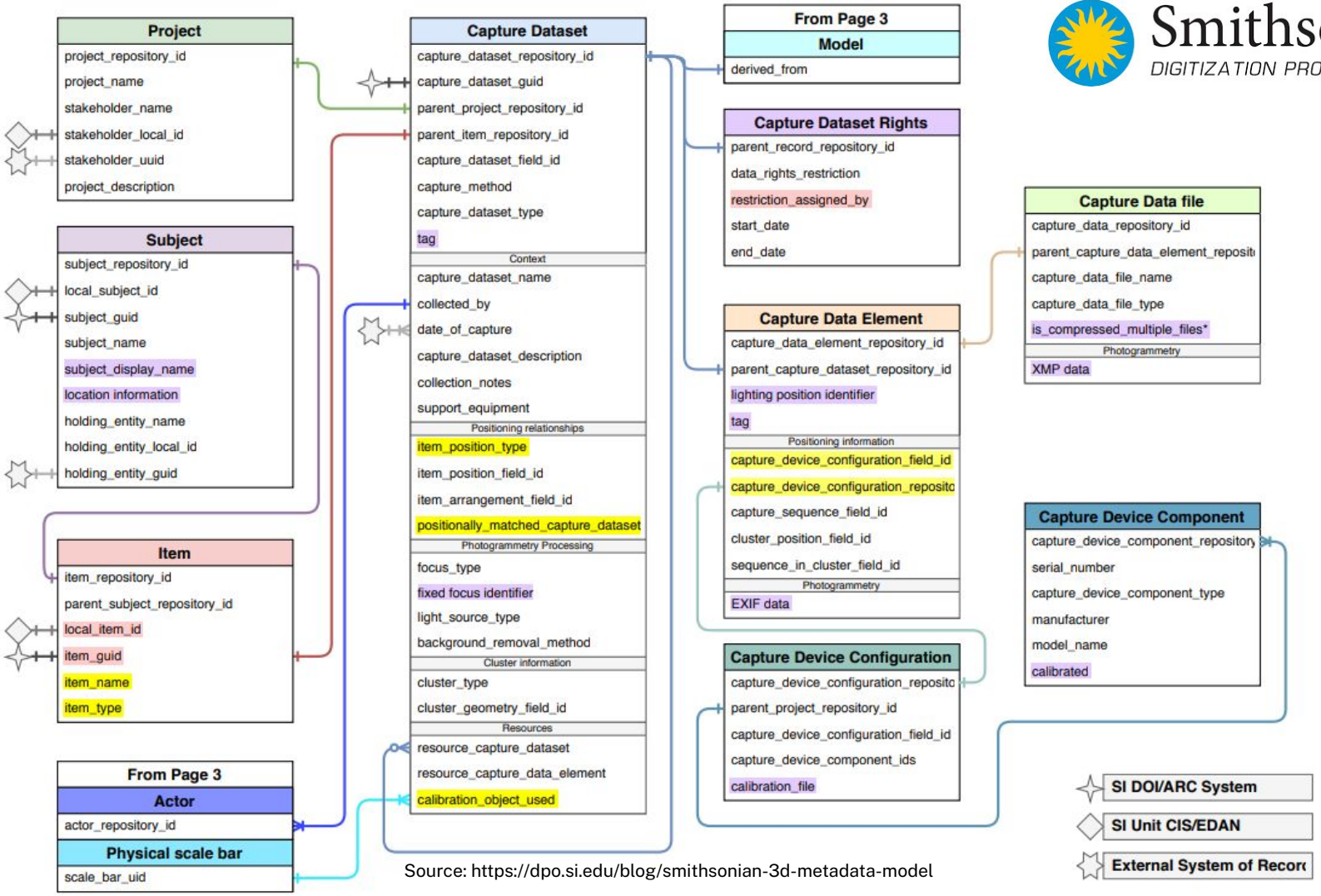
Recommended Formats Statement Library of Congress

Design and 3D

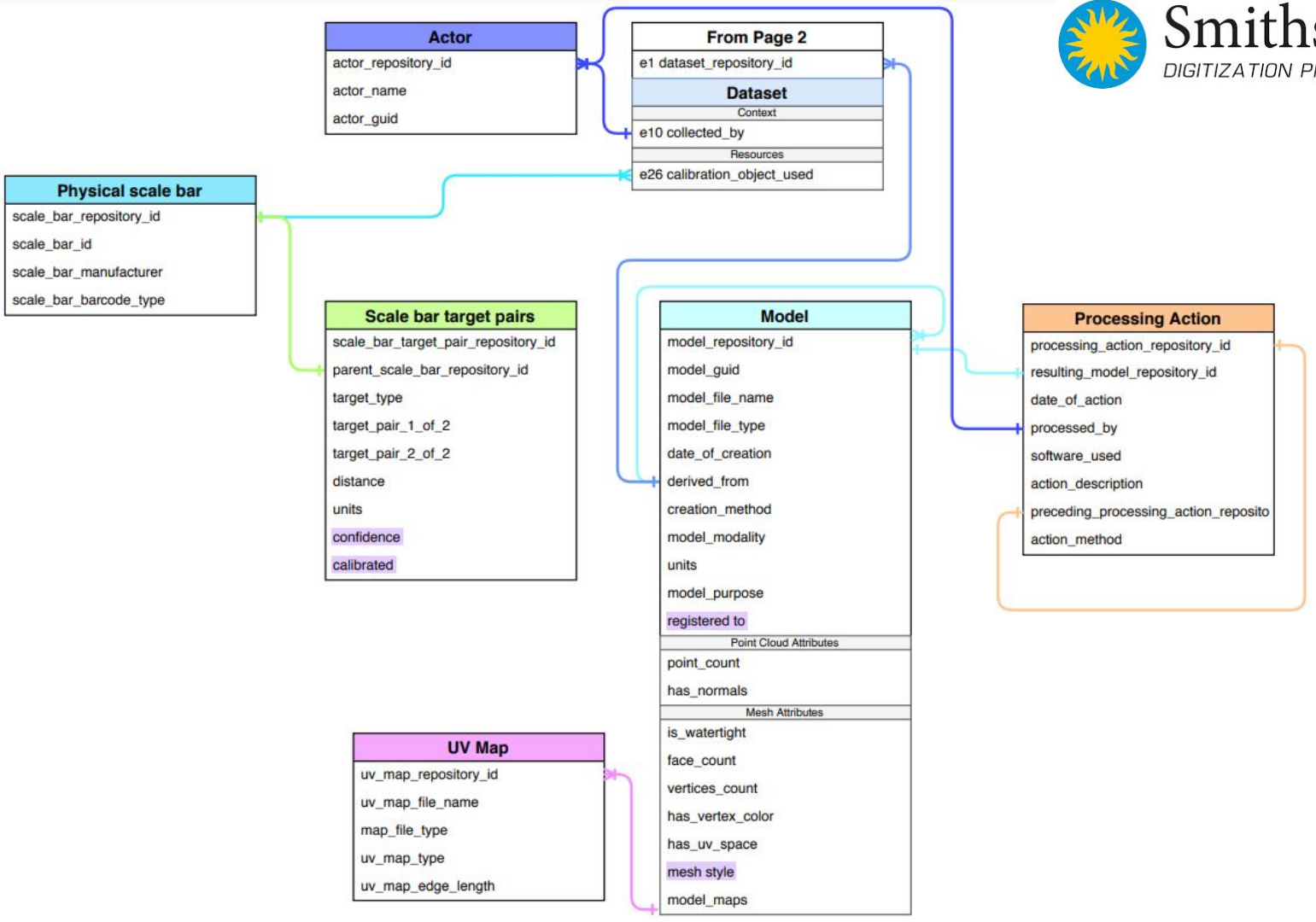
iii. Scanned 3D Objects (output from photogrammetry scanning)

iii. Scanned 3D Objects (output from photogrammetry scanning)

	Preferred	Acceptable
A. Formats		› STereoLithography (*.stl) › Reflectance Transformation Imaging (*.rti) › Polygon File Format (*.ply) › Wavefront (*.obj)
B. Related Materials	› Includes indexes, study guides or other matter if available › Also includes annotations, accompanying tabular or textual matter or other interpretive aids	
C. Metadata	1. As supported by format: a. Title b. Creator c. Creation Date d. Place of publication e. Publisher/producer/distributor f. Contact information 2. Include if available: a. Common embedded schema (e.g., FGD, ISO 19115) b. Language of work c. Other relevant identifiers (e.g., DOI, LCCN, etc.) d. Subject descriptors e. Abstracts f. Key or reference to each data field and technical production information (type of paper, how processed, publisher internal tracking numbers)	Metadata provided separately in external text or XML-based file
D. Technological Measures	Files must contain no measures (such as digital rights management technologies or encryption) that control access to or prevent use of the digital work.	

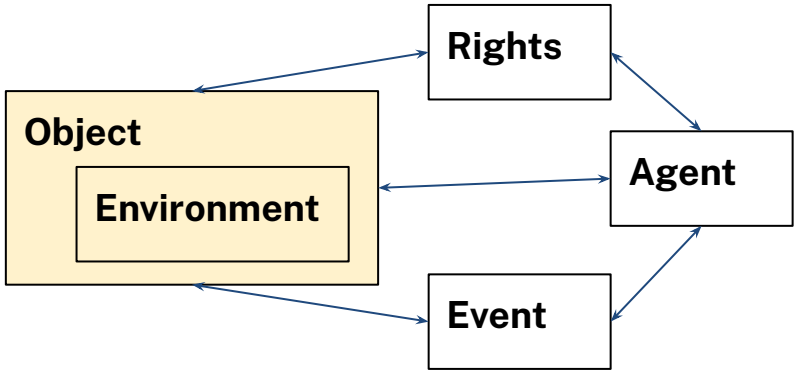


Source: https://dpo.si.edu/blog/smithsonian-3d-metadata-model


Smithsonian
DIGITIZATION PROGRAM OFFICE
Actor
actor_repository_id
actor_name
actor_guid
From Page 2
e1 dataset_repository_id
Dataset
Context
e10 collected_by
Resources
e26 calibration_object_used
Physical scale bar
scale_bar_repository_id
scale_bar_id
scale_bar_manufacturer
scale_bar_barcode_type
Scale bar target pairs
scale_bar_target_pair_repository_id
parent_scale_bar_repository_id
target_type
target_pair_1_of_2
target_pair_2_of_2
distance
units
confidence
calibrated
Model
model_repository_id
model_guid
model_file_name
model_file_type
date_of_creation
derived_from
creation_method
model_modality
units
model_purpose
registered to
Point Cloud Attributes
point_count
has_normals
Mesh Attributes
is_watertight
face_count
vertices_count
has_vertex_color
has_uv_space
mesh style
model_maps
Processing Action
processing_action_repository_id
resulting_model_repository_id
date_of_action
processed_by
software_used
action_description
preceding_processing_action_reposito
action_method
UV Map
uv_map_repository_id
uv_map_file_name
map_file_type
uv_map_type
uv_map_edge_length

PREMIS

Object entity



Intellectual entity

Stop-motion puppet Alice, from *Ma Vie de Courgette*

Representation

- Original TIFFs
- Raw 3D model (.obj)
- Assembled 3D model (.fbx)
- Animated video

File

TIFF, OBJ, FBX, AVI

objectCharacteristics

<mix:BasicImageCharacteristics>
<mix:imageWidth>5894</mix:imageWidth><mix:imageHeight>7768</mix:imageHeight></mix:BasicImageCharacteristics>

+ fixity
+ file name
+ file format
+ creating application
+ unique identifier (ark)

...

Processing Parameters

General	
Cameras	229
Aligned cameras	229
Markers	25
Scale bars	2
Coordinate system	Local Coordinates (m)
Rotation angles	Yaw, Pitch, Roll
Point Cloud	
Points	142,006 of 301,123
RMS reprojection error	0.138628 (0.547687 pix)
Max reprojection error	0.408247 (6.46498 pix)
Mean key point size	3.5873 pix
Point colors	3 bands, uint8
Key points	No
Average tie point multiplicity	3.02263
Dense Point Cloud	
Points	30,895,093
Point colors	3 bands, uint8
Reconstruction parameters	
Quality	High
Depth filtering	Aggressive
Model	
Faces	6,179,018
Vertices	3,090,093
Vertex colors	3 bands, uint8
Texture	8,192 x 8,192, 4 bands, uint8
Reconstruction parameters	
Surface type	Arbitrary
Source data	Dense
Interpolation	Enabled
Quality	High
Depth filtering	Aggressive
Face count	6,179,018
Processing time	19 minutes 53 seconds
Texturing parameters	
Mapping mode	Generic
Blending mode	Mosaic
Texture size	8,192 x 8,192
Enable hole filling	Yes
Enable ghosting filter	Yes
UV mapping time	1 minutes 29 seconds
Blending time	28 minutes 43 seconds
Software	
Version	1.4.4 build 6848
Platform	Windows 64

Linked to the PDF (PDF/A) processing report from Agisoft software in the PREMIS relation. METS section `structMap TYPE="attachment"`

Camera Calibration

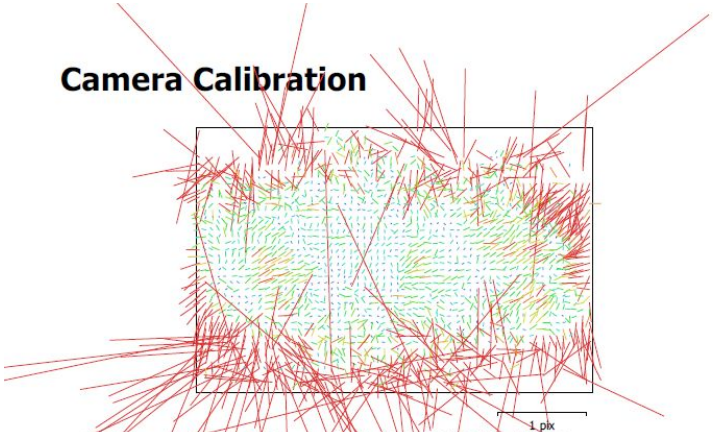


Fig. 2. Image residuals for NIKON D7200 (35mm).

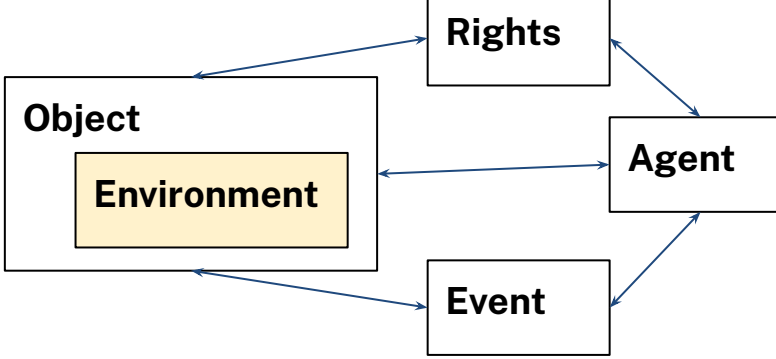
NIKON D7200 (35mm)
98 images

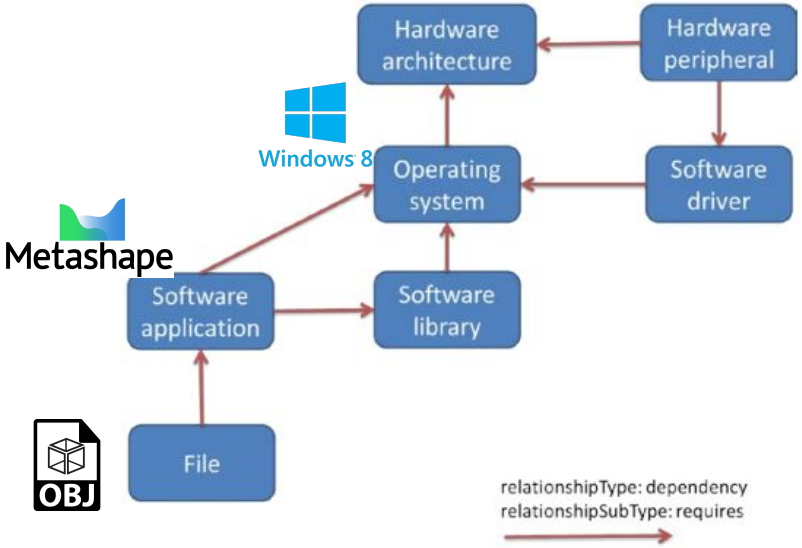
Type	Resolution	Focal Length	Pixel Size
Frame	**6000 x 4000**	**35 mm**	**3.91 x 3.91 µm**

	Value	Error	F	K1	K2	P1	P2
F	**8391.8**	0.74	1.00	0.24	0.22	-0.14	-0.62
K1	**0.137699**	0.00011		1.00	-0.84	0.00	-0.11
K2	**0.485517**	0.00081			1.00	-0.04	-0.13
P1	**0.00310123**	7.9e-06				1.00	0.12
P2	**-0.00423759**	1.2e-05					1.00

Table 2. Calibration coefficients and correlation matrix.

PREMIS Object entity - Environment





Metashape OBJ

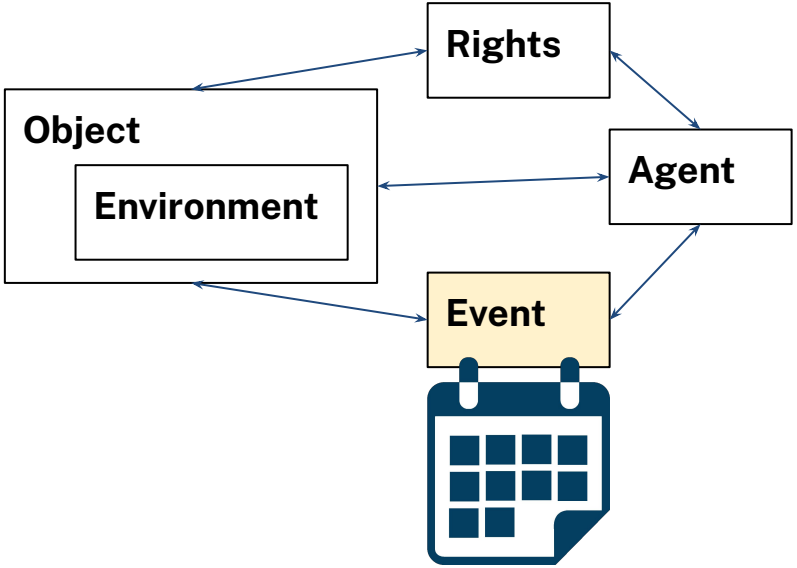
Raw 3D model (.obj) + regenerate 3D model from the TIFF

Parts of an environment stack and dependency relationships between them
Source: PREMIS data dictionary, v3

PREMIS - Environment metadata		
environmentFunction (O, R) [Intellectual Entity of type environment]		Dependencies
environmentFunctionType (M, NR) [Intellectual Entity of type environment]	Software	Operating system
environmentFunctionLevel (M, NR) [Intellectual Entity of type environment]	2	1
environmentDesignation (O, R) [Intellectual Entity of type environment]		
environmentName (M, NR) [Intellectual Entity of type environment]	Metashape	Windows
environmentVersion (O, NR) [Intellectual Entity of type environment]	1.4.4.	8
environmentOrigin (O, NR) [Intellectual Entity of type environment]	Agisoft LLC	Microsoft Corporation
environmentDesignationNote (O, R) [Intellectual Entity of type environment]	64-bit	64-bit
environmentDesignationExtension (O, R) [Intellectual Entity of type environment]		
environmentExtension (O, R) [Intellectual Entity of type environment]		

PREMIS Event entity

capture	message digest calculation
creation	metadata extraction
decompression	metadata modification
decryption	migration
deletion	normalization
digital signature validation	packing
encryption	policy assignment
exporting	quarantine
extraction	recovery
fixity check	rendering
format identification	transfer
imaging	unquarantine
information package creation	virus check
ingestion	validation



Tip : use a controlled vocabulary

Source: AMIA Preservation Metadata PREMIS, 14/07/2021

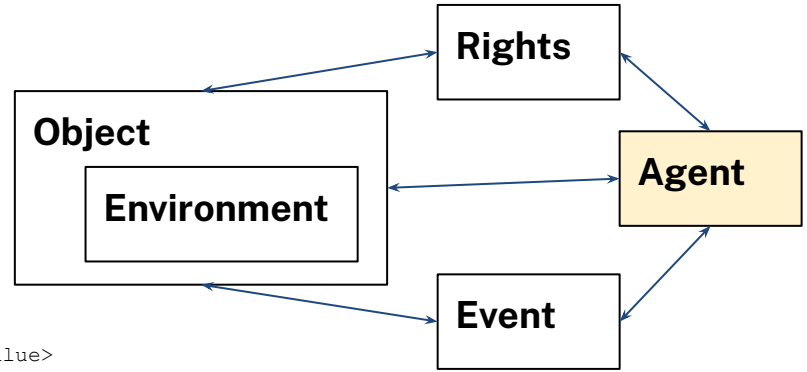
```
<premis:eventIdentifier>
  <premis:eventIdentifierType> UUID</premis:eventIdentifierType>
  <premis:eventIdentifierValue> 77efc04b-6520-4de4-9ad9-c85e13394546 </premis:eventIdentifierValu
  e>
</premis:eventIdentifier>
<premis:eventType> format identification </premis:eventType>
<premis:eventDateTime>2022-08-19T10:15:1+00:00</premis:eventDateTime>
<premis:eventDetailInformation>
  <premis:eventDetail>program=" PRONOM"</premis:eventDetail>
</premis:eventDetailInformation>
<premis:eventOutcomeInformation>
  <premis:eventOutcome/>
  <premis:eventOutcomeDetail>
    <premis:eventOutcomeDetailNote> fmt/1210</premis:eventOutcomeDetailNote>
  </premis:eventOutcomeDetail>
</premis:eventOutcomeInformation>
<premis:linkingAgentIdentifier>
  <premis:linkingAgentIdentifierType> local</premis:linkingAgentIdentifierType>
  <premis:linkingAgentIdentifierValue> PCW-PZ-RENDU1</premis:linkingAgentIdentifierValue>
  <premis:linkingAgentRole> hardware</premis:linkingAgentRole>
</premis:linkingAgentIdentifier>
<premis:linkingAgentIdentifier>
  <premis:linkingAgentIdentifierType> local</premis:linkingAgentIdentifierType>
  <premis:linkingAgentIdentifierValue> Lasco</premis:linkingAgentIdentifierValue>
  <premis:linkingAgentRole> executing program</premis:linkingAgentRole>
</premis:linkingAgentIdentifier>
```

PREMIS Agent entity

Hardware/Organization/Person/Software

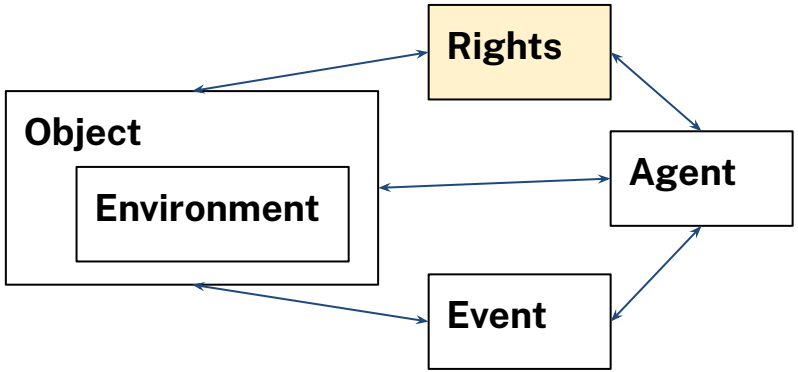
[Tip : use a controlled vocabulary]

```
<premis:agent>
  <premis:agentIdentifier>
    <premis:agentIdentifierType> local</premis:agentIdentifierType>
    <premis:agentIdentifierValue> Rochat, Rebecca</premis:agentIdentifierValue>
  </premis:agentIdentifier>
  <premis:agentName> Rebecca Rochat</premis:agentName>
  <premis:agentType> person</premis:agentType>
</premis:agent>
<premis:agent>
  <premis:agentIdentifier>
    <premis:agentIdentifierType> local</premis:agentIdentifierType>
    <premis:agentIdentifierValue> PCW-PZ-RENDU1</premis:agentIdentifierValue>
  </premis:agentIdentifier>
  <premis:agentName> PCW-PZ-RENDU1</premis:agentName>
  <premis:agentType> hardware</premis:agentType>
</premis:agent>
<premis:agent>
  <premis:agentIdentifier>
    <premis:agentIdentifierType> local</premis:agentIdentifierType>
    <premis:agentIdentifierValue> Agisoft Metashape</premis:agentIdentifierValue>
  </premis:agentIdentifier>
  <premis:agentName> Agisoft Metashape</premis:agentName>
  <premis:agentType> software</premis:agentType>
  <premis:agentVersion> 1.4.4</premis:agentVersion>
  <premis:agentNote> logiciel de numérisation 3D</premis:agentNote>
</premis:agent>
```



Agent when creating the 3D model with the software Agisoft Metashape

PREMIS Rights entity



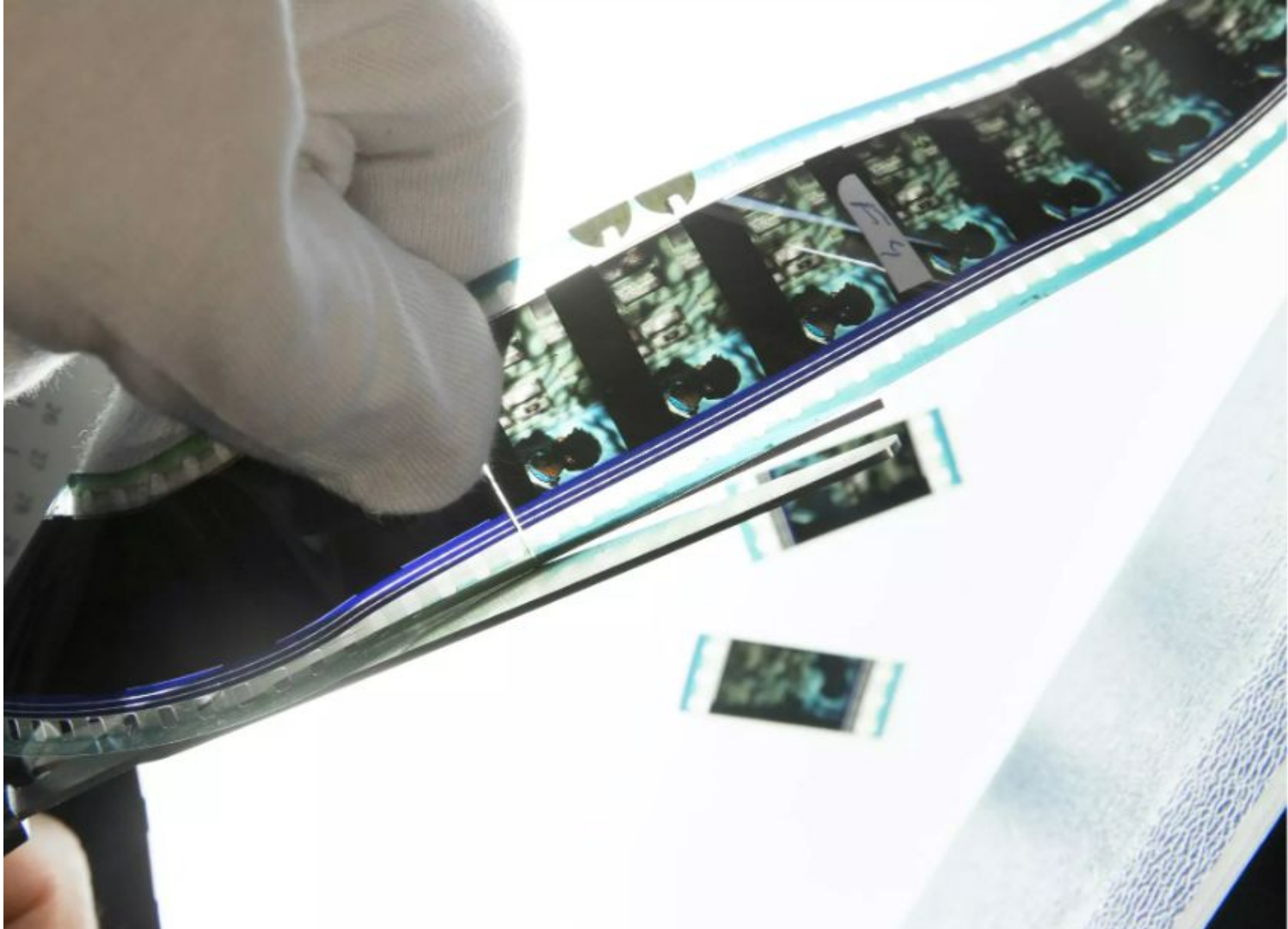
The Cinémathèque suisse decided not to fully use the PREMIS rights entity.

In the package, we will **add general right information** in the *rightBasis* section of PREMIS and a link to the database

```
rightBasis = copyright (from date), right holders + explanation on the rationale of the 3D model
```

Film

object



Organising and Preserving digital moving images objects with METS/PREMIS

Digital audiovisual files :

- Are in many ways similar to any other digital objects, with general applications of PREMIS possible (object's characteristics, events affecting it, agents acting upon it, etc.
- must be preserved digitally to be accessed over time

Furthermore, digital audiovisual files are :

- **Complex** (combining a video track, an audio track and possibly text tracks)
- **Time-based** and at risk of obsolescence
- Characteristics must be handled so that they can be understood by users now and in the future

PREMIS provides a framework for describing audiovisual files and events that occur during their lifecycle

ProRes

Example with a ProRes file:

Whatever the file format, metadata preservation scheme is the same, the change is in the metadata value

PREMIS - Implémentation du dictionnaire pour la CS	Obligatoire Dictionnaire	Modèle CS commun	Example flat video file	Définitions/explications	METS
Unités sémantiques : https://www.loc.gov/standards/premis/v3/premis-3-0-datadictionary-only.pdf					
Unités sémantiques (semantic units)					
1. Object					
1.1 objectIdentifier (M, R)	oui	oui	premis:file		techMD
1.1.1 objectIdentifierType (M, NR)	oui	oui	ARK	ARK	techMD
1.1.2 objectIdentifierValue (M, NR)	oui	oui	n° ARK	n° ARK	techMD
1.2 objectCategory (M, NR)	oui	oui	file	file	techMD
1.3 preservationLevel (O, R) [Intellectual Entity, Representation, File]		oui		niveau de préservation	techMD
1.3.1 preservationLevelType (O, NR) [Intellectual Entity, Representation, File]		oui	bit-level	type de niveau de préservation	techMD
1.3.2 preservationLevelValue (M, NR) [Intellectual Entity, Representation, File]	oui	oui	medium	*A value indicating the set of preservation functions expected to be applied to the object.* p. ex. low/medium/high	techMD
1.3.3 preservationLevelRole (O, NR) [Intellectual Entity, Representation, File]		oui	capability	liste contrôlée: capability/intention/requirement	techMD
1.3.4 preservationLevelRationale (O, R) [Intellectual Entity, Representation, File]		oui	on-site backup, 2 copies of file with checksums	*This optional semantic unit records the reason for applying the preservationLevelValue*	techMD
1.3.5 preservationLevelDateAssigned (O, NR) [Intellectual Entity, Representation, File]		oui	2022-12-16	date/heure	techMD
1.5 objectCharacteristics (M, R) [File, Bitstream]	oui	oui		*Technical properties of a file or bitstream that are applicable to all or most formats.*	techMD
1.5.2 fixity (O, R) [File, Bitstream]		oui		*Information used to verify whether an object has been altered in an undocumented or unauthorized way.*	techMD
1.5.2.1 messageDigestAlgorithm (M, NR) [File, Bitstream]	oui	oui	md5	algorithme MD5	techMD
1.5.2.2 messageDigest (M, NR) [File, Bitstream]	oui	oui	2157a8f94f24f49fd624139346e79a36	checksum md5	techMD
1.5.2.3 messageDigestOriginator (O, NR) [File, Bitstream]		oui	AND (labo) ou md5sum v1.2	*The Agent that created the original message digest that is compared in a fixity check* CS, labo,... cf. DCNC facility list, mais peut aussi être la version du logiciel	techMD
1.5.3 size (O, NR) [File, Bitstream]		oui	67432	size in octets	techMD
1.5.4 format (M, R) [File, Bitstream]	oui	oui		*identification of the format of a file or bitstream where format is defined as the organization of digital information according to preset specifications.*	techMD
1.5.4.1 formatDesignation (O, NR) [File, Bitstream]		oui		désignation du format	techMD
1.5.4.1.1 formatName (M, NR) [File, Bitstream]	oui	oui	quicktime	nom du format (conteneur ou codec?)	techMD
1.5.4.1.2 formatVersion (O, NR) [File, Bitstream]		oui		version du format	techMD
1.5.4.2 formatRegistry (O, NR) [File, Bitstream]		oui			techMD
1.5.4.2.1 formatRegistryName (M, NR) [File, Bitstream]	oui	oui	PRONOM	Registre de référence (PRONOM...)	techMD
1.5.4.2.2 formatRegistryKey (M, NR) [File, Bitstream]	oui	oui	x-fmt/384	identifiant du format dans cette liste de référence	techMD

PREMIS - Implémentation du dictionnaire pour la CS	Obligatoire Dictionnaire	Modèle CS commun	Example flat video file	Définitions/explications	METS
1.5.5 creatingApplication (O, R) [File, Bitstream]		oui			techMD
1.5.5.1 creatingApplicationName (O, NR) [File, Bitstream]		oui	source: mediaInfo/ writing application + writing library	Logiciels de création de fichiers AV: DaVinci Resolve; EasyDcp; Clipster; etc. MediaInfo métadonnée champ <Encoded_Application>	techMD
1.5.5.2 creatingApplicationVersion (O, NR) [File, Bitstream]		oui		n° de version	techMD
1.5.5.3 dateCreatedByApplication (O, NR) [File, Bitstream]		oui	2021-11-08	date de création de l'objet	techMD
1.5.6 inhibitors (O, R) [File, Bitstream]		oui		*Features of the object that inhibit access, use, or migration*	techMD
1.5.6.1 inhibitorType (M, NR) [File, Bitstream]	oui	oui	null	cryptage	techMD
1.5.6.2 inhibitorTarget (O, R) [File, Bitstream]		oui	null	*The content or function protected by the inhibitor.* Liste contrôlée des fonctions/contenus bloqués	techMD
1.5.6.3 inhibitorKey (O, NR) [File, Bitstream]		oui	null	clé de décryptage	techMD
1.5.7 objectCharacteristicsExtension (O, R) [File, Bitstream]		non			techMD
1.6 originalName (O, NR) [Intellectual Entity, Representation, File]		oui	Zahori_Prores4444_4K_Rec709_Gamma24_51Cine.mov	*The name of the object as submitted to or harvested by the repository, before any renaming by the repository.*	techMD
1.7 storage (O, R) [Representation, File, Bitstream]		oui		*For digital representations and files, the storage container should be repeated if there are two or more copies that are identical bit-wise and managed as a unit, except for the medium on which they are stored. To use this repetition, the copies must have a single objectIdentifier and be managed as a single object by the repository.*	techMD
1.7.1 contentLocation (O, NR) [Representation, File, Bitstream]		oui			techMD
1.7.1.1 contentLocationType (M, NR) [Representation, File, Bitstream]	oui	oui	?	*The means of referencing the location of the content.* P. ex. étagère, URI, etc.	techMD
1.7.1.2 contentLocationValue (M, NR) [Representation, File, Bitstream]	oui	oui	ofc_cineforom-tous_fichiers/Z/ZAHORI_FTR_QSA-XX_51_4K-S_25FPS_PR4444	référence de l'emplacement utilisée par le système de stockage, p. ex. n° de stock, chemin du HSM...	techMD
1.7.2 storageMedium (O, NR) [Representation, File, Bitstream]		oui	magnetic tape	support de stockage de l'objet	techMD
1.8 signatureInformation (O, R) [File, Bitstream]		oui			techMD
1.8.1 signature (O, R) [File, Bitstream]		oui			techMD
1.8.1.1 signatureEncoding (M, NR) [File, Bitstream]	oui	oui	Base64		techMD
1.8.1.2 signer (O, NR) [File, Bitstream]		oui	Cinémathèque suisse		techMD
1.8.1.3 signatureMethod (M, NR) [File, Bitstream]	oui	oui	DSA-SHA1		techMD
1.8.1.4 signatureValue (M, NR) [File, Bitstream]	oui	oui	valeur de la signature		techMD
1.8.1.5 signatureValidationRules (M, NR) [File, Bitstream]	oui	oui	règles de validation	*Operations to be performed in order to validate the digital signature*	techMD
1.8.1.6 signatureProperties (O, R) [File, Bitstream]		oui	2021-04-17	*Additional information about the generation of the signature*	techMD

PREMIS - Implémentation du dictionnaire pour la CS	Obligatoire Dictionnaire	Modèle CS commun	Example flat video file	Définitions/explications	METS
1.8.1.6 signatureProperties (O, R) [File, Bitstream]		oui	2021-04-17	*Additional information about the generation of the signature*	techMD
1.9 environmentFunction (O, R) [Intellectual Entity of type environment]		oui			techMD
1.9.1 environmentFunctionType (M, NR) [Intellectual Entity of type environment]	oui	oui			techMD
1.9.2 environmentFunctionLevel (M, NR) [Intellectual Entity of type environment]	oui	oui			techMD
1.10 environmentDesignation (O, R) [Intellectual Entity of type environment]		oui		*An identification of the environment used to render or execute an object*	techMD
1.10.1 environmentName (M, NR) [Intellectual Entity of type environment]	oui	oui	Windows/ VLC/QT player/DaVinci Resolve	*A commonly accepted name used to describe the environment*	techMD
1.10.2 environmentVersion (O, NR) [Intellectual Entity of type environment]		oui	version		techMD
1.10.3 environmentOrigin (O, NR) [Intellectual Entity of type environment]		oui	Microsoft		techMD
1.10.4 environmentDesignationNote (O, R) [Intellectual Entity of type environment]		oui			techMD
1.11 environmentRegistry (O, R) [Intellectual Entity of type environment]		oui			techMD
1.11.1 environmentRegistryName (M, NR) [Intellectual Entity of type environment]	oui	oui	PRONOM		techMD
1.11.2 environmentRegistryKey (M, NR) [Intellectual Entity of type environment]	oui	oui			techMD
1.13 relationship (O, R)		oui			techMD
1.13.1 relationshipType (M, NR)	oui	oui	derivation	cf. vocabulaire contrôlé	techMD
1.13.2 relationshipSubType (M, NR)	oui	oui	compressed from	cf. vocabulaire contrôlé	techMD
1.13.3 relatedObjectIdentifier (M, R)	oui	oui	file		techMD
1.13.3.1 relatedObjectIdentifierType (M, NR)	oui	oui	ARK		techMD
1.13.3.2 relatedObjectIdentifierValue (M, NR)	oui	oui	N° ARK		techMD
1.13.4 relatedEventIdentifier (O, R)		oui			techMD
1.13.4.1 relatedEventIdentifierType (M, NR)	oui	oui			techMD
1.13.4.2 relatedEventIdentifierValue (M, NR)	oui	oui			techMD
1.13.5 relatedEnvironmentPurpose (O, R) [Representation, File, Bitstream]		oui			techMD
1.13.6 relatedEnvironmentCharacteristic (O, NR) [Representation, File, Bitstream]		oui			techMD
1.14 linkingEventIdentifier (O, R)		oui			techMD
1.14.1 linkingEventIdentifierType (M, NR)	oui	oui			techMD
1.14.2 linkingEventIdentifierValue (M, NR)	oui	oui			techMD
1.15 linkingRightsStatementIdentifier (O, R)		oui			techMD
1.15.1 linkingRightsStatementIdentifierType (M, NR)	oui	oui			techMD
1.15.2 linkingRightsStatementIdentifierValue (M, NR)	oui	oui			techMD

2. Event					
2.1 eventIdentifier (M, NR)	oui	oui		*The eventIdentifier is likely to be system generated*	digiProvMD
2.1.1 eventIdentifierType (M, NR)	oui	oui	UUID	*A designation of the domain within which the Event identifier is unique.*	digiProvMD
2.1.2 eventIdentifierValue (M, NR)	oui	oui	c3537f7e-2a05-4178-90d2-7841e7e4d970	*Value of the eventIdentifier*	digiProvMD
2.2 eventType (M, NR)	oui	oui	ingest	liste contrôlée	digiProvMD
2.3 eventDateTime (M, NR)	oui	oui		date et heure (machine readable)	digiProvMD
2.4 eventDetailInformation (O, R)	oui	oui			digiProvMD
2.4.1 eventDetail (O, NR)		oui		*Additional information about the Event*	digiProvMD
2.5 eventOutcomeInformation (O, R)		oui		*Information about the outcome of an Event*	digiProvMD
2.5.1 eventOutcome (O, NR)		oui	p. ex. checksum validé/fichiers archivés/etc. définis par des codes	*A categorization of the overall result of the Event in terms of success, partial success, or failure. A coded way of representing the outcome of an Event may be useful for machine processing and reporting.* Liste contrôlée à définir à l'interne (codes p.ex.)	digiProvMD
2.5.2 eventOutcomeDetail (O, R)		oui		*A detailed description of the result or product of the Event*	digiProvMD
2.5.2.1 eventOutcomeDetailNote (O, NR)		oui		Note précisant 2.5.2	digiProvMD
2.6 linkingAgentIdentifier (O, R)		oui		*Identification of one or more Agents associated with the Event*	digiProvMD
2.6.1 linkingAgentIdentifierType (M, NR)	oui	oui	local	*A designation of the domain in which the linking Agent identifier is unique. Must be an existing agentIdentifierType value*	digiProvMD
2.6.2 linkingAgentIdentifierValue (M, NR)	oui	oui	Lasco	*Value of the linking Agent identifier* p. ex. un nom, un numéro, etc	digiProvMD
2.6.3 linkingAgentRole (O, R)		oui	executing program		digiProvMD
2.7 linkingObjectIdentifier (O, R)		oui		*Information about an Object associated with an Event. Digital provenance often requires that relationships between Objects and Events are documented.*	digiProvMD
2.7.1 linkingObjectIdentifierType (M, NR)	oui	oui	ARK	type d'id pour l'objet p.ex. ISAN	digiProvMD
2.7.2 linkingObjectIdentifierValue (M, NR)	oui	oui	n° ARK	valeur de l'id de l'objet p.ex. n° ISAN	digiProvMD

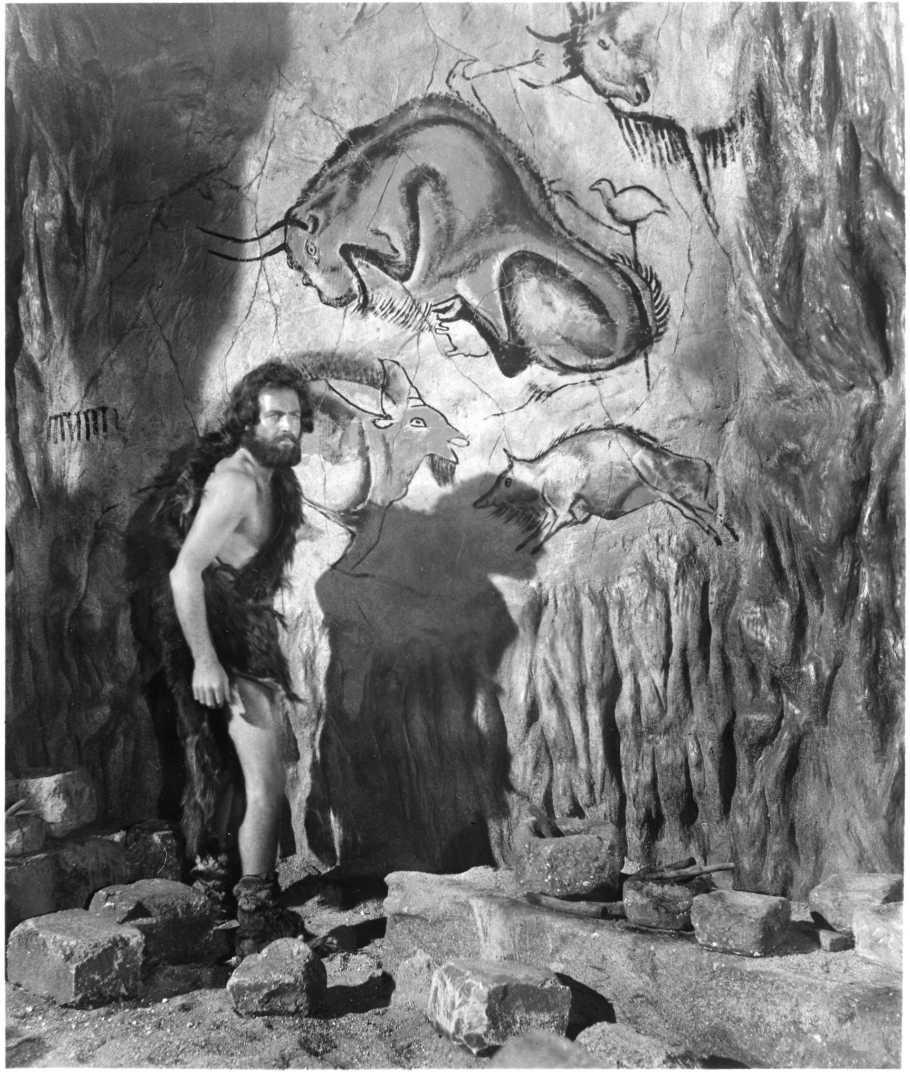
3. Agent					
3.1 agentIdentifier (M, R)	oui	oui		*Each Agent associated with the preservation repository must have a unique identifier to allow it to be related to Events and Rights statements.*	digiProvMD
3.1.1 agentIdentifierType (M, NR)	oui	oui	local	*A designation of the domain in which the Agent identifier is unique*	digiProvMD
3.1.2 agentIdentifierValue (M, NR)	oui	oui	autorité	liste contrôlée LOC pour les organisations	digiProvMD
3.2 agentName (O, R)		oui	nom	nom	digiProvMD
3.3 agentType (O, NR)		oui	personne physique	hardware/software/personne physique/personne morale	digiProvMD
3.4 agentVersion (O, NR)		oui		*The version of the Agent referenced in agentName, if agentType is software or hardware.*	digiProvMD
3.7 linkingEventIdentifier (O, R)		oui			digiProvMD
3.7.1 linkingEventIdentifierType (M, NR)	oui	oui	UUID	type	digiProvMD
3.7.2 linkingEventIdentifierValue (M, NR)	oui	oui	c3537f7e-2a05-4178-90d2-7841e7e4d970	valeur	digiProvMD
3.8 linkingRightsStatementIdentifier (O, R)		oui		*An identifier for a Rights statement associated with the Agent*	digiProvMD
3.8.1 linkingRightsStatementIdentifierType (M, NR)	oui	oui		domaine	digiProvMD
3.8.2 linkingRightsStatementIdentifierValue (M, NR)	oui	oui		valeur	digiProvMD
3.9 linkingEnvironmentIdentifier (O, R)		oui		*[...]it links an Agent to an Object. LinkingEnvironmentIdentifierType and linkingEnvironmentIdentifierValue must therefore match objectIdentifierType and objectIdentifierValue in the related environment Object*	digiProvMD
3.9.1 linkingEnvironmentIdentifierType (M, NR)	oui	oui	ARK	*A designation of the domain within which the linkingEnvironmentIdentifier is unique.*	digiProvMD
3.9.2 linkingEnvironmentIdentifierValue (M, NR)	oui	oui	n° ARK	valeur	digiProvMD
3.9.3 linkingEnvironmentRole (O, R)		oui		*The role of the environment Object associated with this Agent.* *A piece of software or hardware that is captured as an Agent can also* *be preserved in a repository and described as an environment Object.*	digiProvMD

PREMIS - Implémentation du dictionnaire pour la CS	Obligatoire Dictionnaire	Modèle CS commun	Example flat video file	Définitions/explications	Exemple	METS
4. Rights						
4.1 rightsStatement (O, R)		oui				rightsMD
4.1.1 rightsStatementIdentifier (M, NR)	oui	oui				rightsMD
4.1.1.1 rightsStatementIdentifierType (M, NR)	oui	oui		*A designation of the domain within which the Rights statement identifier is unique*		rightsMD
4.1.1.2 rightsStatementIdentifierValue (M, NR)	oui	oui		valeur		rightsMD
4.1.2 rightsBasis (M, NR) A PRENDRE POUR CHAMPS GENERAL RENVOI	oui	oui				rightsMD


IT'S MORE WHAT YOU'D CALL
GUIDELINES
THAN ACTUAL RULES

LASCO

Logiciel d'Archivage, de Sauvegarde et de Conservation des Œuvres



Lasco

Aims:

- Internally developed software to process data packages and implement the OAIS model for long term archival
- Could also help during data migration (e.g. LTO6->LTO9)

Practically :

- A tailor-made tool box to automate the actions required to archive digital files and implement good practices (e.g. OAIS model).

Highly inspired by Archivematica

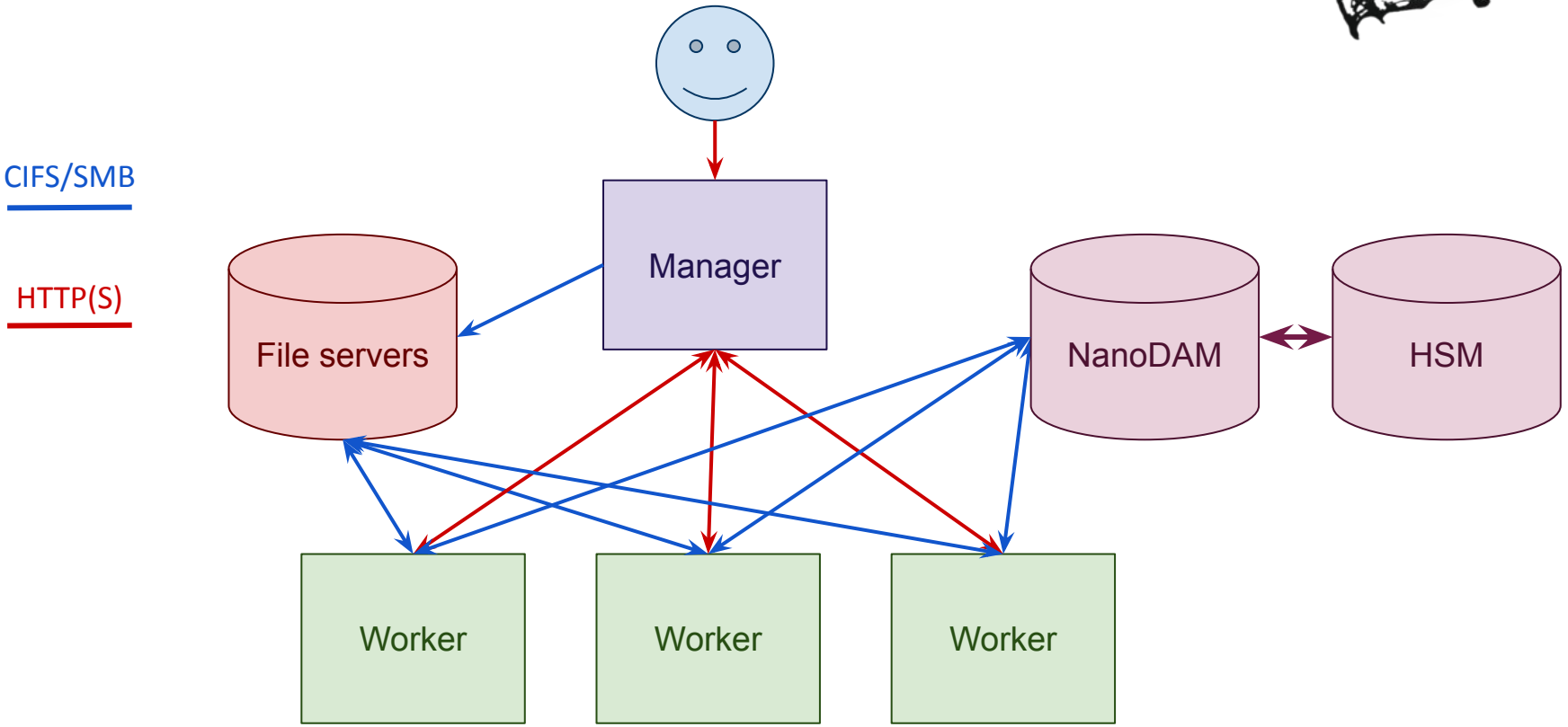
“Not Invented Here” syndrome, partly justified by some constraints

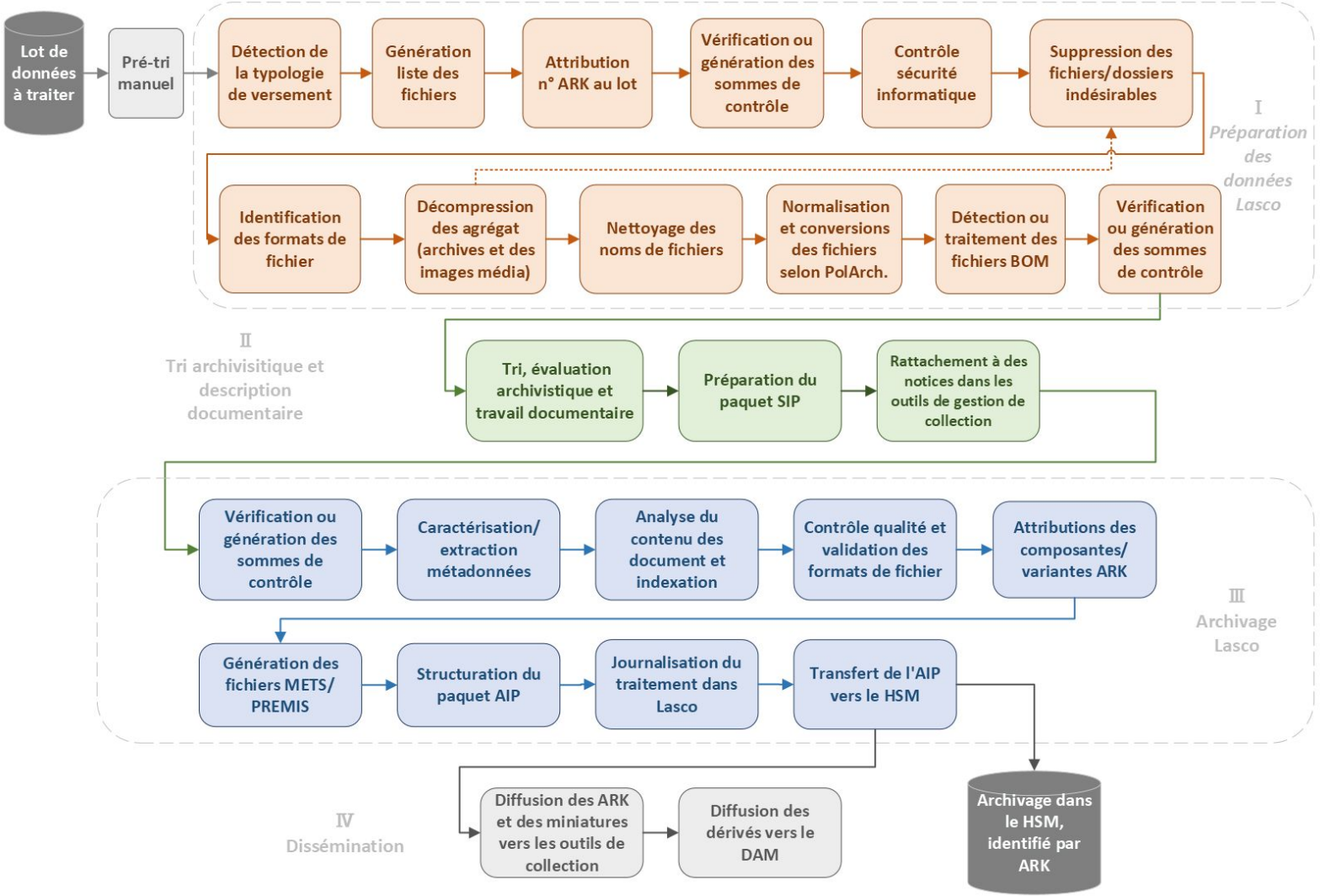
Only one developer (four days a week)

Project history (so far...)

Summer 2020	**Evaluate/analyse the current situation, adapt OAIS to the institution**
Autumn 2020	Basic specifications to evaluation existing software solutions
Autumn 2020	**Testing with Archivematica**
January 2021	**IT department decides to develop the solution internally**
March-July 2021	Prototype
Summer 2021	Development stopped, lack of precise specifications Technical committee → to get all departments on the same page
April 2022	**Main developer quits. Project put on hold**
Summer/Automn 2022	Workshops to write specification for a v0.5
December 2022	Development starts again with a new developer
July 2023	**First go live (v0.5), lacks most functionalities related to OAIS**

Simplified architecture




Lot de données à traiter
Pré-tri manuel
Détection de la typologie de versement
Génération liste des fichiers
Attribution n° ARK au lot
Vérification ou génération des sommes de contrôle
Contrôle sécurité informatique
Suppression des fichiers/dossiers indésirables
I
Préparation des données Lasco
Identification des formats de fichier
Décompression des agrégat (archives et des images média)
Nettoyage des noms de fichiers
Normalisation et conversions des fichiers selon PolArch.
Détection ou traitement des fichiers BOM
Vérification ou génération des sommes de contrôle
II
Tri archivisitique et description documentaire
Tri, évaluation archivistique et travail documentaire
Préparation du paquet SIP
Rattachement à des notices dans les outils de gestion de collection
Vérification ou génération des sommes de contrôle
Caractérisation/ extraction métadonnées
Analyse du contenu des document et indexation
Contrôle qualité et validation des formats de fichier
Attributions des composantes/ variantes ARK
III
Archivage Lasco
Génération des fichiers METS/ PREMIS
Structuration du paquet AIP
Journalisation du traitement dans Lasco
Transfert de l'AIP vers le HSM
IV
Dissémination
Diffusion des ARK et des miniatures vers les outils de collection
Diffusion des dérivés vers le DAM
Archivage dans le HSM, identifié par ARK

```
<?xml version='1.0' encoding='UTF-8'?>
<mets xmlns="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:premis="http://www.loc.gov/premis/v3" xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/version18/mets.xsd" OBJID="ark:/21891/pch2tqb3kh97gh8wg">
  <metsHdr CREATEDATE="2022-03-21T14:56:26" LASTMODDATE="2022-03-30T09:32:26" />
  <dmdSec ID="DMD.1">
      <mdWrap MIMETYPE="text/xml" MDTYPE="DC" LABEL="Dublin Core Metadata">
          <xmlData>
              <dc:title>Frauennot - Fotografien I, Fotografie 11</dc:title>
              <dc:creator>Praesens-Film AG</dc:creator>
              <dc:description>9 x 14 cm, Merida, Yucotan, weitere nicht entzifferbare Beschriftung</dc:description>
              <dc:date>Ohne Datum</dc:date>
              <dc:format>1 Umschlag</dc:format>
              <dc:identifier>https://sv-atom-pz/frauennot-fotografien-i-fotografie-11</dc:identifier>
              <dc:identifier>11</dc:identifier>
          </xmlData>
      </mdWrap>
  </dmdSec>
  <amdSec ID="AMD.1">
          <techMD ID="TECH1.1">
              <mdWrap MDTYPE="PREMIS:OBJECT">
                  <xmlData>
                      <premis:object xsi:type="premis:file" version="3.0">
                          <premis:objectIdentifier>
                              <premis:objectIdentifierType>ARK</premis:objectIdentifierType>
                              <premis:objectIdentifierValue>ark:/21891/pch2tqb3kh97gh8wg/f1r/mdc.master
                              </premis:objectIdentifierValue>
                          </premis:objectIdentifier>
```

COMPLETE EXAMPLE

Project roadmap

November 2023	Release of v0.6
March 2024	Release of v0.7
July 2024	**Release of v0.8**
	Lasco generates METS/PREMIS files
November 2024	Release of v0.9
March 2025	Release of v1.0

As any development project, uncertainty on the roadmap.

Uncertainty also on the LTO6->LTO9 migration

Conclusions

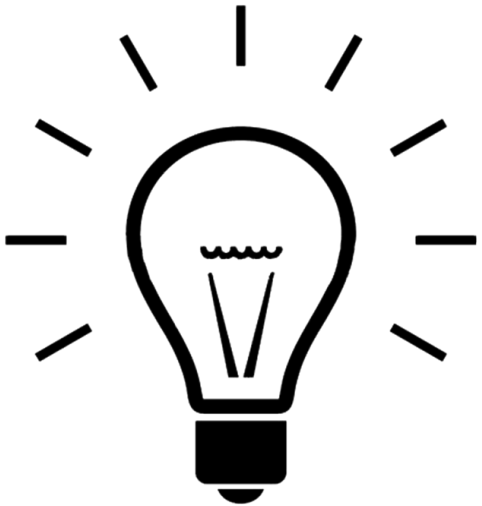
Know and understand your workflows and the preservation metadata they produced (hardware, software)

- Retrieve automatically for PREMIS with mapping

Test your datasets with Archivematica

No over-selection in the data dictionary, keep what you really need

Contact the PREMIS Committee, highly informative and very helpful



The Lord of The Rings: The Fellowship of the Ring, Peter Jackson, 2001. © 2001 New Line Cinema Productions, Inc. All Rights Reserved. Collection Cinémathèque suisse.

References

PREMIS Editorial Committee, & The Digital Preservation Coalition (DPC). (2023, September 12). PREMIS Webinars to introduce the Spanish translation of the "Understanding PREMIS: An Overview of the PREMIS Data Dictionary for Preservation Metadata". Zenodo. https://doi.org/10.5281/zenodo.8337861

What is Digital Preservation. Caplan, Priscilla, 2005
https://journals.ala.org/index.php/ltr/article/view/4224/4809/

AV Cataloging Standards Series: Introduction to PREMIS. AMIA 14 July 2021
https://www.iasa-web.org/event/av-cataloging-standards-series-introduction-premis

References

VERMAATEN, Sally. A Checklist and a Case for Documenting PREMIS-METS Decisions in a METS Profile. D-Lib Magazine, 2010, vol. 16, no 9, p. 2. - https://www.dlib.org/dlib/september10/vermaaten/09vermaaten.html

Community Owned digital Preservation Tool Registry (COPTR), https://coptr.digipres.org/

ClairMeta, Python package for Digital Cinema Package (DCP) probing and checking, https://github.com/Ymagis/ClairMeta