

**Welcome to this live webinar on** Validating with meaning: syntactic and semantic validation of the E-ARK SIP

Start 10:00

21 November 2024

### Audience notes for the Live Webinar



Your **cameras have been turned off** and **microphones muted**.



If you have any technical issues during the event, please use the chat function.



Please **use the Q&A for questions to speakers**. These will be addressed at the end of the event.



**Please note that this webinar is recorded**. No attendee personal information will be captured in these recordings.

# Agenda

---

10:00 – 10:05

## **eArchiving Initiative welcome**

Jaime Kaminski, eArchiving Initiative

10:05 – 10:50

## **Validating with meaning: syntactic and semantic validation of the E-ARK SIP**

Milan Valadou, meemoo

10:50 – 11:00

## **Q&A**



# Validating with meaning: syntactic and semantic validation of the E-ARK SIP

Milan Valadou, meemoo

eArchiving Initiative Training Webinar

# Today's (story) arc

- Who is meemoo?
- A bit of context
- Diving into the 'meemoo SIP'
- Ingesting 'meemoo SIPs'
- Way forward





# meemoo's service provision

- **Digitisation, digital archiving and management** of archival materials
- Make content **accessible and usable**
- Actively **gather and share expertise** on digital archive operations
- **Advise** on digital heritage processes

# Collaboration

182 content partners (CPs) in culture,  
media and government

performing arts

museums

archives

heritage societies

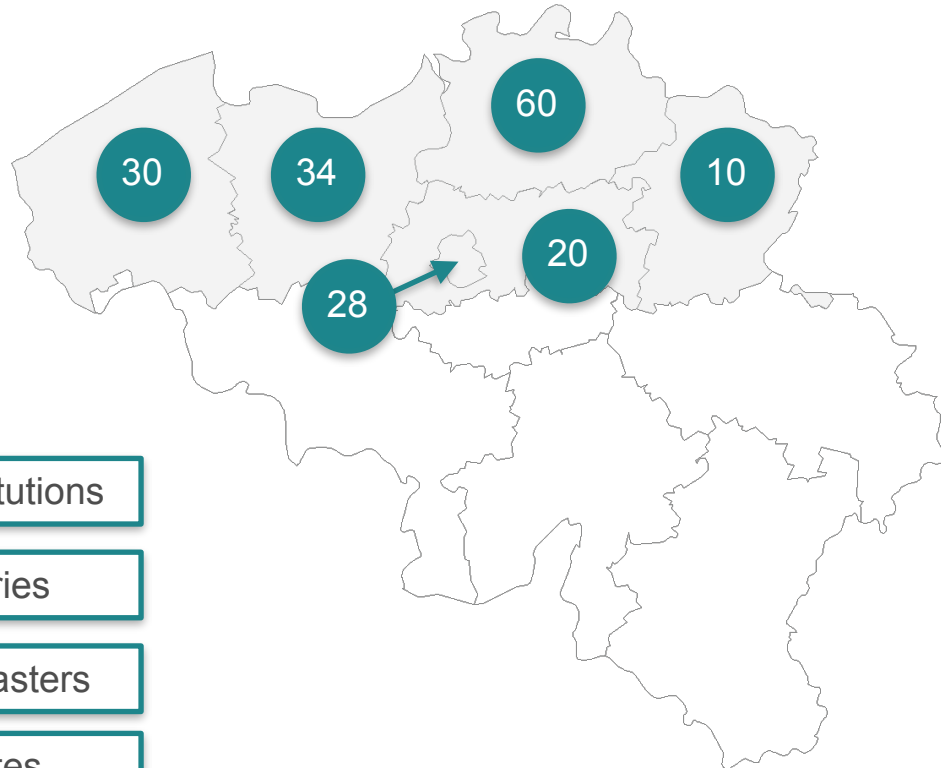
government institutions

heritage libraries

national broadcasters

sector institutes

regional broadcasters



# In figures



> 8 million

objects in our archive system



> 540,000

AV carriers transferred to our archive system



~ 185,000

user accounts at the Archive for Education at end of 2022-2023 academic year

*Interaction*

The Archive for  
Education

hetarchief.be

News of the  
Great War

Art in Flanders

Internal tools

Applications of  
CPs

*Integration*

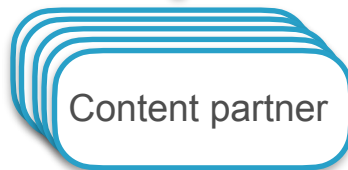
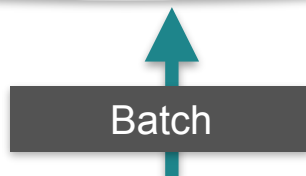
SEARCH

REST API

GRAPHQL

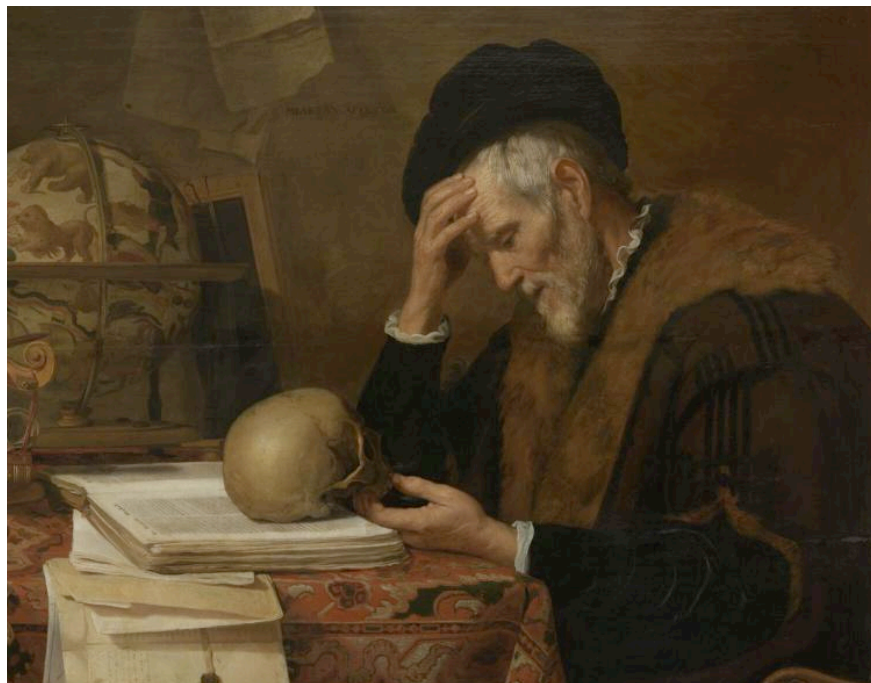
OAI-PMH

*Archive + metadata  
management*

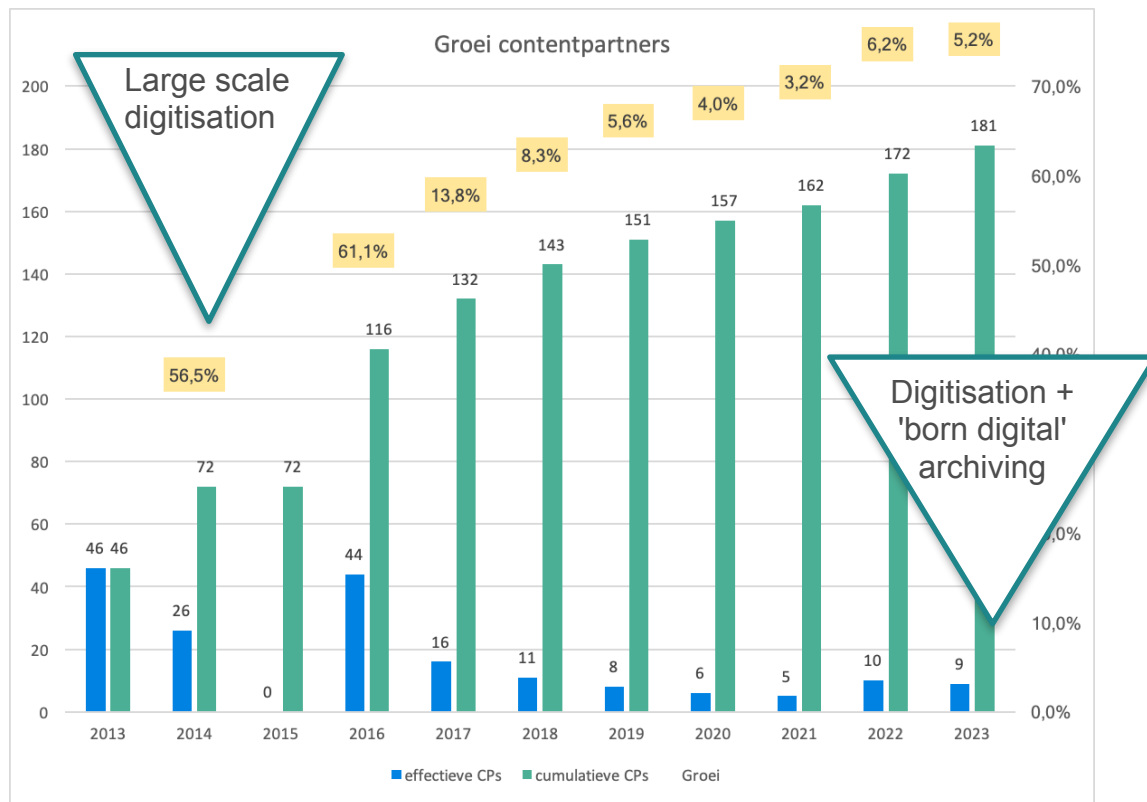


# A bit of context

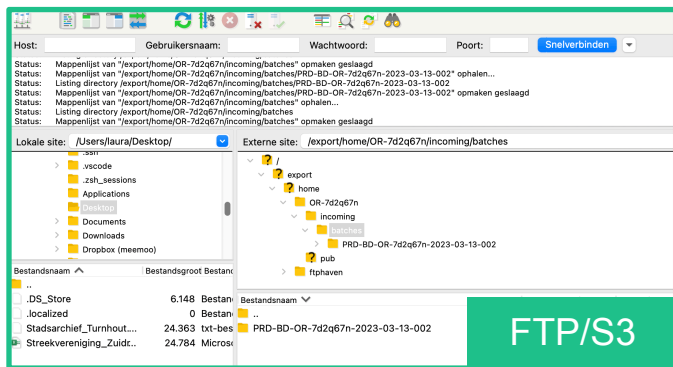
how we came to the realisation that  
we needed a SIP



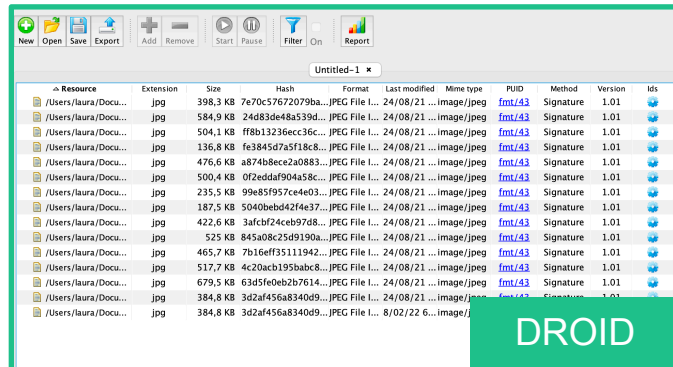
# Rise in number of partners



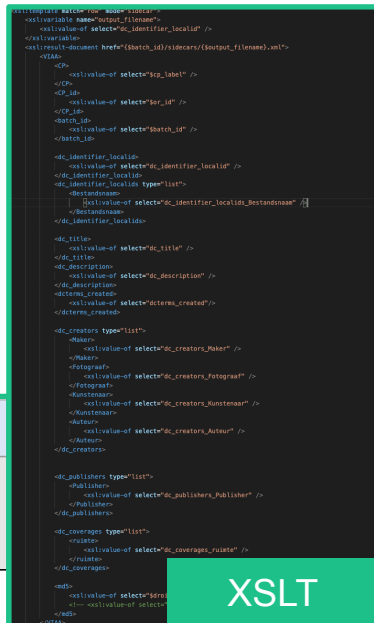
# Influx of digital collections



FTP/S3



DROID

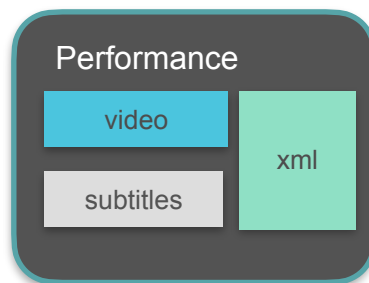
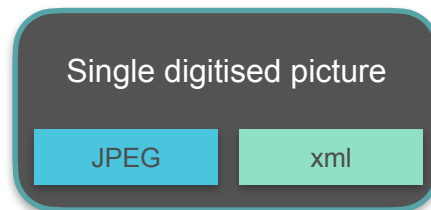
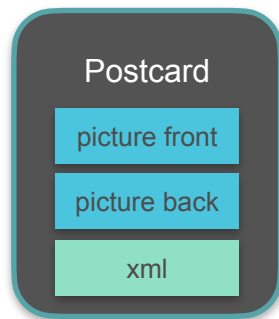


XSLT

dc:identifier_localid	dc:title	dc:description	dcterms_created	dc:identifier_locals_B estandsnaam	dc:titles_reeks	dc:titles_seizoen	dcterms_issued	dc:creators_Acteur
<p>Hoofd lokale CP ID = bestandsnaam zonder extensie</p> <p>Offtherecord_CP_1</p>	<p>Hoofdtitel</p> <p>Off the record</p>	<p>Hoofdbeschrijving</p> <p>Vier jongeren tussen 14 en 18 jaar tonen zich zoals ze zijn. Fei, kwetsbaar, schaamteloos en vol lef. Het podium is een vrijaalbaar waar ze letterlijk en figuurlijk spelen met de vele versies van zichzelf. Als hoofdrolspelers in hun eigen documentaire. Ze zijn de stemmen van het heden, de volwassenen van de toekomst.</p> <p>Off the record is een theaterexperiment van vier jongeren die de touwtjes zelf in handen nemen. Ze vertellen hoe zij hun leven zien en hoe ze denken (niet) te passen in de parameters van de huidige samenleving. Zo maken ze een voorstelling die uit hun hoofden, lijven en harten komt. De productie wordt</p>	<p>Createdatum van het object, EDTF level 0 of 1. Zie developer.meemoo.be (datatypes)</p> <p>2022-02-28</p>	<p>Originele bestandsnaam van het object = hoofd lokale CP ID + extensie</p> <p>Offtherecord_CP_1.jpg</p>	<p>Naam van de voorstelling of productie</p> <p>Off the record</p>	<p>Het seizoen van de voorstelling</p> <p>2021 - 2022; 2022 - 2023</p>	<p>Datum waarop het item is uitgegeven of verzonden. EDTF level 0 of 1. Zie developer.meemoo.be (datatypes)</p>	<p>De acteur in het object in kwestie</p> <p>Ben Herlaar Masina;Ziba Jafari;Mauro Michielsen;Mette Plysier</p>

Descriptive metadata templates

# Increasingly diverse and complex object structures

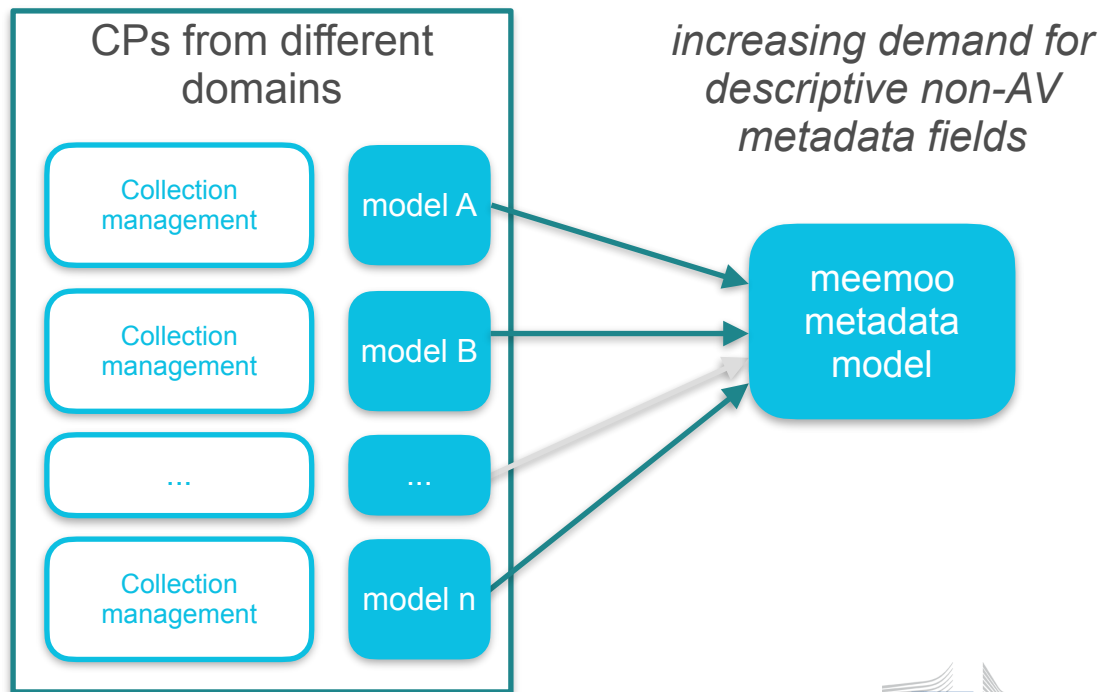


heterogeneous...  
file types  
structure  
object types  
metadata

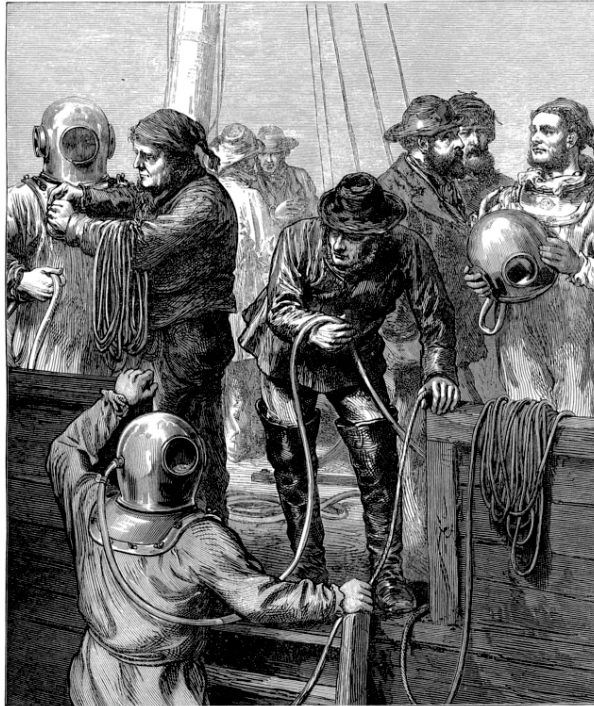


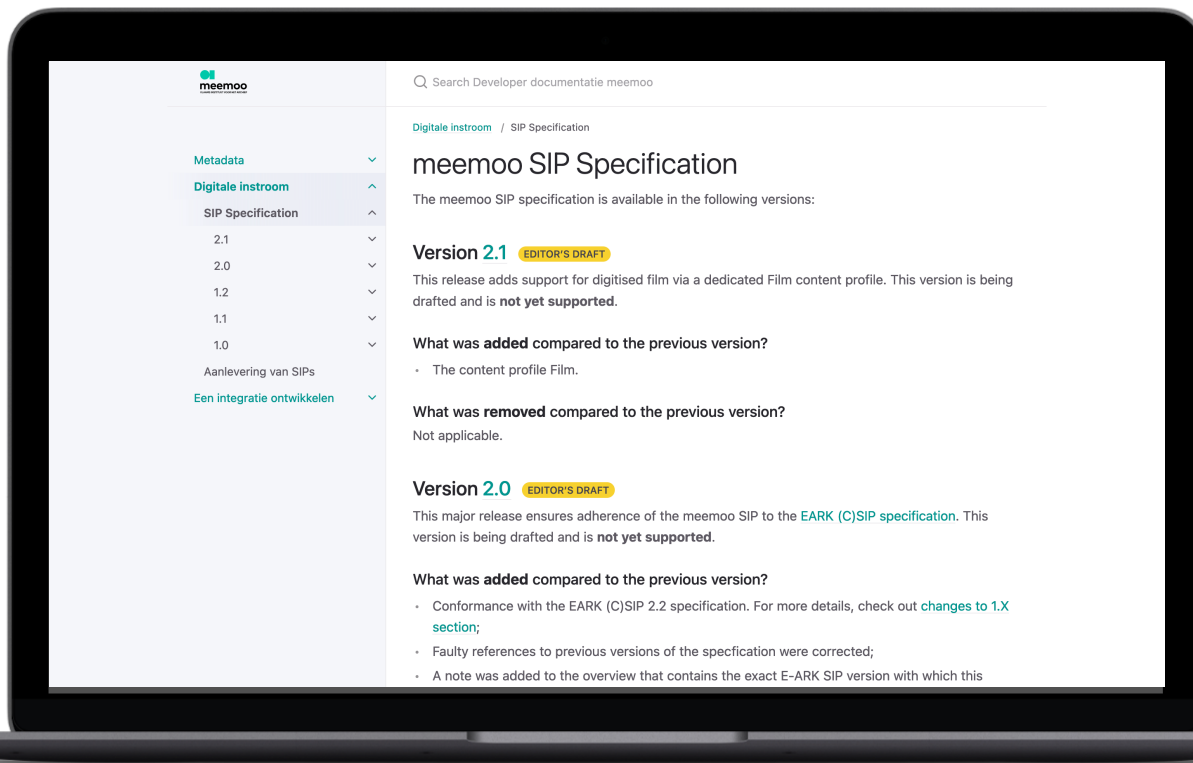
# More heterogeneity in metadata

*diverse metadata intake*



# A (deep) dive into the 'meemoo SIP'





# Rooted in the E-ARK Initiative...

E-ARK offered a new uniform SIP design that:

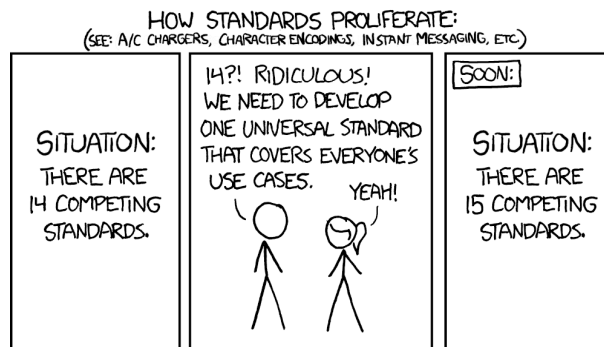
has a **formal and mature specification**

is embedded in **standards**

is technologically aligned with the **media and cultural heritage** sector

is **flexible and extensible**

is rooted in **collaboration** and **community support**



# ... with a layer of (not so) secret meemoo sauce

## Low number of tech-savvy partners:

- flexibility can lead to complexity

- needed more rules regarding object structure and descriptive metadata

- ~ The E-ARK specification, but with **extra requirements and padding**

  - 'terminology' + 'core concepts' + 'use cases' sections

  - narrow definition of the hierarchical structure

  - requiring the use of PREMIS XML

  - conceived the concept of 'content profiles'

# Content profiles

	1 Intellectual Entity 1 Representation # Files	# Intellectual entities # Representations # Files	1 Intellectual entity # Representations # Files
Limited metadata (DC + schema.org)	Basic		
Bibliographic metadata (MODS)		Bibliographic	
High-level metadata (schema.org)			Material artwork
			Film

meemoo Jira Jouw werk Projecten Filters Dashboards Teams Plannen Apps Aanmaken Zoeken 9+ 2 0 MP

Issues Apps Delen Issues exporteren LIJSTWEERGAVE DETAILWEERGAVE 1 van 31

project = "META" AND component = "SIP" Opnieuw instellen Filter opslaan BASIS JQL

Ordenen op 17

SIP 2.0: Closing the EARK gap  
META-274 MP

Voeg lijst van rollen toe voor 'creator', 'contributor' en 'publisher' in de SIP spec  
META-265

Updaten en verhelderen van de "Recommended Formats Statement" in de SIP-spec  
META-257 MS

Verwijder 'MEEMOO-PID' als toegelaten 'premis:objectIdentifierType' op representation-level  
META-249

Migratie van SIP samples en voorbeelden  
META-229

Analyse SIP audiocombo  
META-223 MS

Fix @csip: (OTHER)CONTENTINFOMATIONTYPE  
META-215 MS

UUID verduidelijken in docs  
META-206

Mogelijkheid tot linken naar of queryen van lijst van metadavelden per profiel  
META-201

SIP samples uitbreiden met toegelaten local ID's  
1-31 van 31

META-274  
SIP 2.0: Closing the EARK gap

Bijvoegen Onderliggende issue toevoegen Link issue Aanmaken Link goals

Beschrijving  
We willen EARK compliant zijn en deze EPIC gaat het doen.

Subtaken 12% klaar

META-211	uitklaren mets/@OBJID en data root folder	MS	WORDT BEOORDEELD
META-215	Fix @csip:(OTHER)CONTENTINFOMATIONTYPE	MS	GEREED
META-275	[CSIPSTR4/CSIPSTR12] De METS filenames zijn case sensitive	MS	SELECTED FOR DEVELOPMENT
META-276	[CSIP0] Attribueet 'OAISPACKAGETYPE' moet aanwezig zijn in 'metsHdr' van de representation METS files	MS	SELECTED FOR DEVELOPMENT
META-277	[CSIP2] Content category is een vaste lijst met waardes	MS	SELECTED FOR DEVELOPMENT
META-278	[CSIP107] Het attribueet '@LABEL' van de structmap div in de representation METS moet naar de folder wijzen waar de f...	MS	SELECTED FOR DEVELOPMENT
META-279	[CSIP17/24/27/29/31/38/41/43/66/110] De commons-ip validator heeft het lastig met 'f' in relatieve paden	MS	BACKLOG
META-280	Duid in de meemoo SIP spec aan met welke E-ARK SIP versie die compliant is	MS	SELECTED FOR DEVELOPMENT

Confluence content  
noemde op

2024-04-22 call-up TEAM EXPERTISE	Bijgewerkt op 25 apr 2024	RV
2024-04-04 meeting TEAM EXPERTISE	Bijgewerkt op 4 apr 2024	AS

Voeg een opmerking toe...

Pro-tip: druk op IM om een opmerking te plaatsen

Details

Uitvoerder  
MP Mattias Poppe

Melder  
MS Miel Vander Sande

Ontwikkeling  
1 branch  
2 commits 17 dagen geleden  
1 pull-aanvraag MERGED  
1 build

Releases  
Ongedefinieerde o...

Componenten  
SIP

Epic Name  
eark-gap-close

Start date  
28 mrt 2024

Content Partner  
Geen

Prioriteit  
Highest

Meer velden Herstelversie...

via code / documentation

Q Type to search

<> Code Issues Pull requests 2 Actions Security Insights Settings

Filters is:pr is:closed SIP

Labels 9 Milestones 0 New pull request

Clear current search query, filters, and sorts

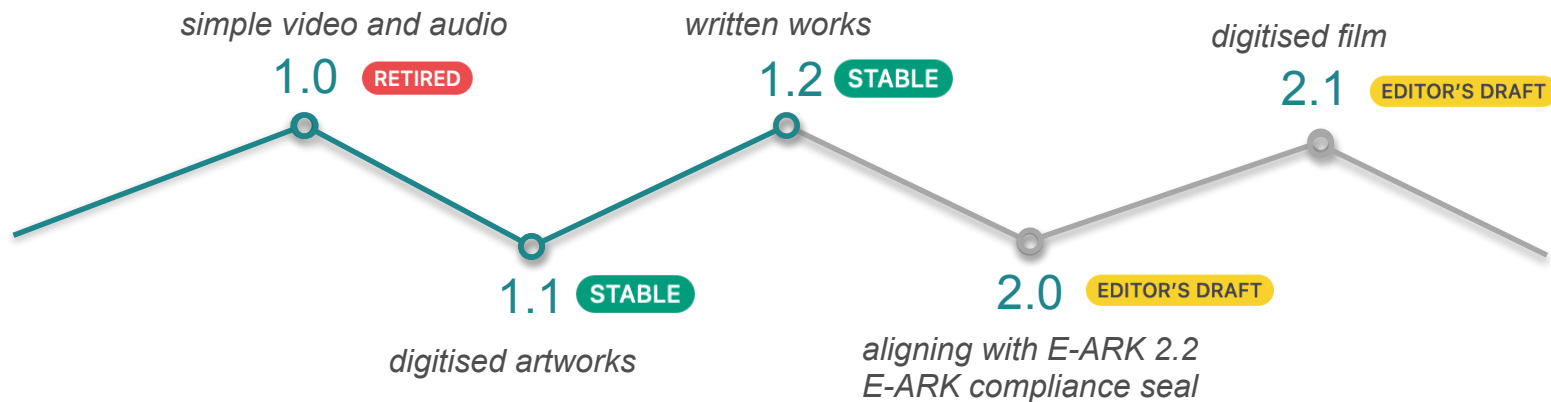
<input type="checkbox"/>	1 Open ✓ 26 Closed	Author	Label	Projects	Milestones	Reviews	Assignee	Sort
<input type="checkbox"/>	<b>Meta 274 sip 2 0 closing the eark gap</b> #130 by mielvds was merged 2 weeks ago							55
<input type="checkbox"/>	<b>META-266 - Fix (C)SIP namespace</b> #125 by spacid was merged on Mar 4 • Approved							25
<input type="checkbox"/>	<b>DEV-2286: External documentation for S3</b> ✓ #119 by maartends was merged on Jan 29 • Approved							12
<input type="checkbox"/>	<b>Meta 171 sip versie 1 2</b> #110 by mielvds was merged on Nov 2, 2023							
<input type="checkbox"/>	<b>Meta 230 aanvullen sip voor metadata give 3 d</b> #100 by mielvds was merged on Sep 20, 2023							
<input type="checkbox"/>	<b>SIP spec: art form en art medium moeten van type lang string zijn en niet string</b> #91 by mielvds was merged on Jun 13, 2023 • Approved							
<input type="checkbox"/>	<b>META-204 Remake bags</b> #72 by spacid was merged on Apr 6, 2023 • Approved							
<input type="checkbox"/>	<b>Add unique language constraint to spec</b> #65 by mielvds was merged on Mar 27, 2023 • Approved							17
<input type="checkbox"/>	<b>META-153 Fixes contractions obligation and cardinality Representation PREMIS</b> #60 by nvanderperren was merged on Feb 2, 2023 • Approved							1
<input type="checkbox"/>	<b>META-141 Fixes structMap requirements and examples</b> #59 by mielvds was merged on Feb 2, 2023 • Approved							29
<input type="checkbox"/>	<b>Meta 92 review samples en use cases</b>							12



# Support level through publication statuses

EDITOR'S DRAFT	This version of the specification is still being developed.
RELEASE CANDIDATE	This version of the specification is ready for <i>public review</i> and <i>implementation</i> . The text can still undergo changes and meemoo does not give any guarantees that material can be supplied in this way.
STABLE	This version of the specification will no longer change and has been fully implemented. Meemoo guarantees that material supplied in this way can be archived.
DEPRECATED	This version of the specification is outdated and will no longer be supported in the future. Its use is not recommended by meemoo.
RETIRED	This version of the specification is no longer supported by meemoo. Material can no longer be supplied in this way.

# Versioning



# Towards a governance model

Release cycle to avoid *ad hoc* alterations

Connect with partners, gather input and cover their needs

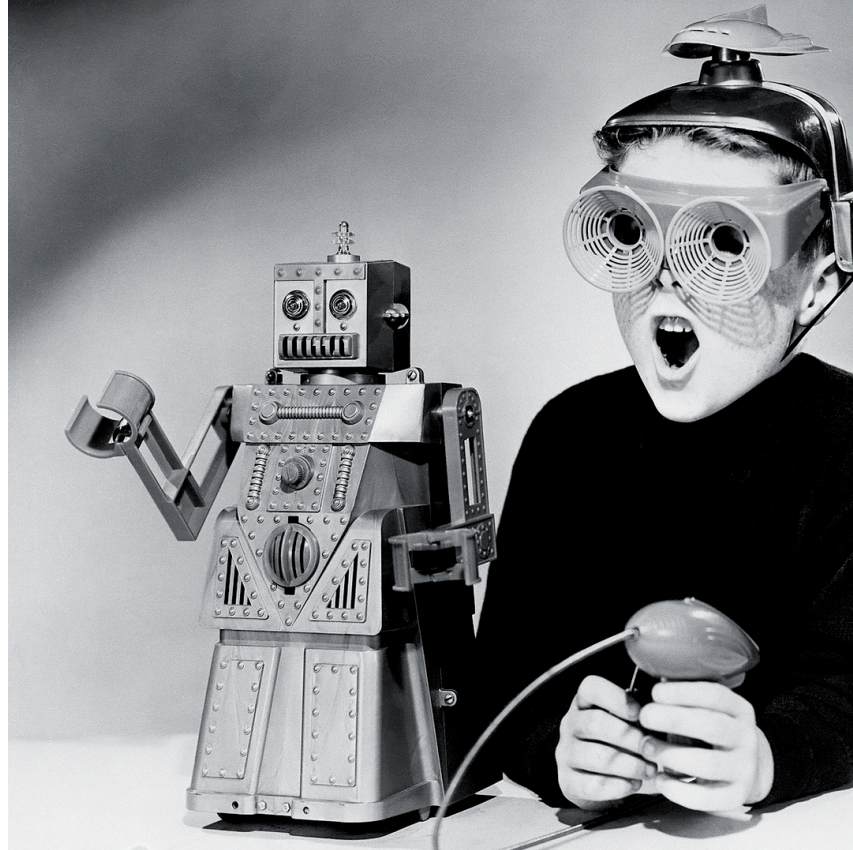
Predictable support for both sides

Ongoing work

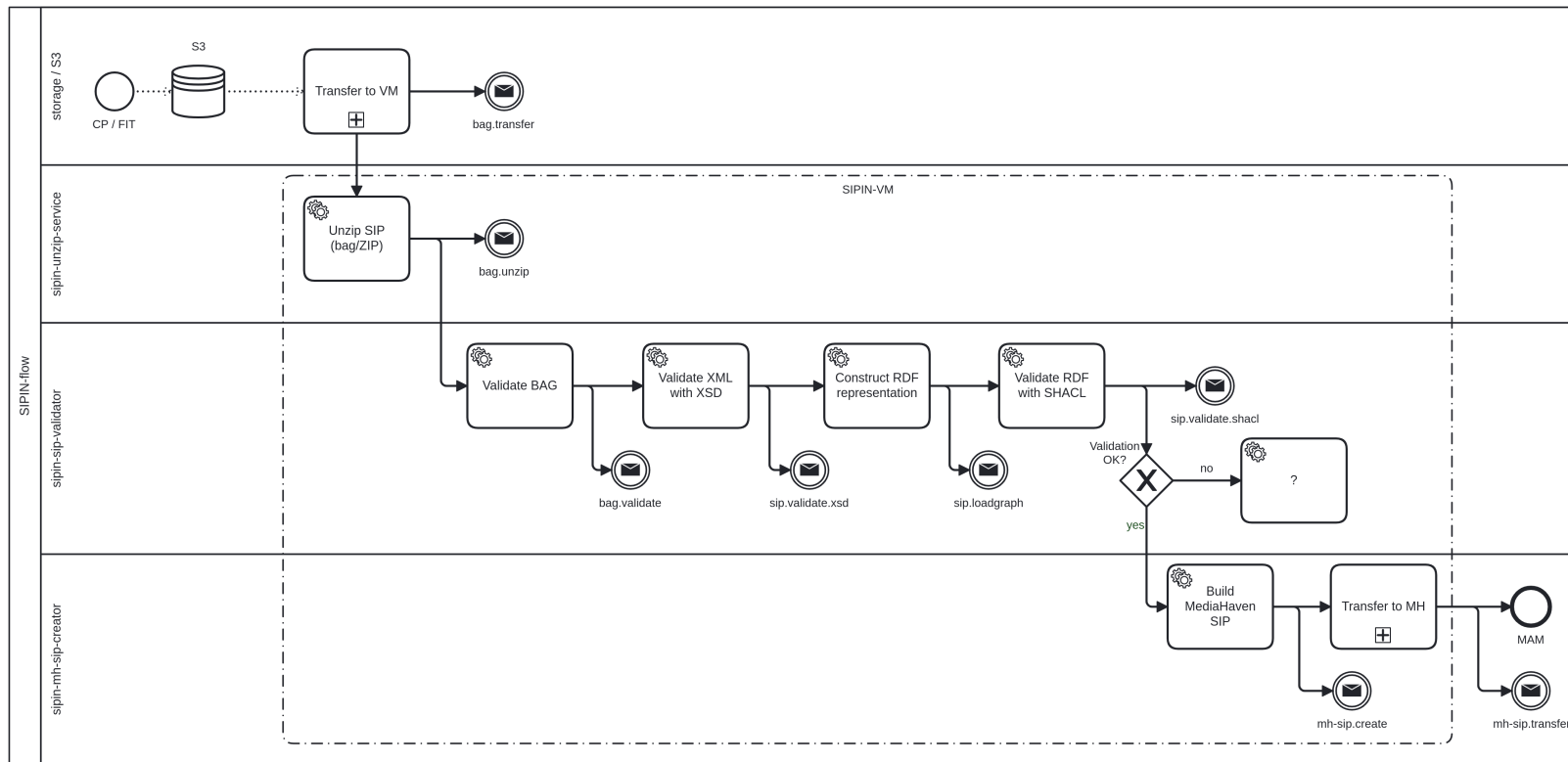


# Ingesting 'meemoo SIPs'

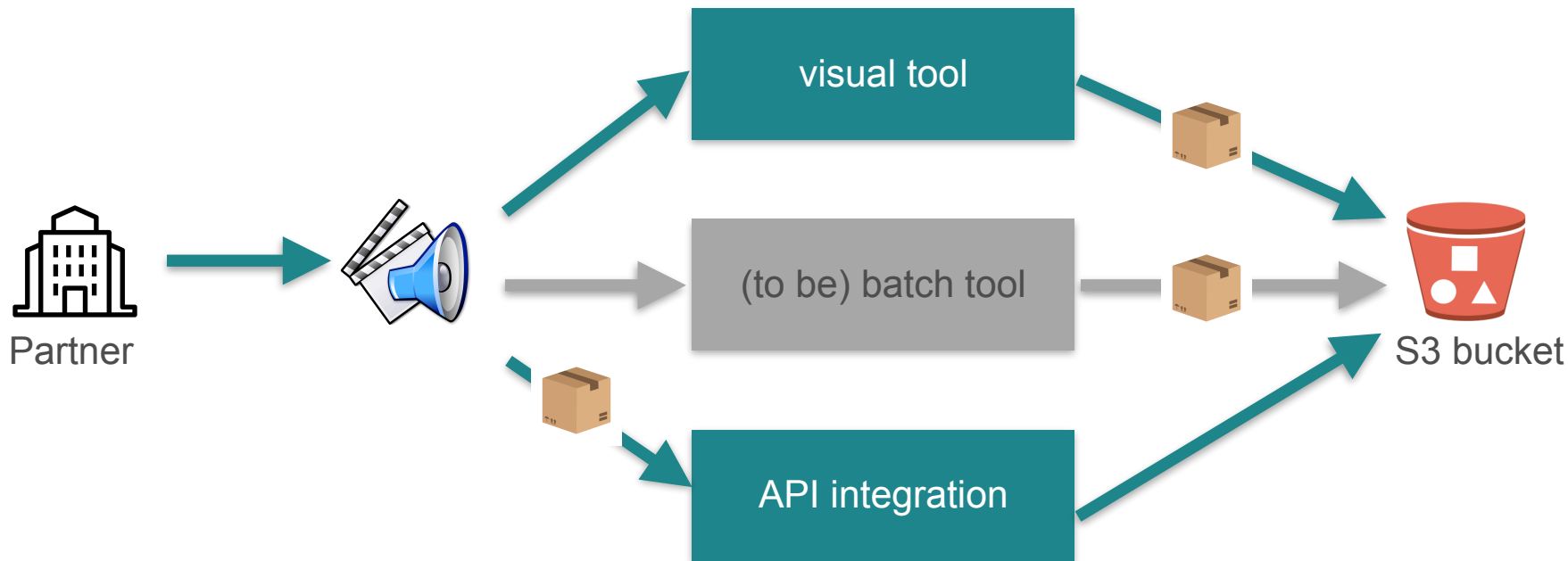
How we transform it to a graph for semantic validation



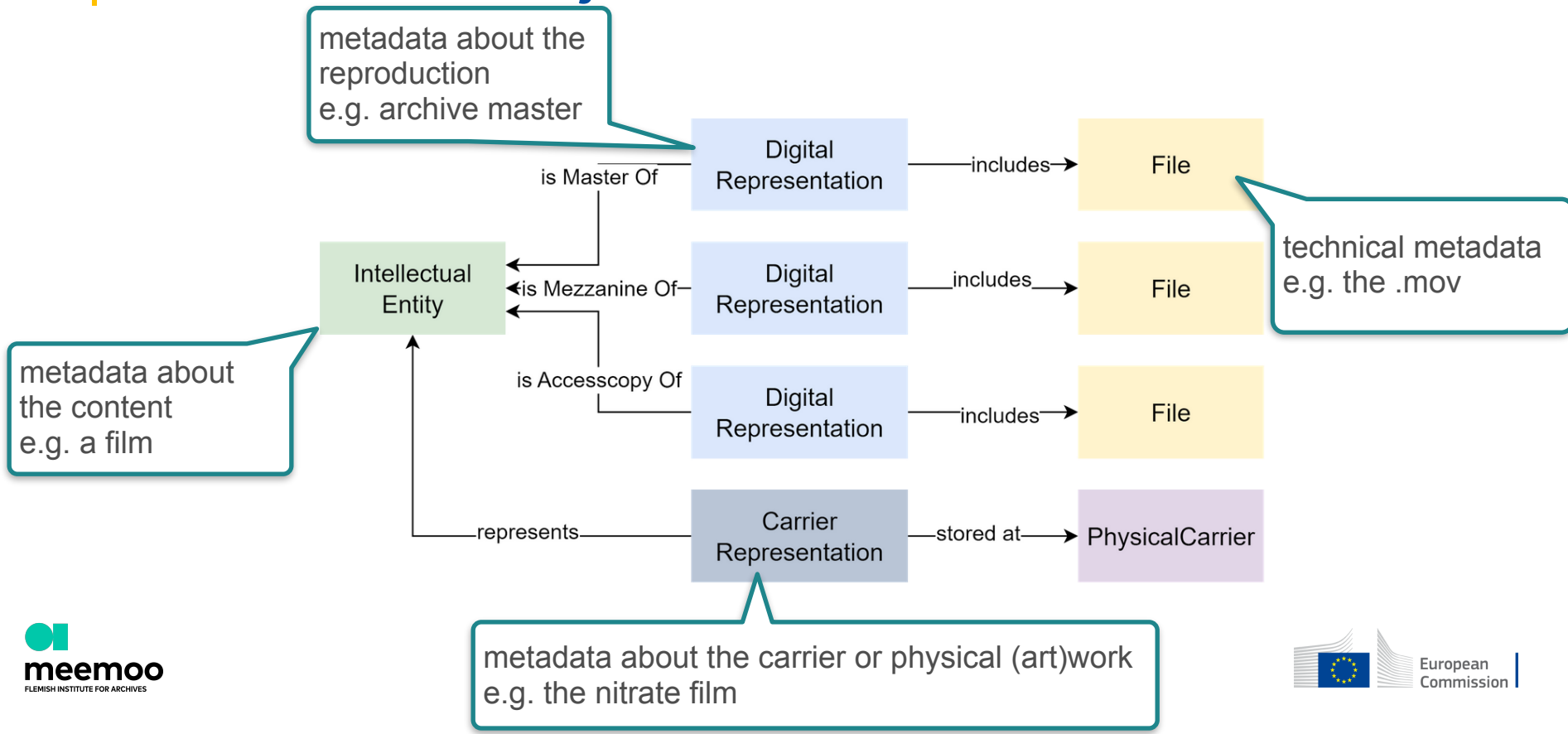
# meemoo's SIPIN flow



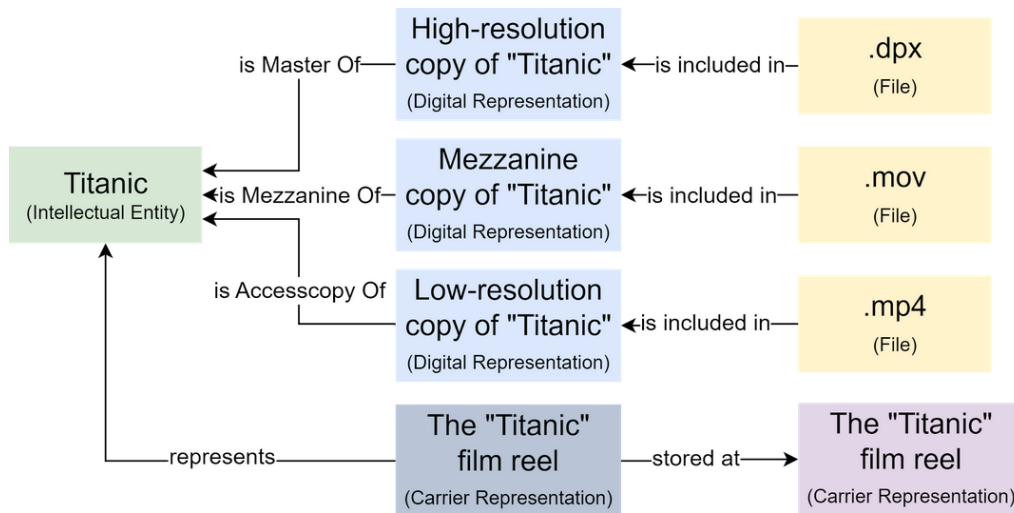
# Step 1: Creation and Delivery



# SIP creation: object structure breakdown

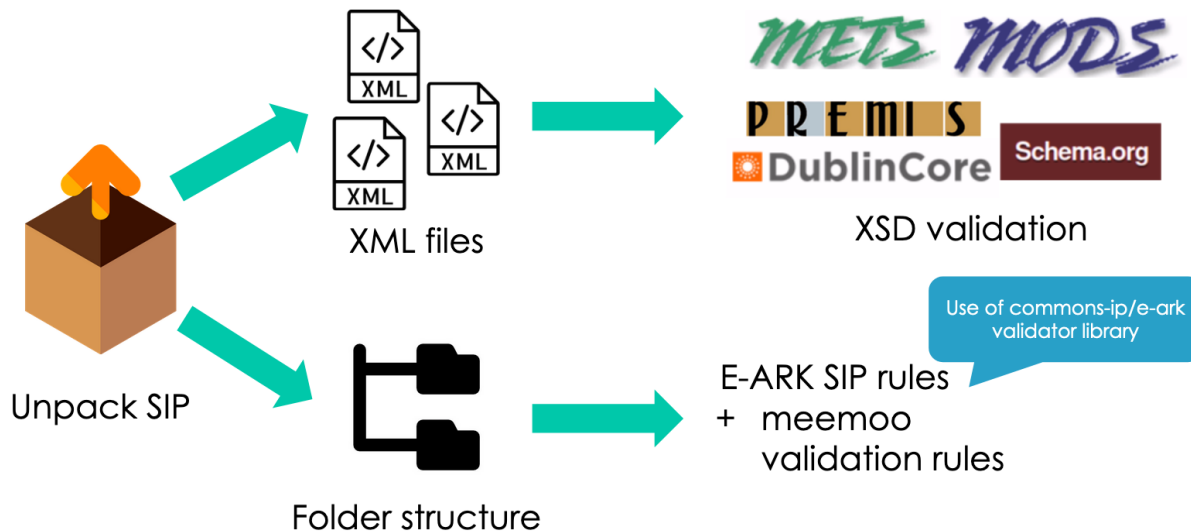


# SIP creation: object structure breakdown

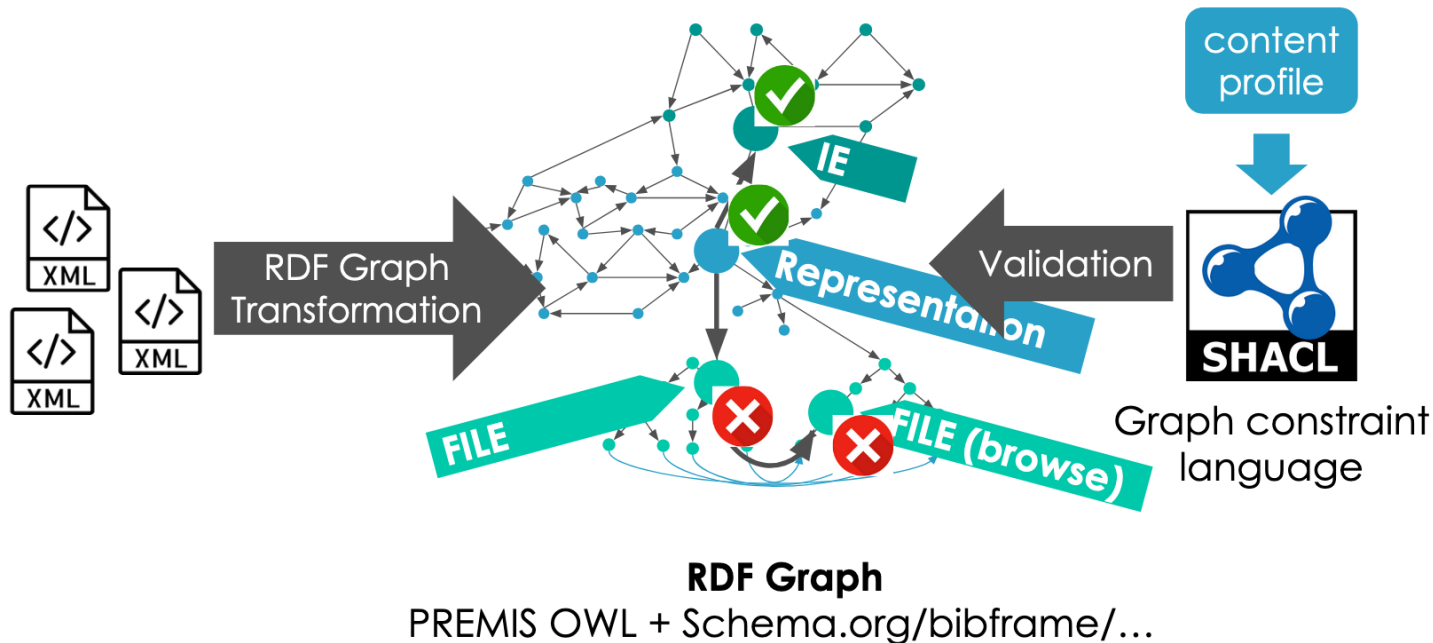




## Step 2: Syntactic validation



# Step 3: Semantic validation



# Step 3: Semantic validation

**Transformation of XML data into RDF with SPARQL**

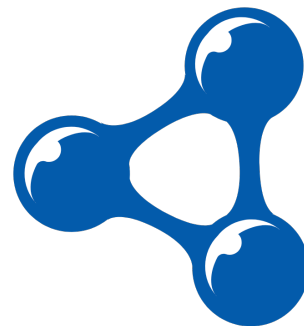
**Content profiles are translated into SHACL models**

Expressive constraints with relations between entities

Human-readable rules

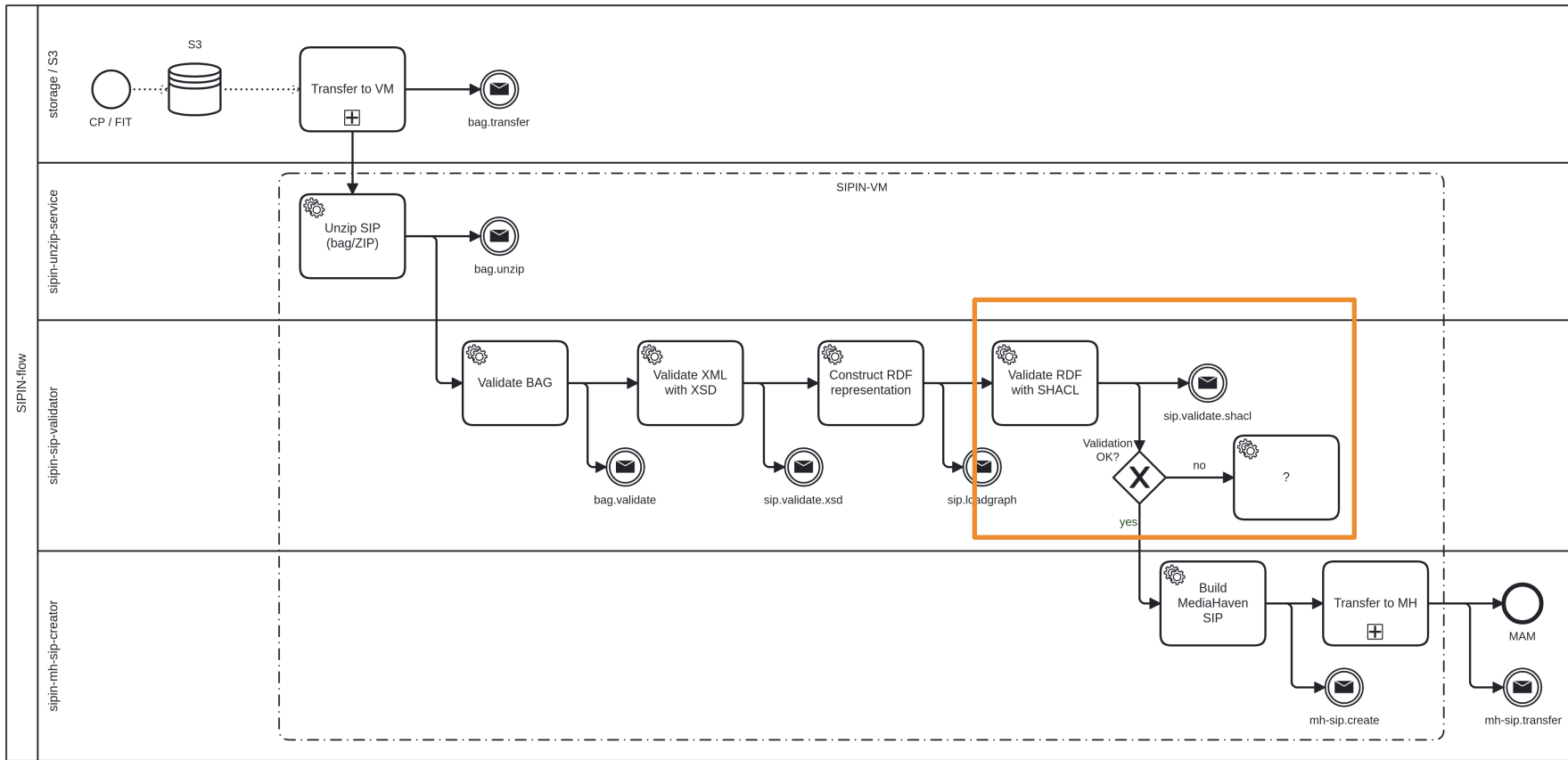
Aligned with our linked data models

Based on ontologies and semantic web standards

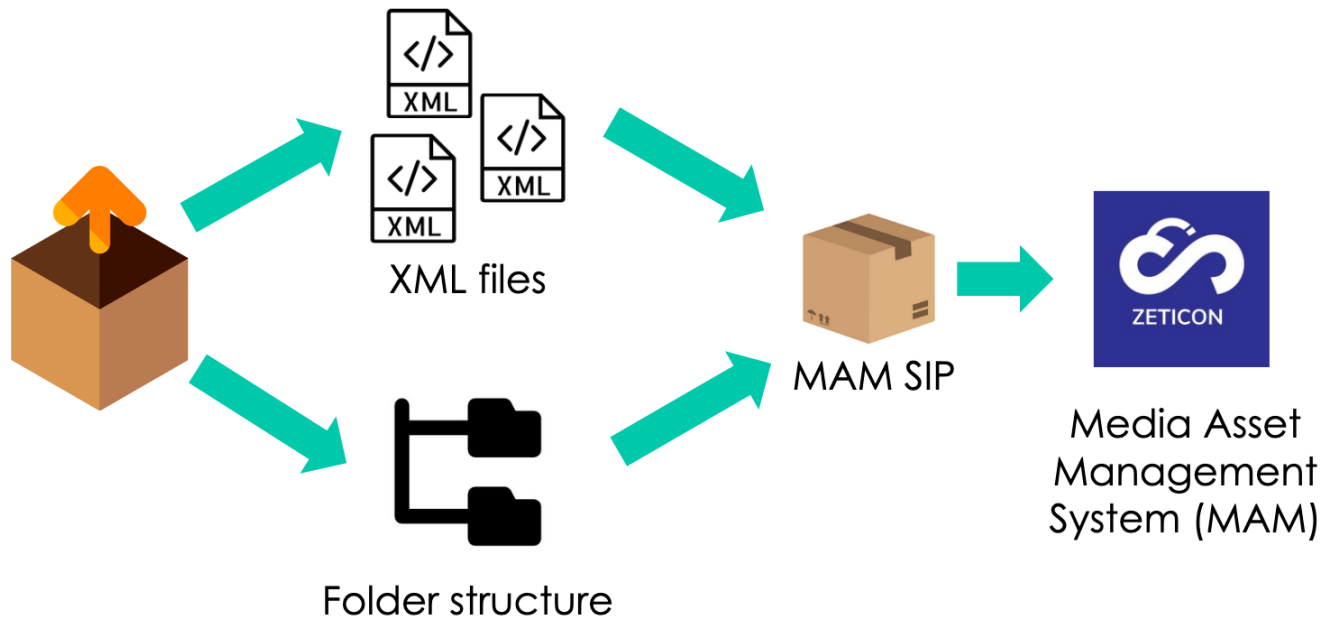


# Time to pray to the demo gods





## Step 4: Archiving



# Way forward

Exciting times ahead



# Improving our documentation

Our SIP specification can **still be a hurdle**

## Improving the accessibility

specific and less formal 'how to' sections / tutorials for users

information about common pitfalls

better documented samples and contexts



# Further integrating the E-ARK tooling

## Reuse the **E-ARK tooling**

commons-ip for creation

commons-ip/eark-validator for validation

contribute to repositories



## Apply for the **eArchiving conformance seal**

Participation in the **E-ARK AIP workgroup** using JSON-LD



# Behind the scenes

Preparing the transition to the 'meemoo SIP' 2.0

Analysis of our current validation methods

- works well, if you know the SIP inside-out :)

- confusing validation reports

- decouple syntactic and semantic validation?

Small steps towards a 'SIP lite' in RDF/Linked Data...

# Piqued your interest?

- [meemoo's general website](#)
- [meemoo's developer documentation](#)
  - [SIP specification](#)
  - [its accompanying Github repository](#)
- [meemoo's \(linked\) data models](#)
  - [their accompanying Github repository](#)
- [our sipin-sip-validator Github repository](#)



## Thank you for listening!

Miel Vander Sande  
([miel.vandersande@meemoo.be](mailto:miel.vandersande@meemoo.be))

Mattias Poppe  
([mattias.poppe@meemoo.be](mailto:mattias.poppe@meemoo.be))

Milan Valadou  
([milan.valadou@meemoo.be](mailto:milan.valadou@meemoo.be))

## Contact



<https://e-ark4all.eu/>



[info@e-ark-foundation.com](mailto:info@e-ark-foundation.com)



[@EU\\_eArchiving](https://twitter.com/EU_eArchiving)



<https://www.linkedin.com/company/eu-earchiving-initiative>



<https://www.youtube.com/@e-ark>