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

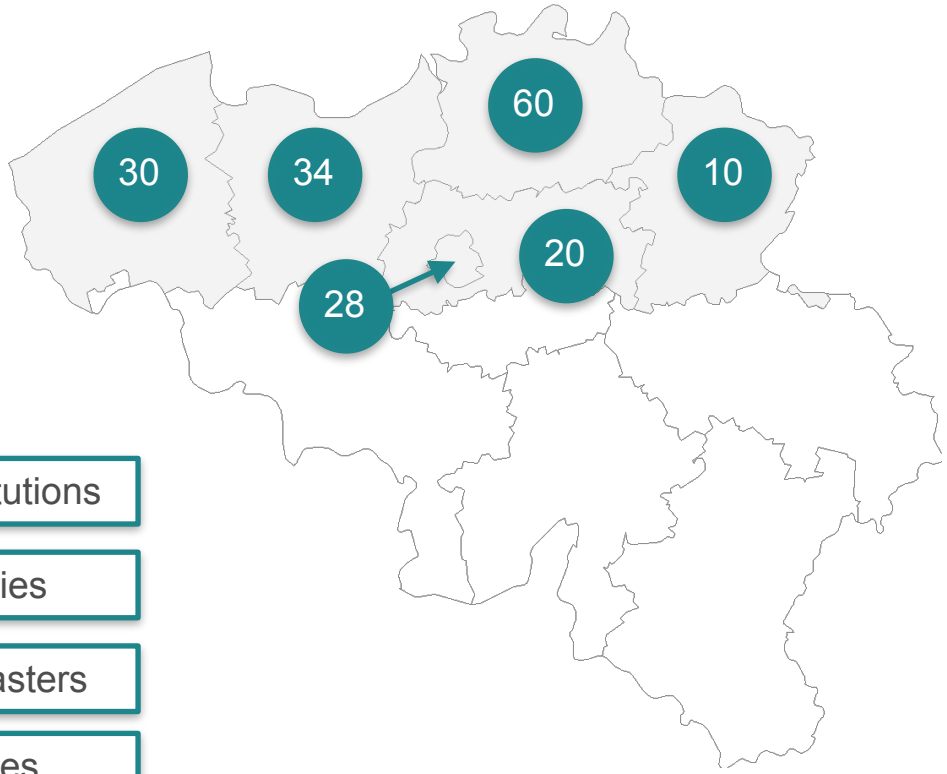
meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

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
- government institutions
- museums
- heritage libraries
- archives
- national broadcasters
- heritage societies
- sector institutes
- regional broadcasters

meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

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

meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

Interaction

The Archive for Education

hetarchief.be

News of the Great War

Art in Flanders

Internal tools

Applications of CPs

Integration

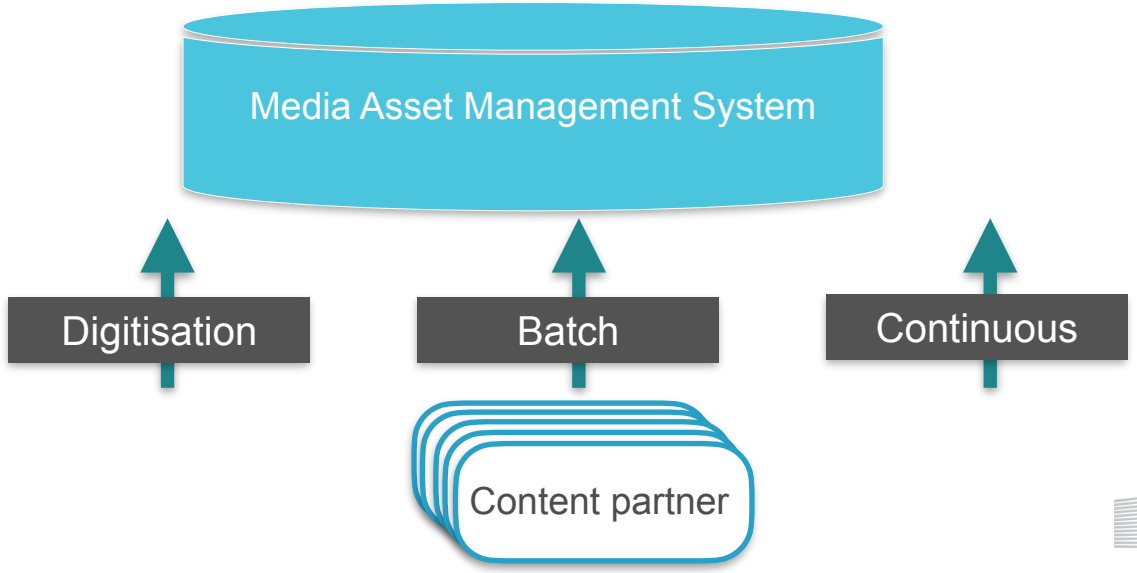
SEARCH

REST API

GRAPHQL

OAI-PMH

Archive + metadata management

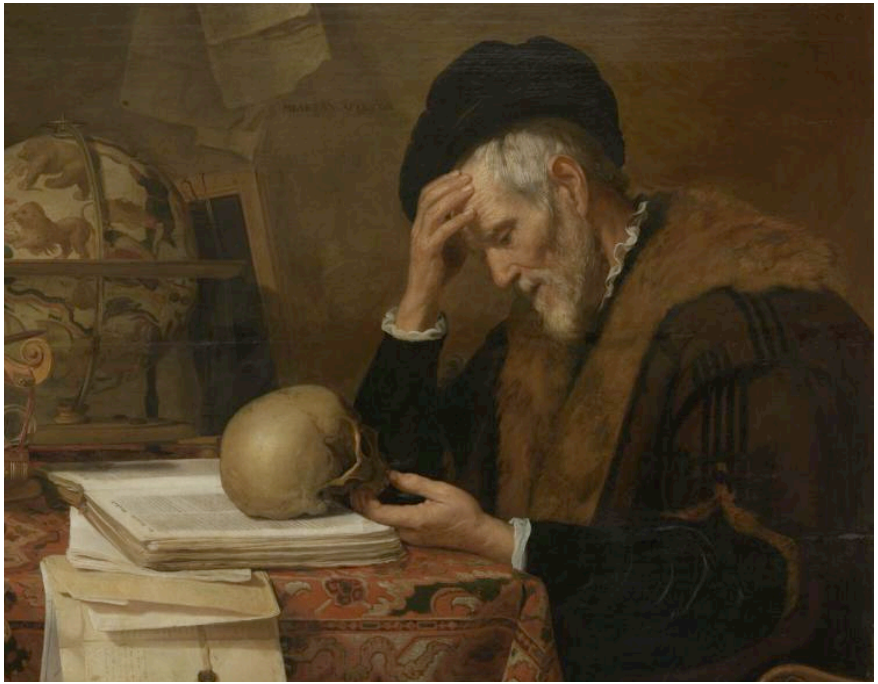


meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

A bit of context

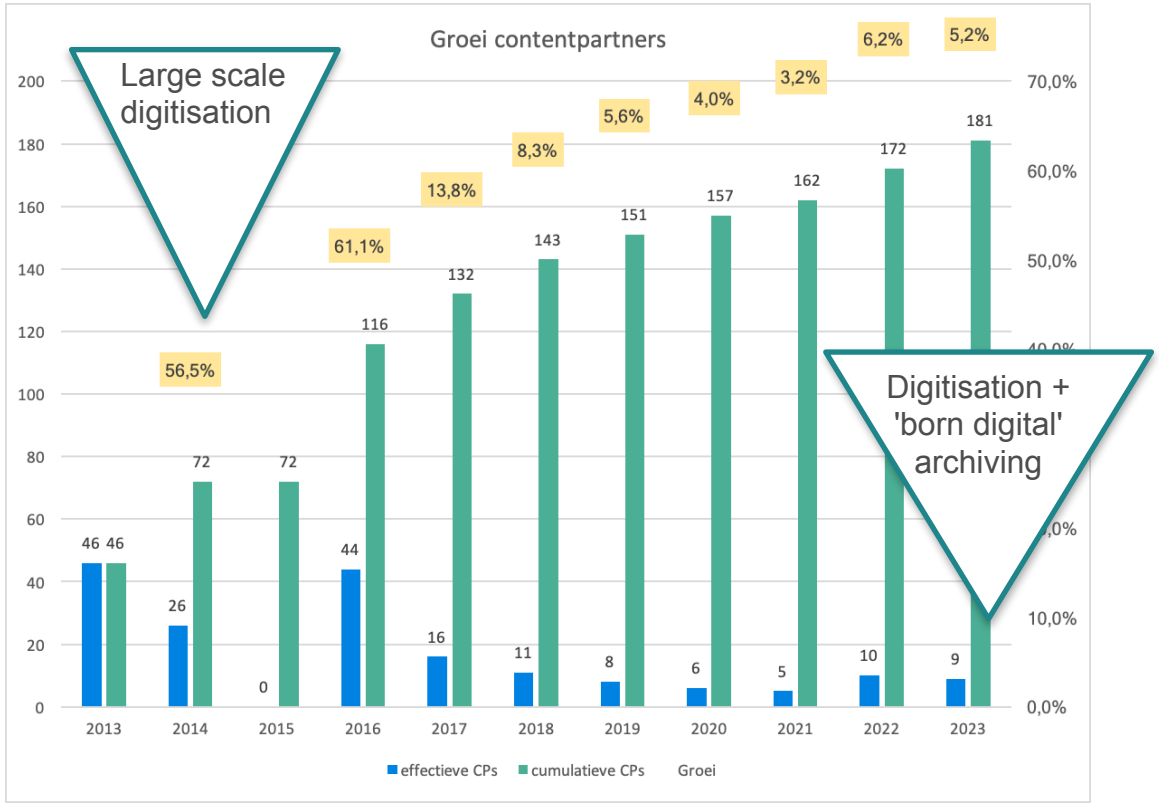
how we came to the realisation that we needed a SIP



meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

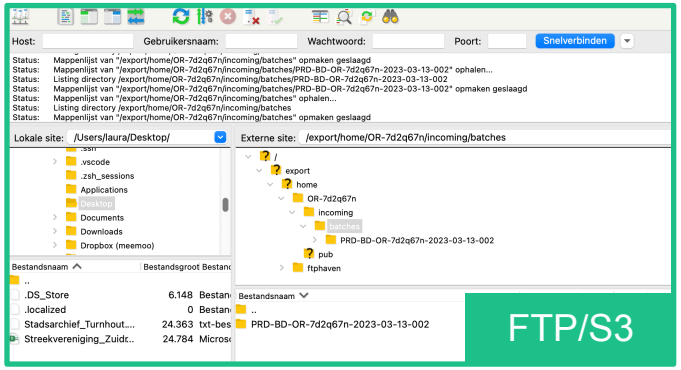
Rise in number of partners

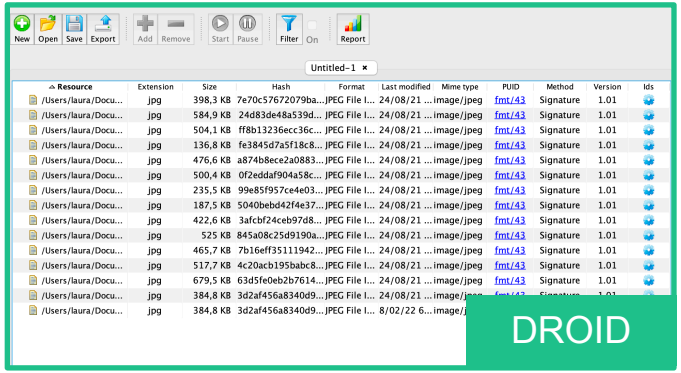


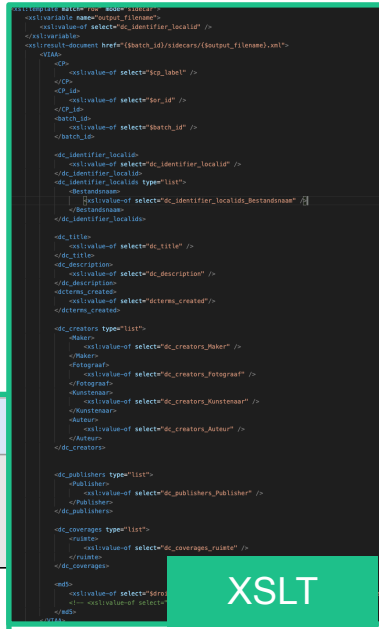
meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

Influx of digital collections







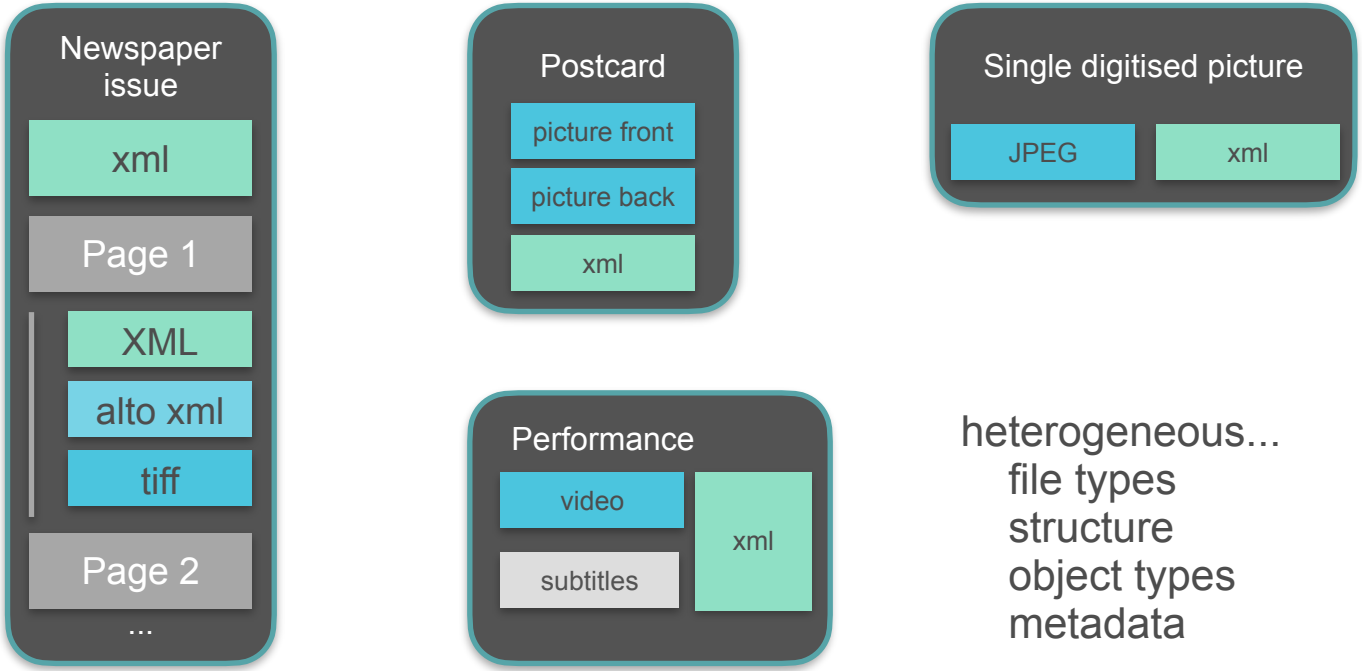
dc_identifier_localid	dc_title	dc_description	dcterms_created	dc_identifier_localids_Bestandsnaam	dc_titles_reeks	dc_titles_seizoen	dcterms_issued	dc_creators_Acteur
Hoofd lokale CP ID = bestandsnaam zonder extensie	*Hoofdtitel*	*Hoofdbeschrijving*	*Creatiedatum van het object, EDTF level 0 of 1. Zie developer.meemoo.be (datatypes)*	*Originele bestandsnaam van het object = hoofd lokale CP ID + extensie*	*Naam van de voorstelling of productie*	*Het seizoen van de voorstelling*	*Datum waarop het item is uitgegeven of uitgezonden. EDTF level 0 of 1. Zie developer.meemoo.be (datatypes)*	*De **acteur** in het object in kwestie*
Offtherecord_CP_1	Off the record	Vier jongeren tussen 14 en 18 jaar tonen zich zoals ze zijn. Fel, kwetsbaar, schaamteloos en vol lef. Het podium is een vrijplaats 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. 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	2022-02-28	Offtherecord_CP_1.jpg	Off the record	2021 - 2022;2022 - 2023		Ben Herlaar Masina;Ziba Jafari;Mauro Michielsen;Mette Plysier

Descriptive metadata templates

meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

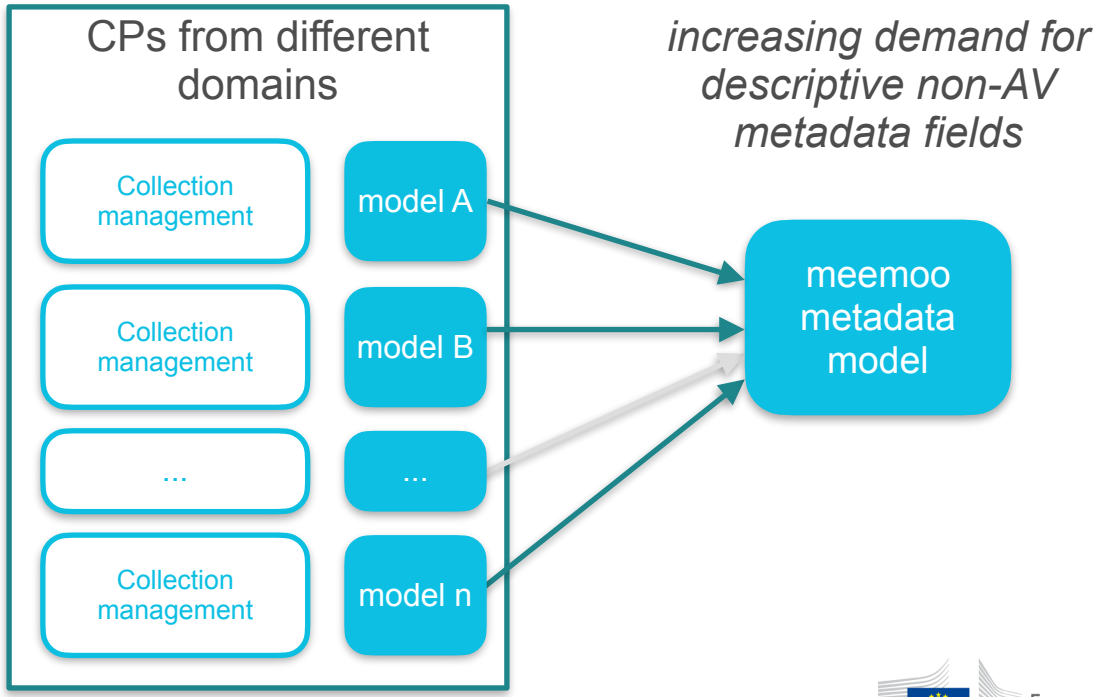
Increasingly diverse and complex object structures



heterogeneous...
- file types
- structure
- object types
- metadata

More heterogeneity in metadata

diverse metadata intake

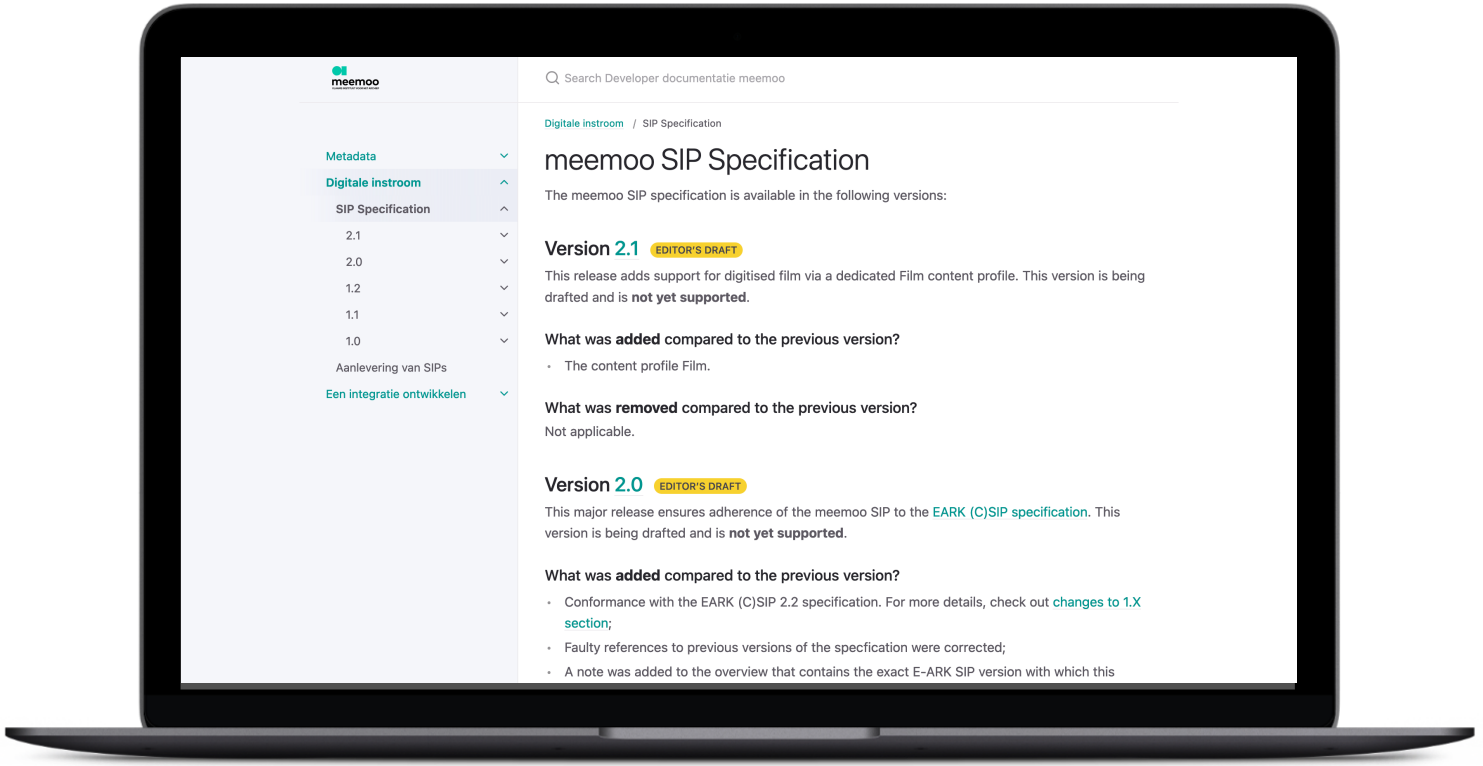


increasing demand for descriptive non-AV metadata fields

A (deep) dive into the 'meemoo SIP'

meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission



meemoo
FLEMISH INSTITUTE FOR ARCHIVES

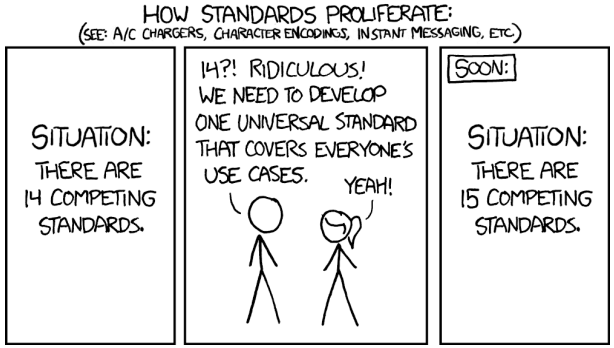
https://data.hetarchief.be/id/sip

European Commission

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**

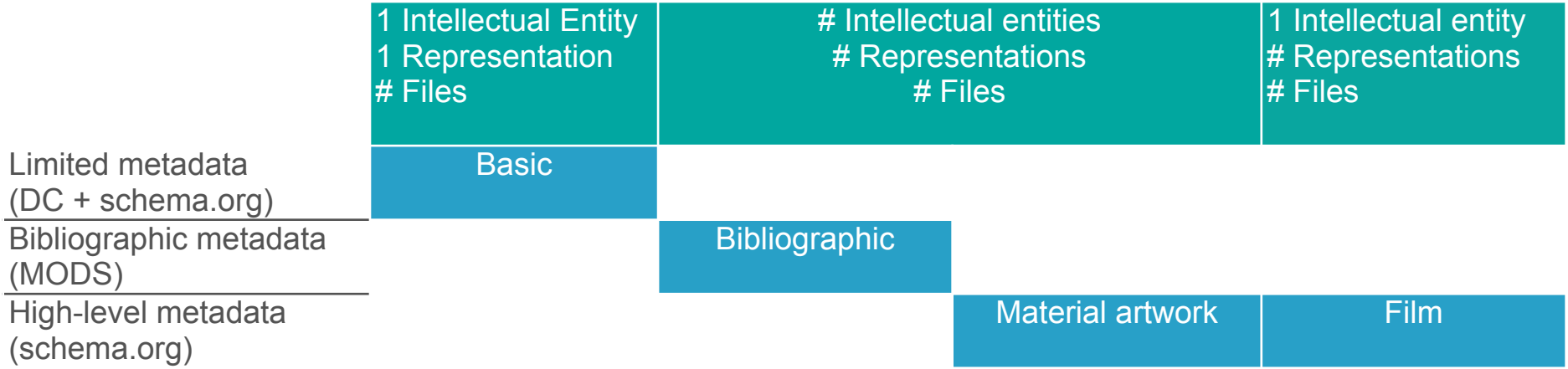
'terminology' + 'core concepts' + 'use cases' sections

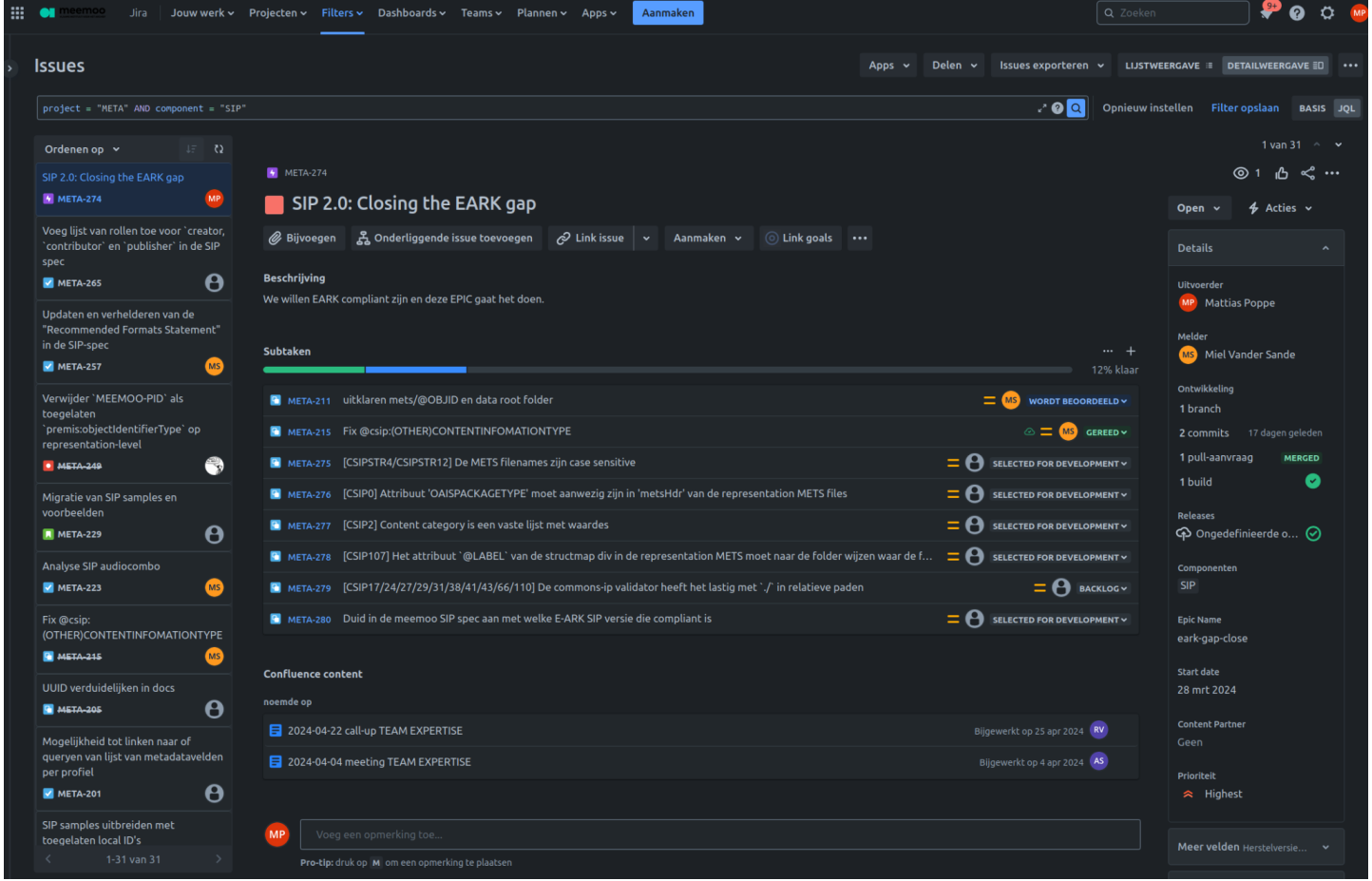
narrow definition of the hierarchical structure

requiring the use of PREMIS XML

conceived the concept of 'content profiles'

Content profiles

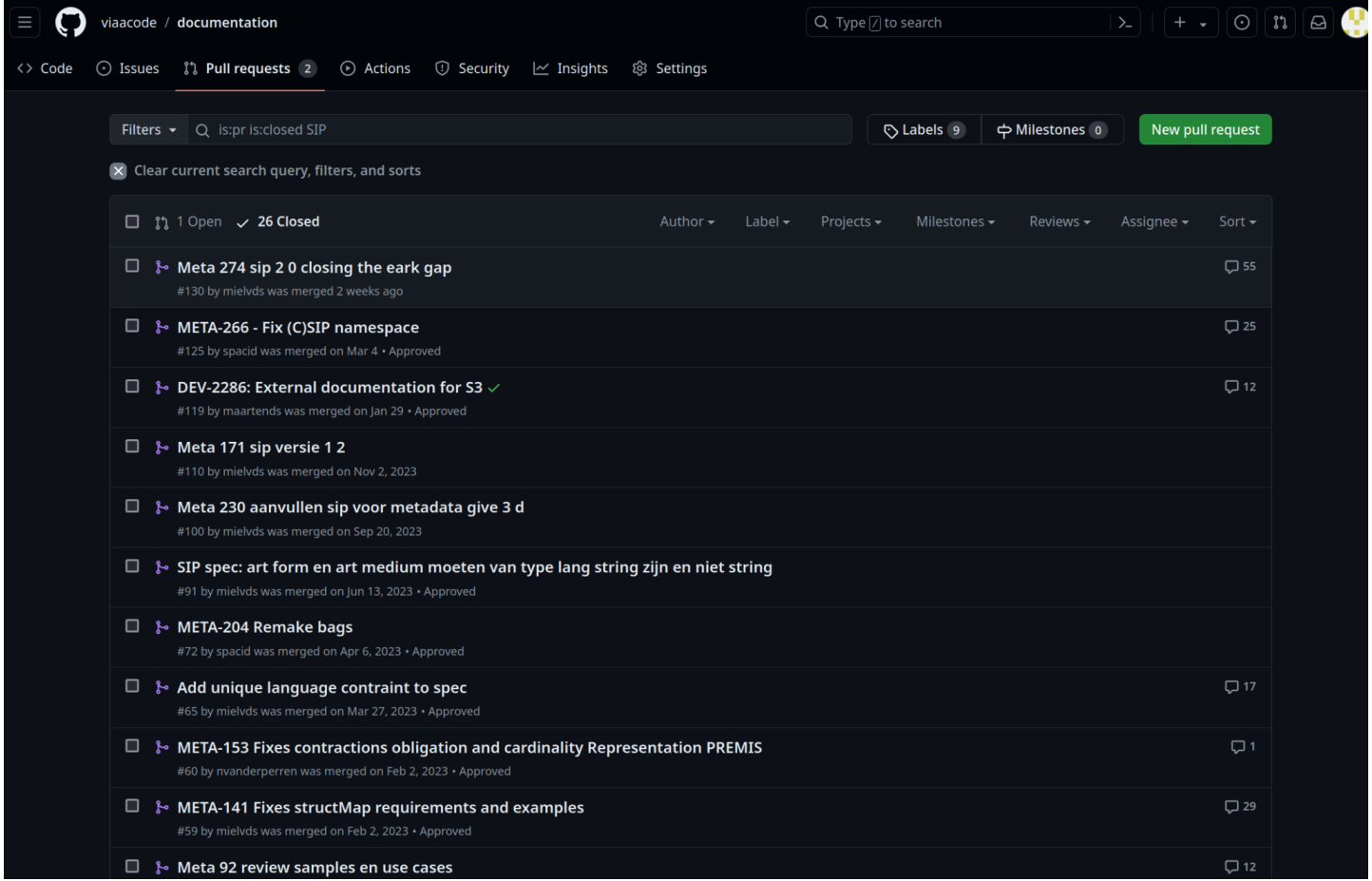




meemoo
FLEMISH INSTITUTE FOR ARCHIVES

Open editorial process using Github PRs and JIRA

European Commission



meemoo
FLEMISH INSTITUTE FOR ARCHIVES

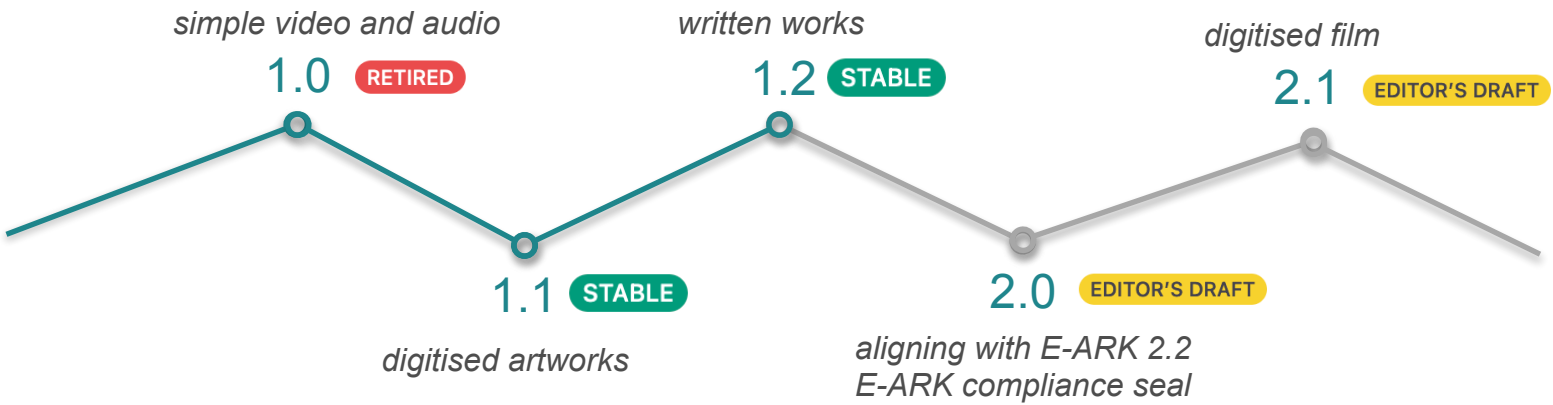
Open editorial process using Github PRs and JIRA

European Commission

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 *public review* and *implementation*. 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



meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

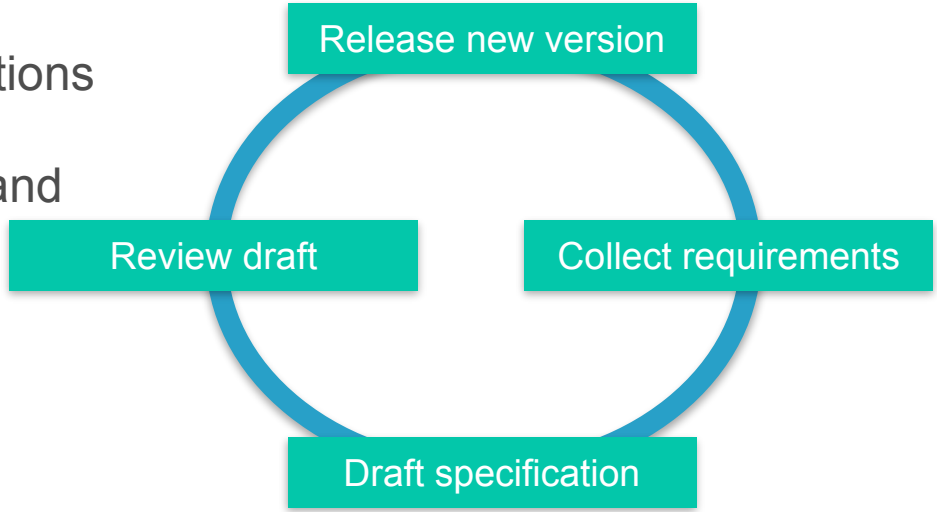
Towards a governance model

Release cycle to avoid *ad hoc* alterations

Connect with partners, gather input and cover their needs

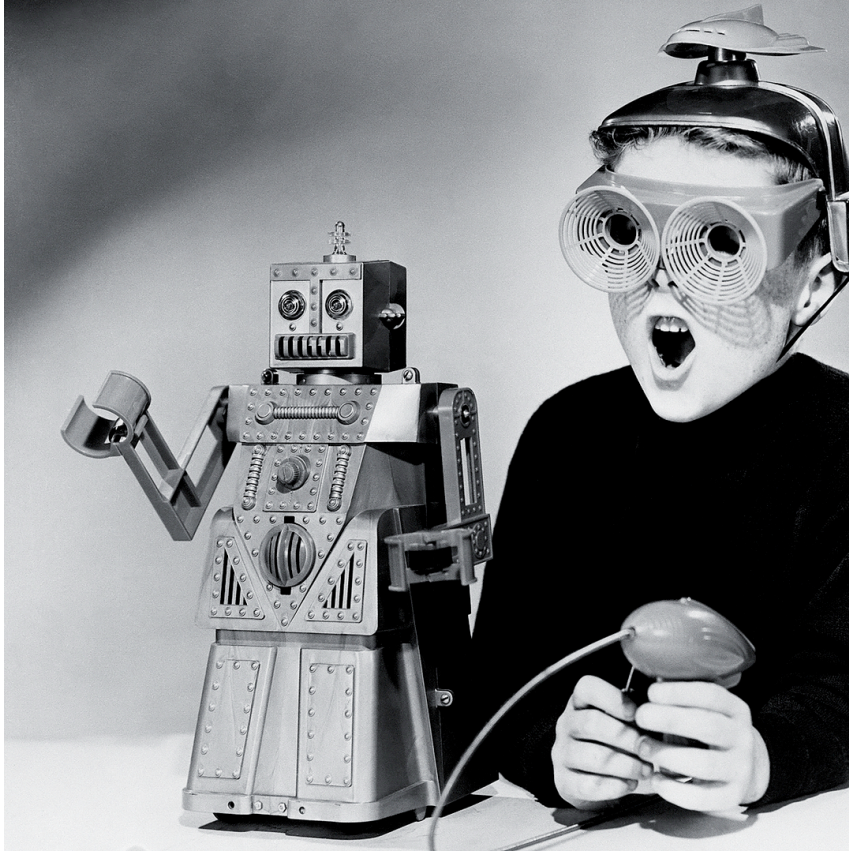
Predictable support for both sides

Ongoing work



Ingesting 'meemoo SIPs'

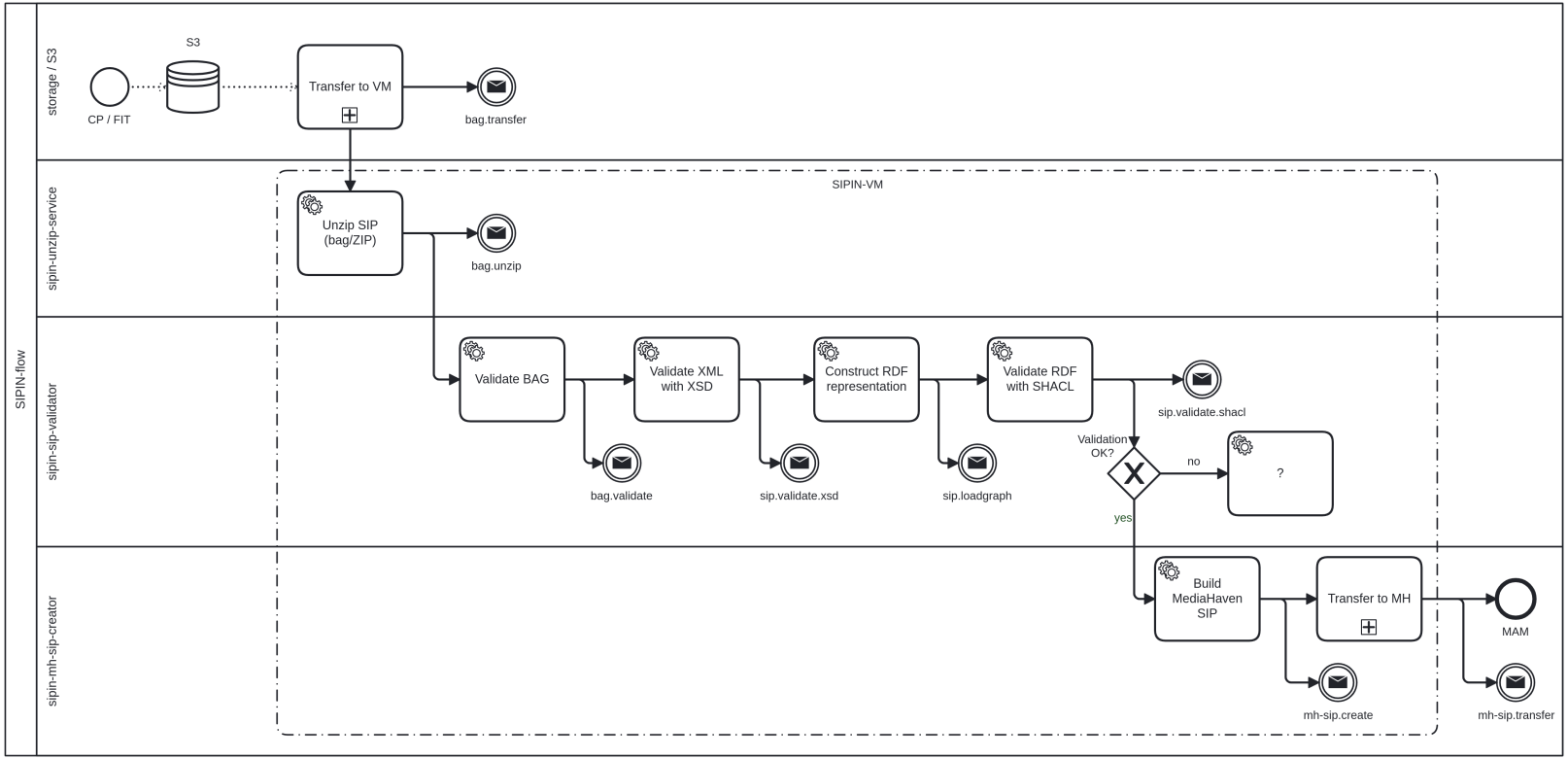
How we transform it to a graph for semantic validation



meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

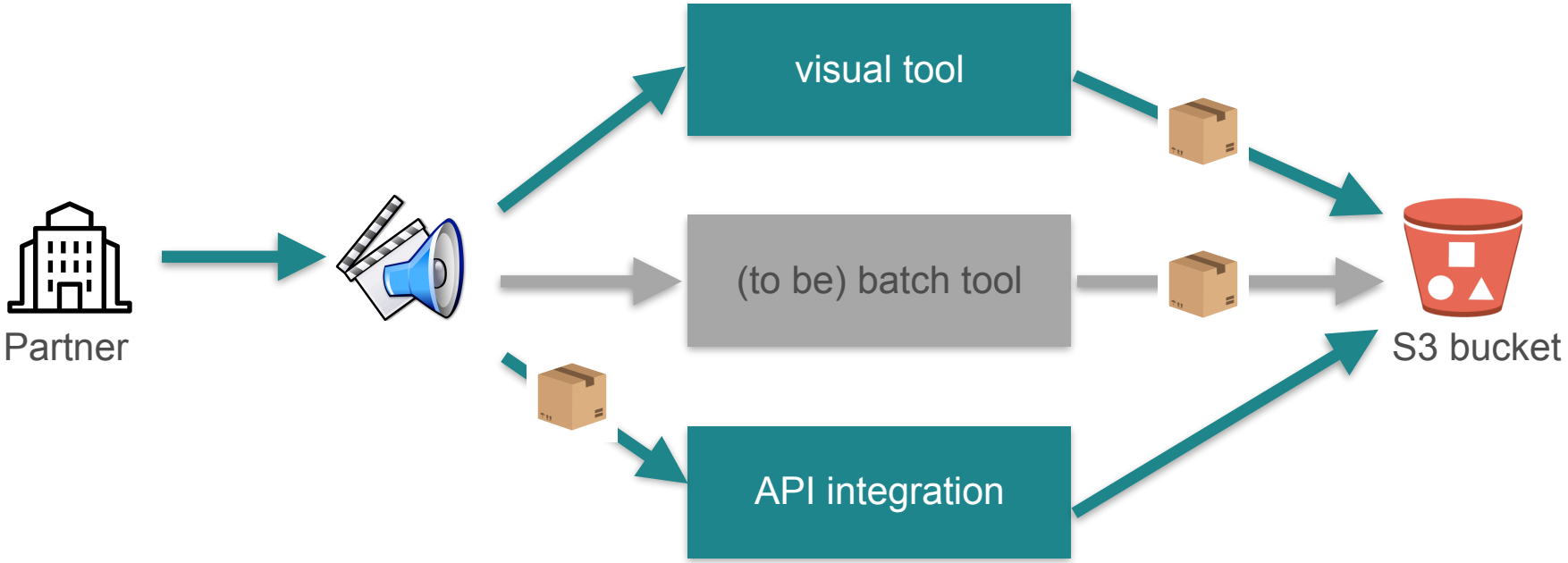
meemoo's SIPIN flow



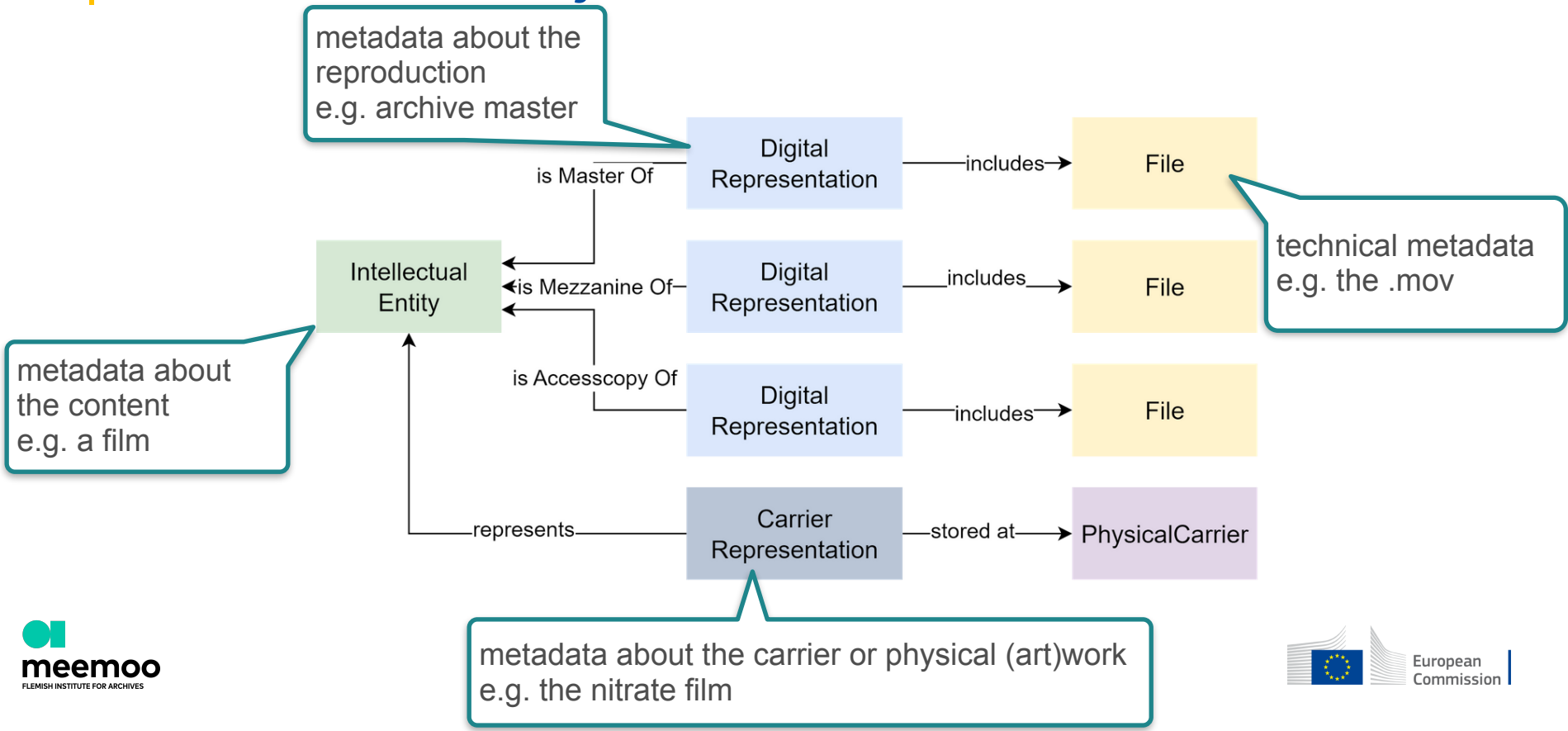
meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

Step 1: Creation and Delivery



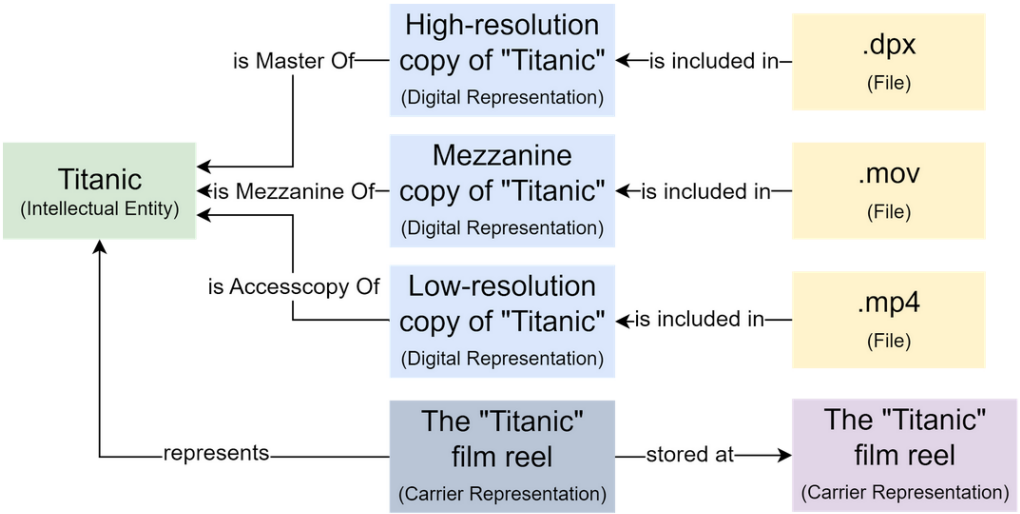
SIP creation: object structure breakdown



meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

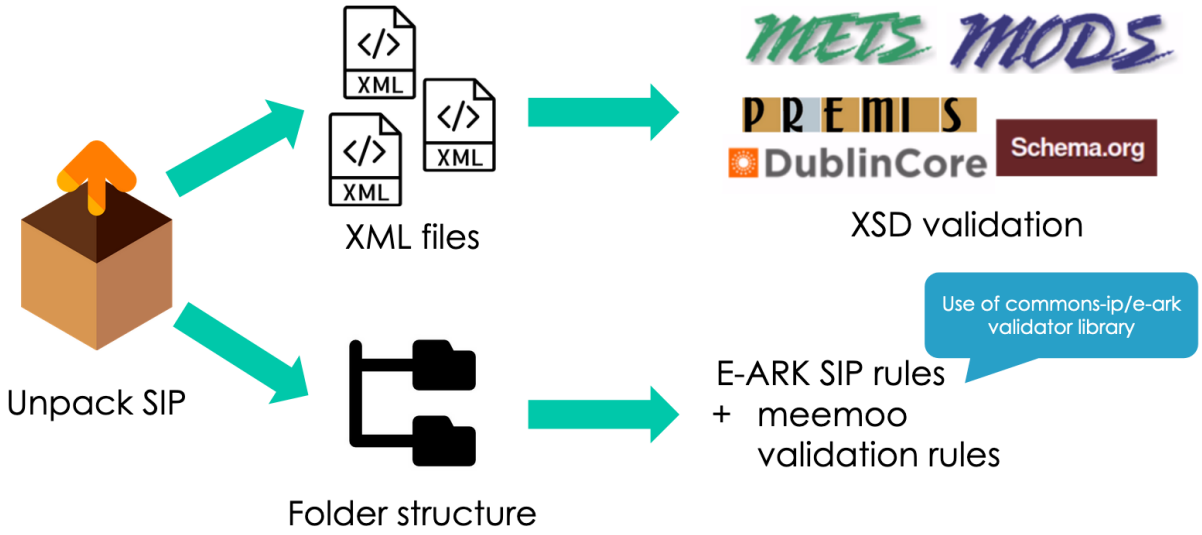
SIP creation: object structure breakdown



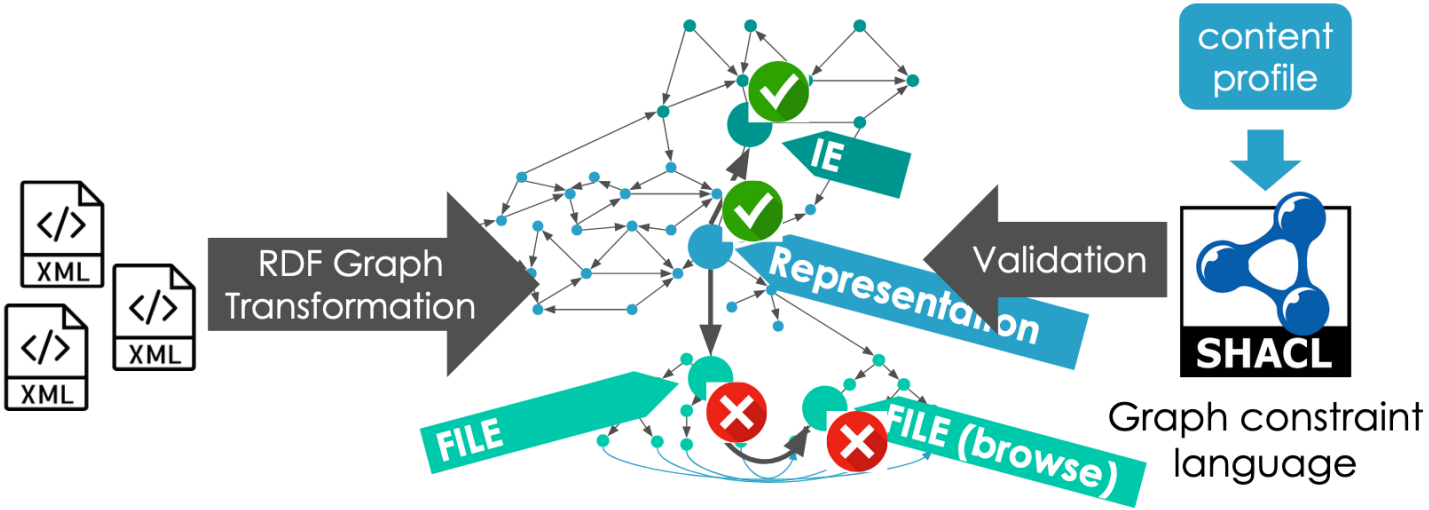
meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

Step 2: Syntactic validation



Step 3: Semantic validation



RDF Graph

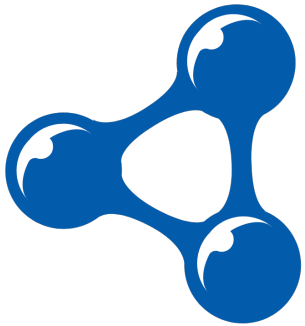
PREMIS OWL + Schema.org/bibframe/…

Step 3: Semantic validation

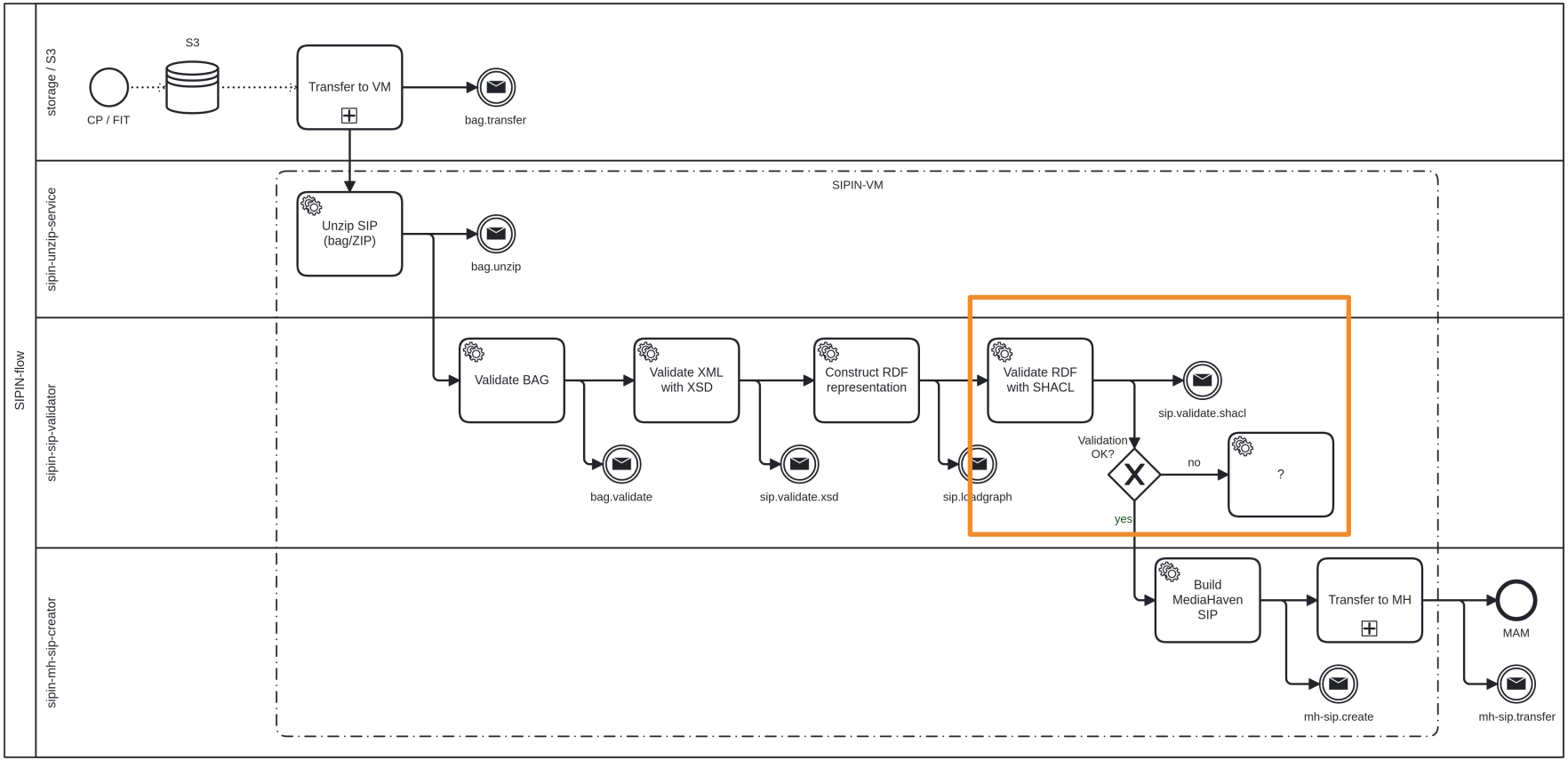
Transformation of XML data into RDF with SPARQL

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



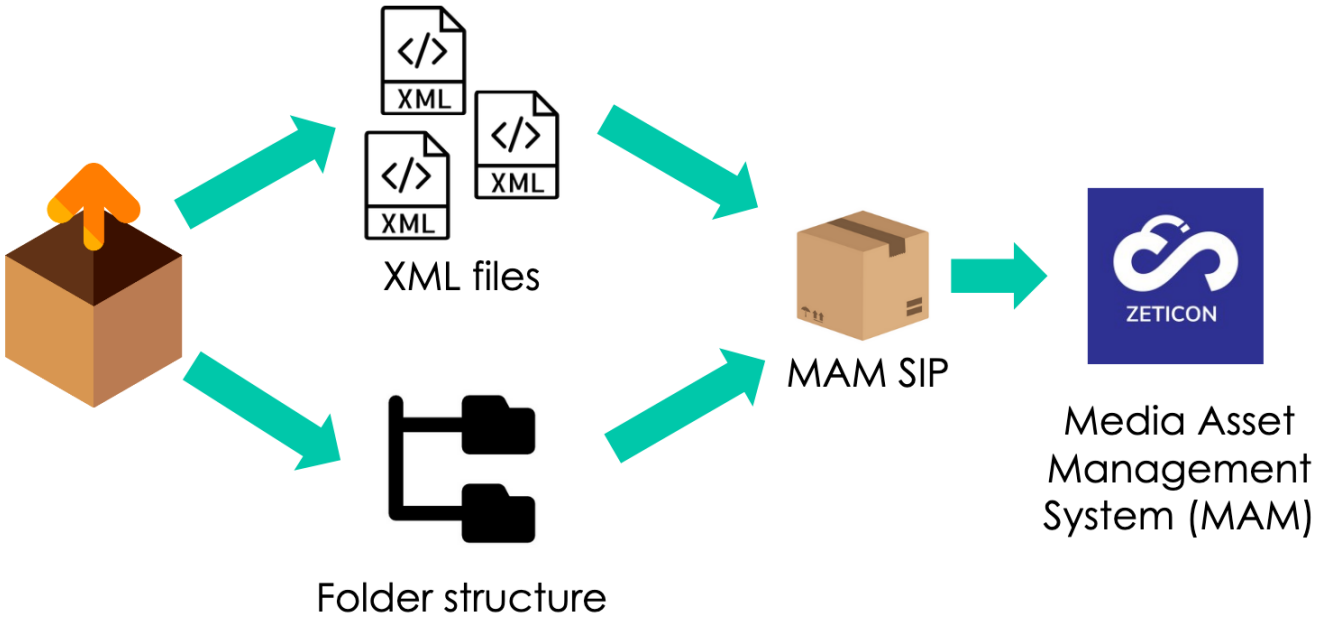
Time to pray to the demo gods



meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

Step 4: Archiving



Way forward

Exciting times ahead

meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

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

meemoo
FLEMISH INSTITUTE FOR ARCHIVES

European Commission

Thank you for listening!

Miel Vander Sande
(miel.vandersande@meemoo.be)

Mattias Poppe
(mattias.poppe@meemoo.be)

Milan Valadou
(milan.valadou@meemoo.be)

Contact

https://e-ark4all.eu/

info@e-ark-foundation.com

@EU_eArchiving

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

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