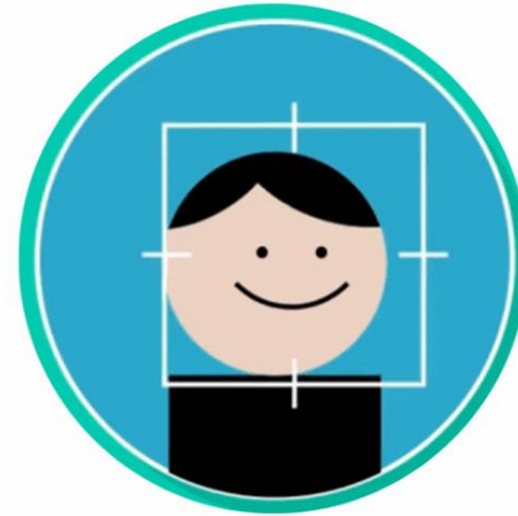
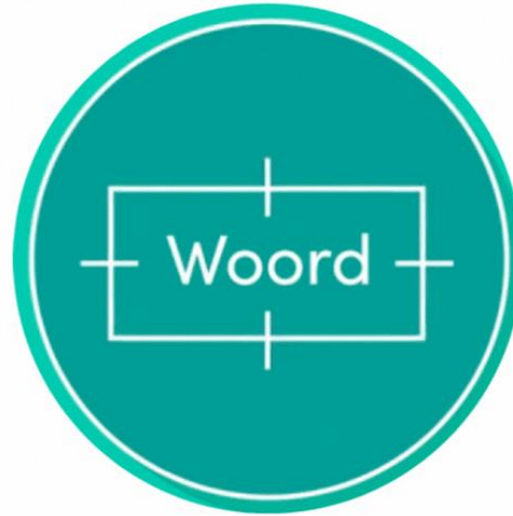
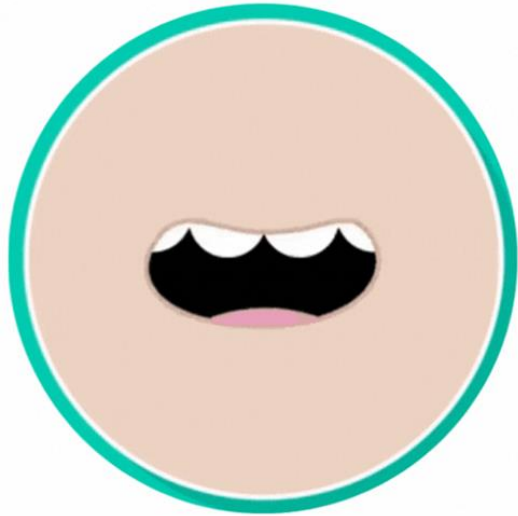


Let's talk about AI.



At **meemoo** we're here for the **archive**.

We help **cultural, media and government**
organisations with advice and practical support,
and want to make archival materials **accessible**
and **usable**.

Services

- **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

Diverse group of Flemish content partners

50

performing arts

12

government institutions

45

museums

7

heritage libraries

24

archives

3

national broadcasters

19

heritage societies

2

sector institutes

10

regional broadcasters

Collaboration

We work with our partners in a complementary way:

- They have content-related **expertise** about their domain and its materials.
- They **manage** and **select** the physical and digital collections
- They **contextualize, share, research...**



Economies of scale: lower cost, better results



Digitisation, digital infrastructure and processes: **expert knowledge and size matter**



Mass storage in three secure locations and **metadata enrichment/management** for our partners



Unlocking **large quantities** of **diverse** materials



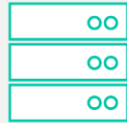
Advice and **knowledge-sharing** with the entire cultural sector

In figures



+180

content partners



> 6 million

objects in our
archive system



> 540,000

audiovisual carriers
digitised and stored
in our archive system



> 185,000

user accounts at The
Archive for Education at
end of 2021-2022
academic year



> 16,000

free downloads
from artinflanders.be
in 2021



nearly 50

training and other
events for cultural,
media and
government
organisations in 2022

What is the GIVE-project?

The Coordinated Initiative for Flemish Heritage Digitization - the abbreviation makes sense in dutch.

Four Tracks:

- **Masterpieces**
245 masterpieces in 59 collections:
manuscripts and 3D-, 2D- en
gigapixelphotography
- **Newspapers** (collaboration with the VEB)
630.000 newspaper pages in 8 heritage
institutions
- **Glass plates**
180.000 glass plates from 33 heritage institutions
- **Metadata enrichment using AI**
160.000 hours of audio en video from 127
organisations



Metadata: challenges for manual approach



Numbers!

Archival materials
from **127**
contentpartners


170.000
hours of audio


124.000
hours of video

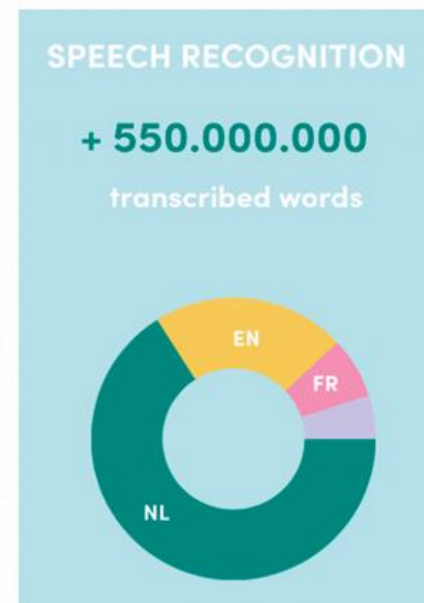
SECTORS



FACE RECOGNITION

 **3.300.000** detected individuals

2.518 unique recognised individuals

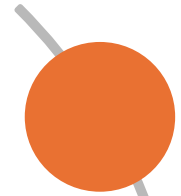


ENTITY RECOGNITION

6.500.000 detected individuals, places and events in text

-  1.7 million
-  3.7 million
-  1.1 million

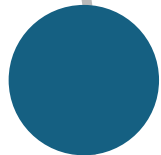
Buy or Build?



Speech-To-Text

- Mature tooling in the market
- Cost efficient solution, easy to integrate

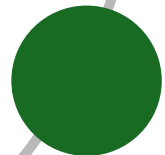
⇒ **BUY**



Named Entity Recognition

- Mature tooling in the market
- Cost efficient solution, easy to integrate

⇒ **BUY**

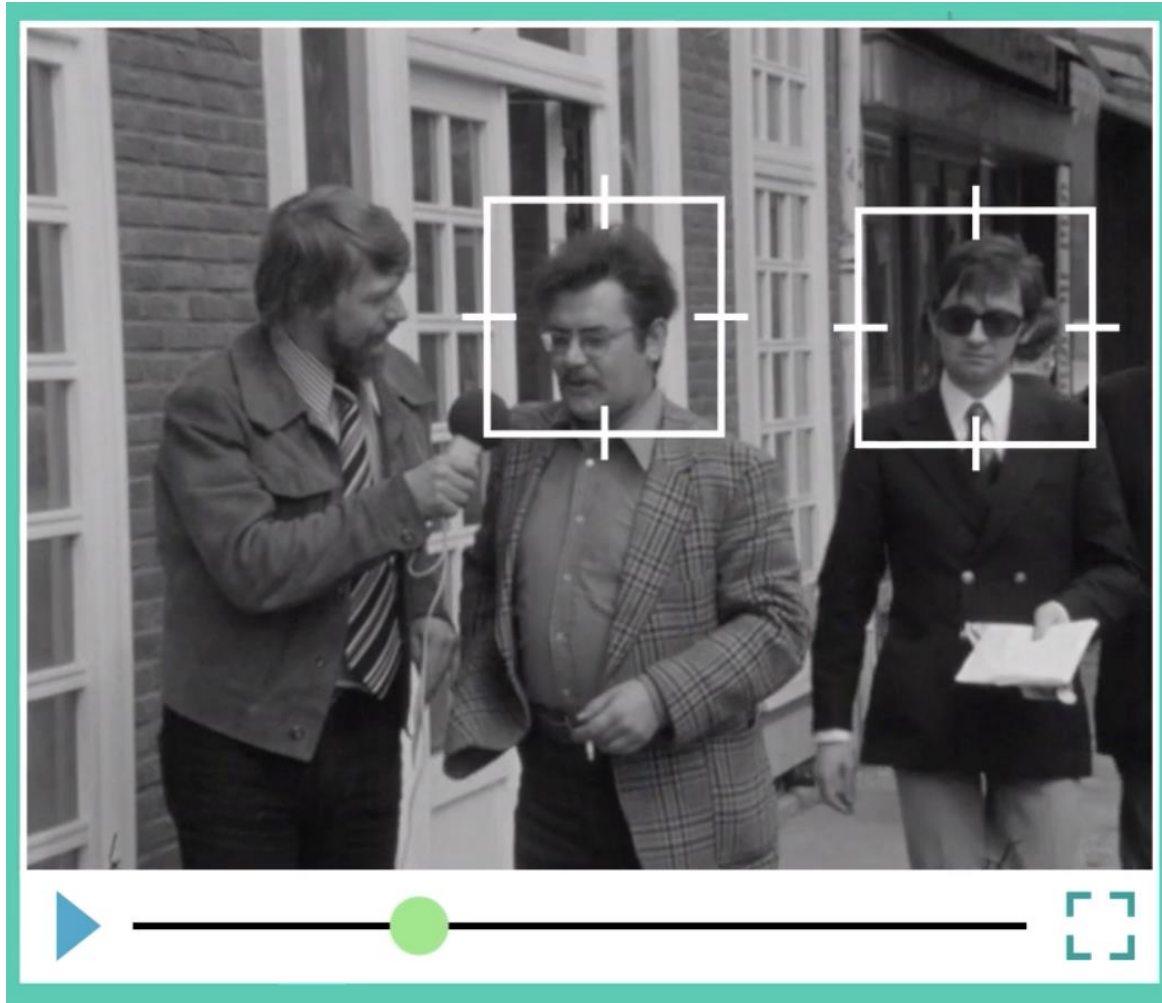


Face detection and recognition

- Partly available in the market (see: what we want)
- Higher cost per hour
- Pre existing experience in previous project (FAME)

⇒ **BUILD**

Zoom in on face recognition



Detecting and grouping **faces**

Match faces to known **individuals**
(based on confidence score)

Identify **unknown** faces

Manage a set of individuals we want
to recognise and **collaborate on**
identifying unknowns

Involving partners

Call for participation to 120 partners,
+/- 10 partners in working group



Data Protection Impact Assessment

RISICO-ANALYSE							
2	Verwerkingsactiviteit	Beschrijving risico		Schadetype	Toelichting type	Kans op schade	Ernst van de impact
3	Vermeld de vooropgestelde activiteit	Beschrijf het mogelijke risico, eventuele voorbeelden staan in BIJLAGE 4	Vermeld de bron van het risico	Kies het toepasselijke schade type	Beschrijf hier waarom je een bepaald type schade hebt aangeduid	Kies hier de kans dat de schade zich voordoet	Kies hier de impact van de schade
4				▼		▼	
5				▼		▼	
6				▼		▼	
7				▼		▼	
8				▼		▼	
9				▼		▼	

+ ☰ Gegevens DPIA ▾ Verwerking ▾ Risico-analyse ▾ Detail Risico ▾ BIJLAGE 1 Voorbeelden GEGEVENS ▾ BIJLAGE 2 Voorbeelden GEGEVENS ▾

Ethical aspects

- Collaboration with Knowledge Center for Data and Society (interuniversity group, funded within Flemish AI research track)
- Multiple stakeholder workshops, focusing on face recognition
 - Shared insights and principles



Face recognition - parameters

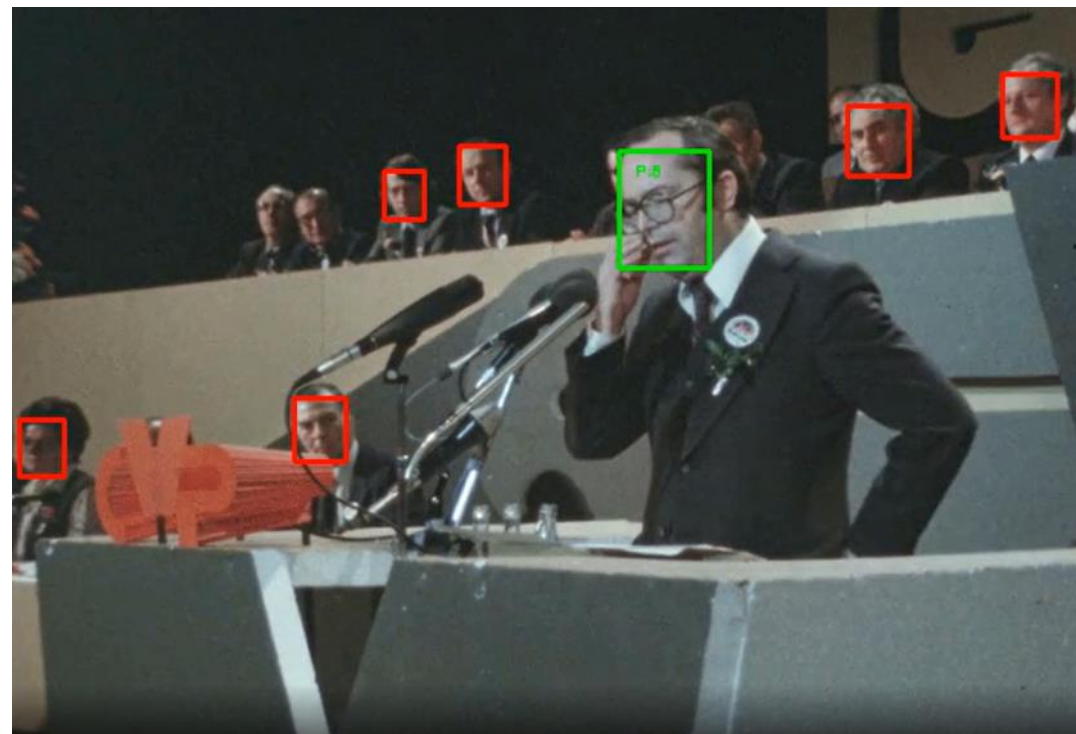
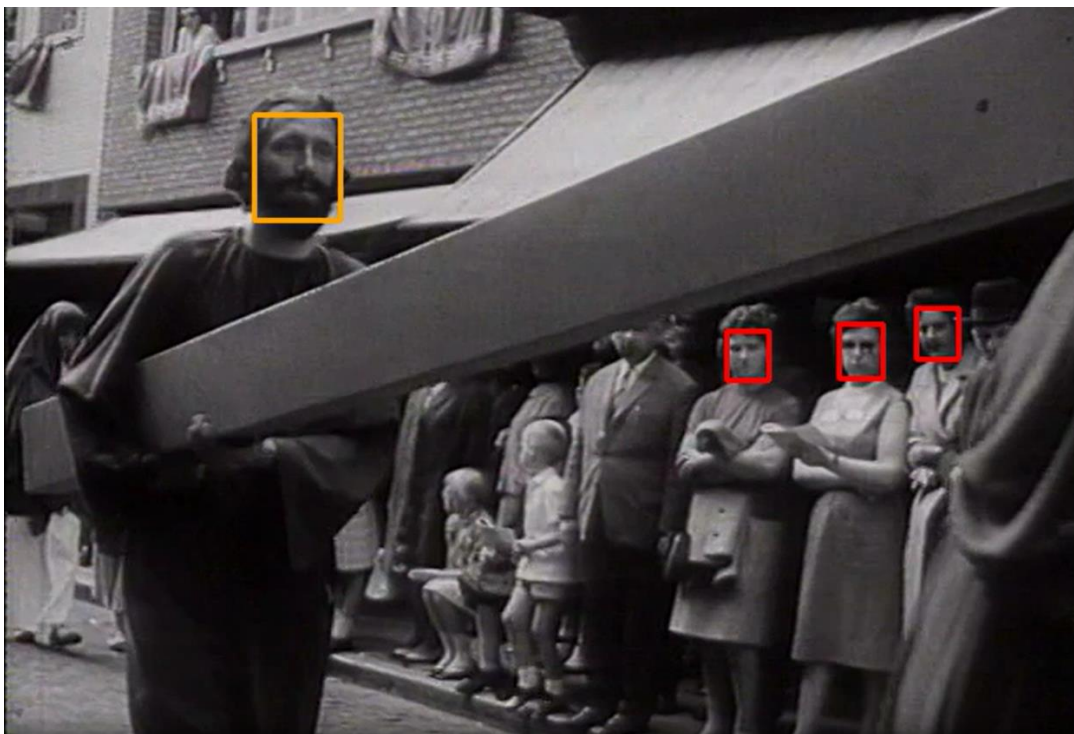
Face 'quality'

- Box-size
- Pose (landmarks)
- Lighting
- Blurriness
- Confidence

Time on screen

- Duration of face visibility in one shot (tracker)
- Number of appearance of a person (trackers per person)
- Total screen time of 1 person in a video

Working group: parameters



Oranje faces are not visible for long enough
Red faces: no metadata because they too small

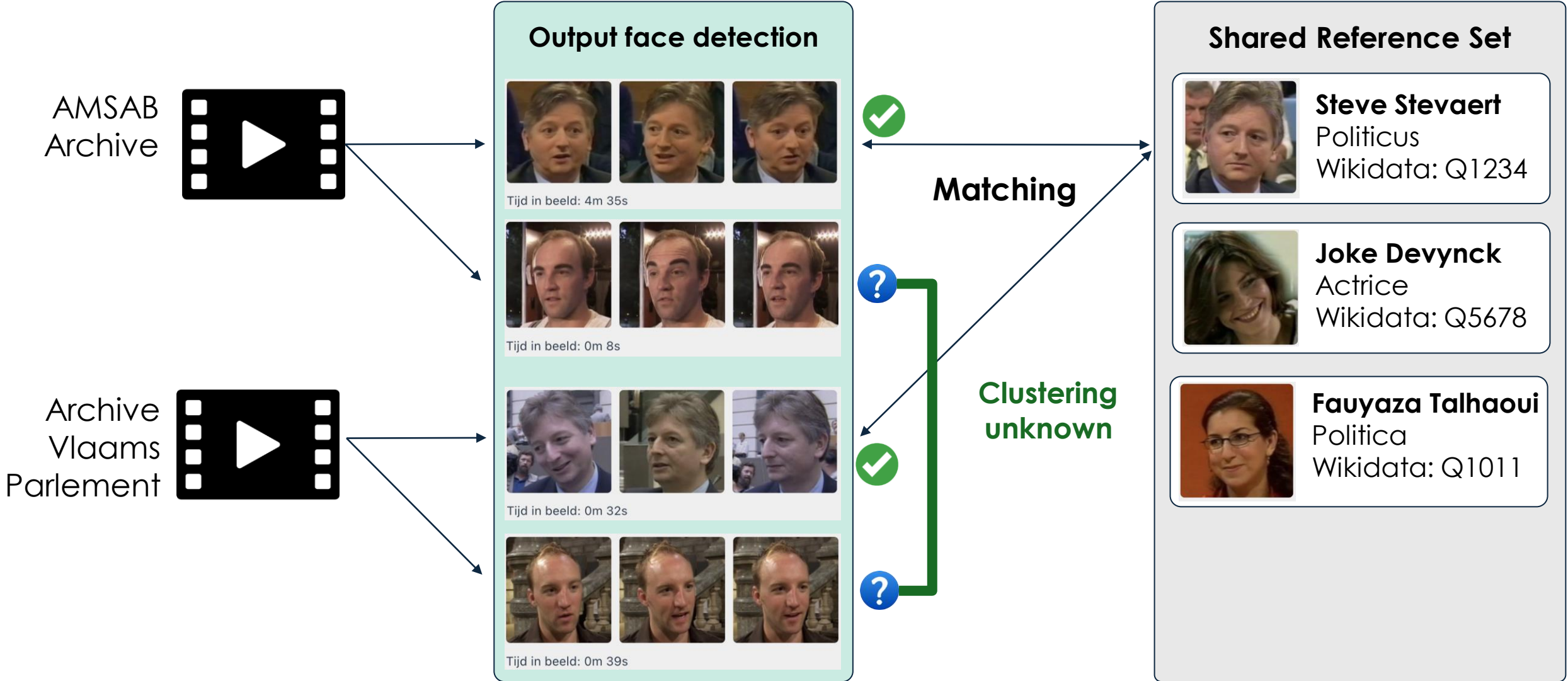
05:48

http://give-face-qas.private.cloud.meemoo.be/task_results/baba7b5ef1d2459180c2ea8ffa49d02cf2aed7b29ba41758f43c7e1eae457a1b6a6377d6b4d4e2c8b032c69a4597bbc_af862059c93c6475d5320919c7c48b7d

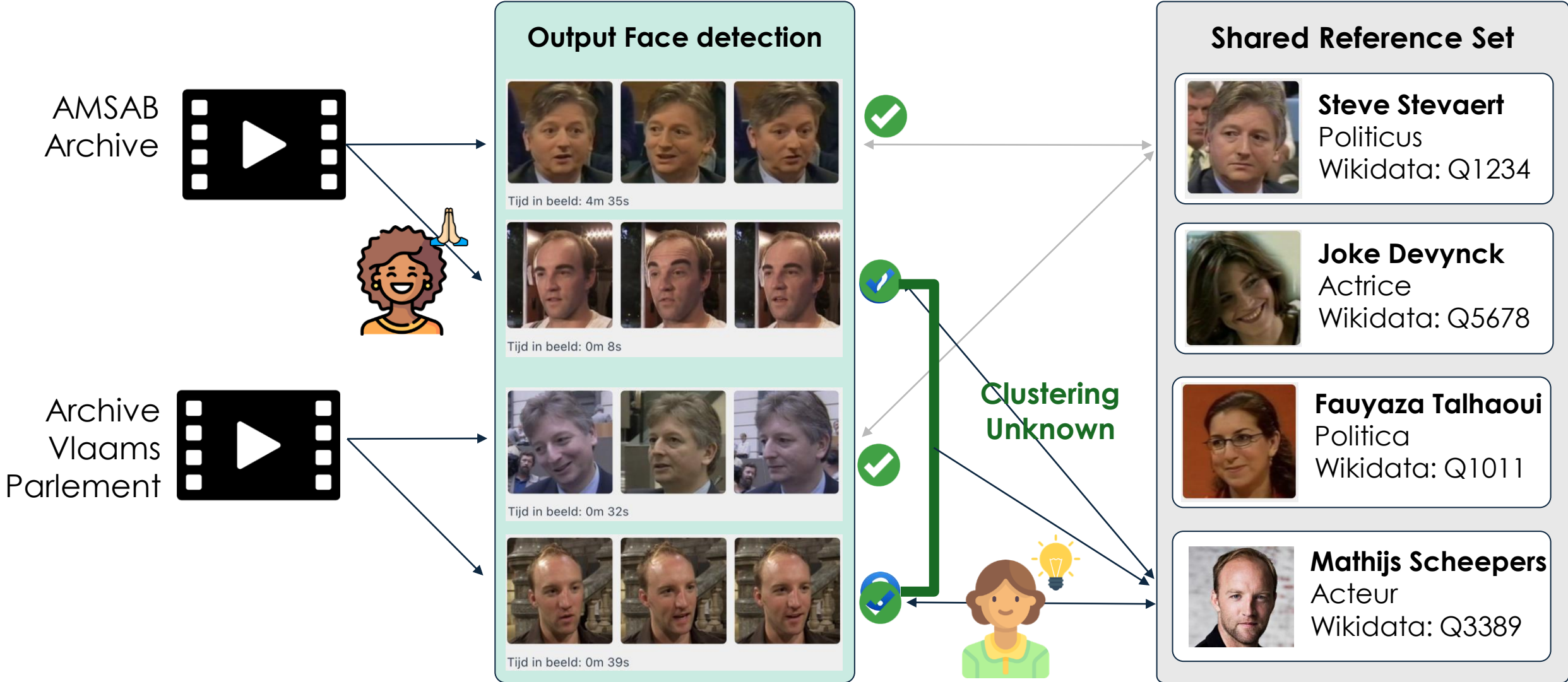
2:00 http://give-face-qas.private.cloud.meemoo.be/task_results/99496d03b4904e1fa23c6b0a203d36db8651dc32f64048bb995ff493d1bc395c28cb9a1f48db84a5c9b10192aba6e9d2c_af862059c93c6475d5320919c7c48b7d

http://give-face-qas.private.cloud.meemoo.be/task_results/99496d03b4904e1fa23c6b0a203d36db8651dc32f64048bb995ff493d1bc395c28cb9a1f48db84a5c9b10192aba6e9d2c_af862059c93c6475d5320919c7c48b7d

Matching faces to the reference set



Unknown? Call an archivist!









Frequently detected but unknown

GiVE Referentieset [Herkende personen](#) [Onbekende personen](#) Matthias Priem (ADV) ▾

Onbekende persoon

Eigen materiaal
Aantal video's: 4
Tijd in beeld: **16m 9s**

[Nieuwe persoon aanmaken](#)

Frame rotos	Titel	Duurtijd	Tijd in beeld	MAM
	De Zevende Dag, TV1: 16 december 2001 / De Zevende Dag, TV1: 10 februari 2002 / De Zevende Dag, TV1: 17 februari 2002 / De Zevende Dag, TV1: 10 maart 2002 (resolutie Voorwaarts)	3h 8m 30s	9m 46s	3f4km06c4k 
	Vlaams-Waals PVG - VNOS Naaimachines 1-3	1h 3m 8s	4m 54s	r785h8f63q 
	Kopies VNOS - VRT archiefbeelden voor VNOS 07/09/1999: Vlaamse en federale regering	1h 15m 10s	1m 13s	gq6qz22s80 

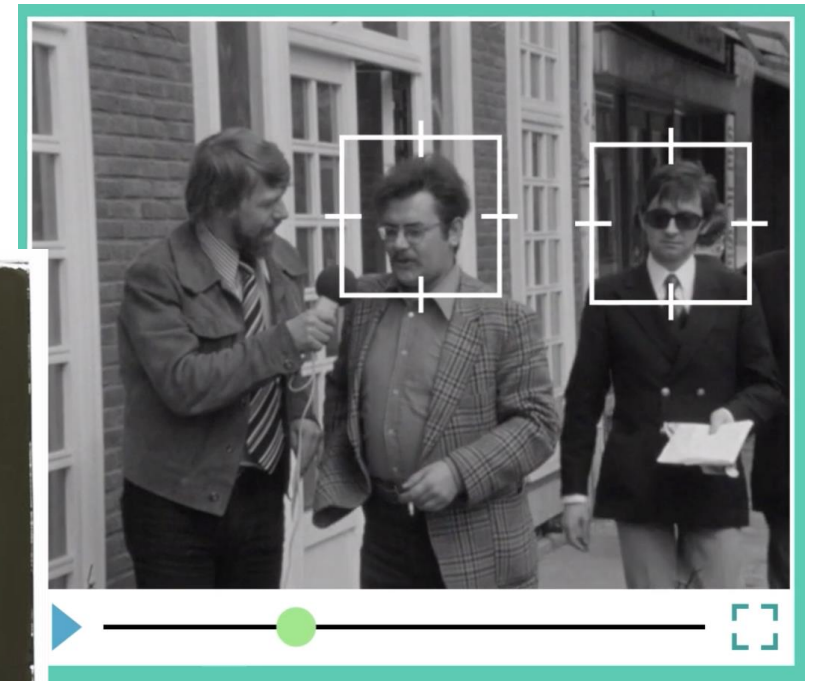
Reference Set

- What is it?
 - A list of people + photo's + links to wikidata + links to authorities of partners
 - Shared management, make use of each others knowledge (musea, archives, media)
 - Transparency: who changed what when?
- Crucial element:
 - It crosses the bridge between detection and matching
 - **Ethical: adding somebody to the reference set is a decision to recognize this person. This is a human decision.**

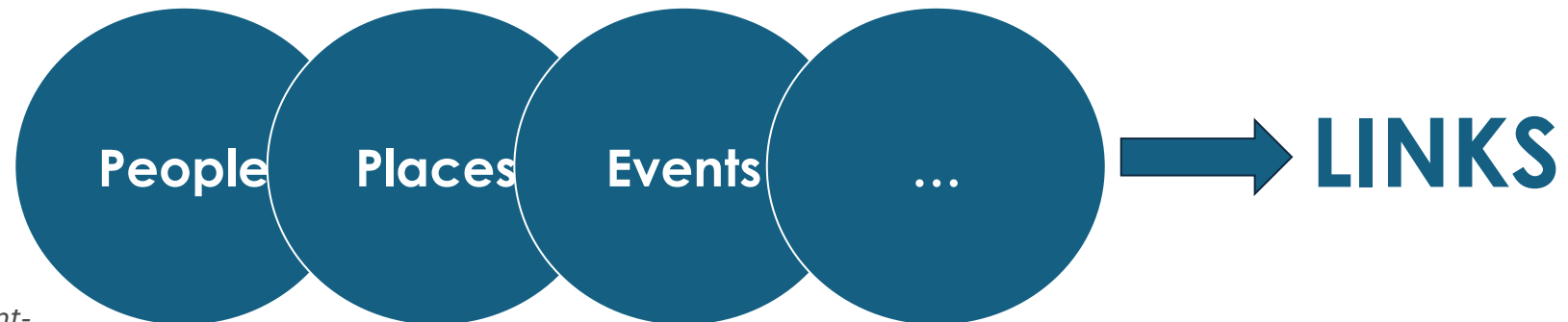
Metadata Mayhem



Gazette van Gend, 27-11-1872, En annonces ofte waerschouwings-blad:...", Universiteitsbibliotheek Gent-Boekentoren, Publiek domein

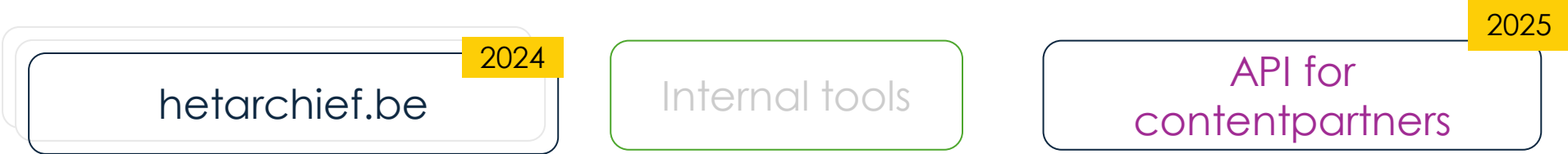


Bronvermelding: Rabindranath Tagore, 1920, T1192, Rabindranath Tagore in de Gasthuisstraat 38 te Antwerpen bij Dr. Arthur de Groodt, Letterenhuis, Publiek domein



Linking and using

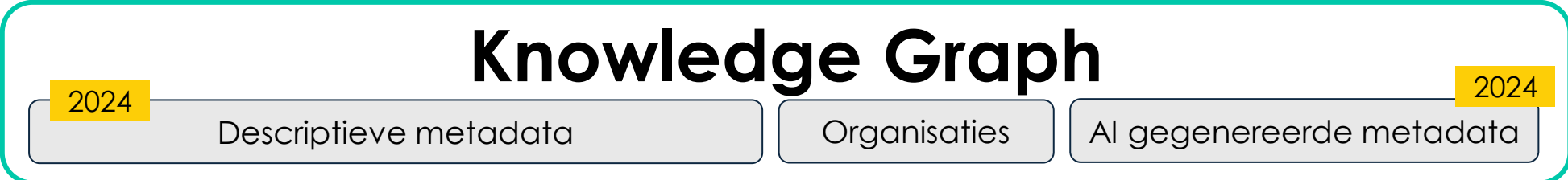
interaction
(platforms)



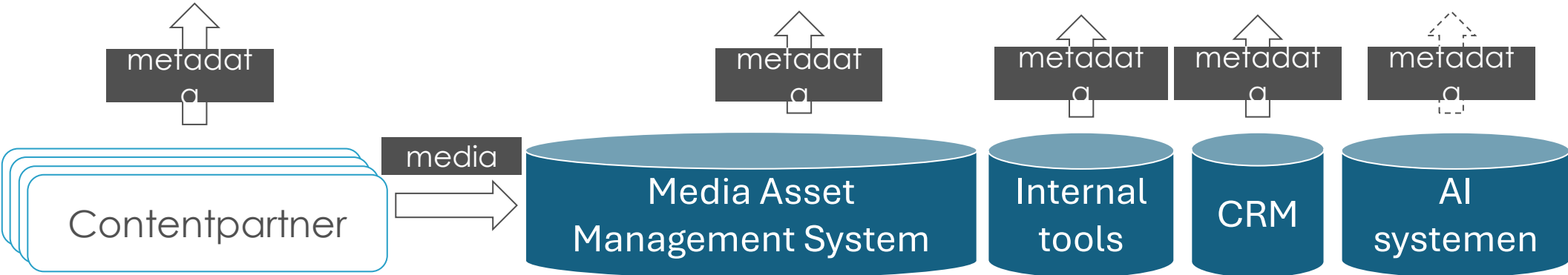
integration



metadata
management



sources



Conclusions

- Partners and meemoo very happy with the project output
- See: ingreatdetail.be / totindetail.be to learn more
- Success factors:
 - Using mature techniques
 - Analysis and benchmarking
 - Integrating legal and ethical track
 - Involving partners: participation and information



Video on using AI for archival metadata

Way forward

- **Follow-up projects already started:**
 - **SHARED AI project:** Applying and expanding the GIVE processes for/with regional media + VRT
 - **Visual Name Authority:** Expanding the reference set tool *as a shared authority*, Work on processes, legal and ethical issues + extend the tool
 - **Hermes:** embed the AI flow in a more flexible software framework, test and execute new techniques (voice identification, audio categorization, summary generation...)

The End!

nico.verplancke@meemoo.be