MIMEHIST: Annotating Eye's Jean Desmet Collection

AMIA, Portland, November 29, 2018

Universiteit van Amsterdam

eye

Hogeschool van Amsterdam

Beeld en Geluid

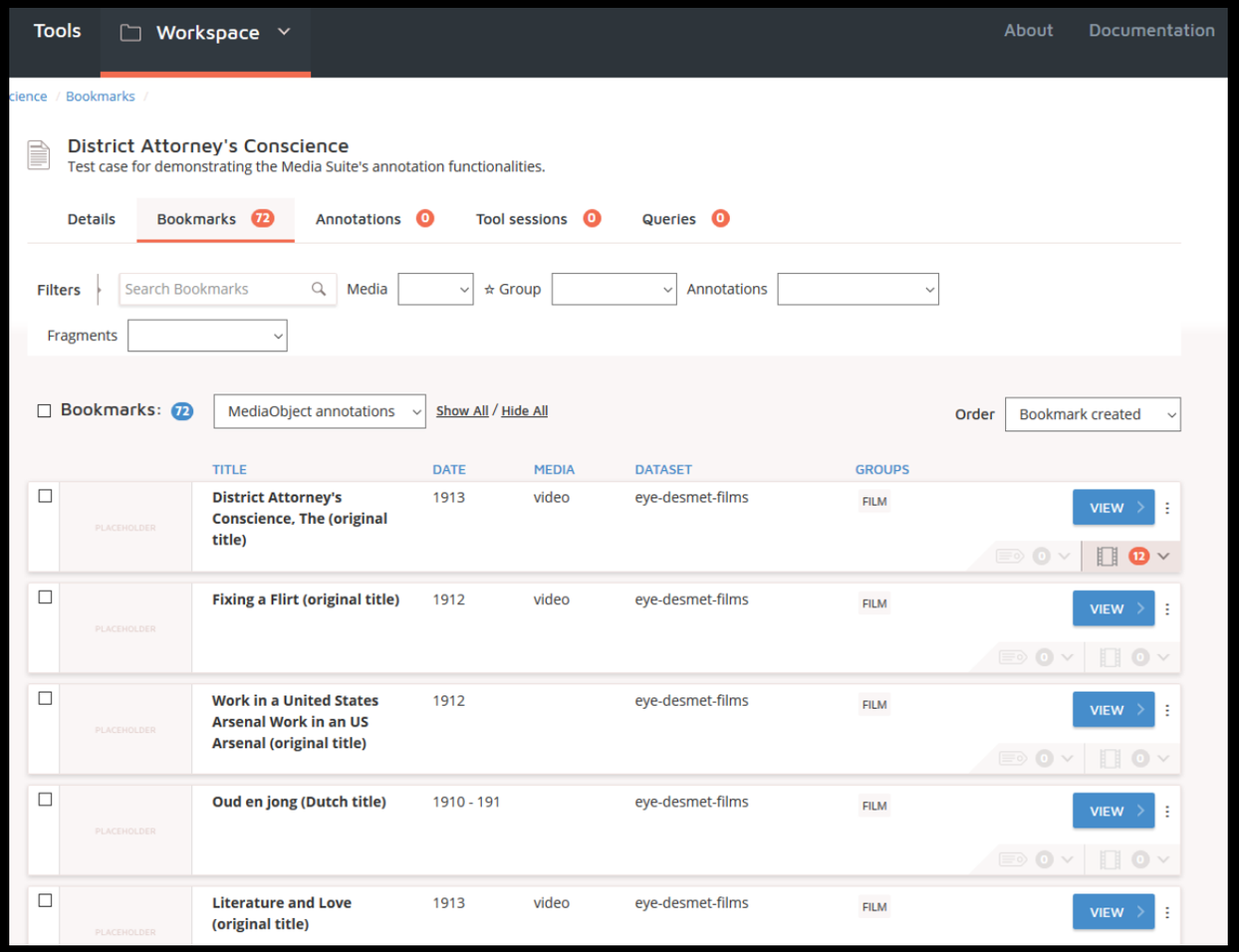
CLARIAH
Common Lab Research Infrastructure for the Arts and Humanities

What we do in MIMEHIST

Jean Desmet (1875-1956)

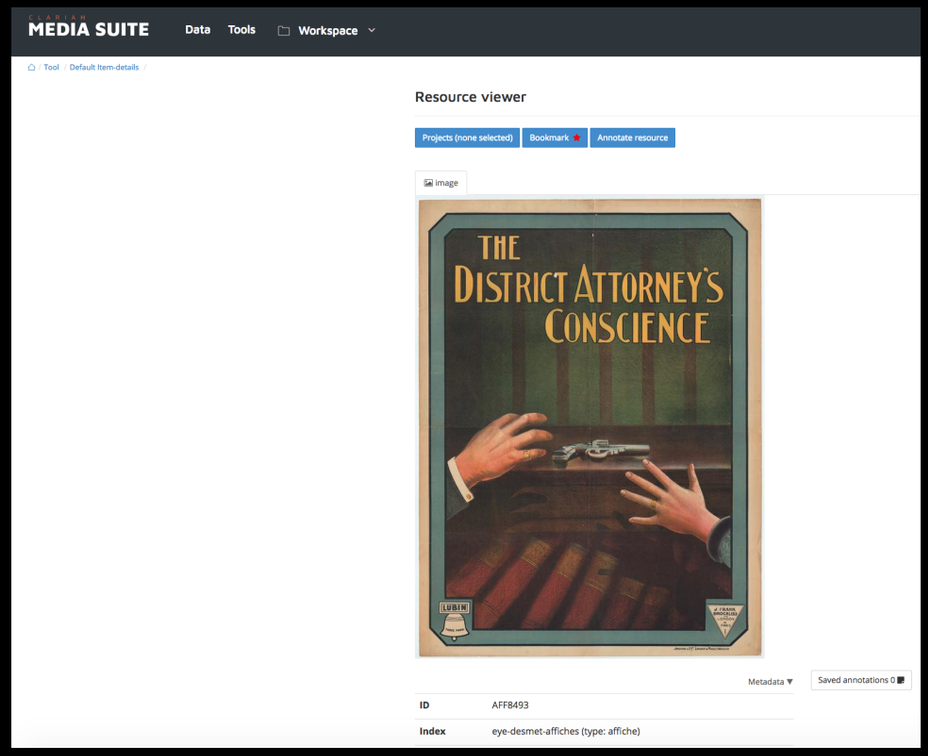
- CLARIAH Mediasuite Pilot Project: UvA, EYE Filmmuseum, NISV
- Desmet Collection and EYE's "Defragmentation" problem
  - Video annotation as a solution.
- Allow for linking and annotating materials in personal user spaces
  - Embed Desmet Collection in the CLARIAH Media Suite
  - Automated analysis of business archive

Bookmarking and Corpus Building



- Corpus building in personal user spaces
- EYE's complete Desmet Collection Metadata
- Films, posters and business archive embedded via EYE server

Segmentation and Annotation



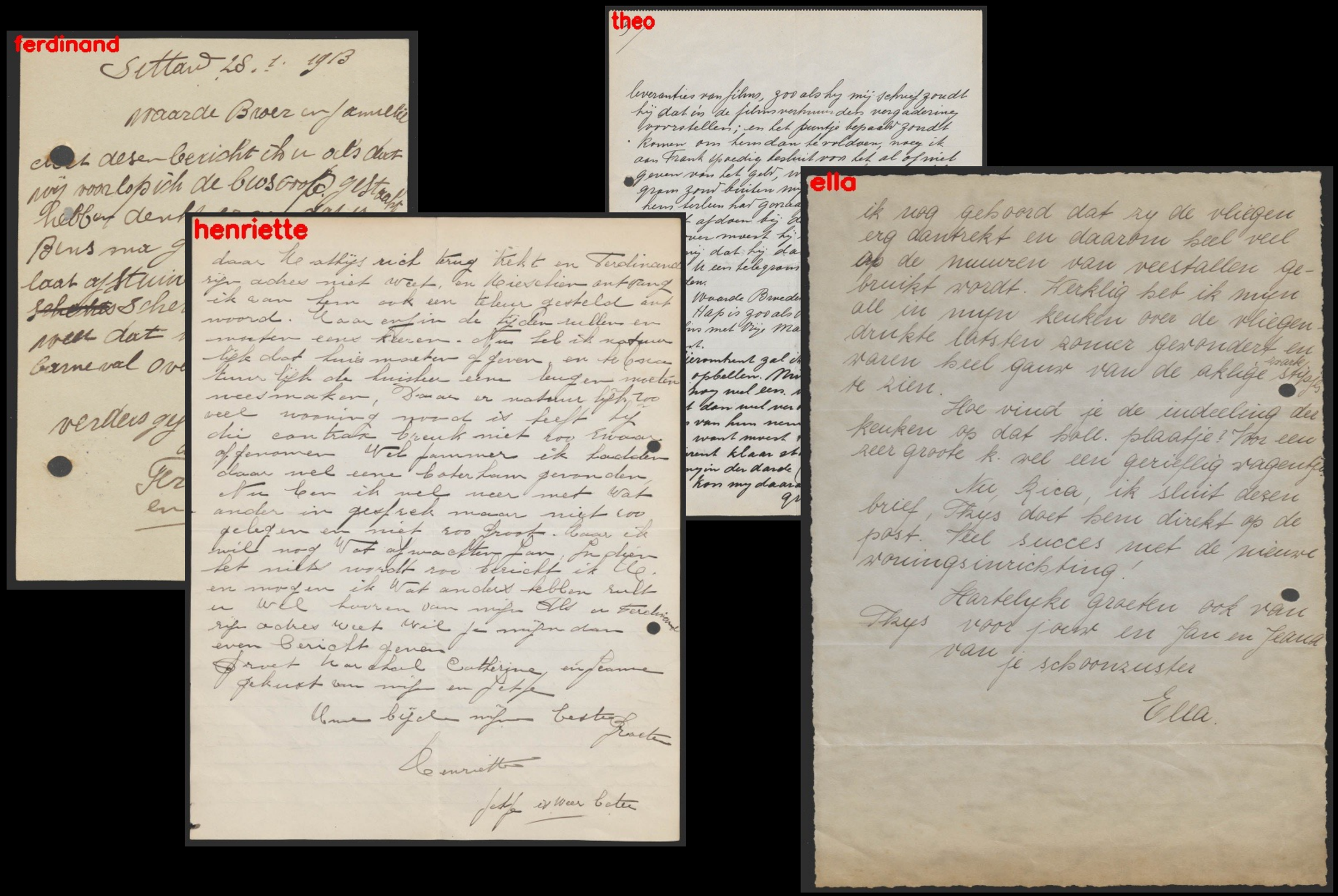
The District Attorney's Conscience (Lubin, US,1912)

- Going beyond Eye's archival finding aid
- Mediasuite annotation functionalities for linking films and related documents
- Explore a film's distribution and exhibition history through e.g. posters, letters

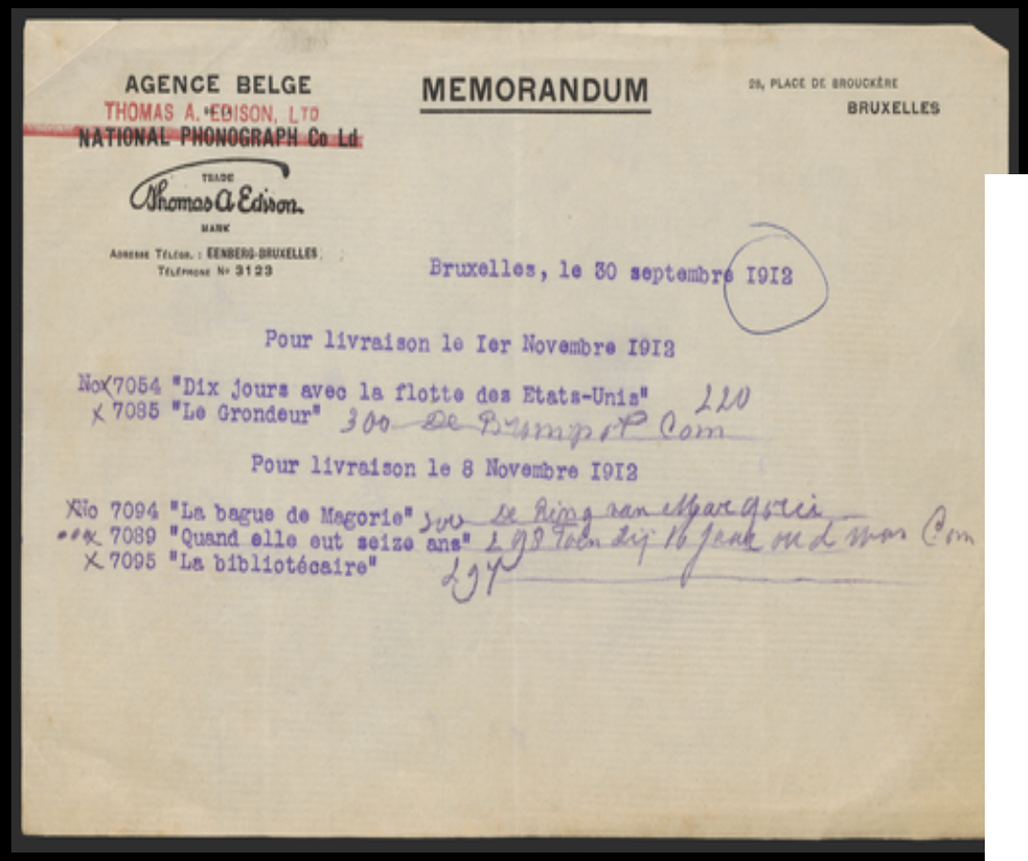
Automated analysis of business archive

- Materially extremely heterogeneous business archive
- EYE's metadata is an excellent starting point, but only folder-specific
- Unlock business archive with OCR and AI methods
- Data included in the CLARIAH Mediasuite with EYE's Metadata

Handwriting identification



Similar layouts


AGENCE BELGE
THOMAS A. EDISON, LTD
NATIONAL PHONOGRAPH Co Ltd

MEMORANDUM

29, PLACE DE BROUCKÈRE
BRUXELLES

Thomas A Edison

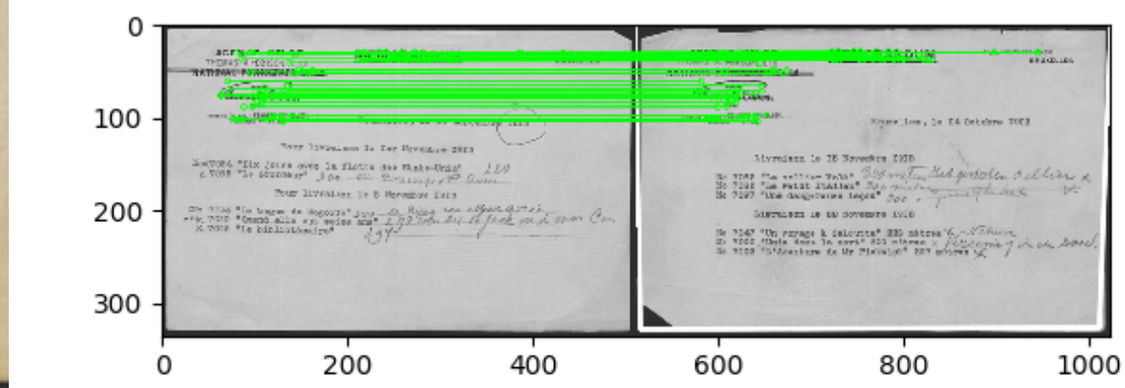
Bruxelles, le 30 septembre 1912

Pour livraison le 1er Novembre 1912

7054 "Dix jours avec la flotte des Etats-Unis" 120
7095 "Le Grondeur" 300

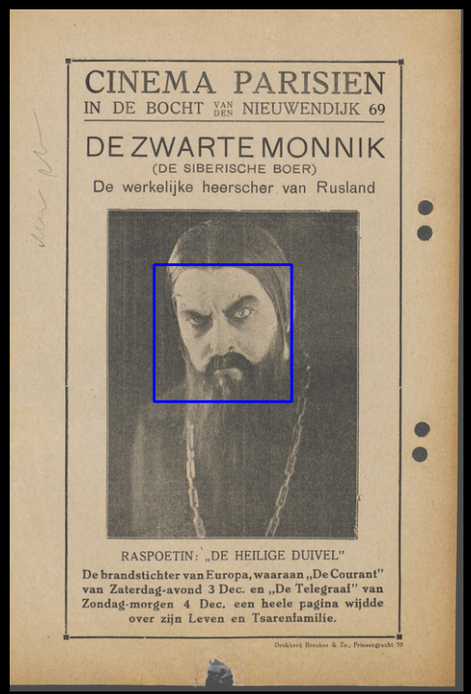
Pour livraison le 8 Novembre 1912

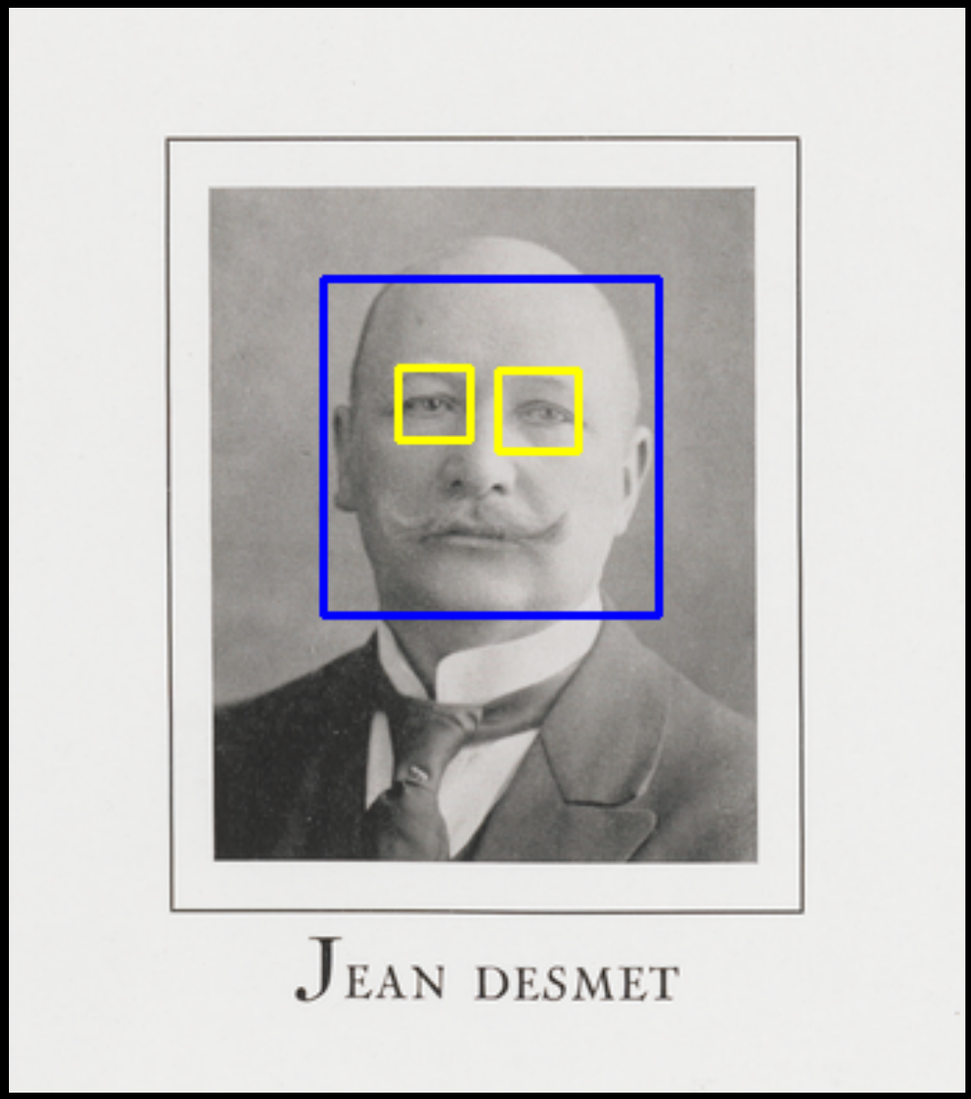
7094 "La bague de Magorie"
7089 "Quand elle eut seize ans"
7095 "La bibliotécaire"



- Similarity index comparison + SIFT (scale-invariant feature transform)
- 27758 documents into 815 clusters,
  - particular individuals, specific institutions, cinemas, production companies, government institutions, grocery stores, electricians, tax forms, rental contracts, film lists, etc.

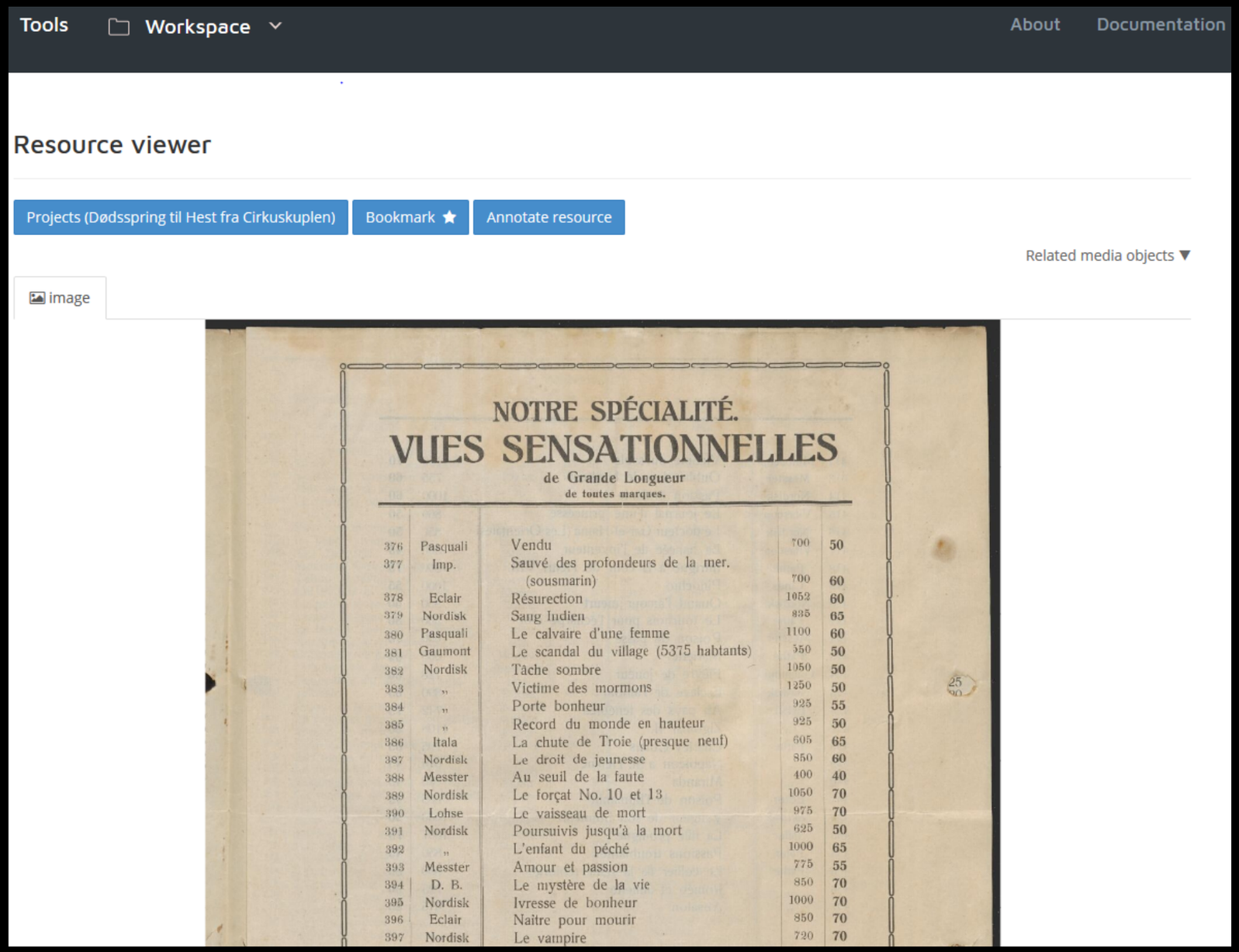
Face detection





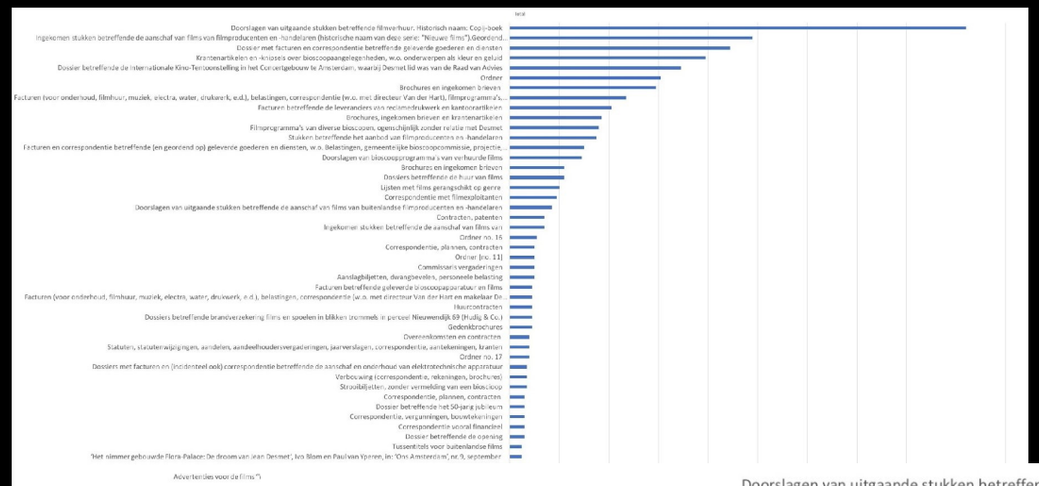
- Frontal face
- Face profile
- Eyes
- Upper body

Business Documents in Media Suite



- 127.000 business documents in Media Suite, OCR data added

Business archive corpus building 1



- Only one of these folders has one of the keywords in its name
- Type-written sources come from 136 different folders (largest category: 184)

Test Cases

- Test Case 1: *Dødsspring til Hest fra Cirkuskuplen* (Dir.: Schnedler-Sørensen, Prod.: Nordisk Films Kompagni, 1912)
  - create a “historical-critical” edition primarily with what is available in the Media Suite

MIMEHIST results

- A significant and widely used historical collection added to the CLARIAH infrastructure
- Unlocking Desmet's business archive and gain new insights into distribution/exhibition history with data enrichments
- Experiment with combining methods: OCR *and* visual classification can support each other

Thanks for your attention!

c.g.olesen@uva.nl