Distributed Annotation ‘n’ Enrichment

Nanne van Noord

Challenges for (heritage) AV collections

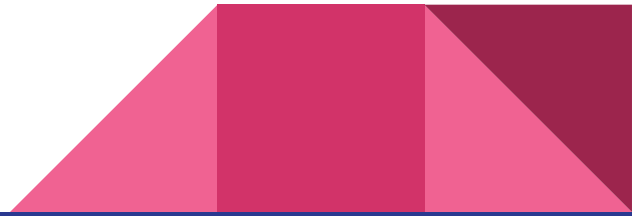
- Large-scale datasets with unique properties
- Manual exploration infeasible
- Audiovisual processing (AVP) tools not or limited available to
  - Heritage institutions
  - (Digital) Humanities scholars
- Cloud-services cannot be used due to
  - IPR / Copyright agreements
  - Costs (of storage)

Interest Group AV-analysis

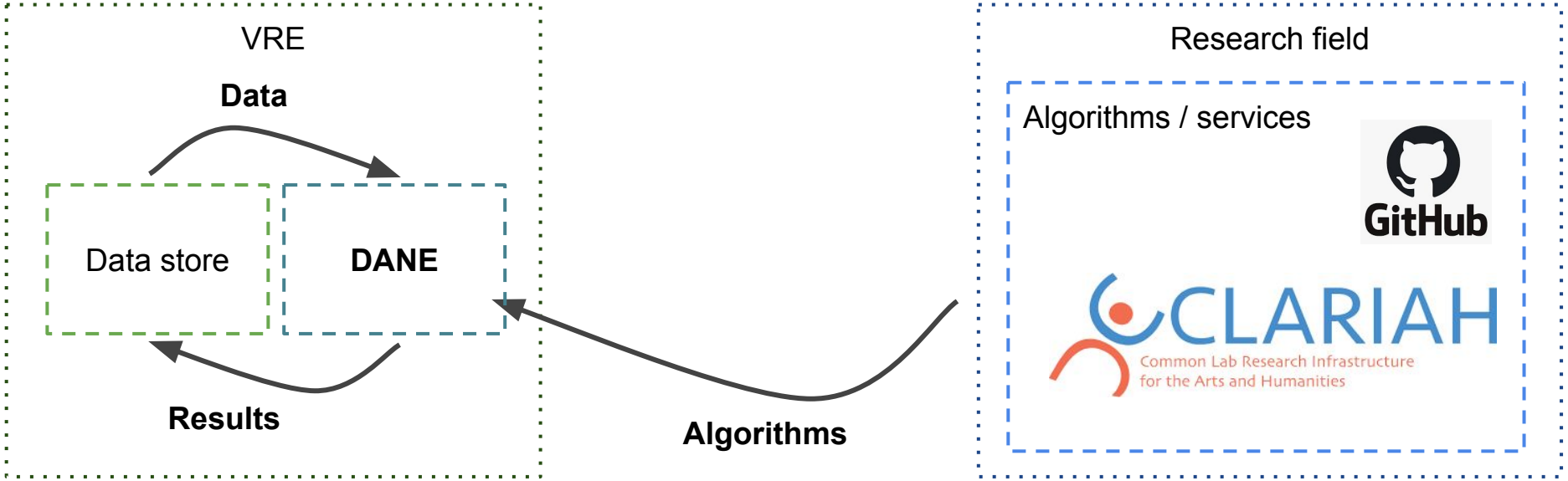
- The IG focuses on analysis tools for audiovisual content such as speech recognition and computer vision
  - These tools enable “semantic access” to content that typically consist of pixels and/or samples and typically only have limited metadata
- Goal is to make these tools available for the CLARIAH infrastructure, specifically multimedia data repositories, in a flexible and sustainable manner
  - The tools should be applicable to both (very) large ‘institutional’ data collections and smaller data sets from individual researchers
  - The tools should adhere to requirements of scholarly research in terms of functionality and quality (tool criticism)

Components

- Individual tools:
  - Speech recognition
  - Speaker identification
  - Computer Vision
  - NLP tools that use previous as input such as
    - Named Entity Recognition
    - Topic Modelling
    - Sentiment Analysis
- **Processing pipeline**:
  - Deploy tools efficiently on (high performance) computer clusters
  - Enable robust handling of input and output data
- **(Set-up that is easily portable within CLARIAH and to cloud infrastructures)**



Distributed Annotation ‘n’ Enrichment (DANE)



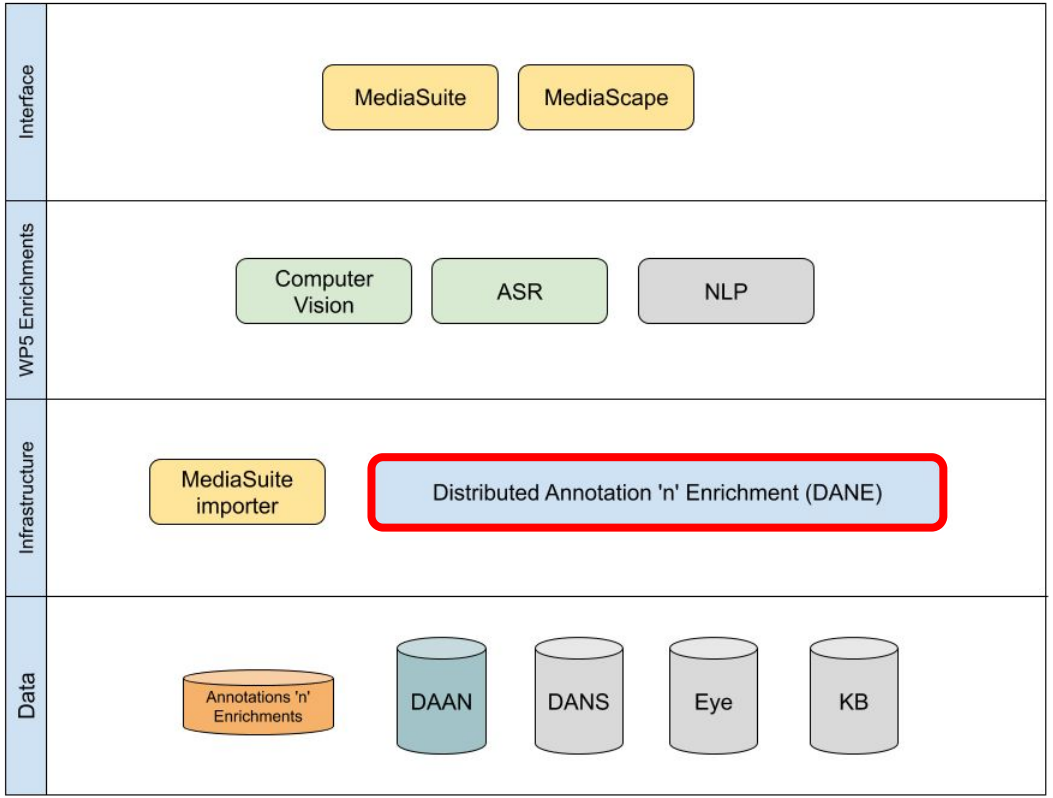
https://github.com/CLARIAH/DANE

DANE context

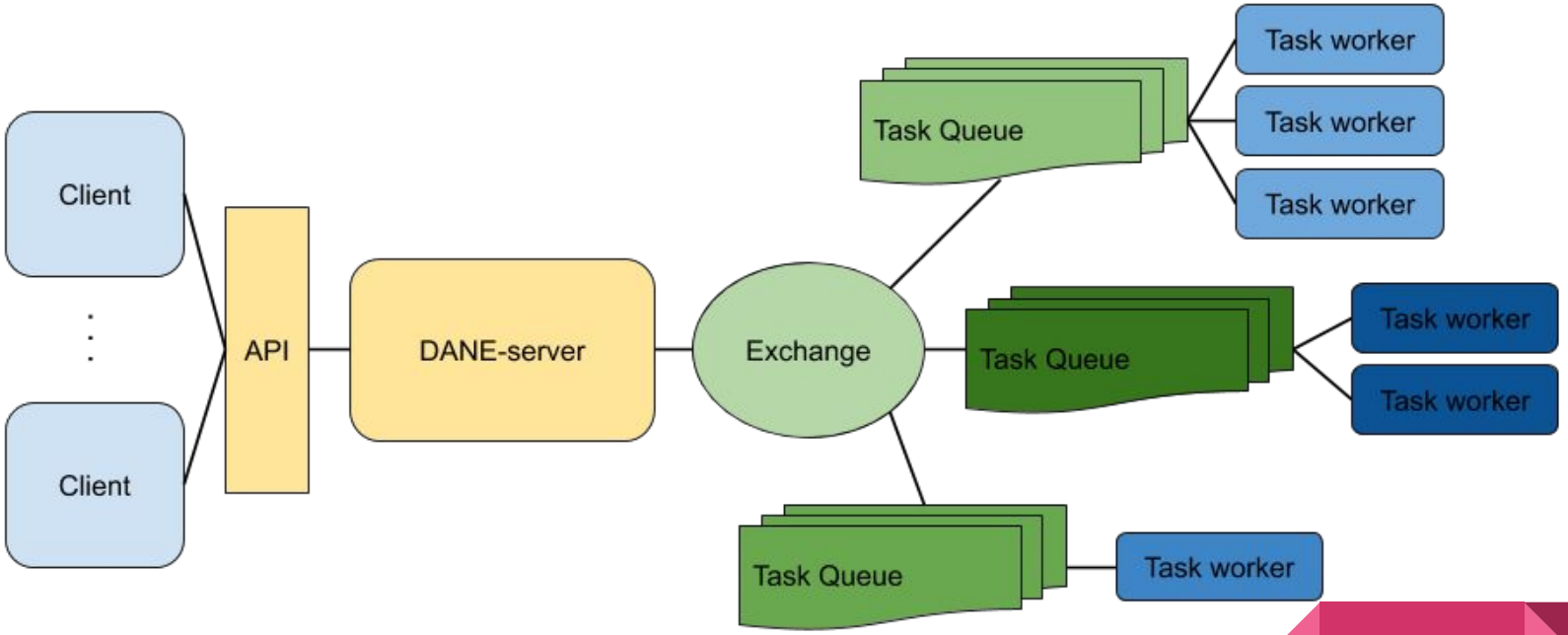
Goal: Make rich metadata, extracted from unstructured GLAM data, available to end-users.

Requirements:

- Access to data
- Access to compute resources
- Modular design
- Available to end-user



Enrichment infrastructure



```
task: {
      key: "DOWNLOAD",
      state: 200,
      msg: "Succes",
      priority: 1,
      args: {
            "foo": "bar"
      },
}
```

```
target: {
      id: oai:openimages:1160937,
      url:
"https://www.openbeelden.nl/files
/11/60/1160939.1160936.Kittens_1.
mp4",
      type: "Video"
},
creator: {
      id: "openbeelden",
      type: "Organization"
}
```

```
result: {
      generator: {
            id: "90d3285",
            type: "Software",
            name: "DOWNLOAD",
            homepage:
            "git@github.com:beeldengeluid/
            download-worker.git"
      },
      payload: {
            file_path:
            "/DANE-data/1160939.1160936.Ki
            ttens_1.mp4",
            file_type: "video",
            Content-Type: "video/mp4",
            Content-Length: 24190176
      }
}
```

```
"created_at": "2021-02-14T13:25:07",
"updated_at": "2021-02-14T13:25:07"
```

DANE

OVERVIEW WORKERS

DOCUMENT OVERVIEW

target id

creator id

SEARCH

id	target	type	creator
5cb417907c2251ba4a74f7cfb6730d9889dcb952	oai:openimages.eu:1246946	Video	openbeelden
40622f3d3be81d764340ab8cb1154b66f9cf6245	oai:openimages.eu:1046913	Video	openbeelden
e7cbe58df98c2aaff451a43d96cc0d317cdb4cbd	oai:openimages.eu:1249376	Video	openbeelden
451ec07723d1ce0c3a62f6c3fb6244b4fc31ce84	oai:openimages.eu:1039472	Video	openbeelden
9870c2c94dece0a85694d1185c73fe83664c5f34	oai:openimages.eu:1251874	Video	openbeelden
77718675b7def94fb8a41f7e88cd151c96d3a397	oai:openimages.eu:1047967	Video	openbeelden
34d138f0bc5269ead008aebf2485e7c9fe74b60e	oai:openimages.eu:1248737	Video	openbeelden
8609d61b712ebbc860ff02125fe6c4920b7c2f7a	oai:openimages.eu:1047784	Video	openbeelden
9aeef76864fc183de4ad17f76a74bd5caff01832	oai:openimages.eu:1248230	Video	openbeelden
16a8611b97c04b3fa519370f0c54c44d663020ca	oai:openimages.eu:1041338	Video	openbeelden

Rows per page: 10 1-10 of 100

d49aa0c9e4e4f49c5deccb169cea02bb8eed7144

oai:openimages.eu:1253631 - https://www.openbeelden.nl/files/12/53/1267331.1253635.WEEKNUMMER373-HRE00015554_1424000_1475960.mp4

Video

by openbeelden Organization

VIEW JSON DELETE

Tasks assigned to this document:

17ab356dd7710b789417919460b4c1acc8e913b5

DOWNLOAD

200 Success

CREATED AT: 2020-12-03T16:17:21 LAST UPDATED AT: 2020-12-03T16:17:28

RETRY FORCE RETRY RESET DELETE

8e7165e7ee768a8f96fbd8459d731bfa2eb9f54b

SHOTDETECTION

200 Success

LAST UPDATED AT: 2021-01-08T18:37:32

RETRY FORCE RETRY RESET DELETE

6242ad96b80fd8be66e4ae3c4f8797a913979a0b

SCENECLASSIFICATION

200 Success

DANE API 1.0

[Base URL: /DANE]
http://localhost:5500/DANE/swagger.json

API to interact with DANE

document Document operations

POST /document/

DELETE /document/{doc_id}

GET /document/{doc_id}

GET /document/{doc_id}/tasks

documents Batch operations on Documents

task Task operations

result Result operations

DELETE /result/{result_id}

GET /result/{result_id}

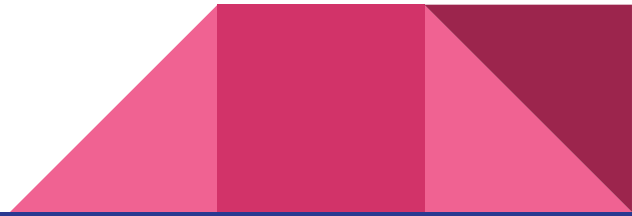
workers Worker operations

search Search operations

Models

Recent trends

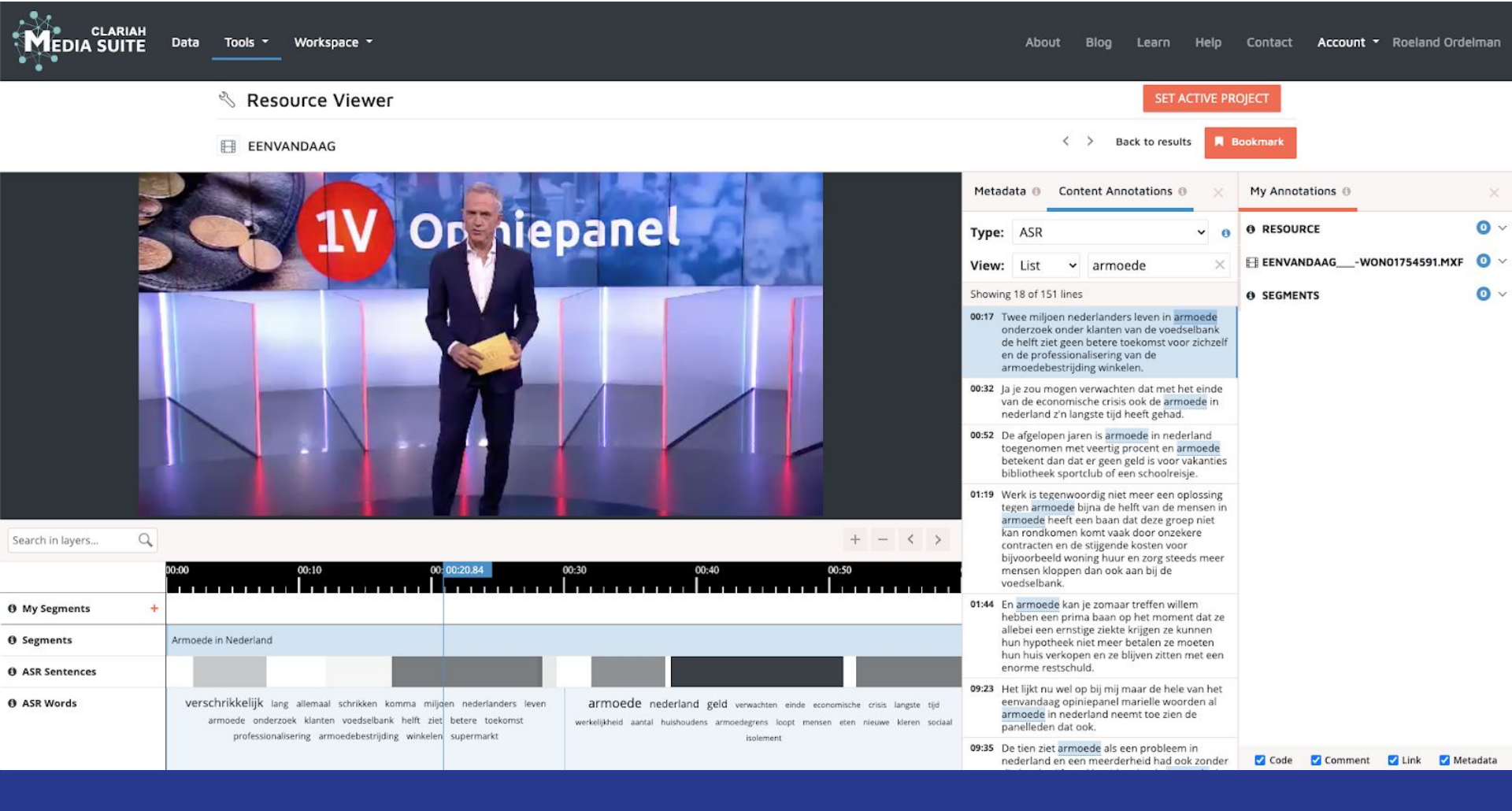
- Deep Learning ‘revolution’
  - Massive progress in terms of performance
  - Increasingly used in industry
- Large-scale digitisation efforts of Audiovisual collections
  - Netherlands Institute for Sound and Vision
    - TV: 650K programs - 200K hours (avg of 20m)
    - Radio: 435K programs - 500K hours (avg >1 hour)



Recent trends

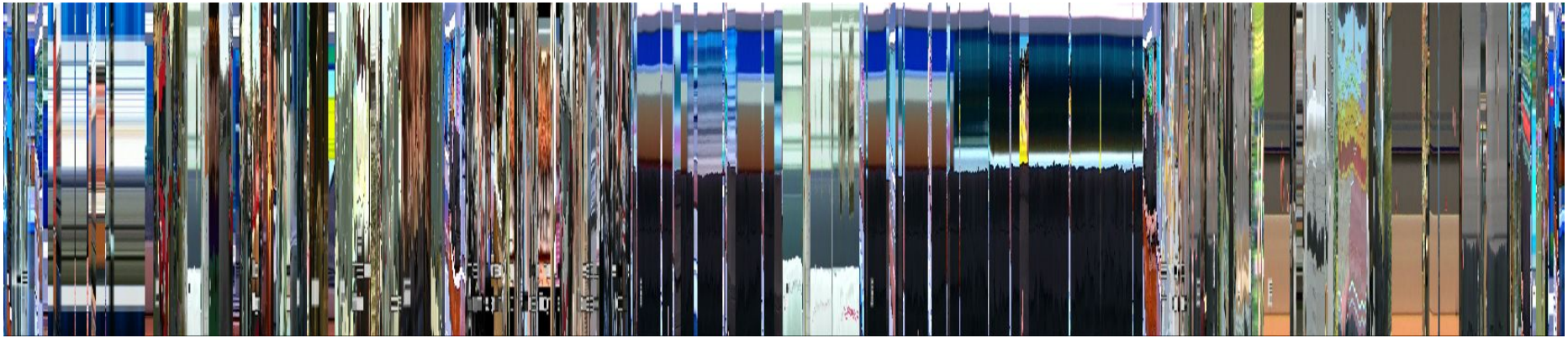
- Deep Learning ‘revolution’
  - Massive progress in terms of performance
  - Increasingly used in industry
- Large-scale digitisation efforts of Audiovisual collections
  - Netherlands Institute for Sound and Vision
    - TV: 650K programs - 200K hours (avg of 20m)
    - Radio: 435K programs - 500K hours (avg >1 hour)
  - Youtube8m [1]: 8 million videos - 500K hours (avg of ~4m)

[1] https://research.google.com/youtube8m/


CLARIAH MEDIA SUITE
Data
Tools
Workspace
About
Blog
Learn
Help
Contact
Account
Roeland Ordelman
Resource Viewer
SET ACTIVE PROJECT
EENVANDAAG
Back to results
Bookmark
1V
Opiniepanel
Search in layers...
00:00
00:10
00:20.84
00:30
00:40
00:50
My Segments
Segments
Armoede in Nederland
ASR Sentences
ASR Words
verschrikkelijk lang allemaal schrikken komma miljoen nederlanders leven armoede onderzoek klanten voedselbank helft ziet betere toekomst professionalisering armoedebestrijding winkelen supermarkt
armoede nederland geld verwachten einde economische crisis langste tijd werkelijkheid aantal huishoudens armoedegrens loopt mensen eten nieuwe kleren sociaal isolement
Metadata
Content Annotations
My Annotations
Type: ASR
View: List
armoede
Showing 18 of 151 lines
00:17 Twee miljoen nederlanders leven in armoede onderzoek onder klanten van de voedselbank de helft ziet geen betere toekomst voor zichzelf en de professionalisering van de armoedebestrijding winkelen.
00:32 Ja je zou mogen verwachten dat met het einde van de economische crisis ook de armoede in nederland z'n langste tijd heeft gehad.
00:52 De afgelopen jaren is armoede in nederland toegenomen met veertig procent en armoede betekent dan dat er geen geld is voor vakanties bibliotheek sportclub of een schoolreisje.
01:19 Werk is tegenwoordig niet meer een oplossing tegen armoede bijna de helft van de mensen in armoede heeft een baan dat deze groep niet kan rondkomen komt vaak door onzekere contracten en de stijgende kosten voor bijvoorbeeld woning huur en zorg steeds meer mensen kloppen dan ook aan bij de voedselbank.
01:44 En armoede kan je zomaar treffen willem hebben een prima baan op het moment dat ze allebei een ernstige ziekte krijgen ze kunnen hun hypotheek niet meer betalen ze moeten hun huis verkopen en ze blijven zitten met een enorme restschuld.
09:23 Het lijkt nu wel op bij mij maar de hele van het eenvandaag opiniepanel marielle woorden al armoede in nederland neemt toe zien de panelleden dat ook.
09:35 De tien ziet armoede als een probleem in nederland en een meerderheid had ook zonder
RESOURCE
EENVANDAAG___-WON01754591.MXF
SEGMENTS
Code
Comment
Link
Metadata

Time-coded enrichments

Shot detection / Keyframe selection



"shots": [
[0, 124], [126, 232], [234, 281],
[284, 331], [334, 429], [447, 516],
...,
[52224, 52292]
],
"keyframes": [
21, 27, 67, 123, 147, 264, 281, ..., 52051, 52099, 52163, 52291
]

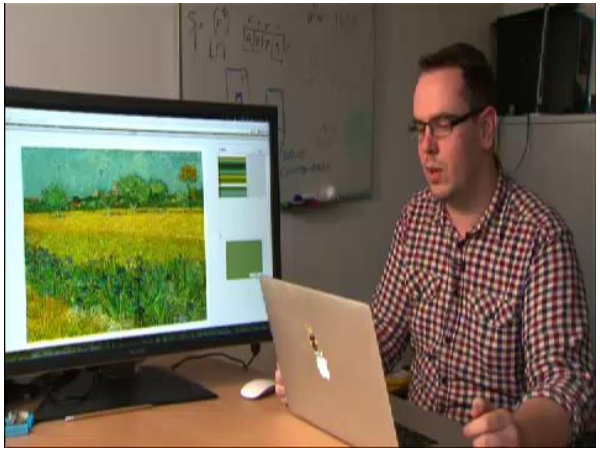
Computer vision enrichments

Text detection

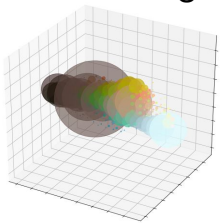
"texts": ["nannevannoord", "onderzoeksprop", "van"]

Object detection

Classification



Colour histogram



Five most dominant colours

Top 5 Objects

monitor	21%
CRT screen	12%
desk	9%
desktop computer	5%
television	4%

Top 5 Places

/o/office	48%
/h/home_office	17%
/o/office_cubicles	6%
/c/computer_room	3%
/a/art_studio	3%

Enrichment progress

- Analysis of openbeelden collection done
  - ~10k videos
  - 2.2M keyframes
  - To be made available in *Open Data Lab*
    - Media Suite version with PD data
- Analysis of Desmet collection (EYE) in progress

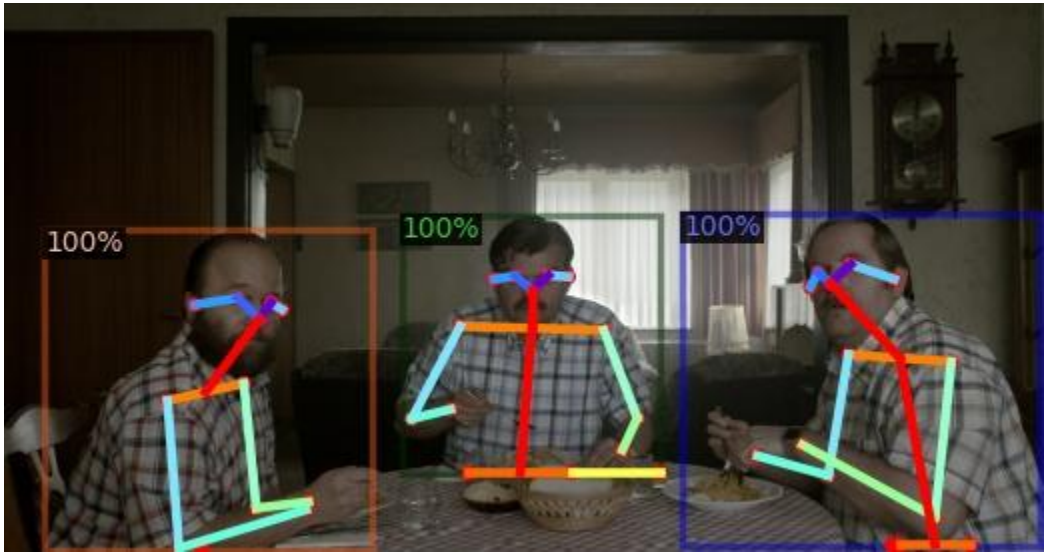
- Todos
  - Integrating automatic enrichments into metadata
  - Interface updates
  - Case studies with researchers

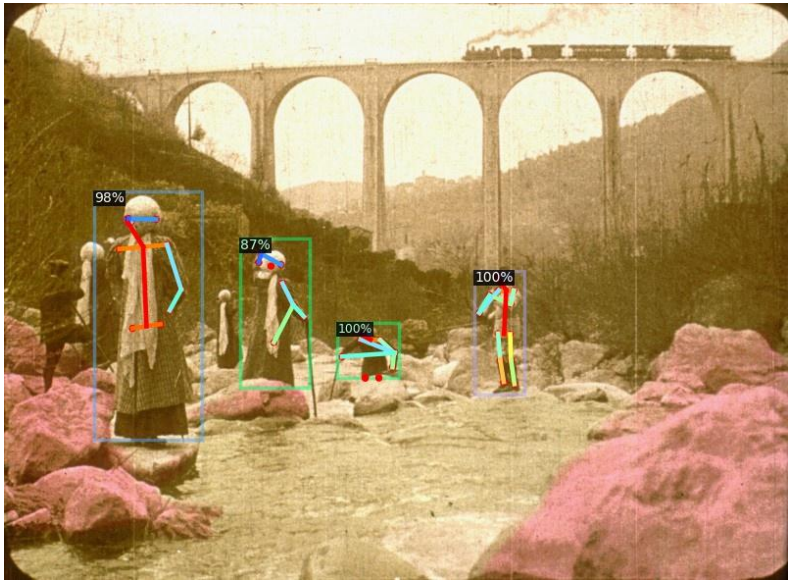
Use case: Pose analysis for genre classification

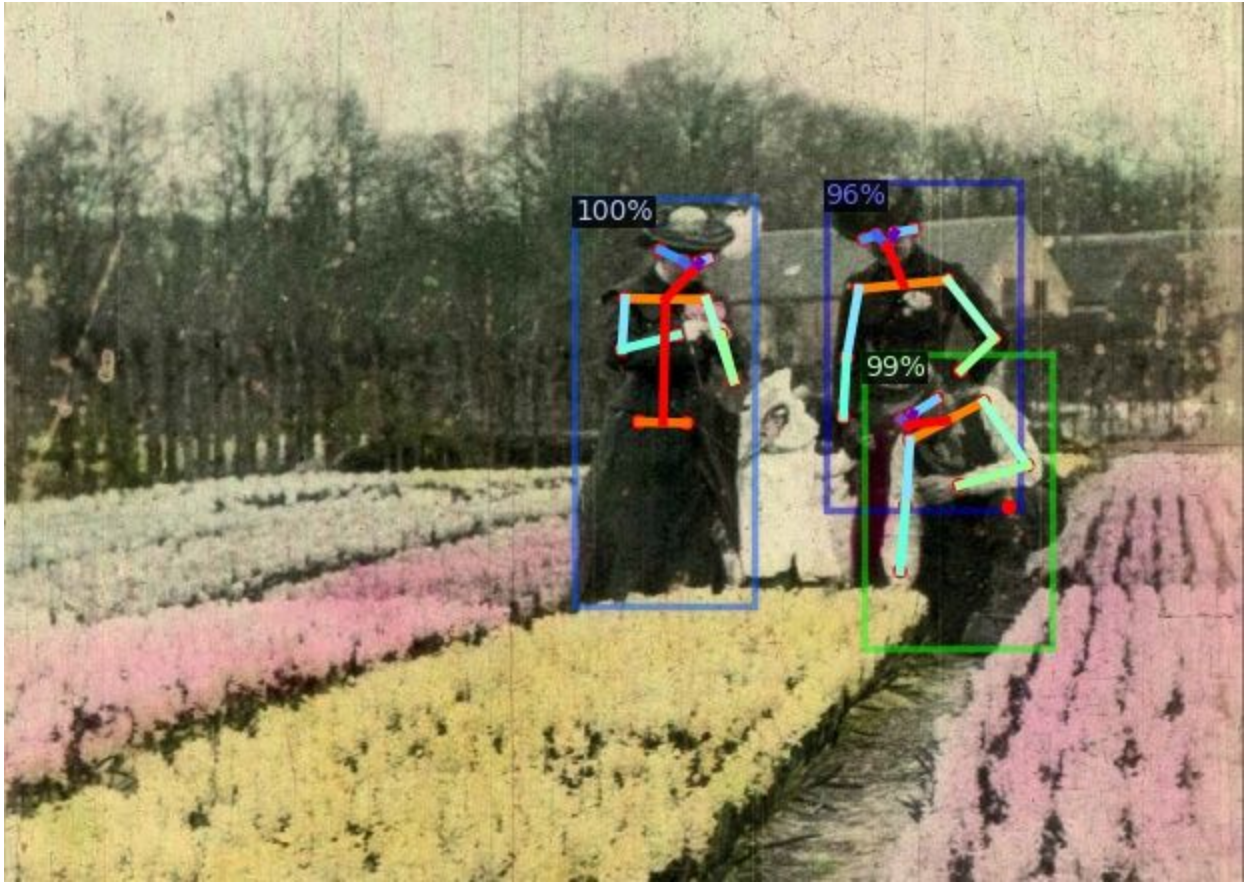
- Theatre actors performed in early cinema
- National ‘traditions’
  - Scandinavian
  - German
  - Italian
- Comedy <> vaudeville

Explore pose recognition to distinguish between genres

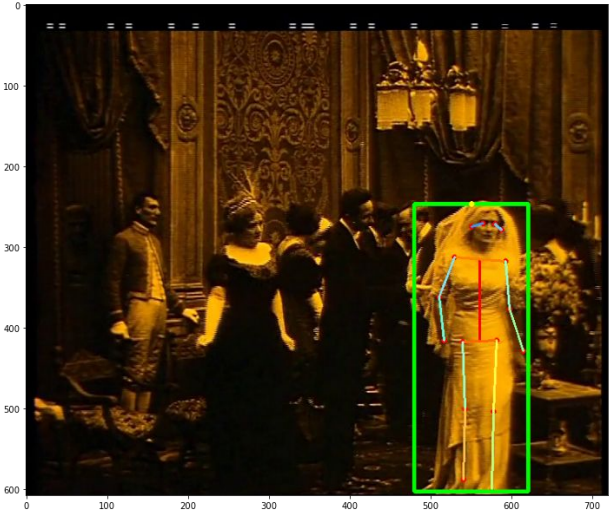
Challenges






100%
96%
99%

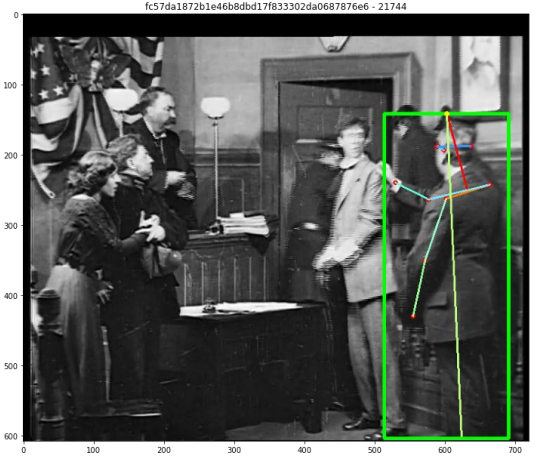
d91eb878774efcaeb3ff2d0a44e6ddfbb370dc57 - 31329

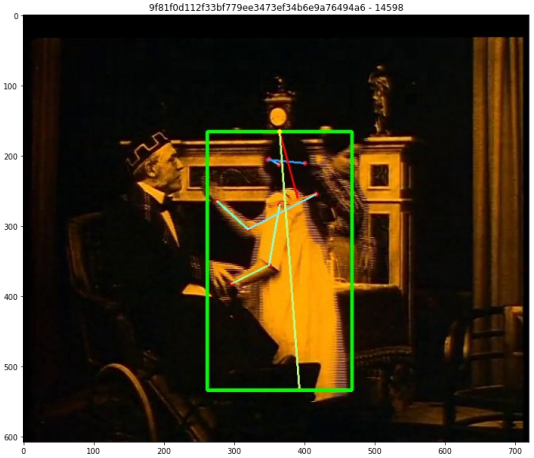


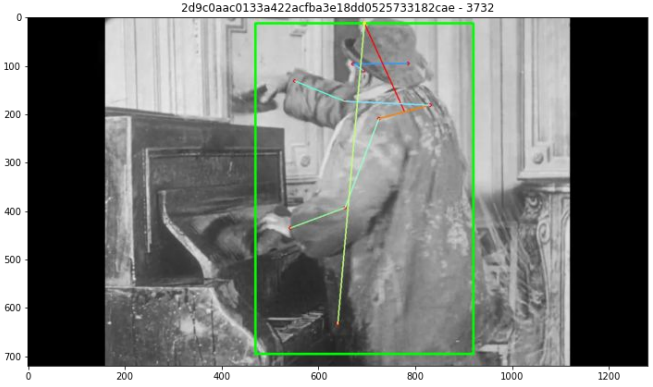
5714d4cad63712437eef3aeee10c3be1526e9a6d - 25502

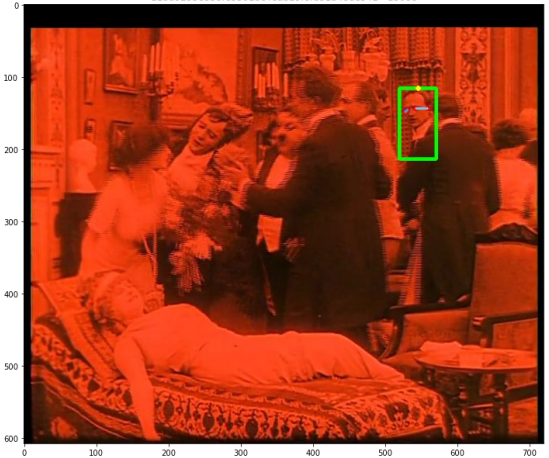
7b648a219e062f3c1ef3f0299f9bbe26077be1c0 - 8519

c5efb50896024209c991cec9ba6f7793d9672b25 - 2219

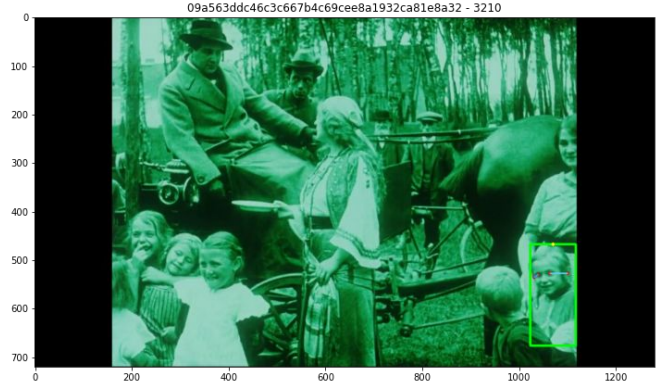

fc57da1872b1e46b8dbd17f833302da0687876e6 - 21744


9f81f0d112f33bf779ee3473ef34b6e9a76494a6 - 14598


2d9c0aac0133a422acfba3e18dd0525733182cae - 3732


11bd9253833bfe5501bc4eab19f8fdb1b48de541 - 13088


f6cca21b93168374872c13ca584f24e243c672e0 - 426


09a563ddc46c3c667b4c69cee8a1932ca81e8a32 - 3210

Use case: Tracing re-use in audiovisual collections

The main goal of this project is tracing the re-use of images and speech throughout the archive.

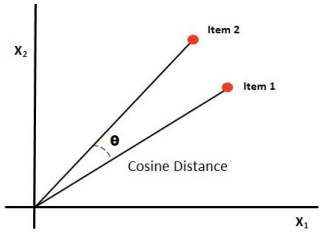
- Initial focus on television re-use
  - Analyse 1 full week of broadcasting
- Generic that it can be broadened to any multimodal framing analysis

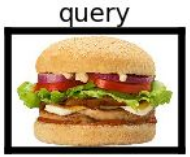

0.99


kro ncrv
OP
npo
SVEN KOCKELMANN &


Bron: VPRO

Dense vector similarity



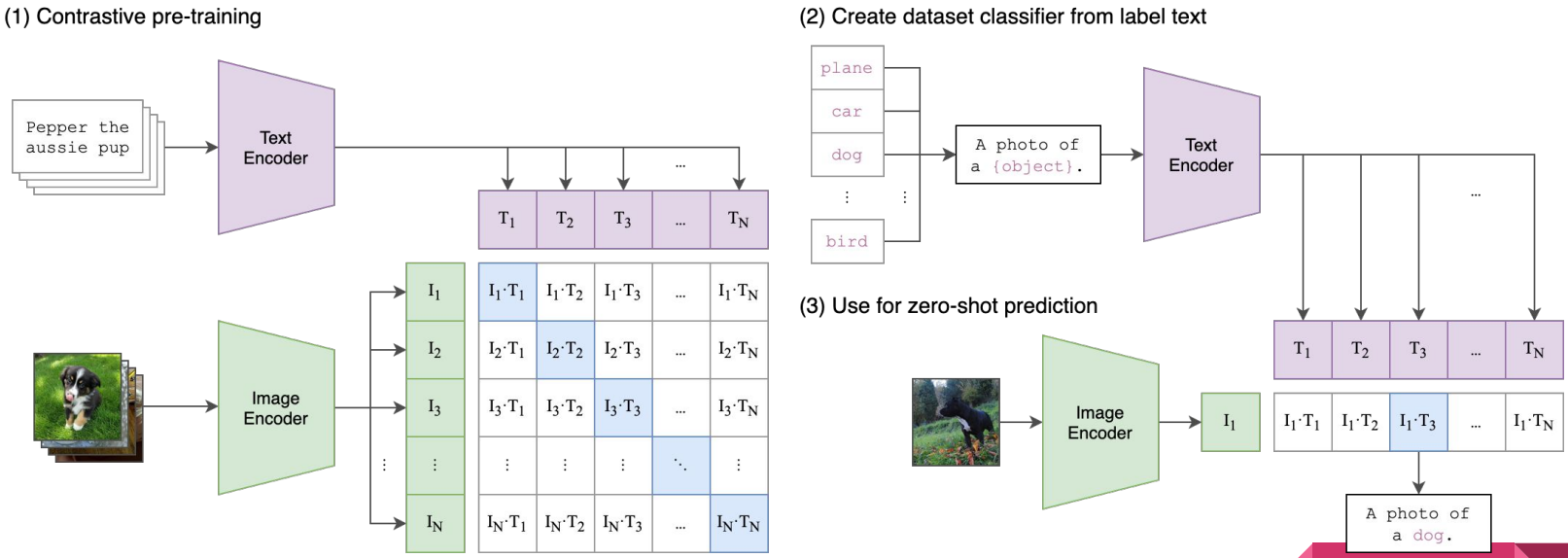



```
"query": {
          "script_score": {
            "query": {
                "match_all": {}
                },
            "script": {
              "source":
"cosineSimilarity(params.query_vector, 'features') + 1.0",
              "params": {"query_vector": query_vector}
            }
          }
        }
```

Video Indexing

	Size	Docs
dane-colourhistogram	872Mi/1.70Gi	2.20M
dane-imageembedding	37.8Gi/75.5Gi	5.89M
dane-poseembedding	5.93Gi/11.9Gi	10.1M

CLIP



Radford et al. (2021) Learning Transferable Visual Models From Natural Language Supervision. Arxiv:2103.00020

Multi-modal retrieval

DANE-bot APP 10:31 AM

Querying for "a school"...

Result for query: "a school" found in https://openbeelden.nl/media/631278

Frame occurs at 00:00:04 in https://www.openbeelden.nl/files/06/31/631292.631277.WEEKNUMMER360-HRE00018B17.mp4#t=4.44

Multi-modal retrieval

A close-up of a car

A wide shot of a car

Birds eye view of a car

Multi-modal retrieval: Failure case

DANE-bot APP 10:45 AM

Querying for "a photo of a bunch of carrots"...

Result for query: "a photo of a bunch of carrots" found in https://openbeelden.nl/media/986907

Frame occurs at 00:03:20 in https://www.openbeelden.nl/files/09/86/986916.986911.1513_WASALAMN-NAT00Z04XMD.mp4#t=200.28

Multi-modal retrieval: Bias?

DANE-bot APP 10:39 AM

Querying for "A European woman"...

Result for query: "An African woman" found in https://openbeelden.nl/media/27849

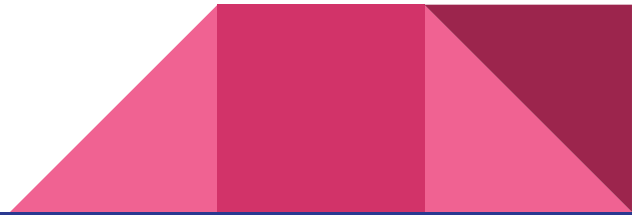
Frame occurs at 00:00:46 in https://www.openbeelden.nl/files/27/27966.27848.WEEKNUMMER531-HRE0000D514.mp4#t=46.92

Software & Documentation

Core components are in development (beta state) and fully open-source:

- DANE 'client' + core logic code: https://github.com/CLARIAH/DANE
- DANE server code: https://github.com/CLARIAH/DANE-server
- Documentation: https://dane.readthedocs.io/en/latest/intro.html

Current workers are not open-sourced (yet), but as shown here hardly any extra code is needed to turn a piece of code into a worker.

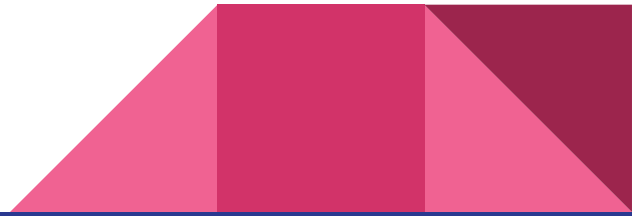


Concluding remarks

- By making tools available in VRE we can have a real influence on DH workflow
- In-house hosting offers flexibility and customisation potential
- Close collaboration with DH scholars

Open challenges

- Personalised classifiers
- Incorporating corrections
- Level of granularity



Thank you!

CLARIAH
Common Lab Research Infrastructure for the Arts and Humanities