



Img source: Rebecca Wilson, Solid State Logic SL 9000J STUDIO A, The Cutting Room Recording Studios, NYC

Mixing computers and cultural heritage  
challenges and opportunities in cross-disciplinary  
collaborations

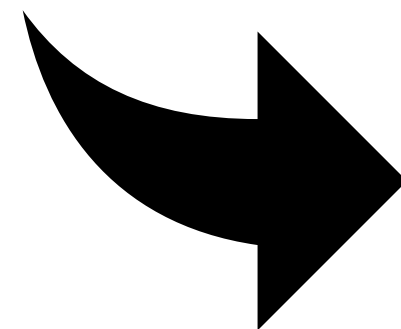
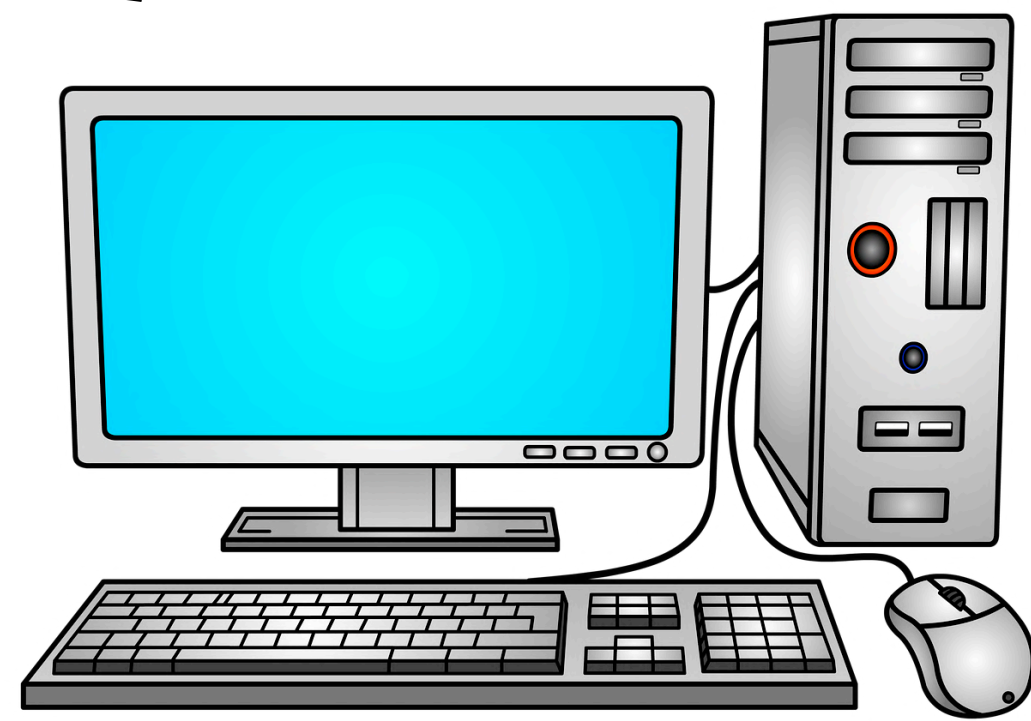
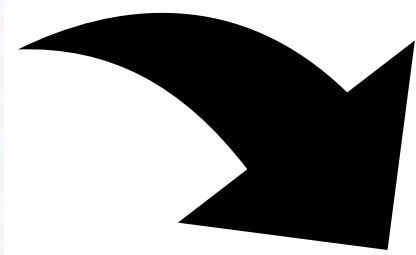
[Marieke.van.Erp@dh.huc.knaw.nl](mailto:Marieke.van.Erp@dh.huc.knaw.nl)



merpeltje









# Cross-disciplinary challenges

---

- Data challenges
- Technology challenges
- Cultural/User challenges





# DINSDAG

Kopje ossestaartsoep

Knakworstjes

Witte bonen met ananas

Lofsla

Witte bonen met ananas

1 of 2 blikken witte bonen in tomatensaus, 1 uitje, 1 eetlepel mosterd, 2 eetlepels keukenstroop, 12 knakwostjes, 1 klein blikje ananasblokjes.

Rasp het uitje en meng dit door de bonen met tomatensaus. Doe de groente in een ingevette vuurvaste schaal. Roer de mosterd door de stroop en giet hier de helft van over de bonen. Dek de schaal af met een deksel of folie en zet hem 20 minuten in een warme oven. Verwarm inmiddels de worstjes in warm water en leg deze warm op de

dinsdag

6 ossestaartsoep

HUt \*orstjes

l 0( \* bonen met ananas

\t t e bonen met ananas

Va0,1 2 blikken witte bonen In 1 uitje

1 eetlepel ?lWd- 2 eetlepels keuken-

12 knakwostjes, 1 klein

„ ftaLananasDlokJes- SoJrJi het uitJe

en meng dit ,Qoe h bonen met

tomatensaus. Nir,;e groente in een

ingevette ?fd h ste schaal. Roer de

mos- Je hni?or de stroop en giet hier

2\* van over de bonen. Dek af met





# Algorithmic transparency

KB newspaper archive (100M articles, 6 months server logs)  
study:

- Reran 1M real user queries through the (black box) search engine
- Counted how often each document appears in top 10 (100, 1000)
- Analysed correlation technical document features with retrieval counts

Findings:

- 96% of the articles never make it into the top 10 (76% never in top 100)
- Engine discriminates against very short and very long documents
- Best scoring articles contain long list of names (local elections, swimming diplomas, ...)

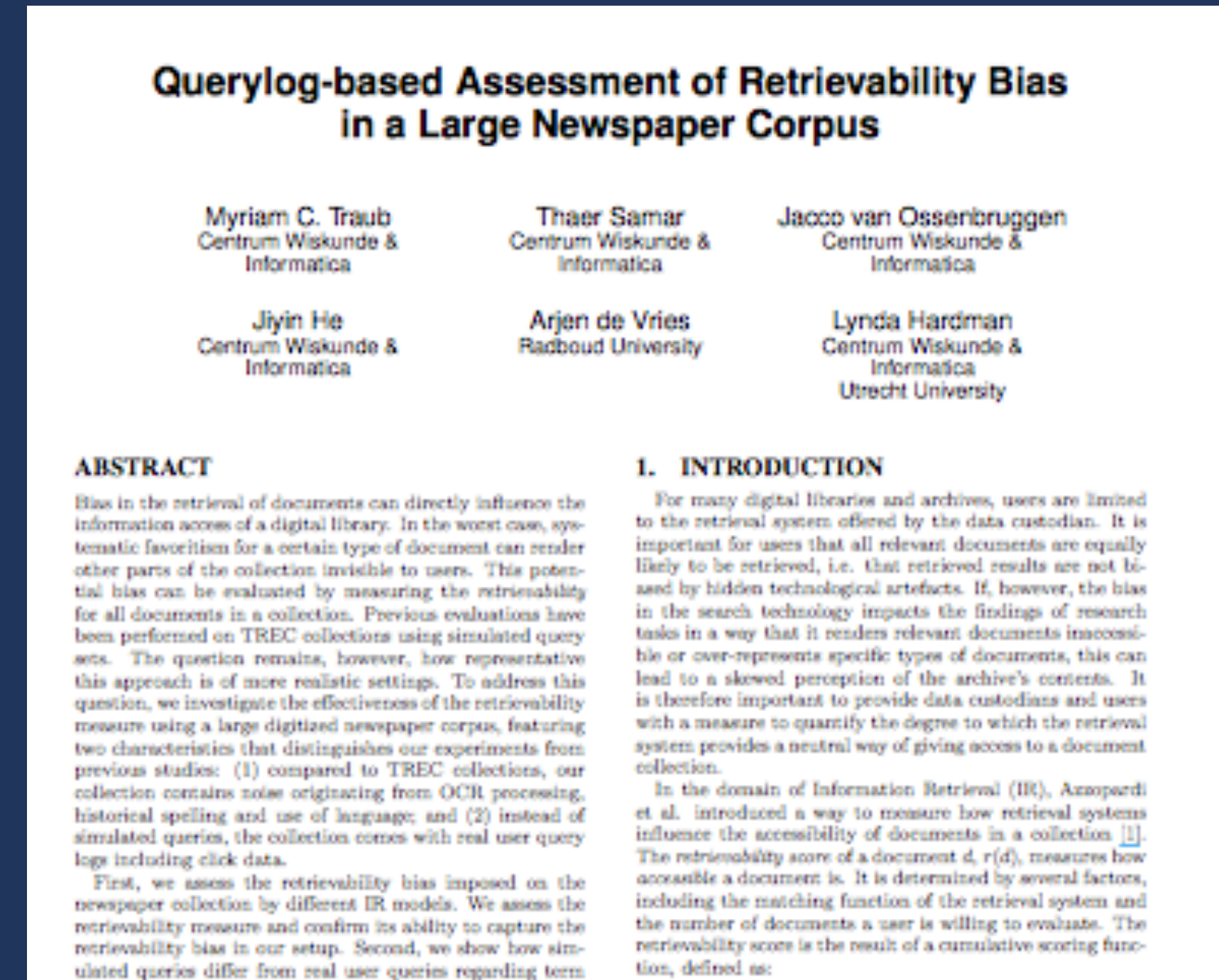






Image source: [https://static.ah.nl/static/recepten/img\\_006188\\_890x594\\_JPG.jpg](https://static.ah.nl/static/recepten/img_006188_890x594_JPG.jpg)

# Constructing a Recipe Web from Historical Newspapers

Marieke van Erp  
Melvin Wevers  
Hugo Huurdeman



@merpeltje  
@melvinwevers  
@timelessfuture



# Butter, salt & pepper

---

- Analysis of food customs:
  - historians
  - dieticians
  - ethnologists
- 1945 - 1995 Parool, Volkskrant, NRC & Trouw
- Dataset and code available through: <https://github.com/DHLab-nl/historical-recipe-web>
- Winner National Library - Rijksmuseum - Network Digital Heritage HackaLOD Hackathon
- You & other researchers are invited to work with us on case studies around food culture



Image source: [https://assets3.thrillist.com/v1/image/1623749/size/tl-horizontal\\_main\\_2x.jpg](https://assets3.thrillist.com/v1/image/1623749/size/tl-horizontal_main_2x.jpg)





## Newspapers as a source for recipes

- perception of a Dutch food culture formed in the 1950s
- newspapers are producer and messengers of public discourse
- newspapers contain views on daily life and customs
- But:
  - keyword search for 'recepten' imprecise
  - noise from digitisation process







# DINSDAG

Kopje ossestaartsoep

Knakworstjes

Witte bonen met ananas

Lofsla

Witte bonen met ananas

1 of 2 blikken witte bonen in tomatensaus, 1 uitje, 1 eetlepel mosterd, 2 eetlepels keukenstroop, 12 knakwostjes, 1 klein blikje ananasblokjes.

Rasp het uitje en meng dit door de bonen met tomatensaus. Doe de groente in een ingevette vuurvaste schaal. Roer de mosterd door de stroop en giet hier de helft van over de bonen. Dek de schaal af met een deksel of folie en zet hem 20 minuten in een warme oven. Verwarm inmiddels de worstjes in warm water en leg deze warm op de

dinsdag

6 ossestaartsoep

HUt \*orstjes

l 0( \* bonen met ananas

\t t e bonen met ananas

Va0,1 2 blikken witte bonen In 1 uitje

1 eetlepel ?lWd- 2 eetlepels keuken-

12 knakwostjes, 1 klein

„ ftaLananasDlokJes- SoJrJi het uitJe

en meng dit ,Qoe h bonen met

tomatensaus. Nir,;e groente in een

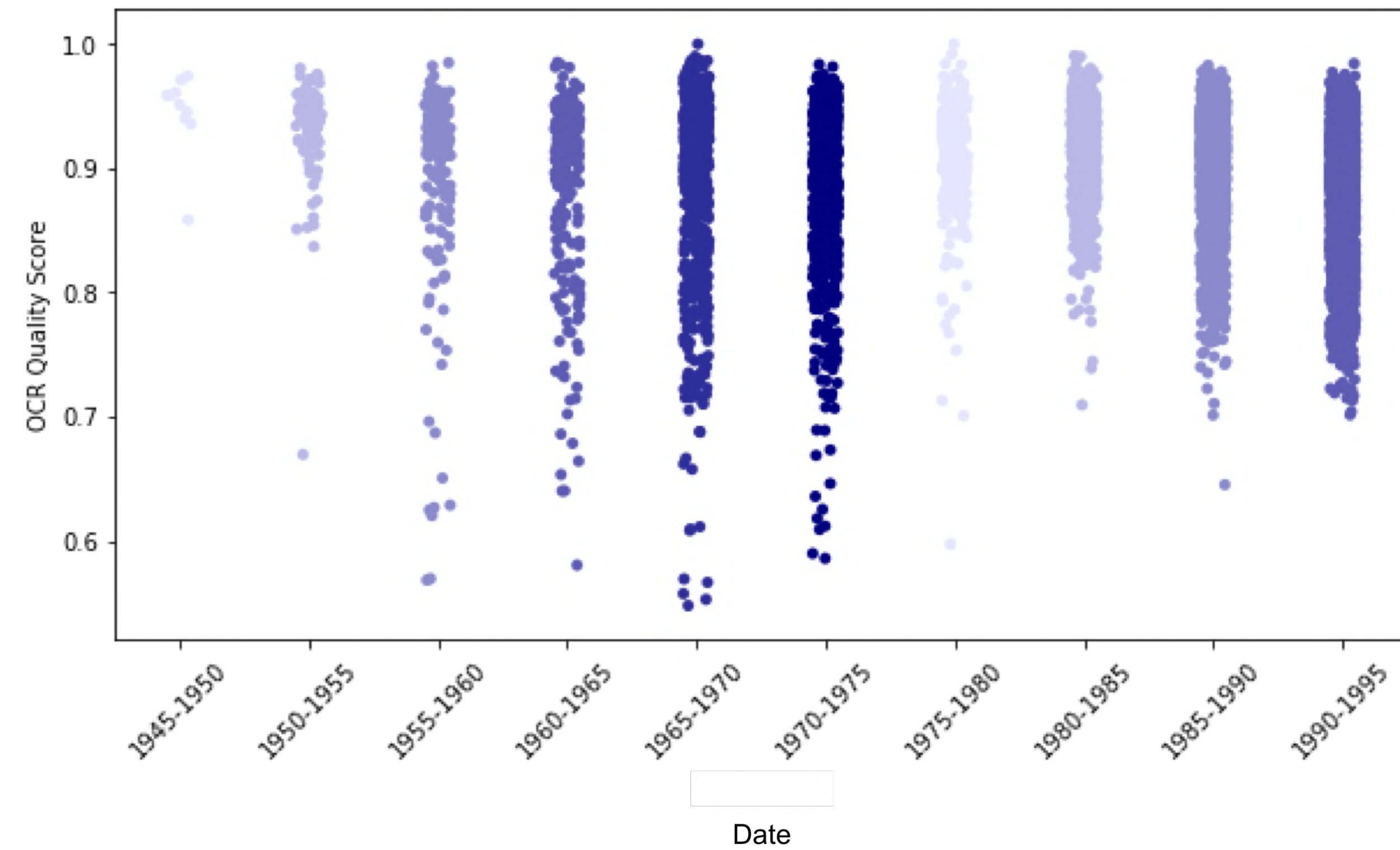
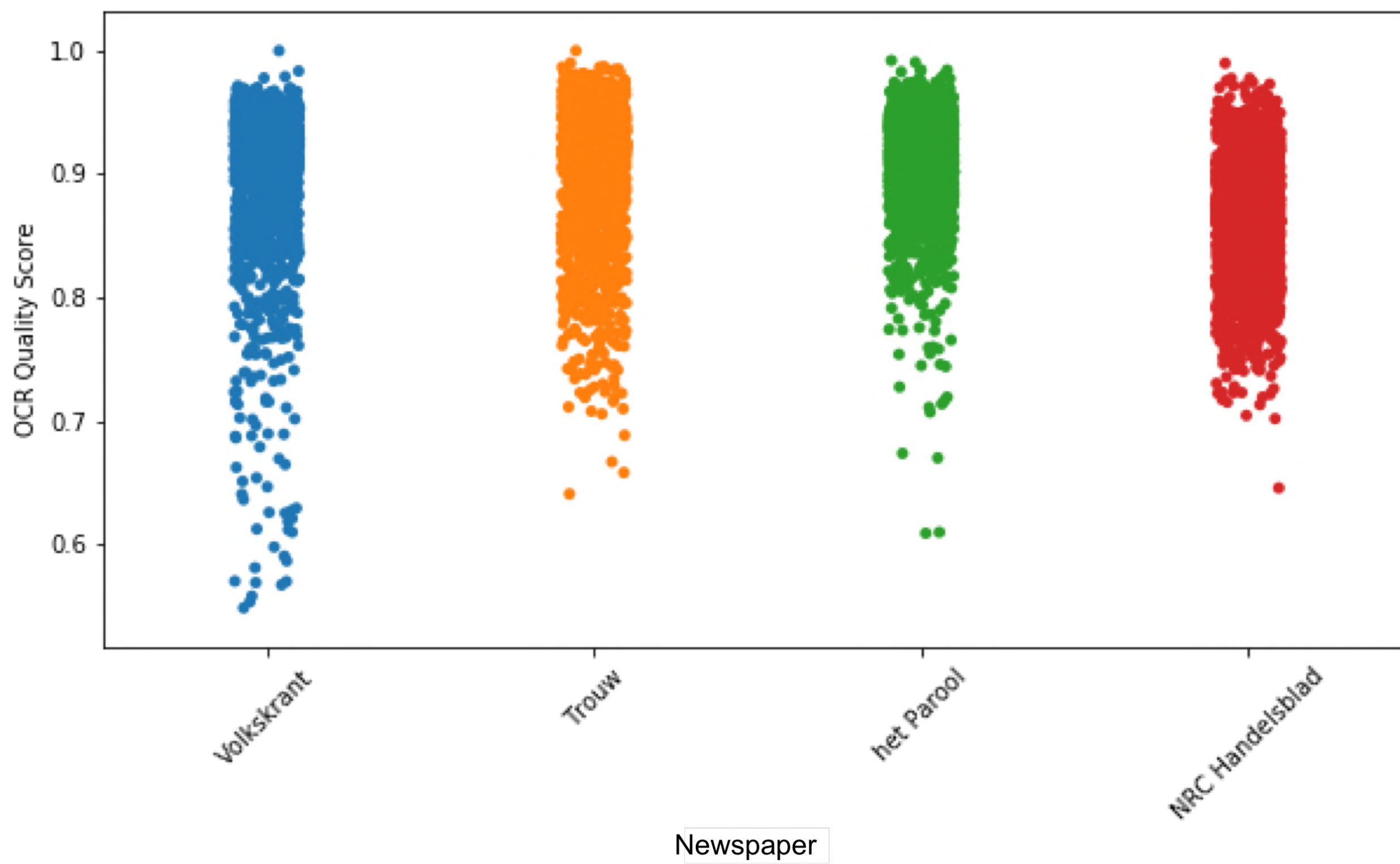
ingevette ?fd h ste schaal. Roer de

mos- Je hni?or de stroop en giet hier

2\* van over de bonen. Dek af met

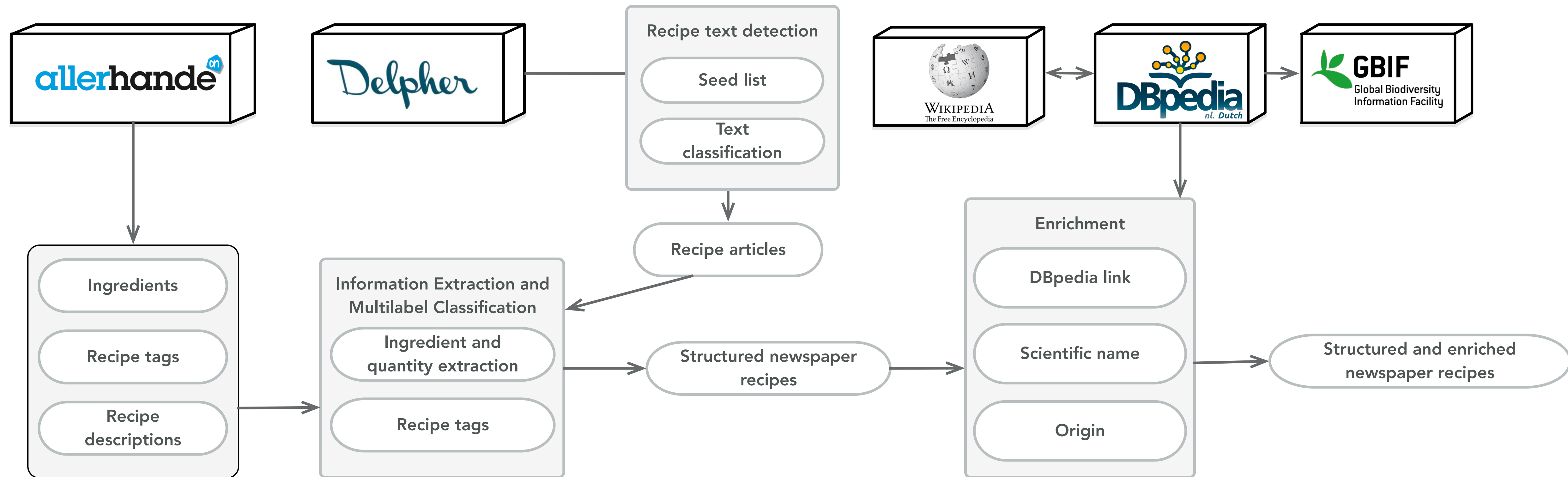


# OCR Quality





# From newspapers to a recipe web







# What & how much?

- articles cannot automatically be segmented
- OCR errors and non-grammatical sentences are a hurdle for standard NLP pipelines
- lexicon-based extraction of ingredients and quantities



# Evaluation

- 100 articles were manually annotated using Recogito
- OCR errors in ingredients or quantities marked separately
- IAA .85 but OCR boundaries difficult:  
j°ar,anen' vs °ar,anen'
- Most precise lexicon: f1 = .67
- More research is needed for out-of-lexicon ingredients

## MAANDAG

Knakworstjes Witte bonen met ananas Lofsla Vruchtenyoghurt Witte bonen met ananas 1 of 2 blik  
-1 eetlepel mosterd - 2 eetlepels keukenstroop - 8 a 12 knakworstjes - 1 klein blikje ananasstukjes.  
tomatensaus. Doe de bonen over in een met boter ingesmeerde vuurvaste schaal. Roer de mosterd  
stroop zachtjes wordt verwarmd. Giet de helft van dit mengsel uit over de bonen. Dek de schaal af  
geheel in de oven. Ongeveer 20 minuten tot alles goed warm is. Verwarm ondertussen de knakwor  
zodat ze niet kapot springen. Leg deze als de bonen warm zijn in de schaal met de ananasstukjes (c  
stroop over uit. Zet de de schaal terug in de oven zonder deksel gedurende 10 minuten.

## Monday

Knackwurst White beans with pineapple. Belgian endive salad. Fruit  
pineapple. 1 or 2 cans (jars) white beans in to-. mato sauce - 1 onie  
2 tablespoons syrup - 8 to 12 knackwursts - 1 small can of pineapp  
mix with the beans in tomato sauce. Put the beans in a with butter s  
mustard through the syrup. This will be easier if the syrup is heated  
mixture over the beans. Cover the dish with a lid or aluminium foil  
mins until it's all warmed up. In the meantime, heat the knackwurst  
don't burst. Put them in the dish together with the pineapple chunk  
are hot and pour on the rest of the syrup. Put the dish back in the o



# Results ingredients extraction

---

	Clean Ingredients				With OCR errors		
	precision	recall	f <sub>1</sub>	correct	precision	recall	f <sub>1</sub>
Allerhande	0.70	0.65	0.67	998	0.70	0.59	0.64
DBpedia	0.60	0.33	0.47	513	0.60	0.30	0.45
WordNet	0.62	0.50	0.56	764	0.62	0.45	0.54
AH-DBP-WN	0.56	0.75	0.66	1,154	0.56	0.68	0.62



# 27,411 new (old) recipes

- 34,479 Tags
- 365,133 ingredients
- >17,000 Links to external sources
- Data and software available at: <https://github.com/DHLab-nl/historical-recipe-web>





*Wat eten we*

## WOENSDAG

Gehaktballen in een pittige saus  
(Fins recept)

Andijvie, Aardappelen

Griesmeelpudding met bessesaus.

Gehaktballen in pittige saus.

Hier hebben we voor nodig: 500 gram rundergehakt, 2 koude gekookte aardappelen, 1 kleine gekookte biet, 2 uien, 1 ei, 80 gram boter, 1 lepel bloem,  $\frac{1}{2}$  dl bouillon of water, een scheut koffieroom, wat mosterd, 1 klein blikje tomatenpuree.

Kook een ui in weinig water gaar. Hebt u een vleesmolen? Haal dan de gekookte aardappelen, de biet, de gekookte ui en het gehakt er tweemaal doorheen. Maak onder de zonges

### Take home message

- OCR errors can impact information extraction
- OCR post-correction is an active research field, but errors will remain
- Focus on most important elements to extract



Acknowledgements:

**KB** Koninklijke Bibliotheek  
National Library  
of the Netherlands

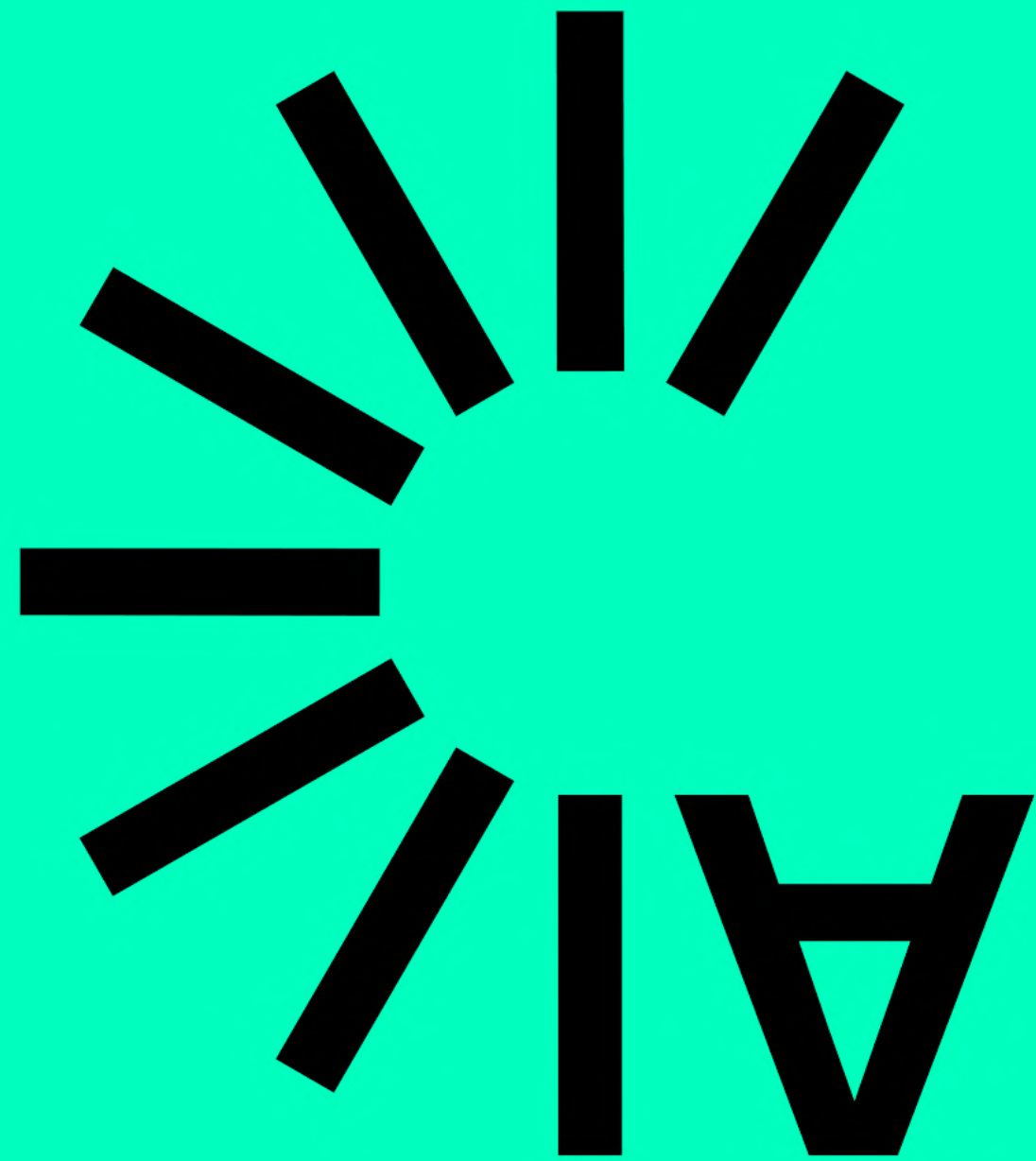


**RIJKS MUSEUM**



Image source: [https://twelvemilesfromalemondotcom.files.wordpress.com/2014/09/img\\_0326.jpg](https://twelvemilesfromalemondotcom.files.wordpress.com/2014/09/img_0326.jpg)





# Cultural AI

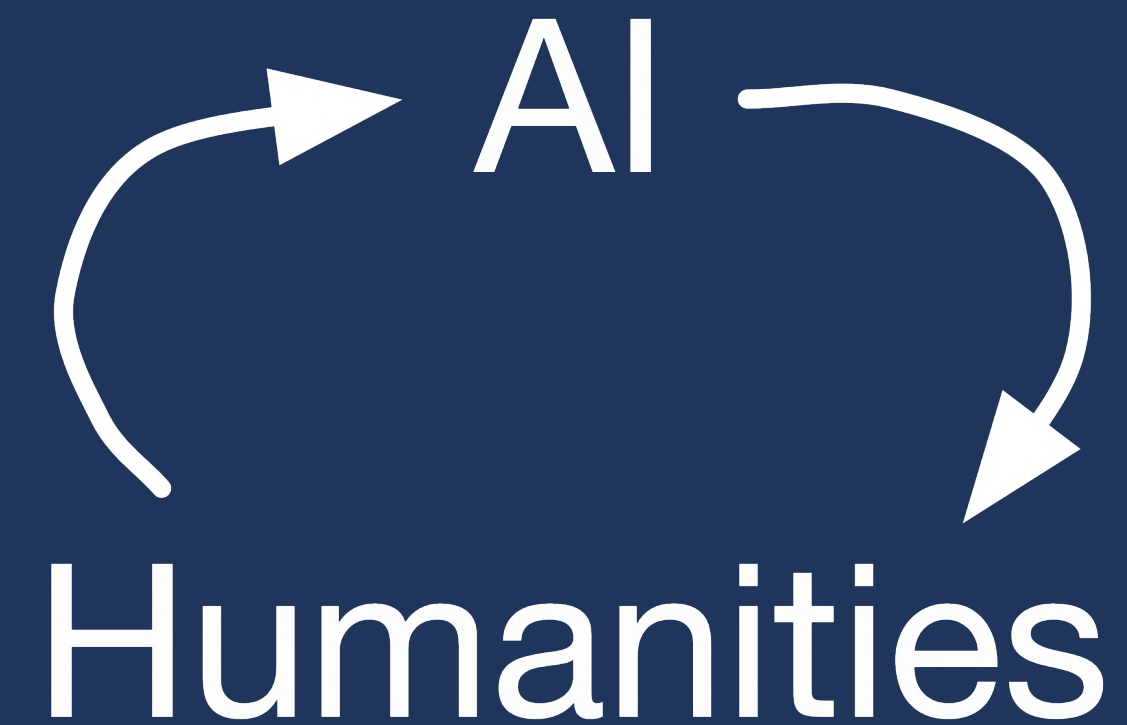
a lab for  
culturally  
valued AI







AI for humanities, and humanities for AI



*“Cultural AI is as much about using AI for understanding human culture as it is about using knowledge and expertise from the humanities to analyze and improve AI technology.”*





## What is Cultural AI?

*“Cultural AI is the study, design and development of AI systems that are implicitly or explicitly aware of the subtle and subjective complexity of human culture.”*

Bias  
Ethics  
Cultural differences  
Perspectives

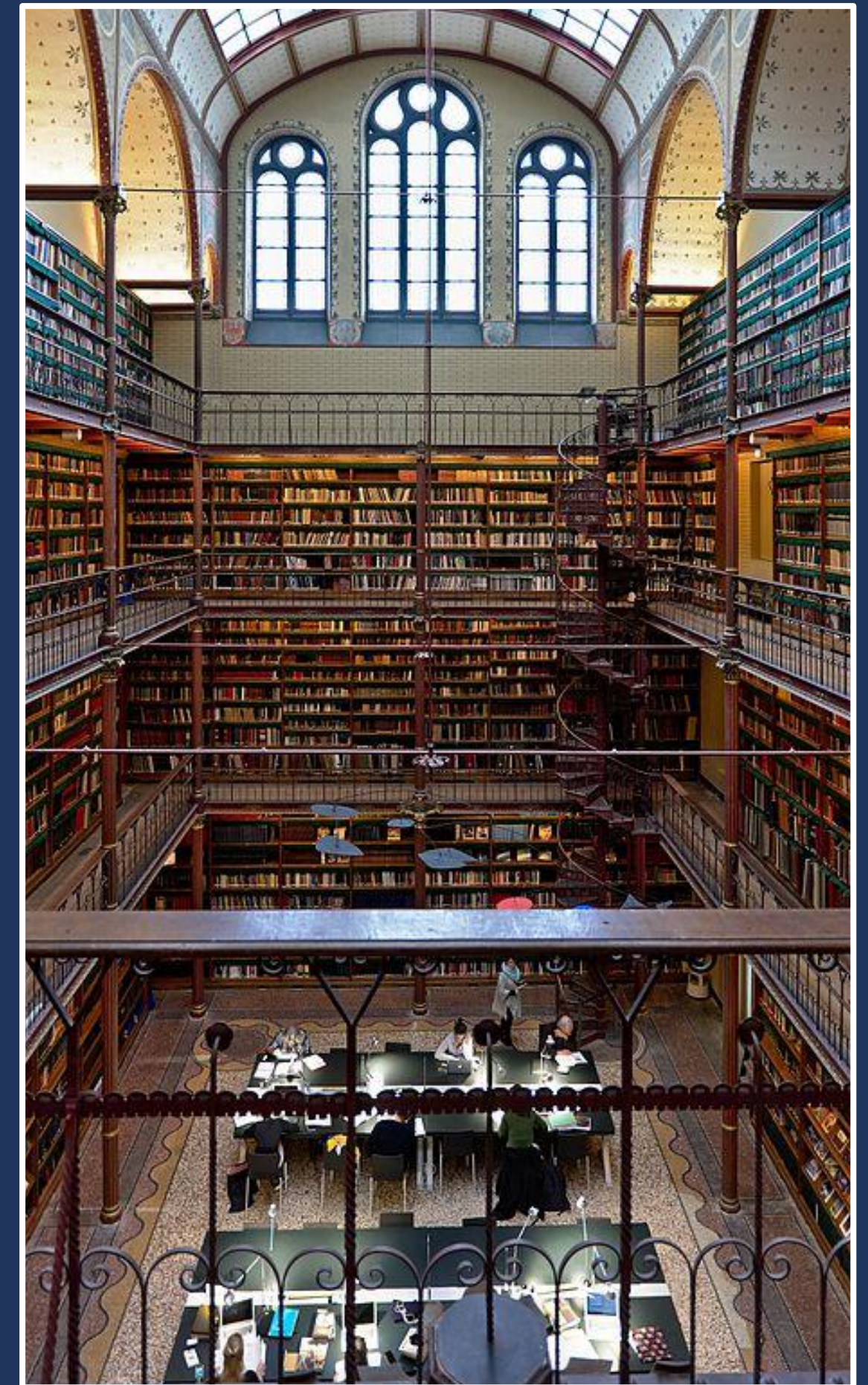






# Lab organisation and mission

- PhD students and researchers spend time at academic AND heritage partners
- Close collaboration through “data sprints”, monthly reading club, joint conference/workshop visits
- Core partners, Associate partners, Affiliate partners.







## Layers of bias

- *Bias in the data*
- *Bias introduced by tool creators*
- *Technological bias*

*We do not want to erase bias, we want to make it visible!*



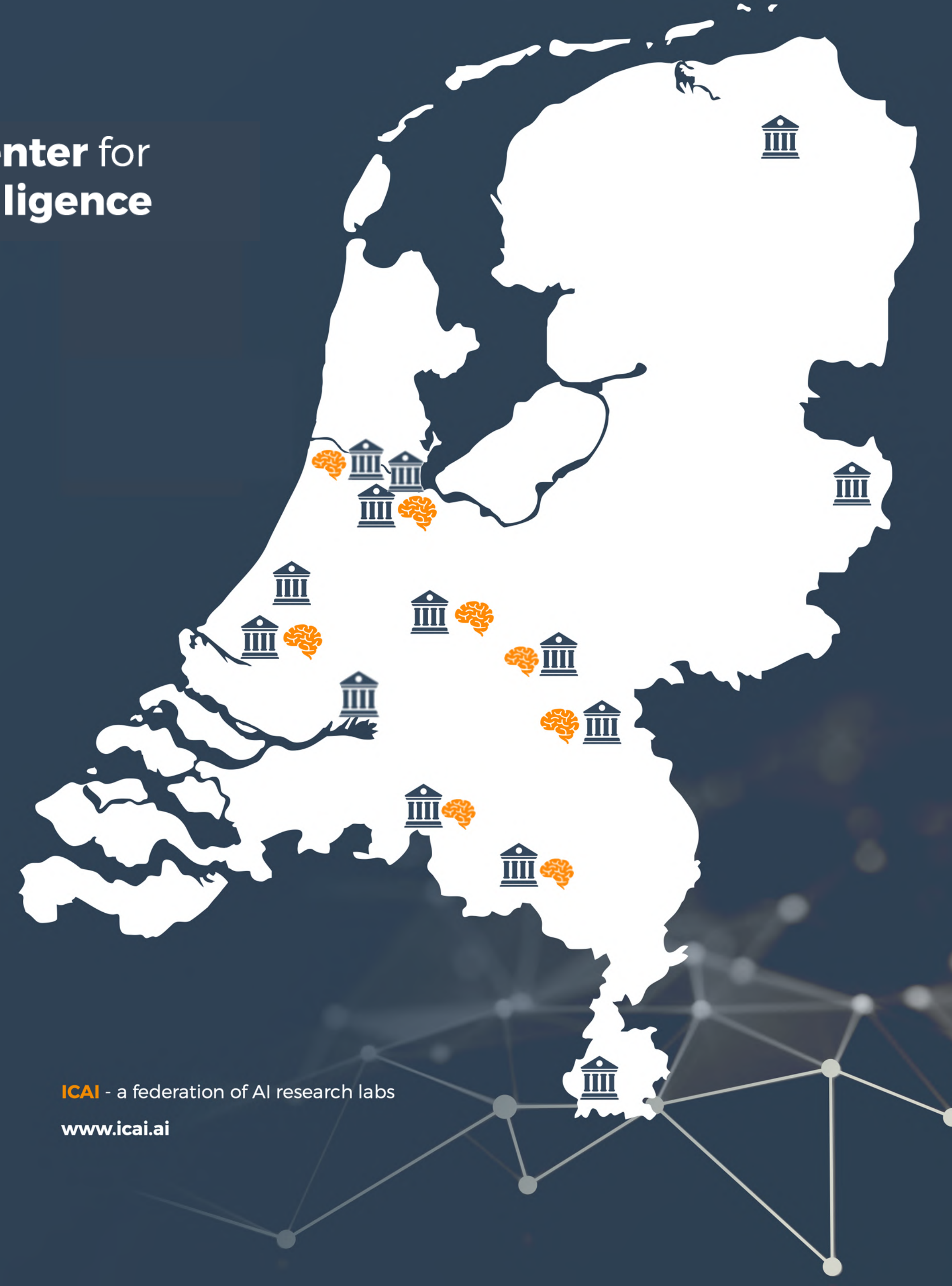


**Innovation Center for  
Artificial Intelligence**

The National Innovation Center for Artificial Intelligence (ICAI) has the mission to **keep** the Netherlands at the **forefront** of **knowledge** and **talent** development in AI.

Creating and nurturing a **national** AI knowledge and talent **ecosystem**.

More info: [icaai.ai](https://www.icaai.ai)



**ICAI** - a federation of AI research labs

[www.icaai.ai](https://www.icaai.ai)





## Current projects in the Cultural AI Lab

- BETTER-Mods (funded by NWO) 2 PhD students
- Culturally Aware AI (funded by NWO) 2 PhD students
- SABIO (funded by NDE) 1 researcher
- RE-FRAME (funded by Sound and Vision) 1 PhD student

### Upcoming:

- Library AI Principles (funded by National Library) 1 PhD student
- Researcher in Residence (funded by National Library) 1 postdoc
- Transparent pipelines (funded by NWO/NLeSc) 1 postdoc
- Responsible AI in public media (funded by NWO) 1 PhD student





- knowledge  
- traditions of  
- learning of

- (native orks)  
- (native orks)

- (native orks)  
- (native orks)

- (native orks)  
- (native orks)

community -  
world wide

community -  
world wide

community -  
world wide

community -  
world wide

community -  
world wide

community -  
world wide

community -  
world wide

community -  
world wide



</



1  
*Tobacco Absolute*  
Jorge Otero-Pailos  
1969

2  
*Coumarin*  
Isabelle Chazot

3  
*The Palace and Gardens of Versailles* by Eliza Douglas  
Marie Clapot

4  
*Amber Gris Tincture*  
Lizzie Marx

5  
*Battle of Waterloo* by Birgit Sijbrands (IFF)  
Caro Verbeek

6  
*Fishmarket (AromaPrime)*  
Chris Tuckley

7  
*Air Pollution (IFF)*  
Peter de Cupere

8  
*Cedarwood*  
Andrea Buettner

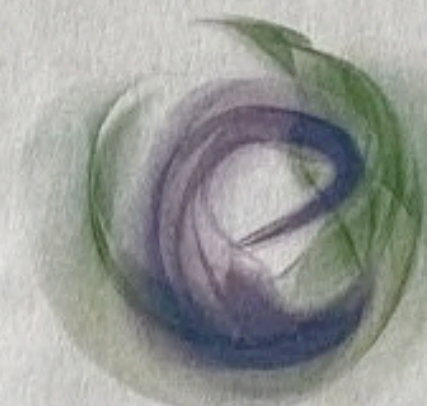
9  
*St Paul's Library* by Sarah McCartney  
Cecilia Bembibre

10  
*5318008 (CPL Aromas)*  
Tasha Marks

*Sample Kit by*

# *Mediamatic*

For the Odeuropa Workshop  
Working with Scent in GLAMs



Odeuropa

Do not open the Sample Kit until the workshop starts on the 20th of May and have a pair of your scissors ready.

In this package you will find a zip-lock bag in which you can dispose of the blotters, to contain their smell.





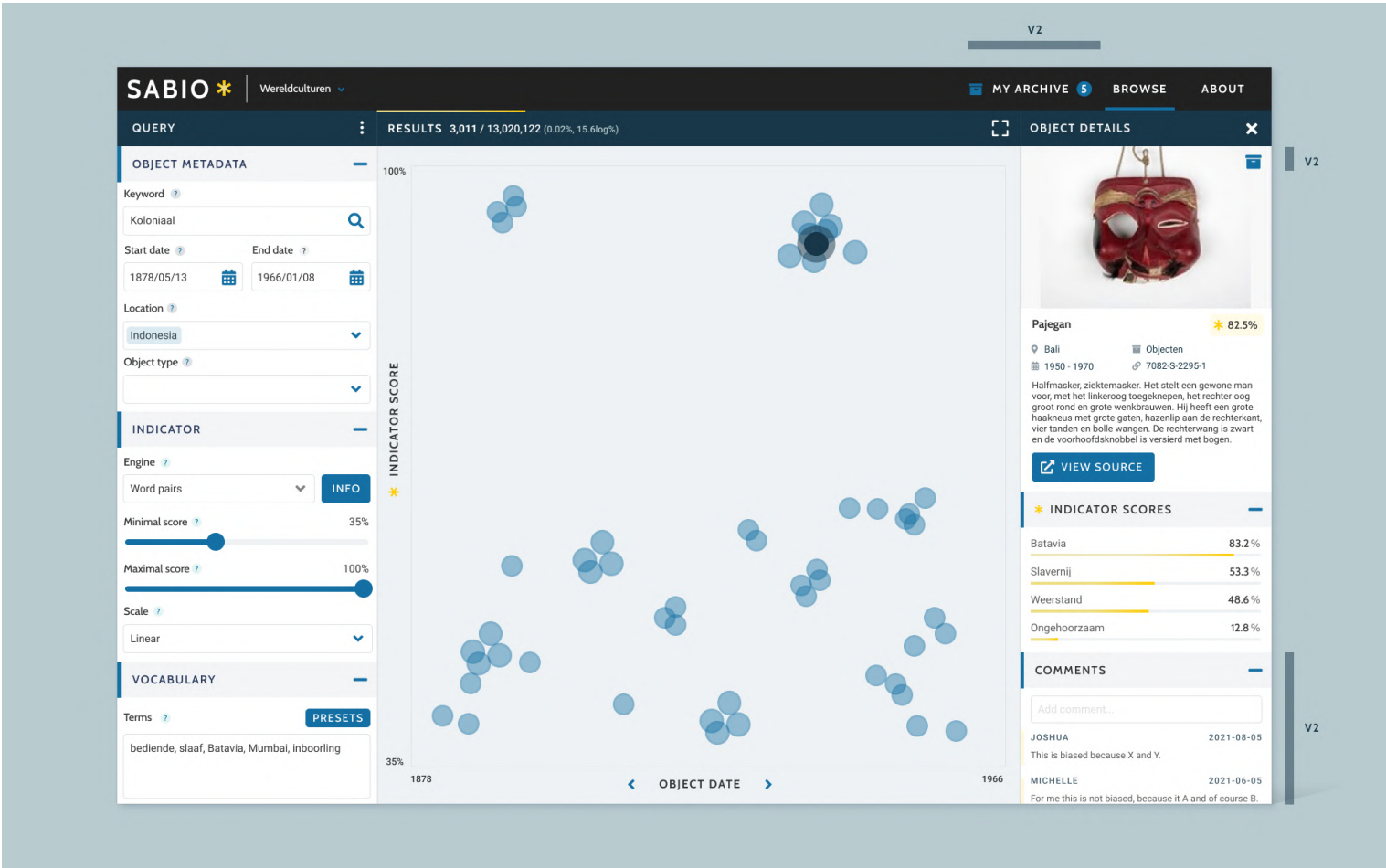
# Conclusions

---

- Custom domains require custom solutions
- Domain expertise is key
- Take time to get to know each other's domain & (research) culture
- Where possible, bring physical elements into online meetings



# Work in progress



Navigating Bias (Valentin Vogelman)



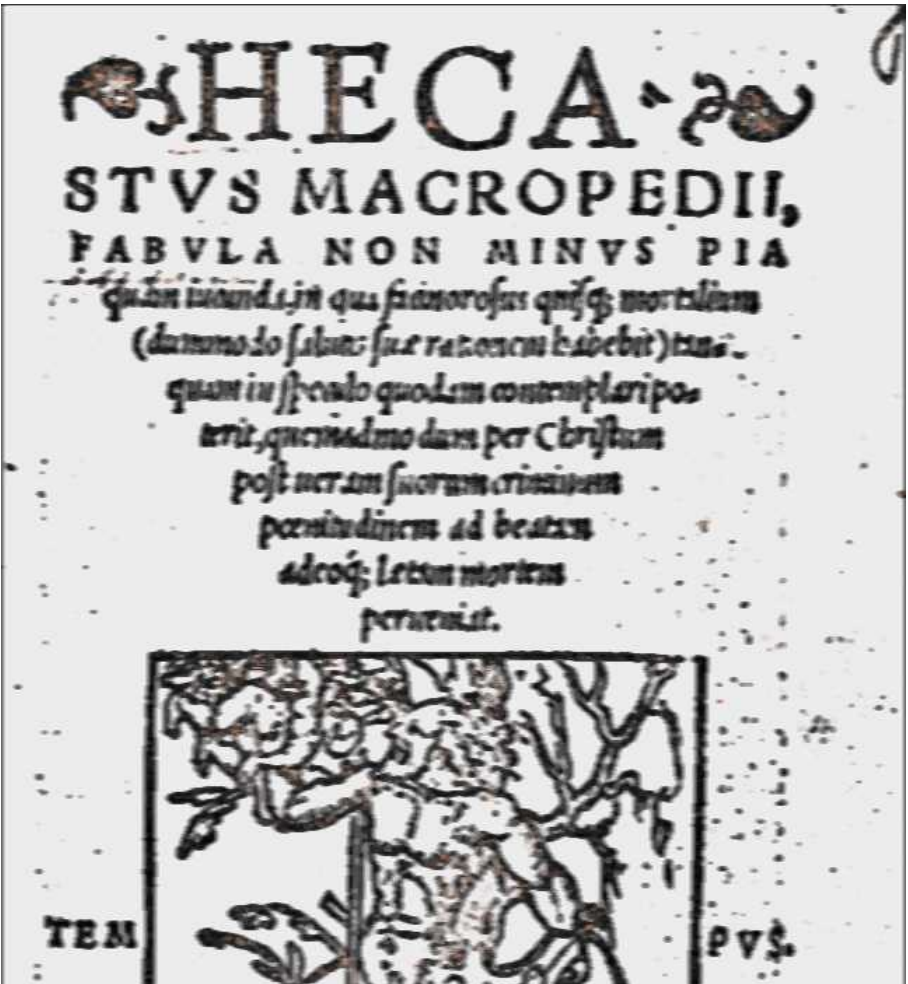
Historical Maritime Knowledge Graph  
(DHLab + HI + VU)



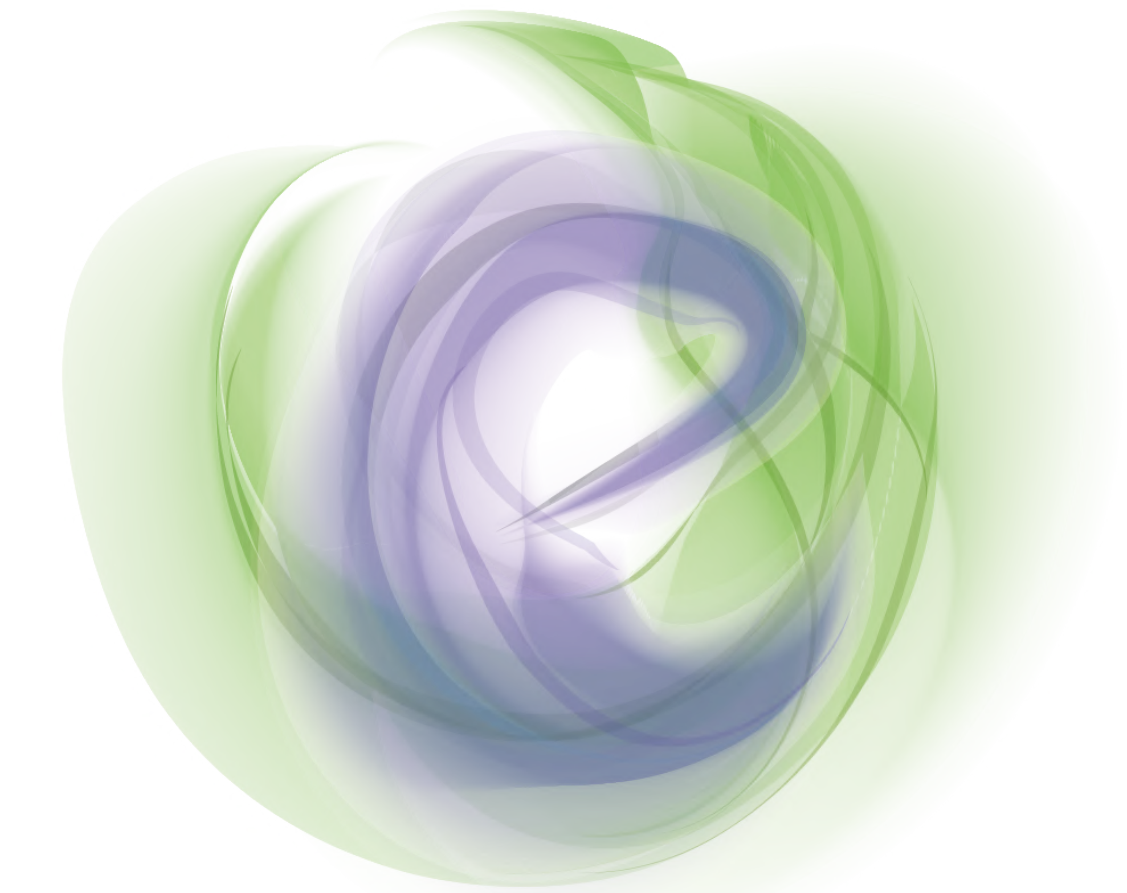
Symbolism on the Semantic Web  
(Bruno Sartini, University of Bologna)



Contentious Terms in Context  
(Ryan Brate, Andrei Nesterov, Valentin Vogelman, Laura Hollink, Jacco van Ossenbruggen, Marieke van Erp)



Translatin  
(Andrea Peverelli, Huygens Institute)



Odeuropa



