



HUMANITIES, ARTS & CULTURE DATA SUMMIT

Jan Müller

National Film and Sound Archive



14-15 March 2018, Canberra
humanities.org.au | [#HACDS2018](https://twitter.com/HACDS2018)

Platforms for HASS

CLARIAH - a case study

Humanities, Arts & Culture Data Summit
15 March 2018

Jan Müller

CEO National Film and Sound Archive



CLARIAH – Big Data, Grand Challenges



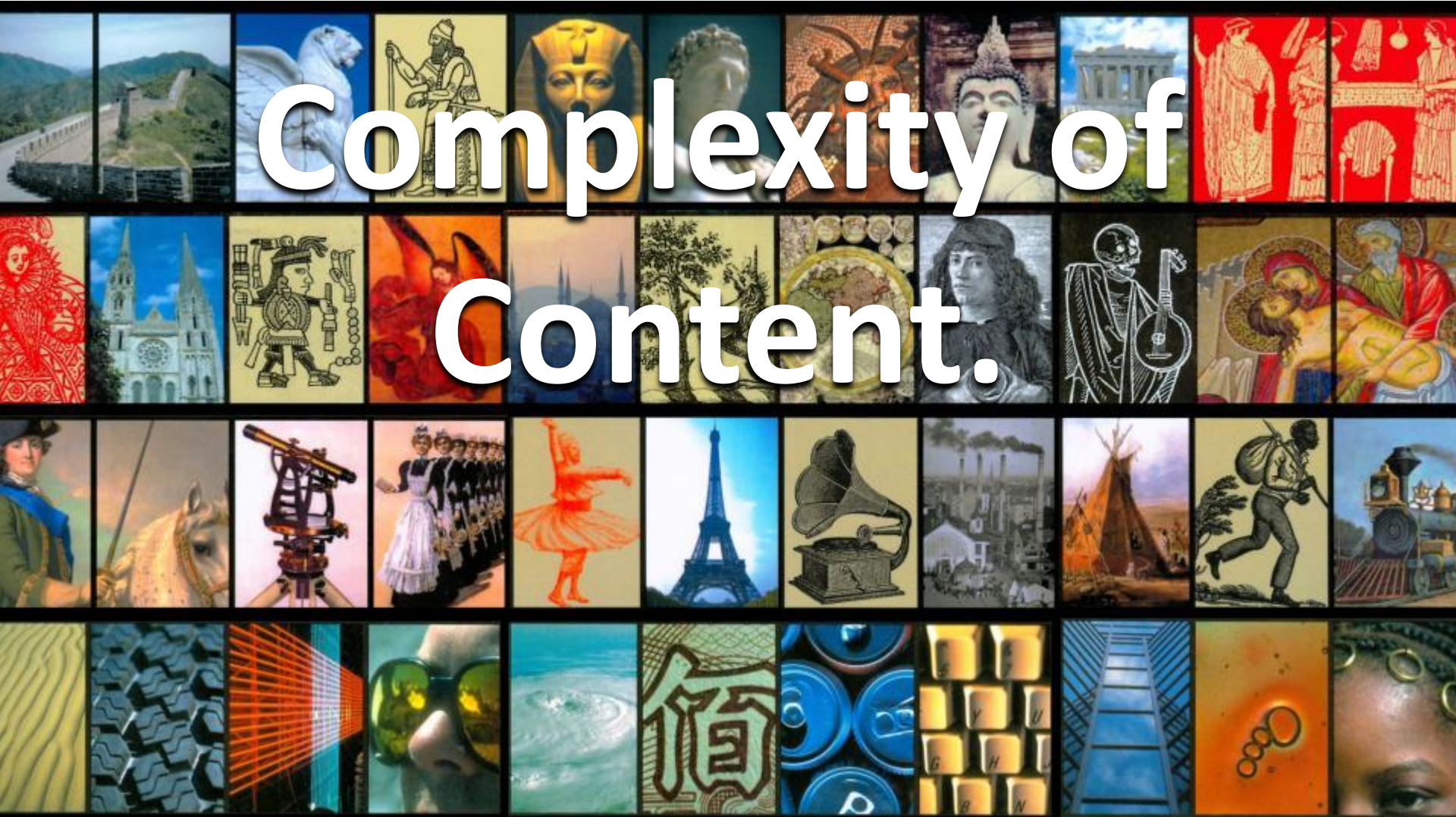
- › An infrastructure for digital humanities
- › A Dutch/European initiative

HUMANITIES

- Investigate human culture
- Analyse and interpret content
- Individual data: images, text, sound, contextual data

The Grand Challenge:

Complexity of Content.



Numerous and Fragmented





Humanities and Big Data



Advancing our Understanding of Human Culture



Why do we need a research infrastructure?

Investigating culture, art, literature and history for present-day society

Example: History of Slavery



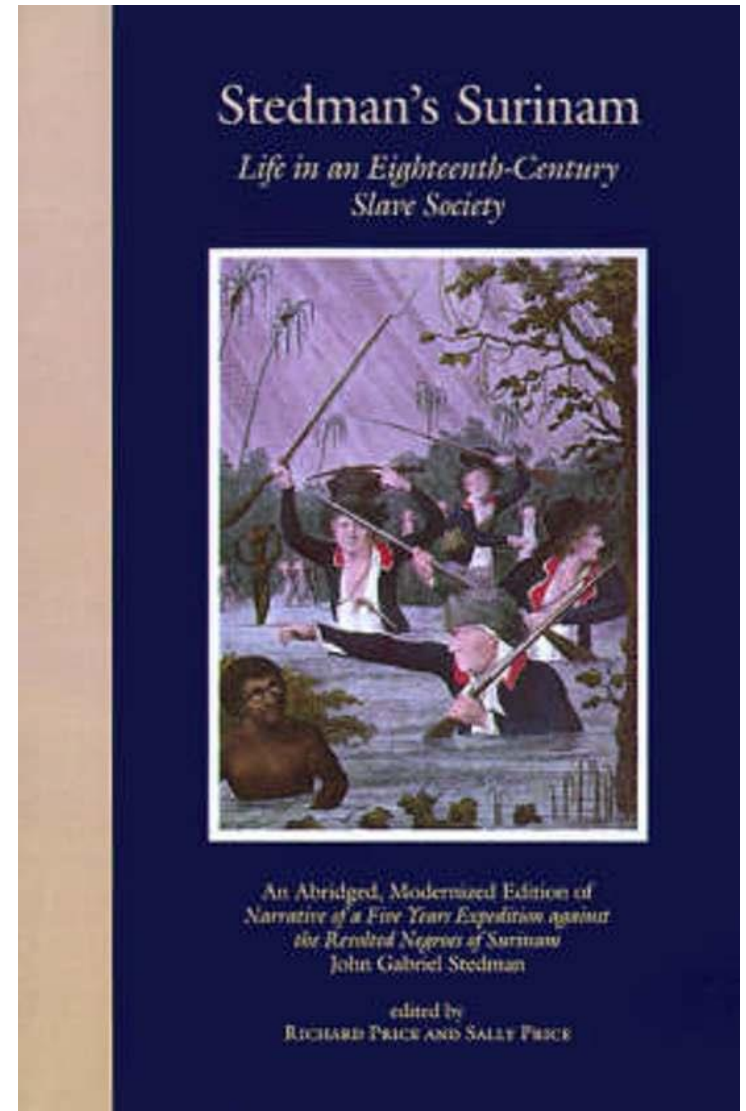
Fierce public debates:

- Racist stereotypes or cultural traditions?
- Commemorate or move on?
- Social and economic deprivation:
‘nature’ or ‘nurture’?

*Empirical knowledge for just and inclusive
policy making*

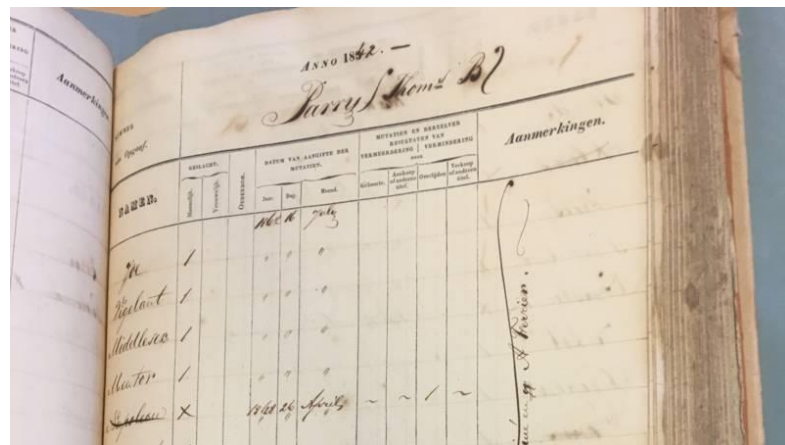
Before CLARIAH: anecdotal evidence

- Select number of sources (sample)
- Manual extraction of data
- Weigh contradictory accounts
- Mentally reconstruct what happened
- Information hidden in unstructured data



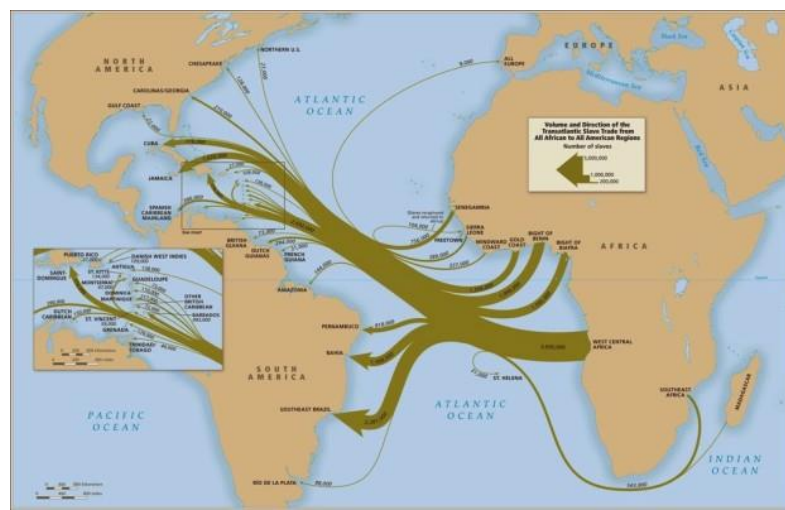
CLARIAH CORE: identify long-term patterns

- Digital databases (voyages, slave registries, written accounts)
CoDeSis project: digitizing 80.000 names from Surinam slave registries
- Combine different datasets
- Tools to query at large scale



CLARIAH PLUS: tracing family histories at scale

- Increase data with factor 10 (3PB in total)
- Enhancing scope (geology, weather, disease)
- Semi-automatic extraction
- Tools to query across different datasets



Towards a more stable and inclusive world

Contribute empirical evidence for the systematic and long-term nature and impact of slavery



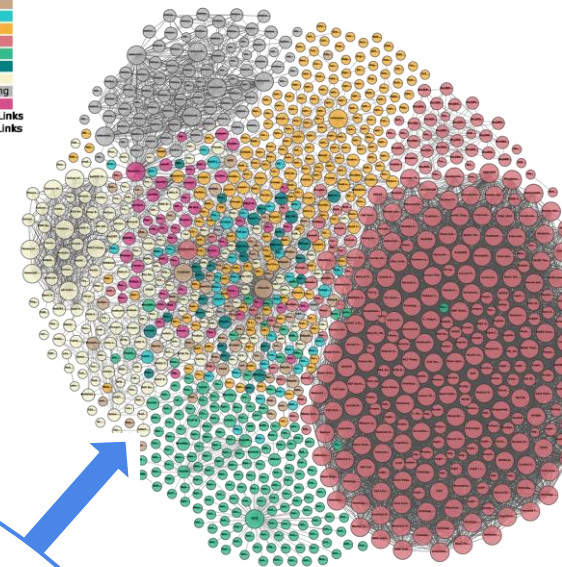
A Virtual Research Infrastructure for Scalable Humanities

- Legend
- Cross Domain
 - Geography
 - Government
 - Life Sciences
 - Linguistics
 - Medicine
 - Publications
 - Social Networking
 - User Generated
 - Incoming Links
 - Outgoing Links

CLARIN



DARIAH-EU



Utrecht University



international institute of social history



/instituut voor de Nederlandse taal/



Data Archiving and Networked Services

DANS



UNIVERSITEIT VAN AMSTERDAM



Universiteit Leiden



Max Planck Institute for Psycholinguistics



Understanding Society



Radboud Universiteit



CLARIAH Roadmap

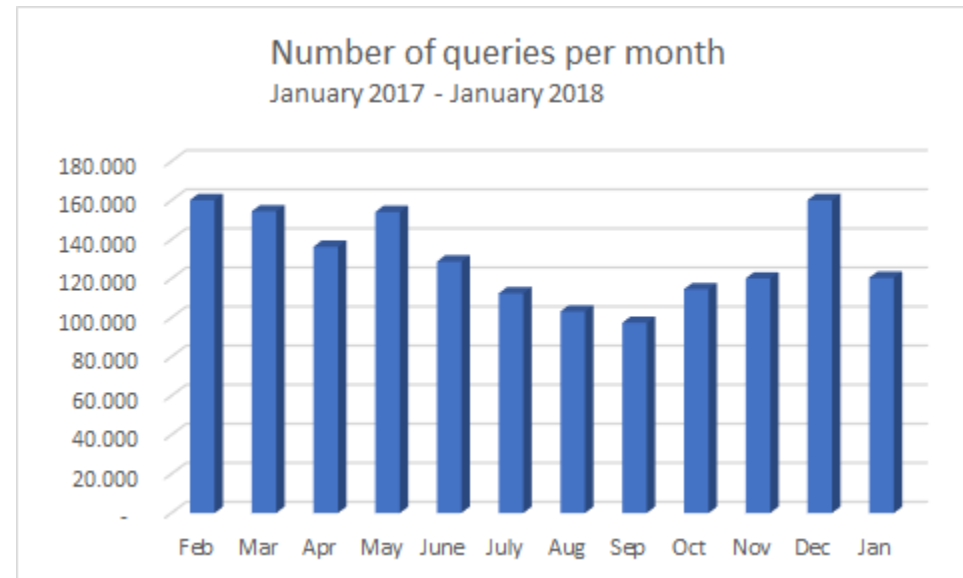


2015-2018	2019-2023	2023-...
CORE	PLUS	FUTURE
Interoperability	Scalability	Prediction
Disciplines: <ul style="list-style-type: none"> • Linguistics • Socio-Economic History • Media Studies 	Disciplines: <ul style="list-style-type: none"> • History • Literary Studies 	Disciplines: <ul style="list-style-type: none"> • All other Humanities • Social Sciences & ...
Data types: <ul style="list-style-type: none"> • Text (language) • Structured data • Audiovisual data 	Data types: <ul style="list-style-type: none"> • Text (content) • Images • Geodata (2D&3D) 	Data types: <ul style="list-style-type: none"> • Device-related data (CCTV, sensors, ...) • ...

Data and Usage CLARIAH CORE



- 130 M Triples
- 17 TB textual data
- 176 TB image data
- 130 TB AV data (audio + RTV)
- Users Corpus Historical Dutch:
 - 80.000 unique page views
 - 4500 sessions (15 mins)
 - 1500 users (10 countries)
 - 67% returning; 33% new



User statistics for Historical Dictionaries Online

Evaluating CLARIAH's success

- By 2023 available data in CLARIAH has increased with a factor of 10
- In 2023 CLARIAH is used by 40% of humanities scholars
- CLARIAH implements strategies for measuring uptake
- CLARIAH benefits society

(policy-making, journalism, education)

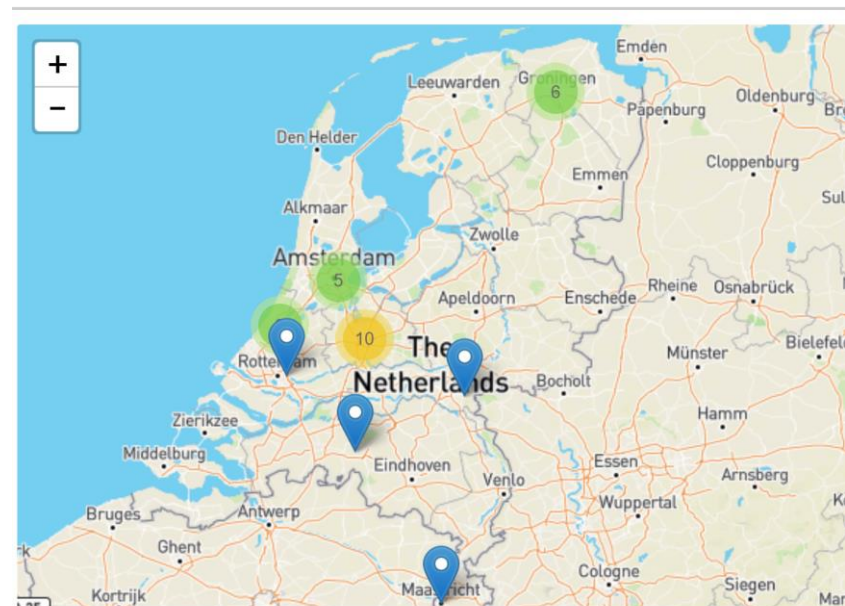


Why CLARIAH-PLUS?

- **Relevant**
 - Big Data of the past
 - Address new questions, in new ways
- **Timely**
 - Digital Humanities are on the rise
 - Critical mass
- **Expert team**
 - Forerunners (NL, EU and world)
 - Previous experience

Digital Humanities Registry - Courses

Course registry 2.1 | [About](#)



Audio Visual Archives:

We help remember.
We keep memories alive.

(by telling stories with our collections and
archives)

Our collections

It's not about preservation and/or digitisation anymore.

It's about what you *do* with your collections!

(and: we're always 'on' and always 'in beta')



Thank You!

NFSA

@JanMullerNFSA

