

State of the CLARIN Infrastructure

Franciska de Jong

Executive Director CLARIN ERIC

f.m.g.dejong@uu.nl

CLARIN Annual Conference

28 September 2021



Where do we stand? 1/5

- New Strategy 2021-2023: Language as social and cultural data
Enhanced infrastructural support for research based on language materials
- 22 + 3 countries
- federated service offer, access to
 - FAIR data
 - chainable tools
- > 60 registered centres
 - 25 certified B-centres
 - network of K-centres
- strongly aligned with the Open Science agenda, focus on **interoperability**



CLARIN

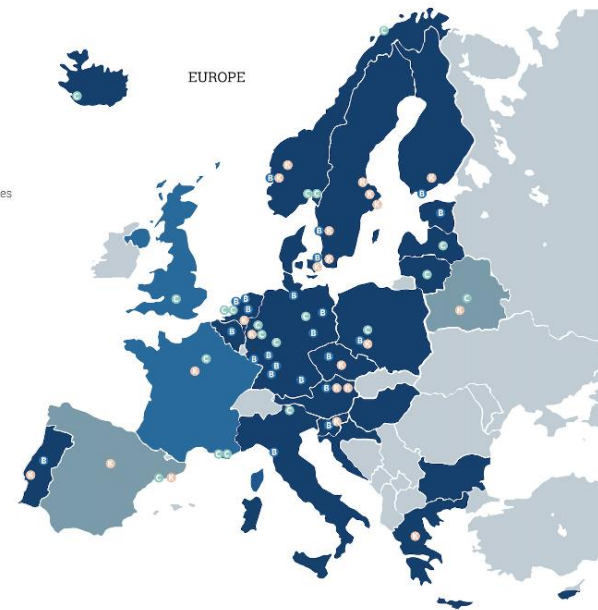
- ERIC members
- Observers
- Countries with participating centres
- B Centre Providing Data
- Centre Providing Metadata
- Knowledge Centre



USA



SOUTH AFRICA



Where do we stand? 2/5

Reinforced collaboration in SSH cluster via project SSHOC

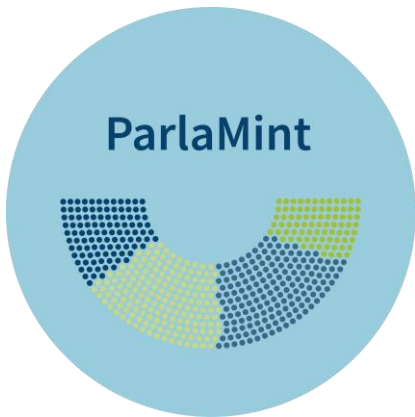
- Collaboration between all SSH research infrastructures; support for users of cultural data, language data, survey data
- Strategic themes:
 - Federation of resources at SSH level
 - SSH Open Marketplace: [link](#)
 - Methodological frameworks for heterogenous data, mixed methods, multimedia
 - FAIR models for sensitive data
 - Innovative services for translation
 - Training and education for researchers and GLAM
- CLARIN participation: Central team & 6 CLARIN nodes



Where do we stand? 3/5

CLARIN Resources Families: a strategic pillar for harmonisation:

- Collaboration across national consortia
- Model for improvement of metadata quality
- Alignment with disciplinary and multidisciplinary agendas involving comparative research across languages and regions



ParlaMint: CLARIN flagship project

- Towards Comparable Parliamentary Corpora
- First data release available:
<https://www.clarin.eu/content/parlamint>
- Parliamentary data covered for 17 languages
- Funding made available for next stage: 2021-2023

Where do we stand? 4/5

CLARIN Ambassador programme

Events aimed at uptake in a wide range of communities of use (partly in SSHOC)

- Webinars (e.g., on how to use tools for interview transcription)
- Tutorials (e.g. for librarians on when/how CLARIN resources can be used)
- Helsinki Digital Humanities Hackathon: ‘Parliamentary Debates in COVID Times’

More opportunities for CLARIN funding & support

- CLARIN training suite: ‘Teaching with CLARIN’
- CLARIN Trainer Network Programme
- Support for virtual events
- Seed grants for Horizon Europe proposal preparation



Where do we stand? 5/5

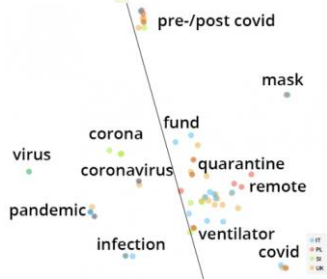
New website

- better geared towards the diversity of services and audiences
- more easy to navigate
- more contemporary look and feel



CLARIN Impact Stories

- Separate website section for use cases: [link](#), [first story](#)
- Context for high quality publications, suited for wider audiences
- To be coupled to format/material of Tour de CLARIN
- Impact stories for research, teaching, expertise, collaboration beyond academia



Where do we want to go? 1/5

Strategy theme: *Language as social and cultural data*

Enhanced infrastructural support for research based on language materials

The position of CLARIN in the landscape is in part fueled by the ever increasing datafication of culture and society: more and more aspects in everyday life are captured as digital and machine-actionable data sets, including language materials.

This is calling for a range of data science methods, including approaches geared towards the processing of language data, as well as guidelines for responsible data science practices.

More and more language data from the cultural heritage domain is becoming available in digital form, which increases the potential for supporting digital humanities agendas.

Where do we want to go? 1/5

Strategy theme : *Language as social and cultural data*

Enhanced infrastructural support for research based on language materials

The position of CLARIN in the landscape is in part fueled by the ever increasing datafication of culture and society: more and more aspects in everyday life are captured as digital and **machine-actionable data sets**, including language materials.

This is calling for a range of **data science methods**, including approaches geared towards the processing of language data, as well as guidelines for **responsible data science practices**.

More and more **language data from the cultural heritage domain** is becoming available in digital form, which increases the potential for supporting **digital humanities agendas**.

Where do we want to go? 2/5

Focus on the potential for impact

- more alignment with (multi)disciplinary research agendas and collaboration with mission-driven research initiatives
 - COST-actions
 - Horizon Europe Pillar 2 projects
 - UN Sustainable Development Goals: [link](#)
- applications /workflows for
 - the development of tools for studying and informing policy responses to global crises (e.g., climate change, migration, COVID-19)
 - the study of mental health conditions and tools for health interventions
- exploring models for innovation and collaboration with industry



Where do we want to go? 3/5

Continuation of CLARIN Research Families initiative

- coordination of metadata curation for more families: [link](#)
- partial automation of the harmonization effort
- call for interest in small project funding: [link](#)



as flagship project: phase II (2021-2023)

Continuation of action lines aimed at wider outreach

- Ambassadors programme with more ambassadors
- capacity at CLARIN Office for virtual events, including the CLARIN Café series
- focus on support for teaching and education

Where do we want to go? 4/5

Reinforced support for CLARIN nodes

- coordination among K-centres
- website as hub for knowledge exchange with landing pages for
 - technical innovation and new methodological frameworks (AI, ML)
 - specific application domains (health, climate transition)

Reinforced collaboration with other infrastructural initiatives

- alignment with dynamics in the landscape of the European Open Science Cloud
 - CLARIN joined EOSC Association as member
 - participation in EOSC Task Forces
- joint planning with SSH cluster: SSH Open Marketplace
- GLAM-sector: Europeana, LIBER
- Platforms for language data: ELG, Humming Face, LDaCA



Where do we want to go? 5/5

Composition Board of Directors

- as of January 2021: new member Francesca Frontini
- as of January 2022: vacancy; Andreas Witt to step back
- 2022: vacancy for executive director ([link](#))

10th Anniversary of CLARIN ERIC (Spring - Summer 2022)

- Book in preparation
- Online events

11th CLARIN Annual Conference 2022: back to face-to-face format?

Funding & support opportunities (www.clarin.eu/funding)

