# SSH RIs as enabler of excellent science in service of society

Franciska de Jong

*executive director CLARIN ERIC*

**ERIC Forum panel @** I C R I

*Brno, 19 October* 2 0 2 2

**Collaboration of the pan-European research infrastructures in SSH and their nodes**

Support for research based on cultural data, language data, survey data, and other relevant digital objects. Objectives:

- Federation of distributed SSH resources
    - data, tools
    - publications
    - training
- Enhanced methodological frameworks and workflows
- SSH Open Marketplace:
    - a common discovery portal
    - included in co eosc catalogue
- Increased potential for synergy and impact, within SSH and across clusters

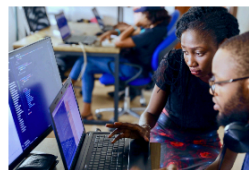# Interconnecting existing and new infrastructures

# SSH and Sustainable Development Goals

- Some SDGs link very directly to objectives of SSH RIs, but no SDG is beyond the scope of SSH.

- Societal impact of excellent research can mostly be attributed to RIs as an indirect and often delayed effect.



## CLARIN and the Sustainable Development Goals

| View | Edit | Delete | Revisions |
|---|---|---|---|

The United Nations' Sustainable Development Goals (SDGs) aim to improve social and economic wellbeing across the globe through joint action. CLARIN ERIC acknowledges the fundamental importance of these goals and actively contributes to their achievement. Finding solutions for the societal challenges that we face today requires innovative methods and cross-border collaboration. CLARIN ERIC offers the resources, technology, and know-how necessary for the development of cutting-edge data-driven approaches as well as a platform for collaboration across countries and disciplines.
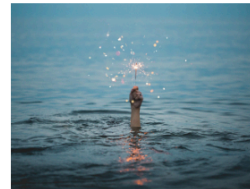
### Goal 4: Quality Education

CLARIN ERIC contributes to SDG 4 in significant ways by helping to increase digital literacy across Europe and beyond. While digital technology has radically transformed the way we live, and will become increasingly important in the future, the pace of digital literacy is lagging behind the pace of technological advancement.

Expand ˅

### Goal 9: Industry, Innovation and Infrastructure

CLARIN is an innovation hub for many of the technologies that are transforming the economy and society more widely: deep learning, automatic speech recognition, machine translation and artificial intelligence (AI). CLARIN responds to the growing need for language data-related services by offering access to an invaluable wealth of tools and data, as well as expertise and training for data-processing tasks.

Expand ˅

### Goal 16: Peace, Justice and Strong Institutions

CLARIN contributes to peace, justice and strong institutions by hosting a large variety of data from parliaments and other public bodies. By providing sustainable access to such data in a structured and annotated form, CLARIN supports democratic values and the transparency of political decisions at national and international levels. Parliamentary data corresponds directly to recent events, including those with a global impact on human health, social life and economics. It is imperative for such data to be available and easily accessible. Moreover, with the recent advances of artificial intelligence, analytics on parliamentary data is becoming a prerequisite for reliable approaches in validating information in contemporary society.

Expand ˅

# SSH-specific opportunities addressing societal needs

- **Relevance for supporting the implementation of public policy**
  - Wide coverage of UN Sustainable Development Goals
  - Responses to COVID-19
  - Cross-disciplinarity of science projects in EOSC eco-system

- **To capture social and cultural dynamics, attention needed for**
  - Diversity of data types, workflows and scholarly cultures
  - Multilinguality, multimodality, terminology, …
  - Longitudinal value of SSH data

# Five European RI clusters

**Aligned agendas and activities**

- common web presence,
  white papers, etc.
- alignment of services to enable
  multi-disciplinary work
- projects aimed at common
  value proposition for ERA and EOSC

**Example project: META-COVID**

- Focus on vocabulary harmonisation across disciplinary domains of  metadata
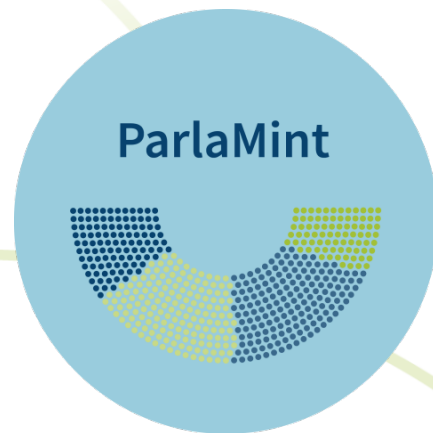  for multilingual heterogeneous data resources related to COVID-1

# CLARIN in data types and 'resource families'

- Newspaper archives
- Literary texts
- Social Media data
- Parliamentary records
- Historical letters
- Oral History data
- Disciplinary libraries
- Institutional archival data
- Broadcast archives
- …

See also the info on the **CLARIN Resource Families** initiative: https://www.clarin.eu/resource-families
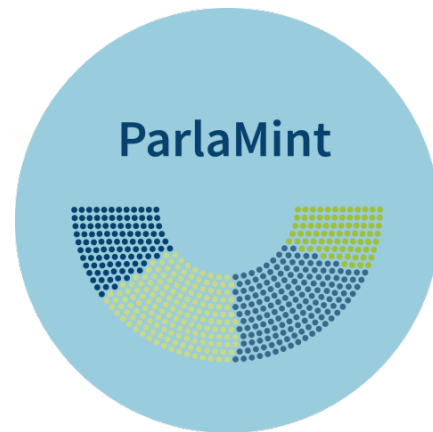
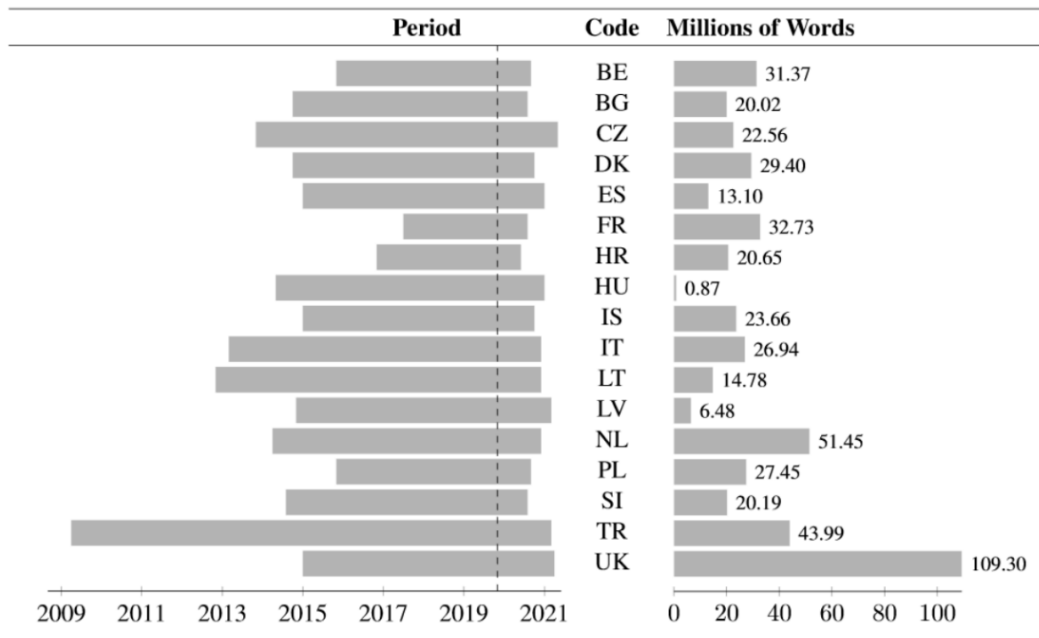# CLARIN in data types and 'resource families'

- Newspaper archives
- Literary texts
- Social Media data
- Parliamentary records
- Historical letters
- Oral History data
- Disciplinary libraries
- Institutional archival data
- Broadcast archives
- …

See also the info on the **CLARIN Resource Families** initiative: https://www.clarin.eu/resource-families

ParlaMint

# ParlaMint I - COVID-19 debates

## Initiative with a duration of 1 year (2020-2021)

**ParlaMint**

| | Period | Code | Millions of Words |
|---|---|---|---|
| | | BE | 31.37 |
| | | BG | 20.02 |
| | | CZ | 22.56 |
| | | DK | 29.40 |
| | | ES | 13.10 |
| | | FR | 32.73 |
| | | HR | 20.65 |
| | | HU | 0.87 |
| | | IS | 23.66 |
| | | IT | 26.94 |
| | | LT | 14.78 |
| | | LV | 6.48 |
| | | NL | 51.45 |
| | | PL | 27.45 |
| | | SI | 20.19 |
| | | TR | 43.99 |
| | | UK | 109.30 |

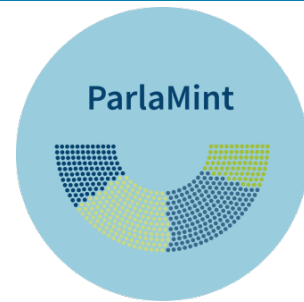2009  2011  2013  2015  2017  2019  2021      0  20  40  60  80  100

Released data available through CLARIN's central discovery platform (via CLARIN-Slovene's repository)

Explorable also through other external platforms, such as
- SketchEngine
- Kontext

**See also**: https://www.clarin.eu/content/parlamint-towards-comparable-parliamentary-corpora

9

# ParlaMint II  -  COVID-19 debates

## Initiative with a duration of 1,5 years (2021-2023)

- Data from 13 more parliaments, period covered extended
- Alignment with speech recordings, translation
- Better documentation
- Improved tutorial.

## Studies of public debate on COVID-19

- Frequency of terms
- Semantic topic detection
- Links with other data types



**A Return of Science? Frequency Patterns**

- Rel. number of different speakers using the lemma "expert" per month
  - between 5% and 25%
  - increases in Spain, Italy and the Netherlands

# META-COVID: An Empirical Case Study on COVID-19 Metadata

*Contribution of TEXTUA -* Universiteit Antwerpen CLIPS | Centrum voor Computerlinguïstiek en Psycholinguïstiek *to the ongoing H2020 project* EOSC Future *in collab with* ecrin

## What

- Development of an ontology of COVID-related topics

## Why

- Aimed at optimising cross-referencing across different scientific disciplines.
- Meant to be relevant for
  - better integration of information from public debate channels and the health domain
  - digital monitoring and analysis of the public debate

## How

- State-of-the-art topic modelling techniques
- Integration of the extracted topics into the ontology.
- Validation of automatically labeling of parliamentary data, social media data and social survey data.