CLARIN
The Language Resources and Technology Infrastructure

Steven Krauwer

CLARIN ERIC / Utrecht University

CLARIN

CLARIN in nine bullets

- **CLARIN** is the Common Language Resources and Technology Infrastructure
- Legal entity with **ESFRI ERIC** status since 2012, **Landmark** since 2016
- that provides easy and sustainable access for scholars in the **humanities and social sciences** (and beyond)
- to **digital language data** (in written, spoken, video or multimodal form)
- and **advanced tools** to discover, explore, exploit, annotate, analyse or combine them, wherever they (= data and tools) are located
- through a **single sign-on** environment
- that serves as an ecosystem for **knowledge sharing**
- strongly committed to the **FAIR principles** (see next slide)
- and: ready for **integration in EOSC** (see next slide)

FAIR and EOSC

- The **FAIR principles** serve (https://www.go-fair.org/fair-principles/):
    - to make your data **F**indable
    - to make your date **A**ccessible
    - to make your data **I**nteroperable
    - to make your data **R**e-usable

- The **European Open Science Cloud (EOSC –** https://eosc-portal.eu/about/eosc**)** is an EU initiative that will offer 1.7 million European researchers and 70 million professionals in science, technology, humanities and social sciences a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines by federating existing scientific data infrastructures, currently dispersed across disciplines and the EU Member States.

CLARIN ERIC in terms of members and centres

A consortium of:

- **20 member countries:**
  AT, BG, CY, CZ, DE, DK, EE,
  FI, GR, HR, HU, IT, LT, LV,
  NL, NO, PL, PT, SE, SI

- **4 observer countries:**
  IS, FR, UK, ZA

- >50 centres, subdivided into
  - Centres providing data
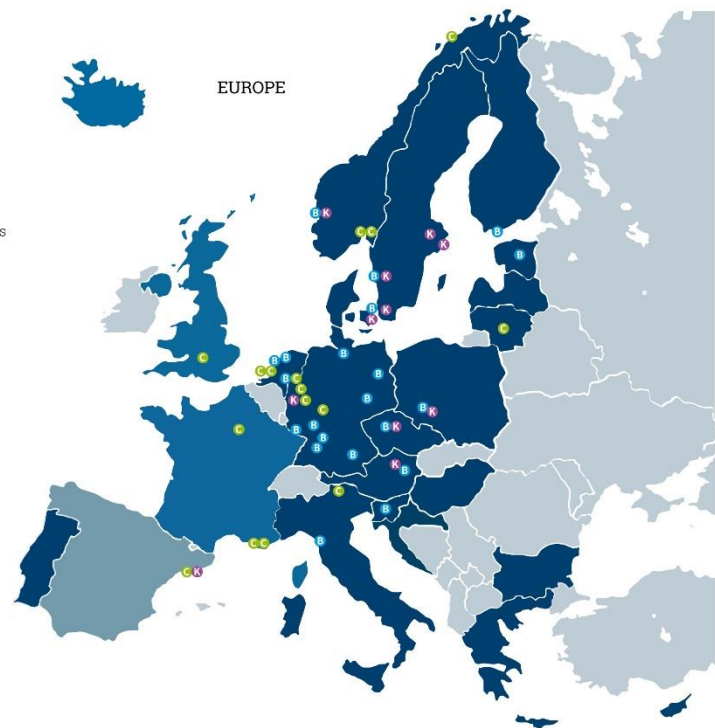 - Centres providing metadata
  - Knowledge centres

EUROPE

CLARIN

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

USA

SOUTH AFRICA

CLARIN in services

CLARIN portal

Get an example-based impression of what's currently available

Depositing services

Store language resources in a sustainable repository at a CLARIN centre

Virtual Language Observatory

Discover language resources using a faceted browser or a map

Easy access to protected resources

Get easy access to protected resources, with your institutional username and password.

Language Resource Switchboard

Explore and analyze language data with a wide variety of tools

Virtual Collections

Create your own digital bookmarks, ideal for citing data sets.

Language Resource Inventory

Submit and access information about language resources relevant to your research.

Content Search (prototype)

Search different corpora with a single search engine

Questions & Answers

Searching for a specific data set or application? Wondering how CLARIN can assist your research? Feel free to contact us!

CLARIN is not just for (computational) linguists! Some roles of language ….

- Carrier of cultural content

- Record of the past

- Main communication instrument within and across societies

- Preserving and disseminating our knowledge

- Instrument to formulate rules for society

- Carrier of information

- Means of human expression

- Focus of cognitive processes

- Component of national or cultural identity

- Object of study

- Object of computer processing

Some roles of language …
… and the corresponding audiences

- Carrier of cultural content: *cultural heritage*
- Record of the past: *archaeology, history*
- Main communication instrument within and across societies: *sociology, anthropology*
- Preserving and disseminating our knowledge: *all disciplines*
- Instrument to formulate rules for society: *law, theology*
- Carrier of information: *media studies, journalism*
- Means of human expression: *literary studies, art, psychology*
- Focus of cognitive processes: *brain studies, psychology, neurology*
- Component of national or cultural identity: *political sciences, social sciences*
- Object of study: *linguistics, language learning*
- Object of computer processing: *language and speech technology*

Why am I presenting CLARIN here?

- Repeat from first slide:
  - Infrastructure … that provides easy and sustainable access for scholars in the **humanities and social sciences** and beyond
  - to **digital language data** (in written, spoken, video or multimodal form)
  - and **advanced tools** to discover, explore, exploit, annotate, analyse or combine them, _wherever they are located_

- Recall the map: Eastern and South-Eastern Europe poorly represented in CLARIN (good players - but few)

- Countries not in CLARIN may be sitting on resources, tools and knowledge that would be of interest to the broader CLARIN community

- And researchers in those countries could benefit from what CLARIN has to offer

- CLARIN would be happy to collaborate with you to make this happen

To conclude

- CLARIN is an infrastructure focused on language in all its manifestations.

- It serves a broad range of audiences, belonging to different disciplines.

- We want to give access to all relevant data and tools - including yours.

- We invite interested parties in CEE and other countries to collaborate with us.

- About getting involved with CLARIN:
    - https://www.clarin.eu/content/get-involved-clarin
    - https://www.clarin.eu/content/information-potential-new-members

- Contact and information:
    - www.clarin.eu
    - Our general email address clarin@clarin.eu
    - Steven Krauwer (steven@clarin.eu)

-ooo-

THANK YOU!