

CLARIN ERIC

K-centre Landscape

Bente Maegaard

bmaegaard@hum.ku.dk

CLARIN K-centre workshop

30 November – 1 December 2020

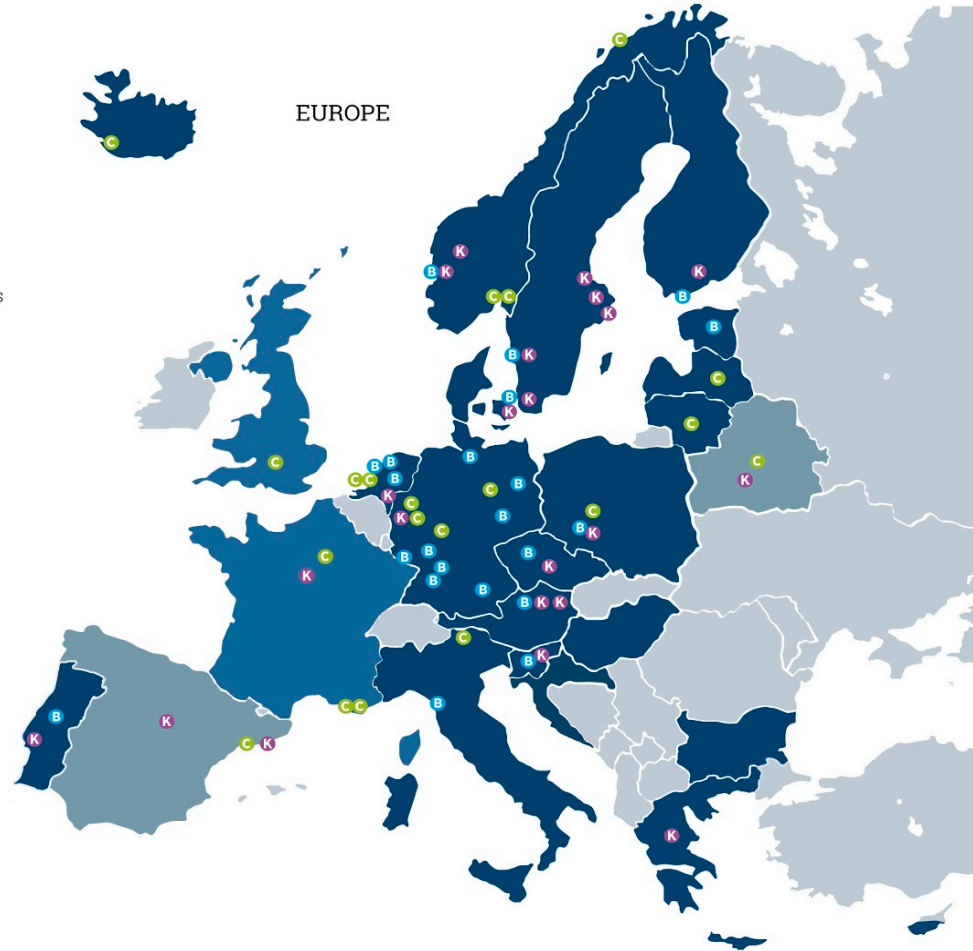
Virtual



CLARIN K-centre Landscape



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



CLARIN K-centres – a cornerstone of the knowledge sharing ecosystem

- **K-centres are institutions that have agreed to share their knowledge and expertise on one or more aspects of the domain covered by the CLARIN infrastructure.**
- The focus of CLARIN is on language resources (in all modalities, from all regions and with any topical orientation) and K-centres serve researchers and educators from any discipline where language plays one of its many roles.
- Most K-centres are in CLARIN countries, but CLARIN K-centres also exist elsewhere, and they all have a virtual presence.
- CLARIN services are free of charge.

CLARIN K-centres

- At present 23 K-centres in 16 countries
- Covering many languages, many types of language resources, many types of methods etc.
- How do you become a K-centre?
 - Application form and guidelines on the website
- What is needed apart from serving users?
 - Keep track of what you have been doing, report once per year
- How do users find appropriate K-centres?

<https://www.clarin.eu/content/knowledge-centres>

- Overview by area of expertise
- Keyword search
- Alphabetical list

K-centre expertise

- [Individual languages](#) (e.g. Danish, Czech, Portuguese), language families (e.g. South Slavic) or groups of languages (e.g. morphologically rich languages, the languages of Sweden)
- [Written text and modalities other than written text](#) (e.g. spoken language, sign language)
- [Linguistic topics](#) (e.g. language diversity, language learning, diachronic studies)
- [Language processing topics](#) (e.g. speech analysis, building treebanks, machine translation)
- [Data types other than corpora](#) (e.g. lexical data, word nets, terminology banks)
- [Using or processing families of language data](#) that will exist for most languages (e.g. newspapers, parliamentary records, oral history)
- [Generic methods and issues](#) (e.g. data management, ethics, IPR, OCR)

List CLARIN K-centres with expertise in modalities. An example: CLARIN SPEECH

Areas of competence

Technical advice on speech analysis relating to all aspects of speech technology, including speech science, speech applications, and speech in interaction.

Audiences served

-

Types of services

awareness, tools, mentoring

Is portal for language(s)

-

Other languages covered

Swedish, English

Modalities covered

Speech, biosignals, audiovisual data, sensor data

Linguistic topics

phonetics, pathology

Language processing

speech analysis, speech modelling, speech processing

Data types

acoustic and language models, dictionaries, vocabularies, pronunciation data, biosignals related to spoken interaction

Resource families

oral history, parliamentary records

Generic topics

deep learning, evaluation, tools, visualization, ASR, legal issue, data management

Other keywords

-

Example: Generic methods and issues, taken across all K-centres

- Critical Data Management; Legal and ethical issues
- linguistic fieldwork
- multimodal and sensor-based methods, including EEG, eye-tracking, articulography, virtual reality, motion capture and av-recording
- deep learning, evaluation, tools, visualization, ASR, legal issue, data management
- deep learning, evaluation of tools (for South Slavic languages)
- OCR, digitisation, visualisation, evaluation of tools
- Data Management, Legal and ethical issues
- Audio and audiovisual recordings

Generic methods and issues (cont.) across all K-centres

- Brain image recording
- Primarily Finite-State Applications, but to some degree also Statistical Methods and Neural Networks
- 1. Grammars; 2. Finite-State Applications; 3. Statistical Methods; 4. Neural Networks
- language policy and planning

Summing up

- Already a very rich collection of K-centres
- Still room for more (cf agenda tomorrow)
- We will continue to follow the developments and to monitor our definition of **Areas of expertise**. You are welcome to send us comments
- You may want to review your own description
- And finally I would like to mention that K-centres are of course not the only source of knowledge in CLARIN:
 - Central team
 - National teams
 - Committees: Legal and ethical, standards, etc.

Thank you!

<https://www.clarin.eu/>

bmaegaard@hum.ku.dk