Introduction to CLARIN
- research infrastructure for language resources
- enabling the study of language as social and cultural data

Franciska de Jong

Dieter Van Uytvanck

INTELE Workshop
13-14 September 2022
Madrid

CLARIN

CLARIN in eight bullets

- **CLARIN** is the Common Language Resources and Technology Infrastructure
- **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 a distributed network of **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
- through a **single sign-on** environment
- that serves as an ecosystem for **knowledge exchange**
- some services **integrated in EOSC** (member of European Open Science Cloud; link)

CLARIN Value Proposition: https://www.clarin.eu/content/value-proposition (link to pdf)

CLARIN in countries and centres

Research Infrastructure CLARIN (= Common Language Resources and Technology Infrastructure)

A consortium of type ERIC; after 10 years:

- 22 members
- 2 observers
- 1 linked party

A distributed network of >60 centres

25 CTS certified data centres,

strong focus on FAIRness & interoperability

- federated login:
- central metadata harvesting for easy discovery:
- chained services:
- language data - in written, spoken, video or multimodal form
- advanced tools - to discover, explore, exploit, annotate, analyse or combine data sets, *wherever they are located*
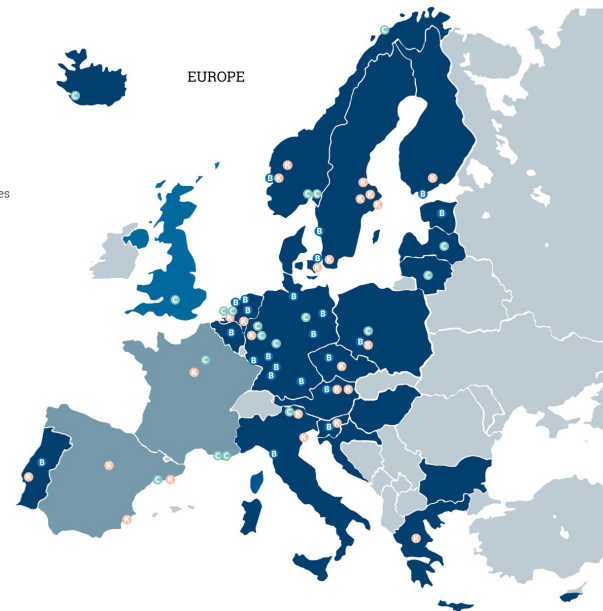
CLARIN in services

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 in services

> 13000 resources in the languages from Spain , including sign language data

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 ed 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!
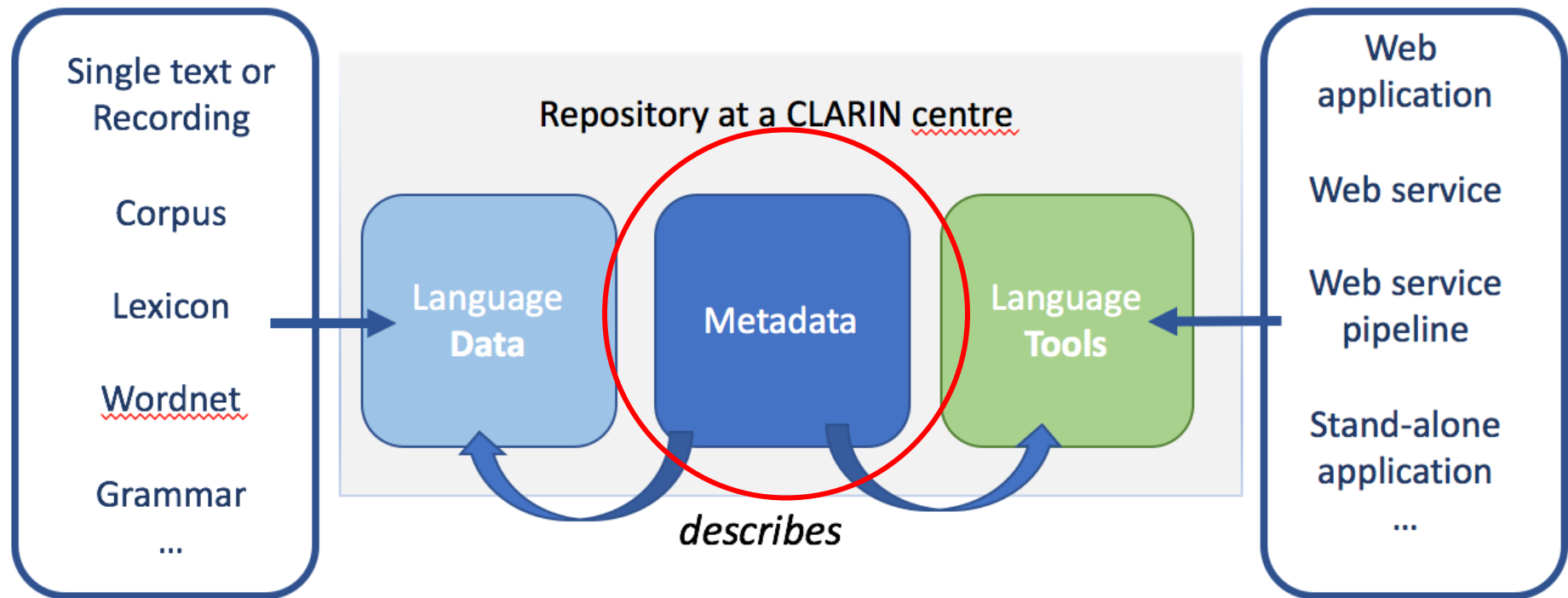
CLARIN in data types

- 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
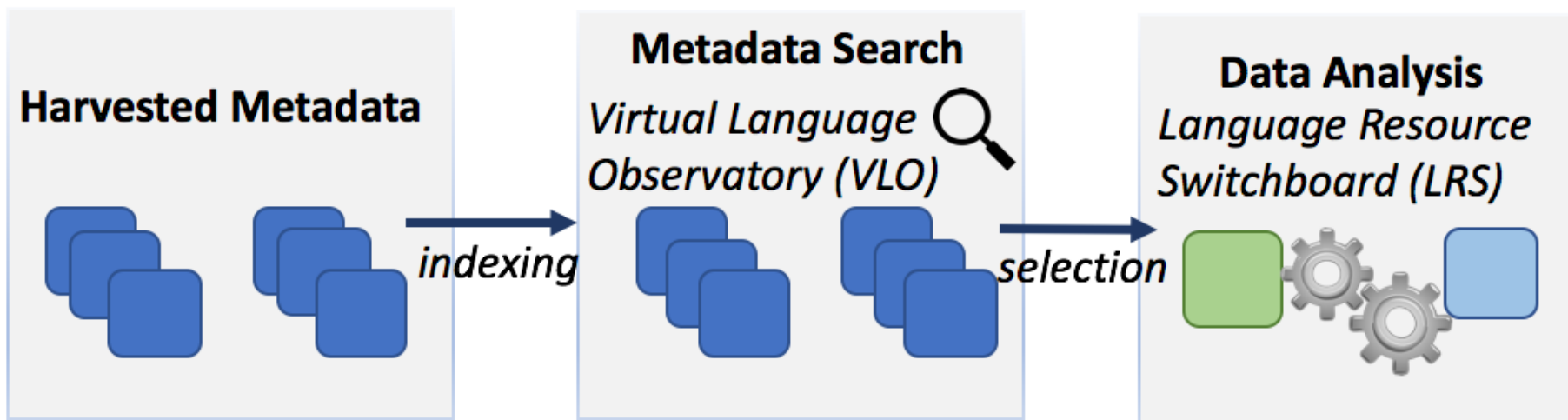
CLARIN in communities of use

- Digital Humanities
- Linguistics and Philology
- Translation and Lexicography
- Literary Studies
- History
- Political and Social Sciences
- Media Studies
- Culture, Folklore, Anthropology
- Speech therapy
- Teachers
- General Public
- ….

The CLARIN data architecture:
a network of distributed centre repositories

The CLARIN data architecture:
central processing of metadata

In the design of CLARIN interoperability is key

- to enable the exchange and harvesting of metadata

- to identify suitable exchange formats for the output of analytic tools for language data

- to the options for supporting comparative research;

ParlaMint, a CLARIN initiative aimed at interoperability of parliamentary data and public debate on COVID-19; data in around 20 languages, including Spanish

ParlaMint

CLARIN and Open Science

- Promoting the sharing and re-use of data through sustainable data registries
  - Academia, Industry, Citizen Science, GLAM sector
- All integrated datasets available in open access for research purposes
- Adherence to the FAIR data principles
  - Findable, Accessible, Interoperable, Re-usable
  - Interoperability through a common metadata framework
- Support for linguistic diversity
  - Data covering more than 1500 languages
  - Tools for many languages
  - Language resources in all modalities
- Strengthening the support for > 500.000 professional SSH researchers in Europe and globally a multitude.

CLARIN: Towards FAIR and Responsible Data Science Using Language Resources." In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, May 2018, 3259-3264.

CLARIN as ecosystem for knowledge exchange

- A network of Knowledge Centres (K-centres)

- Sharing of expertise and best practices
  - Annual Conference
  - Support for workshops and mobility
  - Ambassador network

- Capacity training through life events, online courses and webinars
  - for developers
  - for end-users

- Collaboration with
  - other SSH infrastructures (e.g. CESSDA, DARIAH-EU, ESS, SHARE;
  - Europeana, LIBER, EASSH, ADHO, Time Machine, …

- Membership of EOSC (=European Open Science Cloud)
  - projects: EOSC-hub, EOSC Future, FAIRCORE4EOSC, EOSC Focus;
  - participation in EOSC task forces

Useful Links

- CLARIN Value Proposition (link)

- How to prepare for joining (link)

- Newsflash (link)

- Tour de CLARIN (link)

- CLARIN Funding Hub (link)

- Workshops (link)

- Training Network (link)

- Training Suite (link)
  for 'Teaching with CLARIN'

see you @

www.clarin.eu

CLARIN