

CLARIN ERIC

Darja Fišer

Delavnica Odprta znanost

23. januar 2020



Kaj je CLARIN?

- Common Language Resources and Technology Infrastructure
- Status **ESFRI ERIC** od 2012, **Landmark** od 2016
- Vizija:
 - digitalni **jezikovni viri** in **orodja** za vse (evropske) jezike
 - so **dostopni** prek **enotne prijave**
 - za raziskovalce v **humanistiki** in **družboslovju** ter ostalih ved, ki uporabljajo jezikovne podatke
- Namen:
 - **dolgotrajno** in obsežno **hranjenje** in **dostop** do jezikovnih virov in tehnologij
 - prispevek k **ohranjanju** in **podpiranju** večjezične evropske kulturne dediščine
- Inovacija:
 - zagotavljanje **standardizacije** in **večkratne uporabnosti** virov in orodij

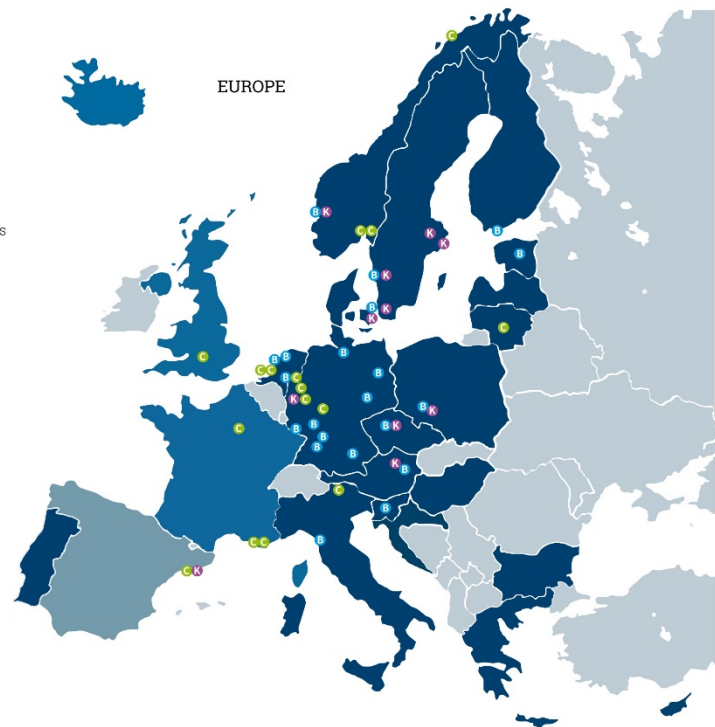
Članice in centri CLARIN ERIC

Konzorcij:

- 20 članic: AT, BG, CY, CZ, DE, DK, EE, FI, GR, HR, HU, IT, LT, LV, NL, NO, PL, PT, SE, SI
- 4 opazovalke: IS, FR, UK, ZA
- > 50 centrov



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



Storitive v CLARIN-u



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!

Jezikovni podatki in orodja v CLARIN-u

- Korpusi
 - časopisna besedila
 - literarna besedila
 - historična besedila
 - družbena omrežja
 - parlamentarne razprave
- Leksikalni viri
 - slovarji
 - glosarji
 - semantični leksikoni
- Orodja
 - tegiranje
 - lematizacija
 - imenske entitete



Kaj CLARIN nudi

- Dogodki
 - Letna konferenca (slovenski delegati in avtorji)
 - Financiranje razvojnih in izobraževalnih delavnic
 - Financiranje raziskovalnih obiskov
- Storitve
 - Strokovno ekspertizo (<https://www.clarin.eu/content/knowledge-centres>)
 - Pregled družin virov in orodij (<https://www.clarin.eu/resource-families>)
 - Učne vsebine (<https://training.parthenos-project.eu>)
 - VideoLectures (<http://videlectures.net/clarin/>)

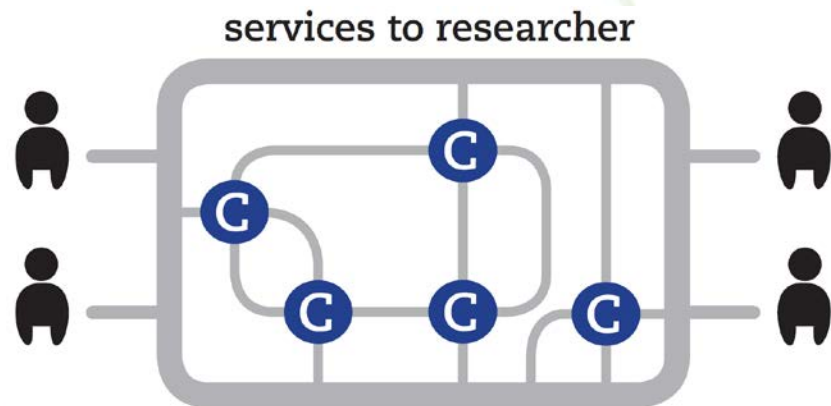
CLARIN centri

- **Distribuirana arhitektura:**

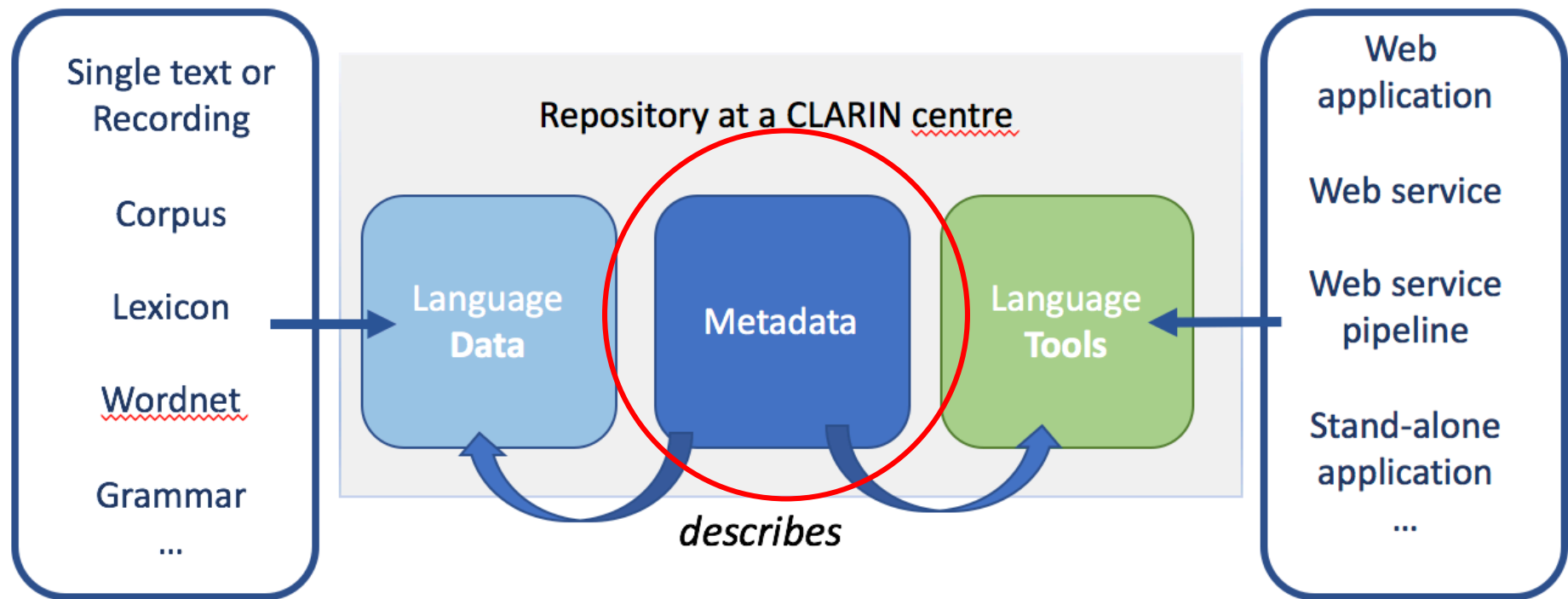
- datoteke, spletne aplikacije in storitve po vsej Evropi
- infrastruktura skupek posameznih centrov
 - 20 certificiranih B centrov
 - > 50 vseh registriranih centrov

- **Orodja in podatki z različnih centrov so interoperabilni**

- omogoče izmenjavo metapodatkov
- omogoča kombiniranje raziskovalnih podatkov
- omogoča združevanje orodij
- podpira komparativne raziskave



Podatkovna arhitektura v CLARIN-u: *rezpozitoriji CLARIN centrov*



CLARIN in odprta znanost

- [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.
- Upoštevanje načel FAIR
 - Trajni registri podatkov omogočajo **deljenje, iskanje, ponovno uporabo** podatkov
 - Integrirani podatki so za raziskovalce na voljo v **odprtem dostopu**
- Podpora jezikovni raznolikosti
 - Podatki v več kot 1500 jezikih
 - Orodja za številne jezike
 - Jezikovni viri v vseh modalnostih
- Spodbujanje odgovorne podatkovne znanosti



Hvala za pozornost!

www.clarin.eu

darja.fiser@ff.uni-lj.si