CLARIN - Infrastructural support for the study of language as social and cultural data

Franciska de Jong
CLARIN ERIC
f.m.g.dejong@uu.nl

Stay tuned to the future, Bologna

24-25 January 2018

CLARIN

- Intro to CLARIN
- Open Science
- Multilinguality and Multidisciplinarity
- Demonstration of impact

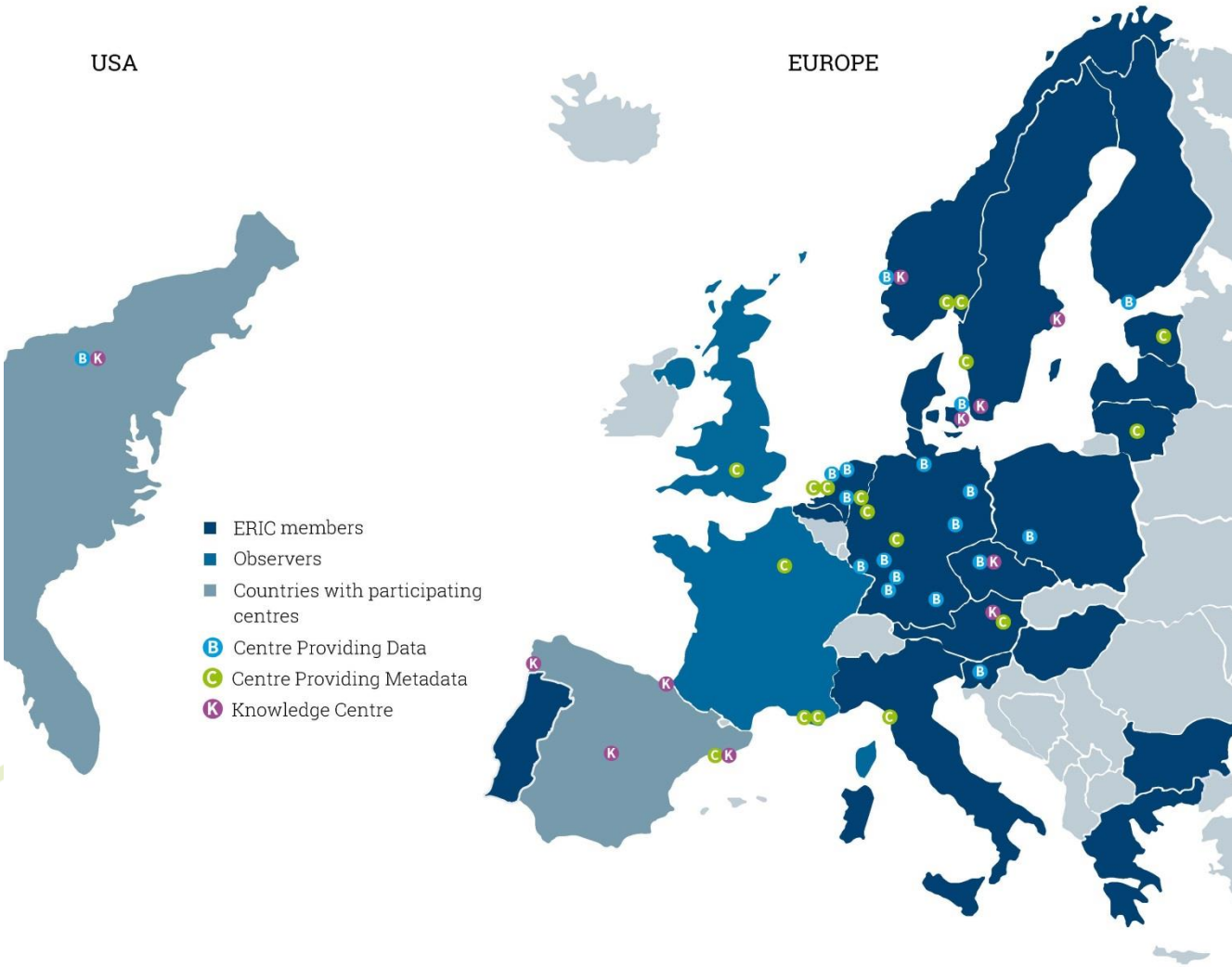
CLARIN in seven bullets

- **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 **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
- and that serves as an ecosystem for **knowledge sharing**.

CLARIN ERIC in members and centres

A consortium of

- 19 members: AT, BG, CZ, DE, DK, DLU, EE, FI, GR, HU, IT, LT, LV, NL, NO, PL, PT, SE, SI
- 2 observers: FR, UK;
- >40 centres



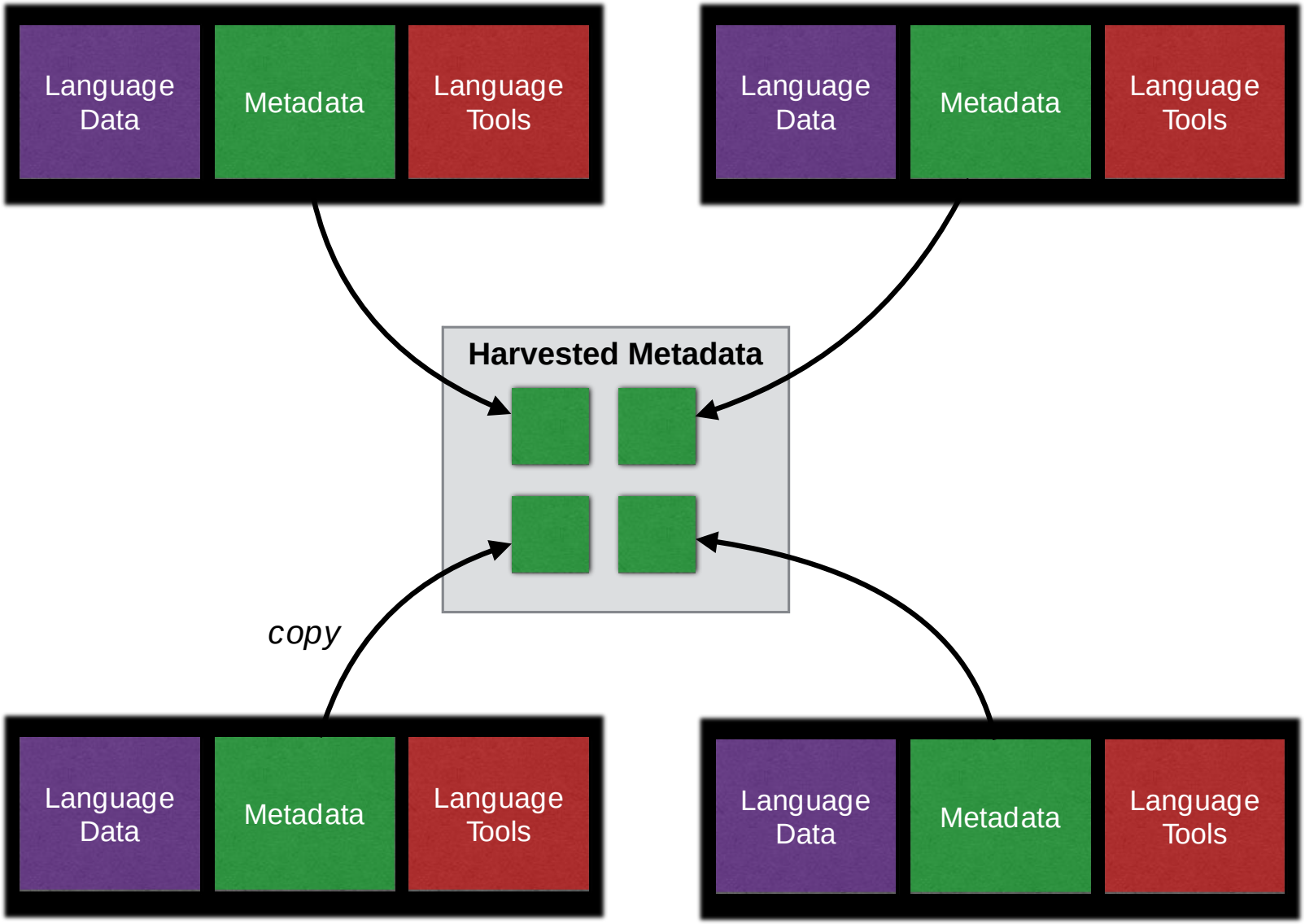
CLARIN in resource types

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

Open Science

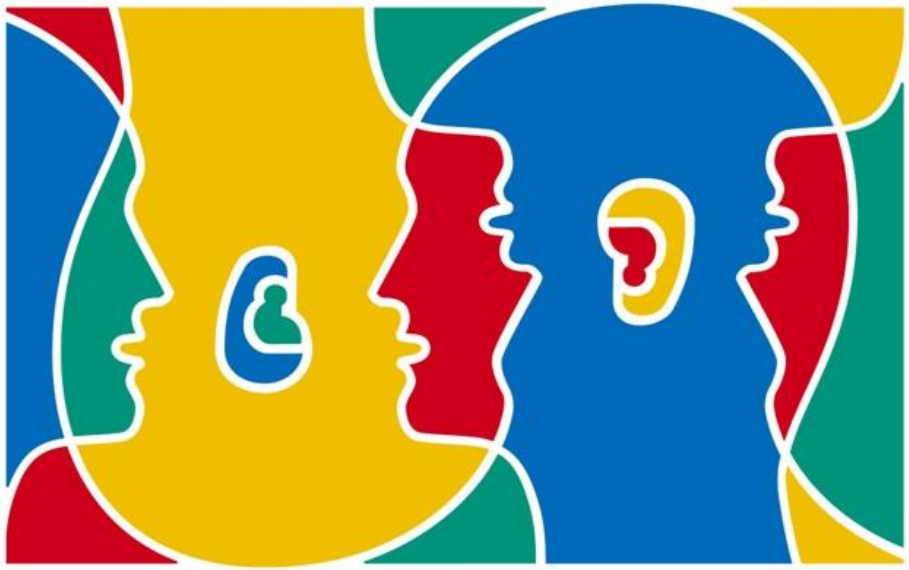
- FAIR data
  - Findable
  - Accessible
  - Interoperable
  - Reusable
- Responsible data science
  - Quality of underlying data
  - Clarity on performance level of analysis tool
  - Transparancy

Harvesting of metadata for Findability



Multilinguality and multidisciplinarity

Europe's multilinguality is key …*



….

- to our understanding of how language affects identity, culture, society
- to our understanding of diversity across boundaries of time and regions
- and therefore for comparative studies

* image from https://www.coe.int/t/dg4/linguistic/jel_en.asp

Multilinguality -> Multidisciplinarity

Europe's **multilinguality** is a basis for **comparative research** of societal and cultural phenomena, that are reflected in language use.

Some examples:

- Migration patterns
- Intellectual history
- Language variation across period and region
- Dynamics in mental health conditions
- Parliamentary discourse
- Customer opinions

Data -> Conclusions -> Decisions

Uptake of the results of data analysis is dependent on:

- transparency of the algorithms applied
- explainability / interpretability of the results
- reproducability

Needed:

- insight in the validity of analysis outcomes
- frameworks for the integrated processing of multiple datatypes.

Demonstration of impact

More than a technical facility

CLARIN provides access to data and tools, and thereby leverages the regional and national investments in data creation, data curation and tool development.

In addition it has the role of knowledge broker:

- organisation of workshops to stimulate multidisciplinary collaboration
- supporter of training nodes
- network of knowledge centres covering a wide range of topics of expertise
- platform for communities of use

Showing impact

- Measurable KPIs
- Involvement in projects with parties from industry
- Involvement in projects around the EOSC agenda
- Role in education at large, inclduing training of new generation of data scientists
- Awards
- Videos with testimonies from users

Authorship attribution

10 mei 2016

Petrus Datheen auteur van het Wilhelmus?

Nieuw onderzoek met computeranalyses

Petrus Datheen

© Beeld RD

Het 'Wilhelmus is het het oudste volkslied ter wereld. Het werd geschreven aan het begin van de Tachtigjarige Oorlog (ca. 1570) toen de Lage Landen in opstand kwamen tegen de repressiepolitiek van de Spaanse koning en zijn 'IJzeren' hertog Alva. Traditioneel wordt de tekst toegeschreven aan Marnix van Sint-Aldegonde. Maar wetenschappers zijn altijd aan die toewijzing blijven twijfelen. Met behulp van computeranalyses is nu een heel andere kandidaat uit de bus komen rollen: Petrus Datheen.

Details: M. Kestemont et. al. In: Proceedings DH2017

see you @

www.clarin.eu
or
f.m.g.dejong@uu.nl