

Digital Objects – towards implementation

the CLARIN use cases



Dieter Van Uytvanck
DO Workshop
5 April 2019
Philadelphia

CLARIN?

- **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
 - to **advanced tools**
 - to discover, explore, exploit, annotate, analyse or combine them, wherever they are located
- through a **single sign-on** online environment

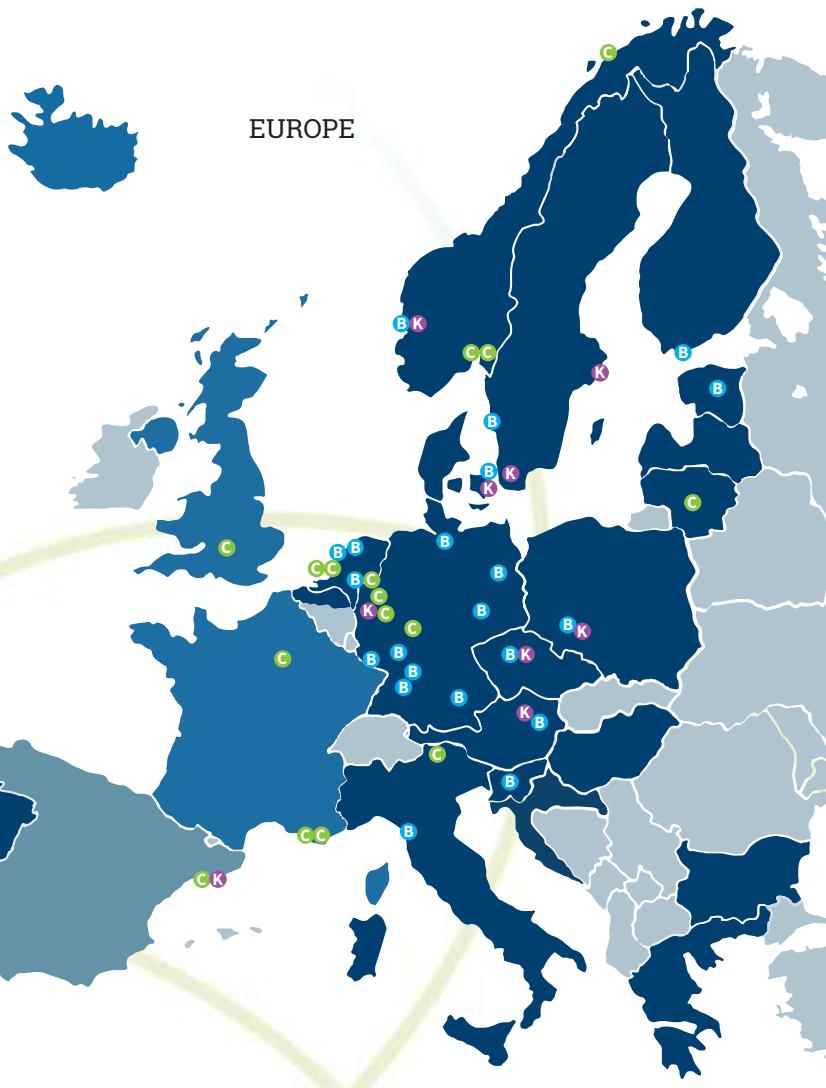
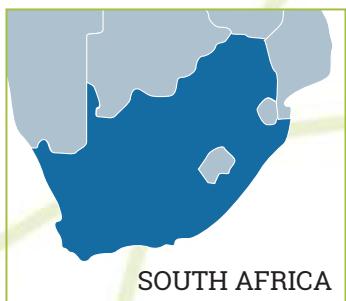
CLARIN ERIC in members and centres

A consortium of:

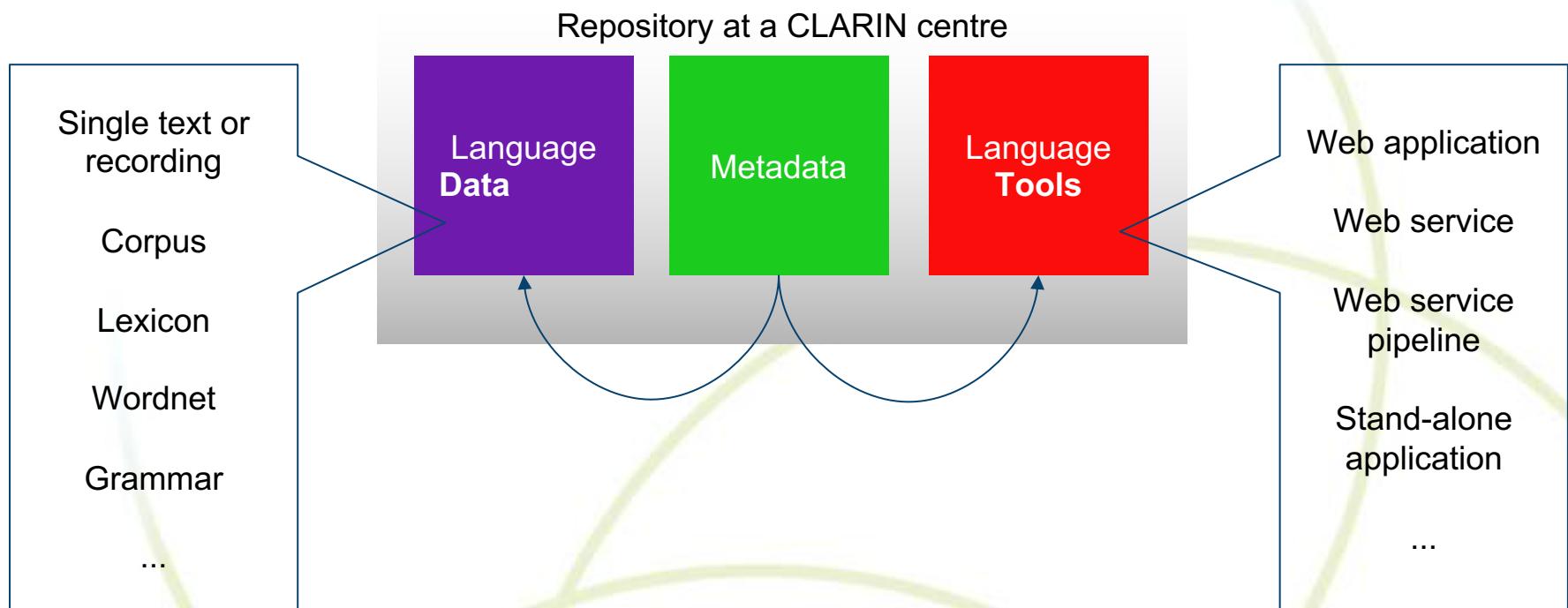
- 20 members:
AT, BG, CZ,
DE, DK, DLU,
EE, FI, GR, HR,
HU, IT, LT, LV,
NL, NO, PL,
PT, SE, SI
- 4 observers:
FR, UK, IS, SA
- >50 centres



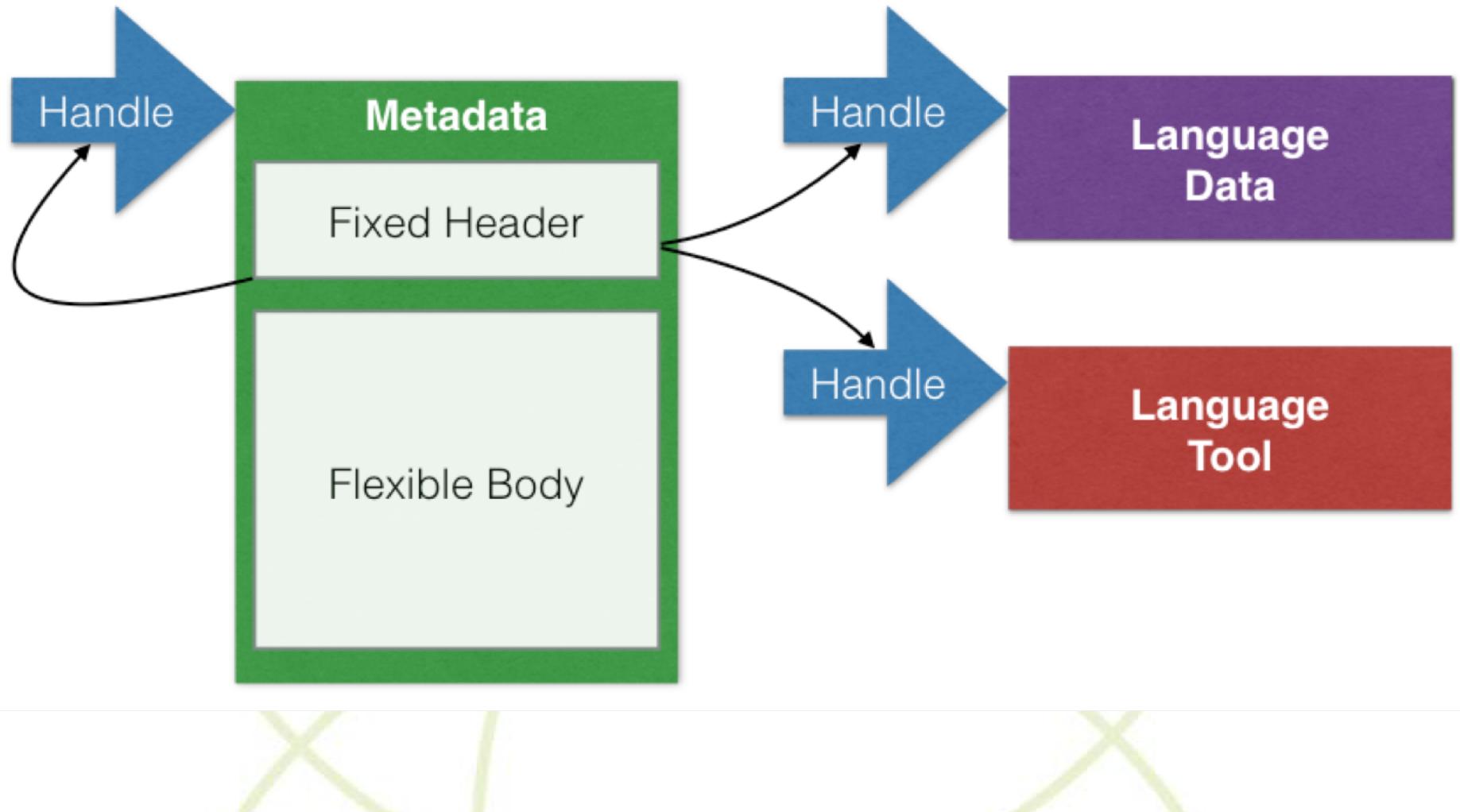
- ERIC members
- Observers
- Countries with participating centres
- B Centre Providing Data
- C Centre Providing Metadata
- K Knowledge Centre



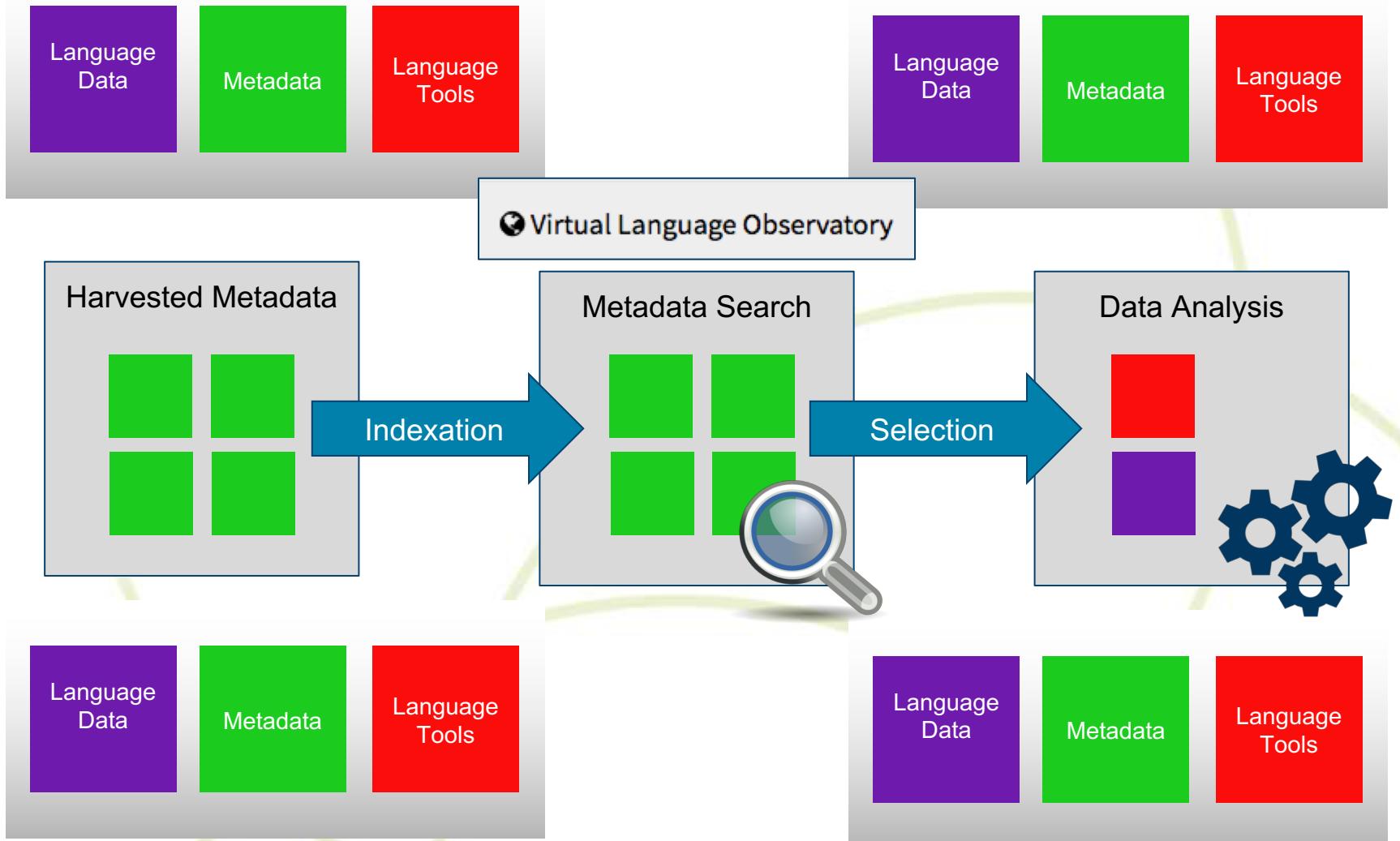
Data architecture at a CLARIN centre



The CLARIN data architecture: Persistent Identifiers (PIPs)

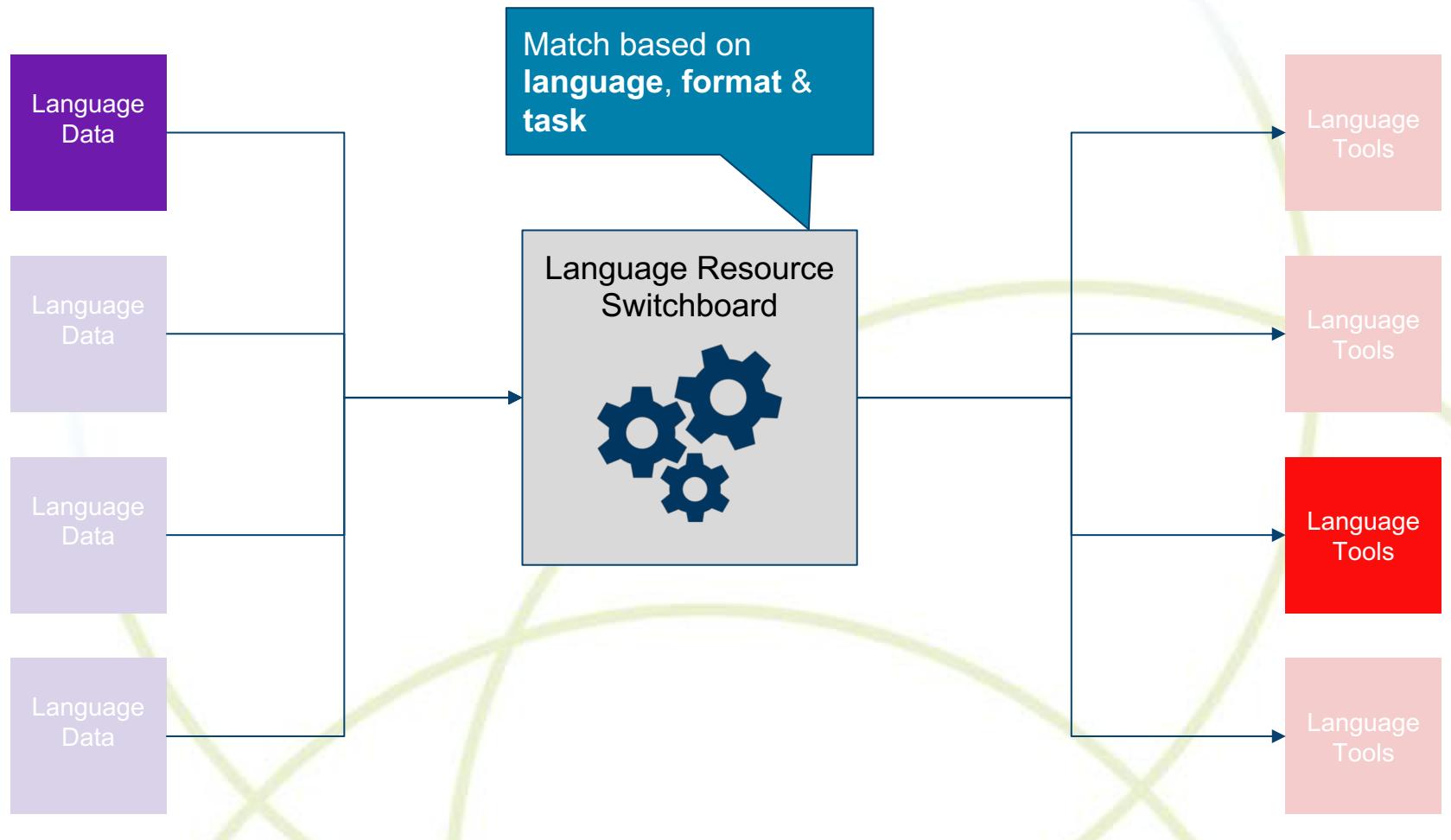


Data discovery and processing

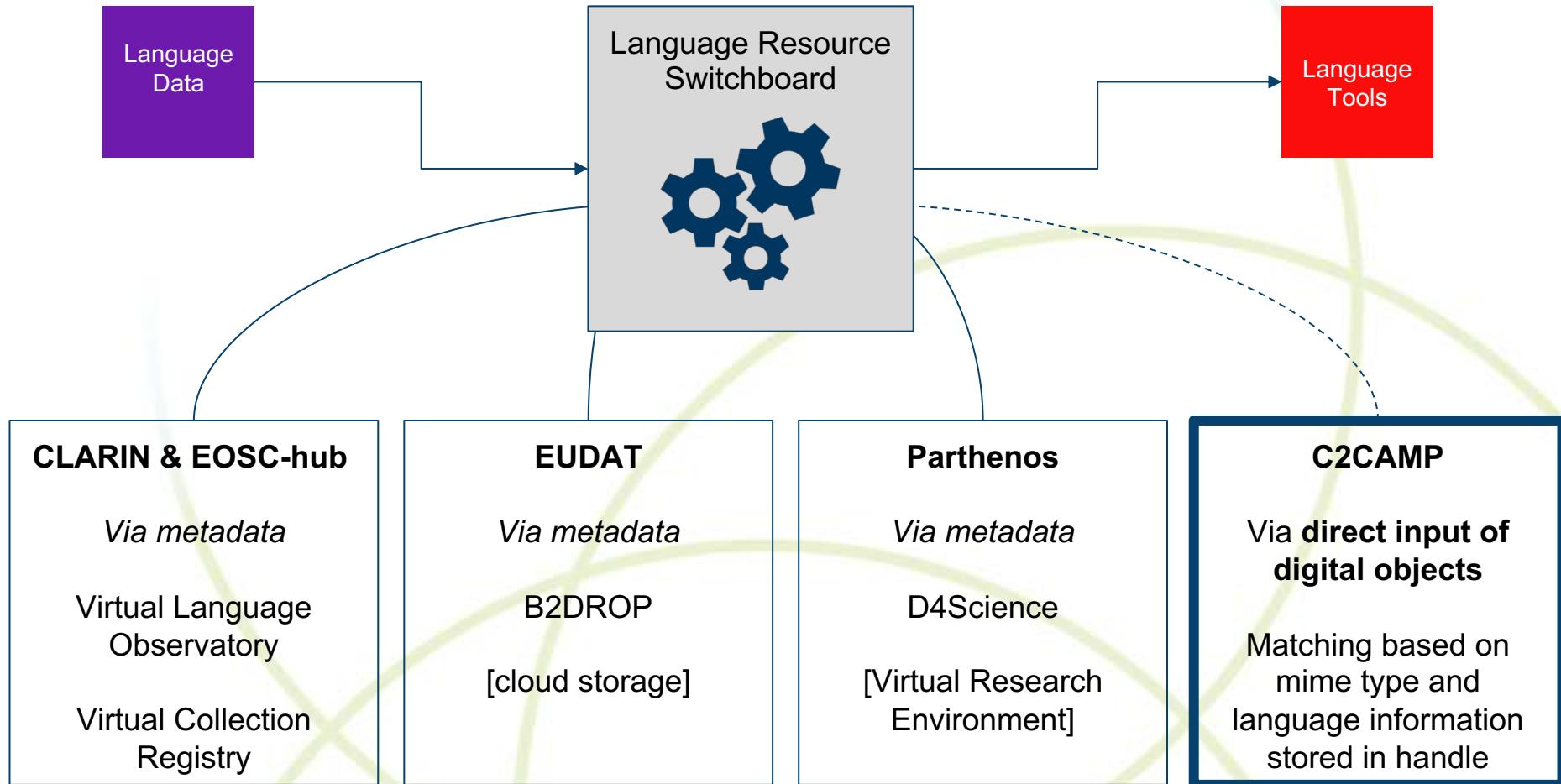


Language Resource Switchboard

switchboard.clarin.eu and clarin.eu/eosc



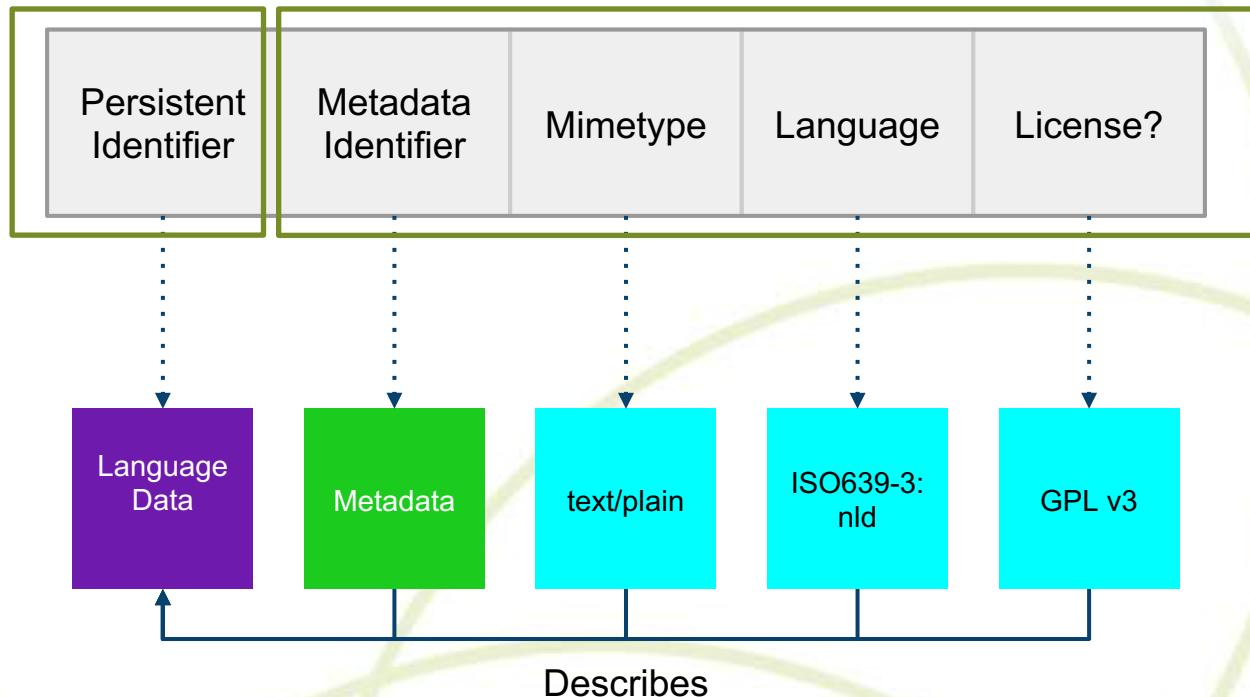
Language Resource Switchboard (switchboard.clarin.eu)



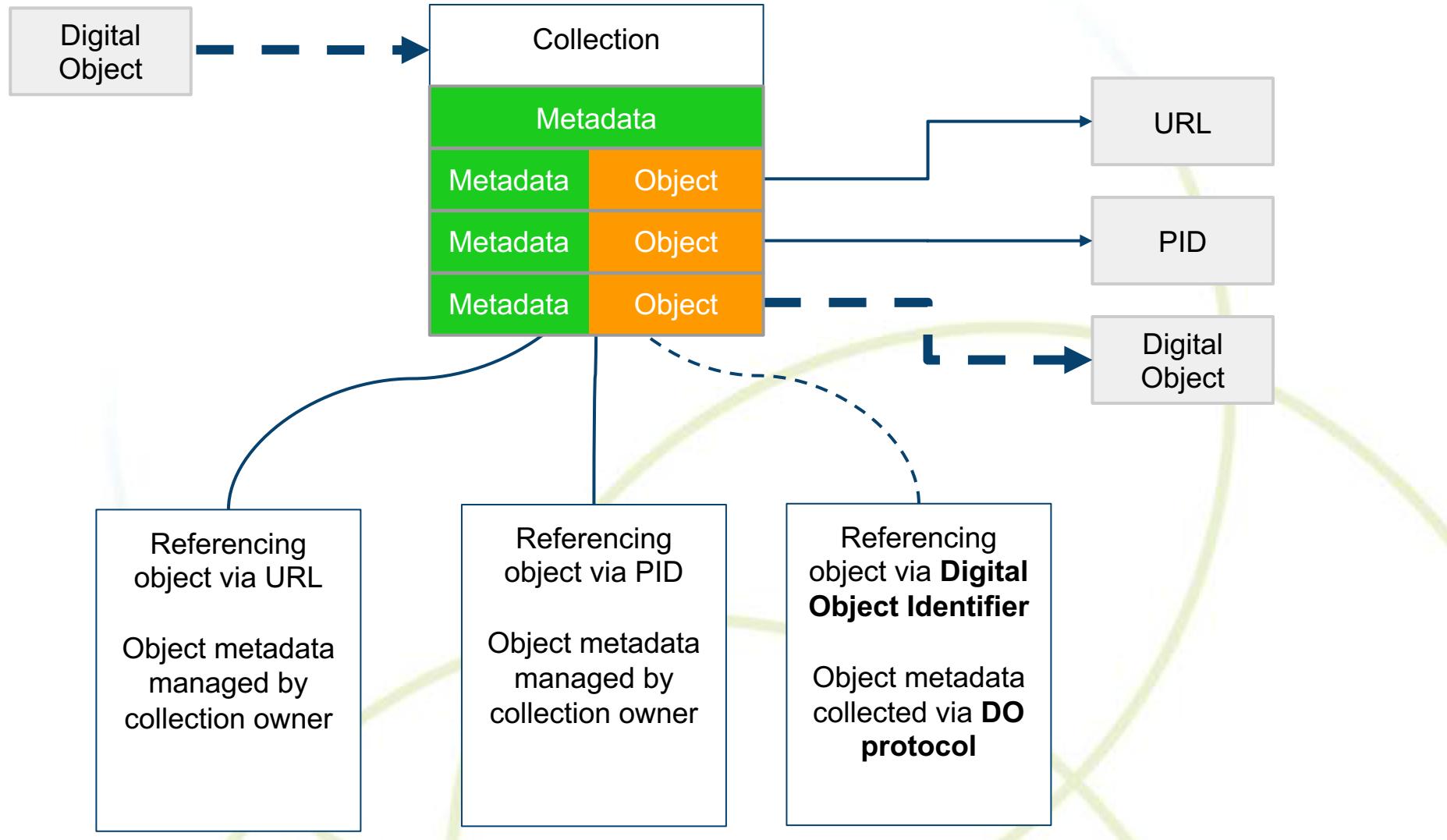
CLARIN Digital Object requirements

PID Kernel Information

Community profile



Virtual Collections (collections.clarin.eu)



Next steps

- Select a DTR [**done**] > the EPIC one
- Register an account and create a prototype profile
- Register a few example handles and make them full-fledged Digital Objects
- Adjust the Switchboard so that it
 - can recognize Digital Objects when a handle is provided as input
 - extract relevant metadata from the DO to feed it into the switchboard as parameter
- Adjust the Virtual Collection Registry
- (if project granted) further steps > DSpace extension etc.

Summary

- CLARIN use-cases
 - Language Resource Switchboard
 - Virtual Collection Registry
- Important DO properties
 - Mime Type
 - Language
 - License
 - Reference to descriptive / user metadata

Digital Object Protocol

