

# Digital Objects & CLARIN use cases

Dieter Van Uytvanck, Willem Elbers

GEDE DO Workshop

26 September 2018

Brussels, Belgium



# CLARIN in six 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** online environment

# CLARIN ERIC in members and centres

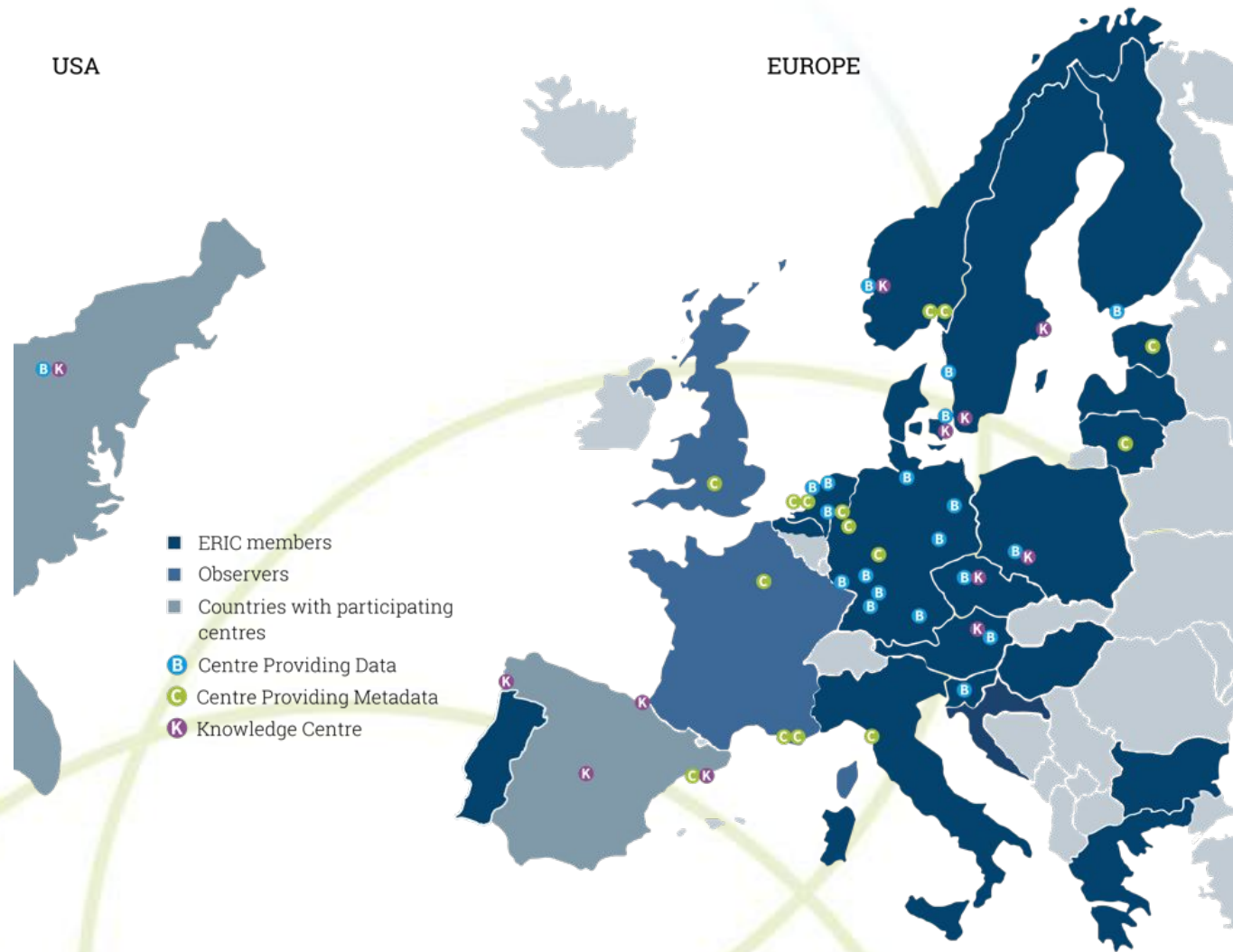
A consortium  
of:

20 members:  
AT, BG, CZ, DE,  
DK, DLU, EE,  
FI, GR, HR, HU,  
IT, LT, LV, NL,  
NO, PL, PT, SE,  
SI

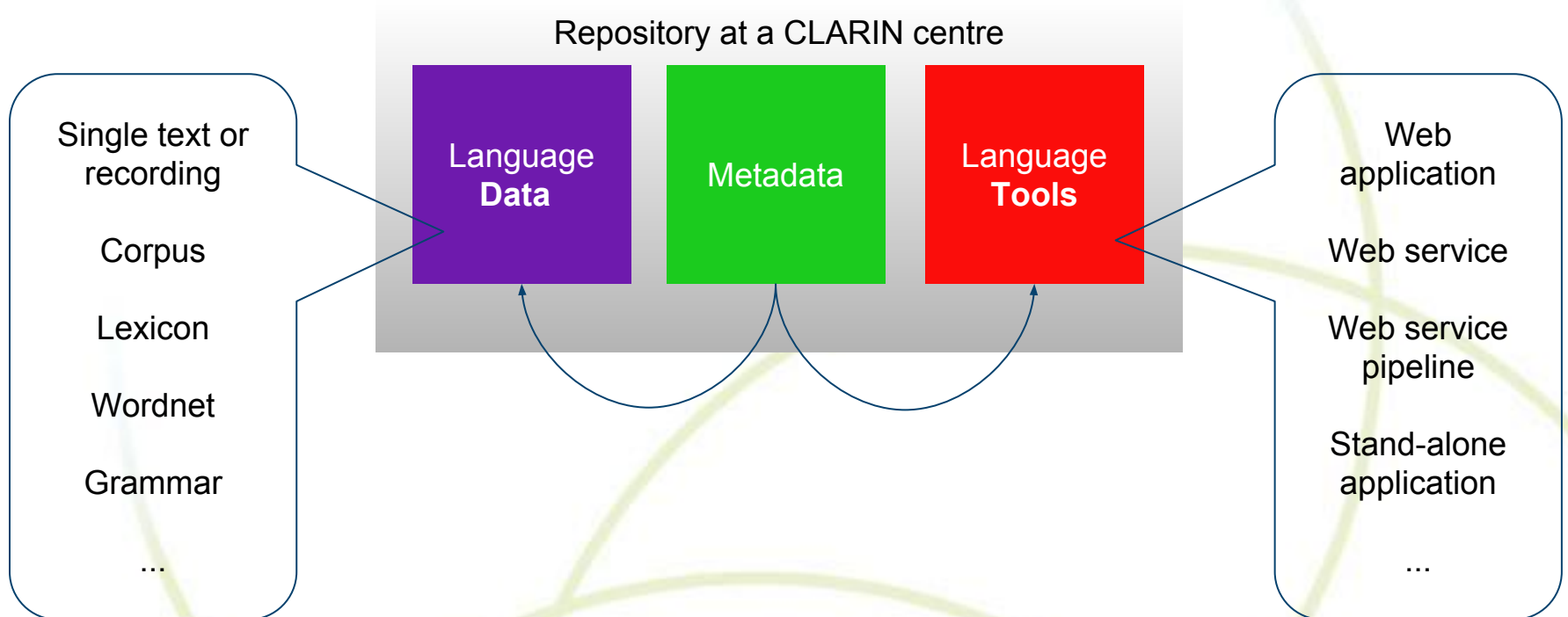
2 observers:

FR, UK;

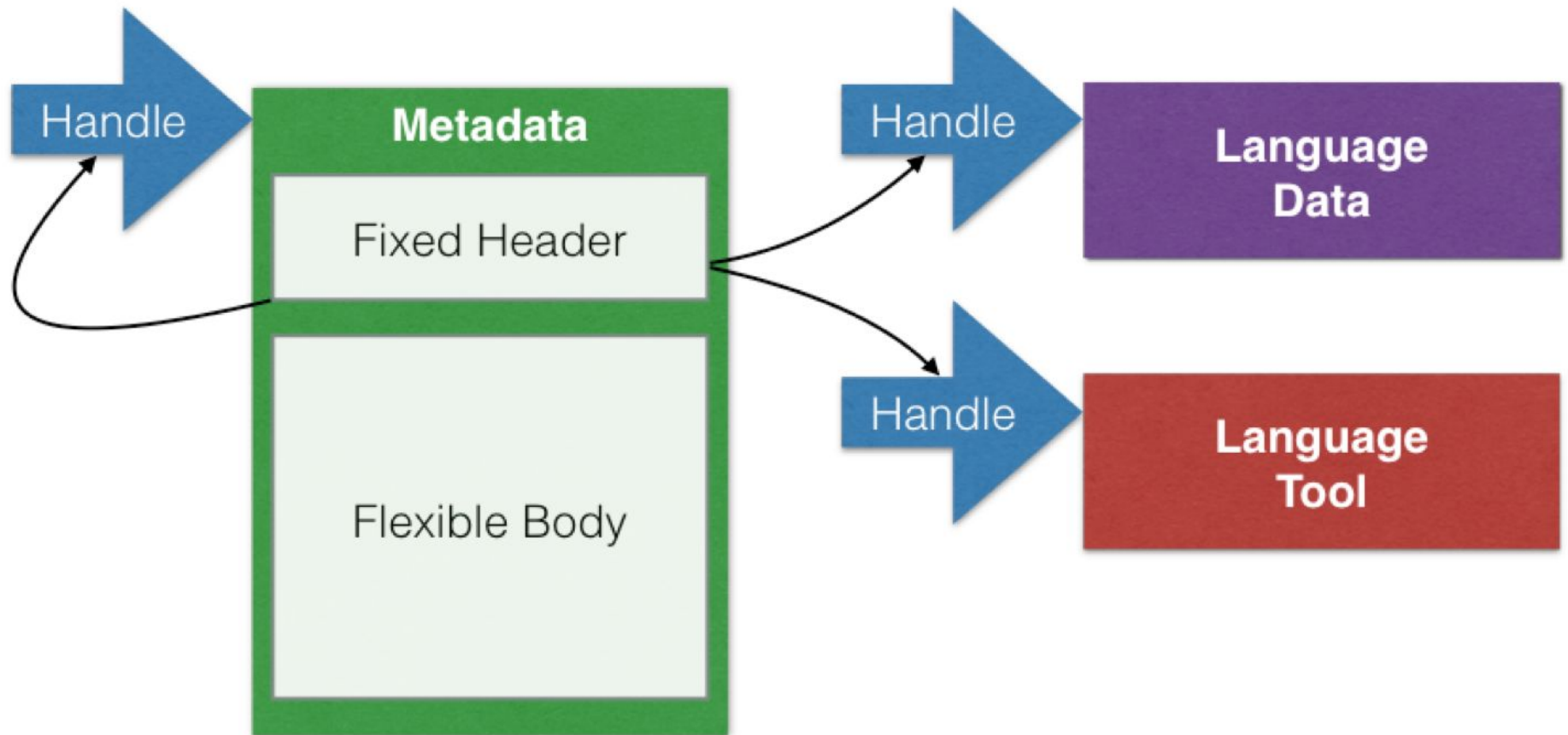
>40 centres



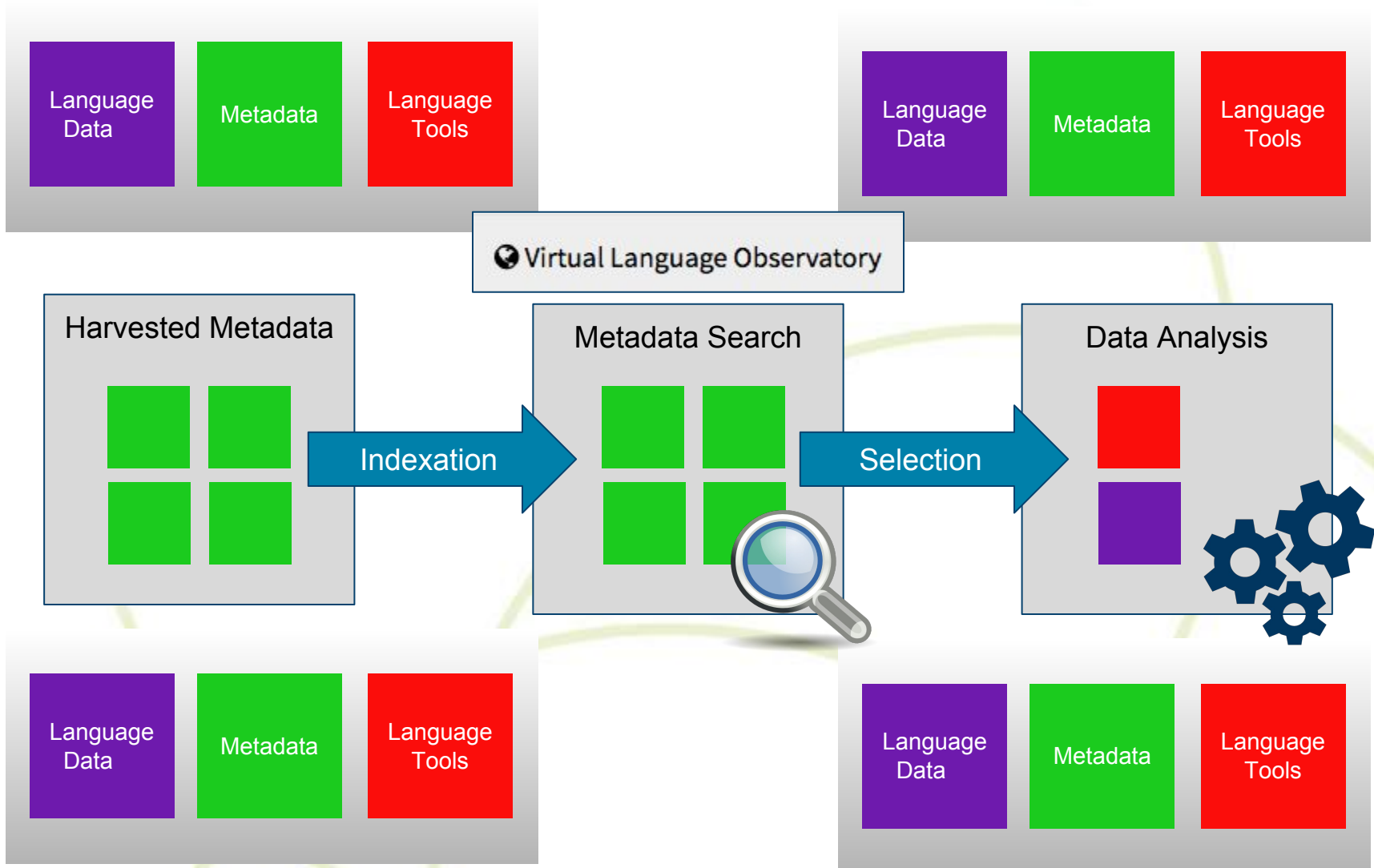
# Data architecture at a CLARIN centre



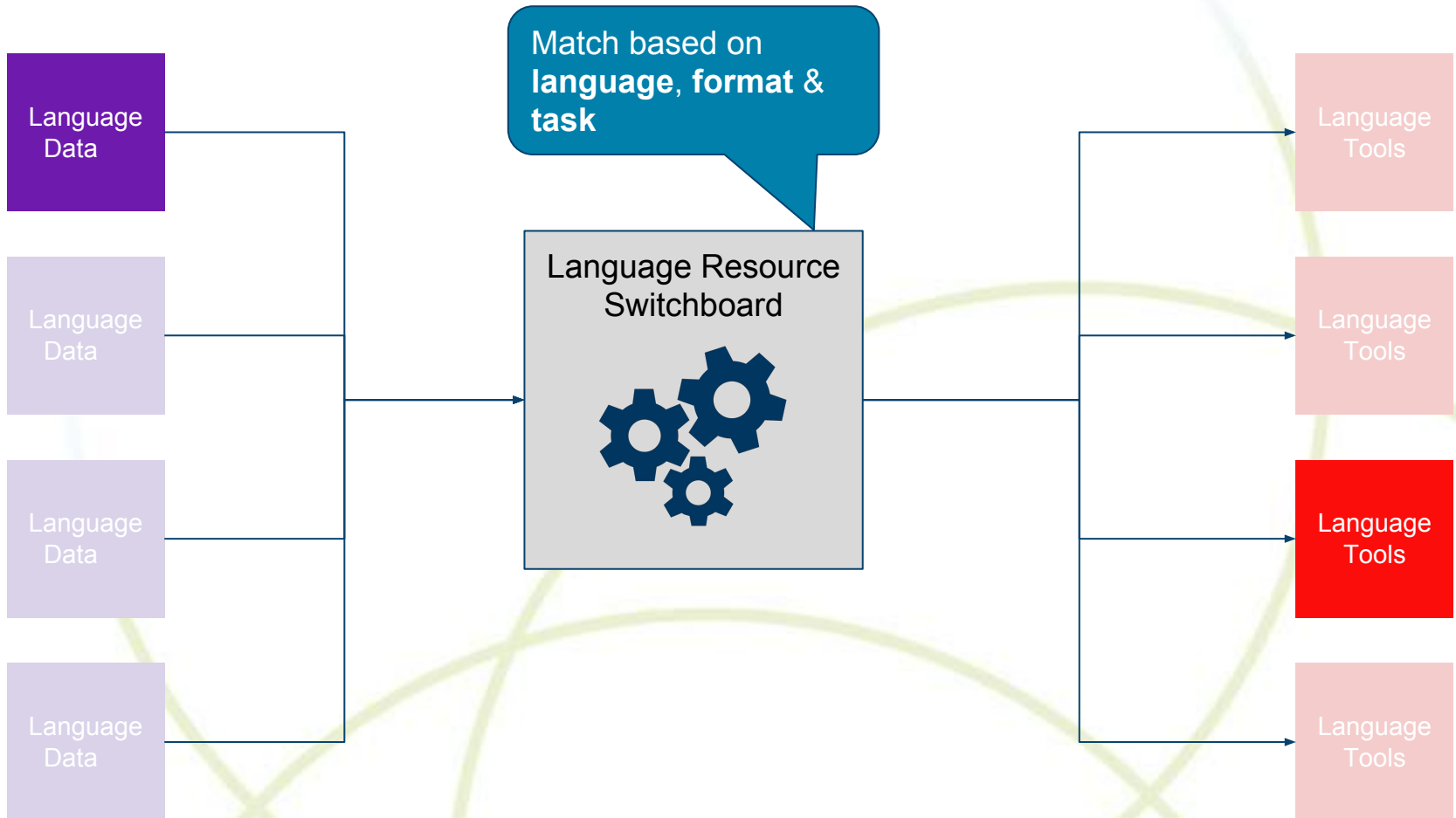
# The CLARIN data architecture: Persistent Identifiers (PIDs)



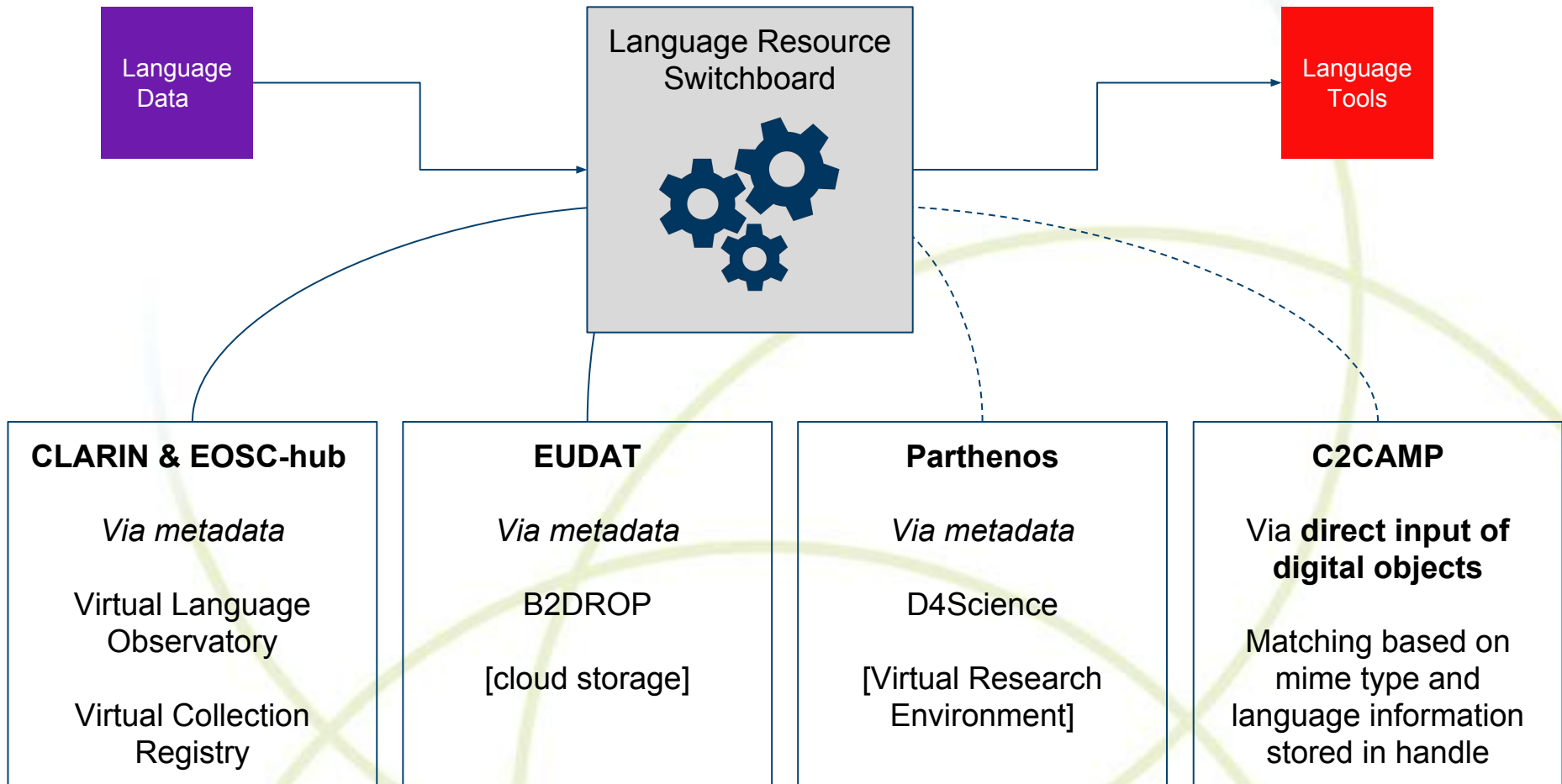
# Data discovery and processing



# Language Resource Switchboard

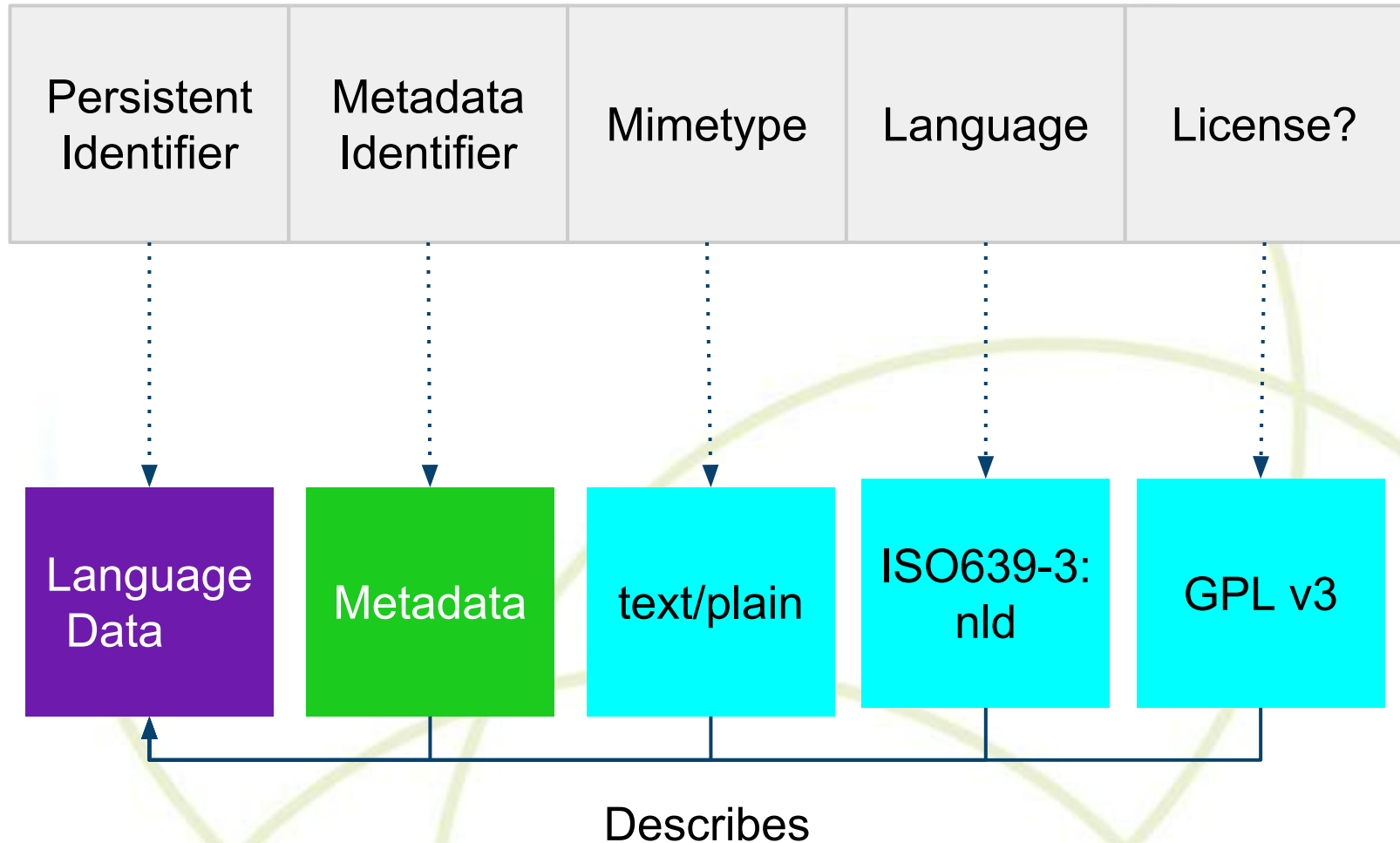


# Language Resource Switchboard

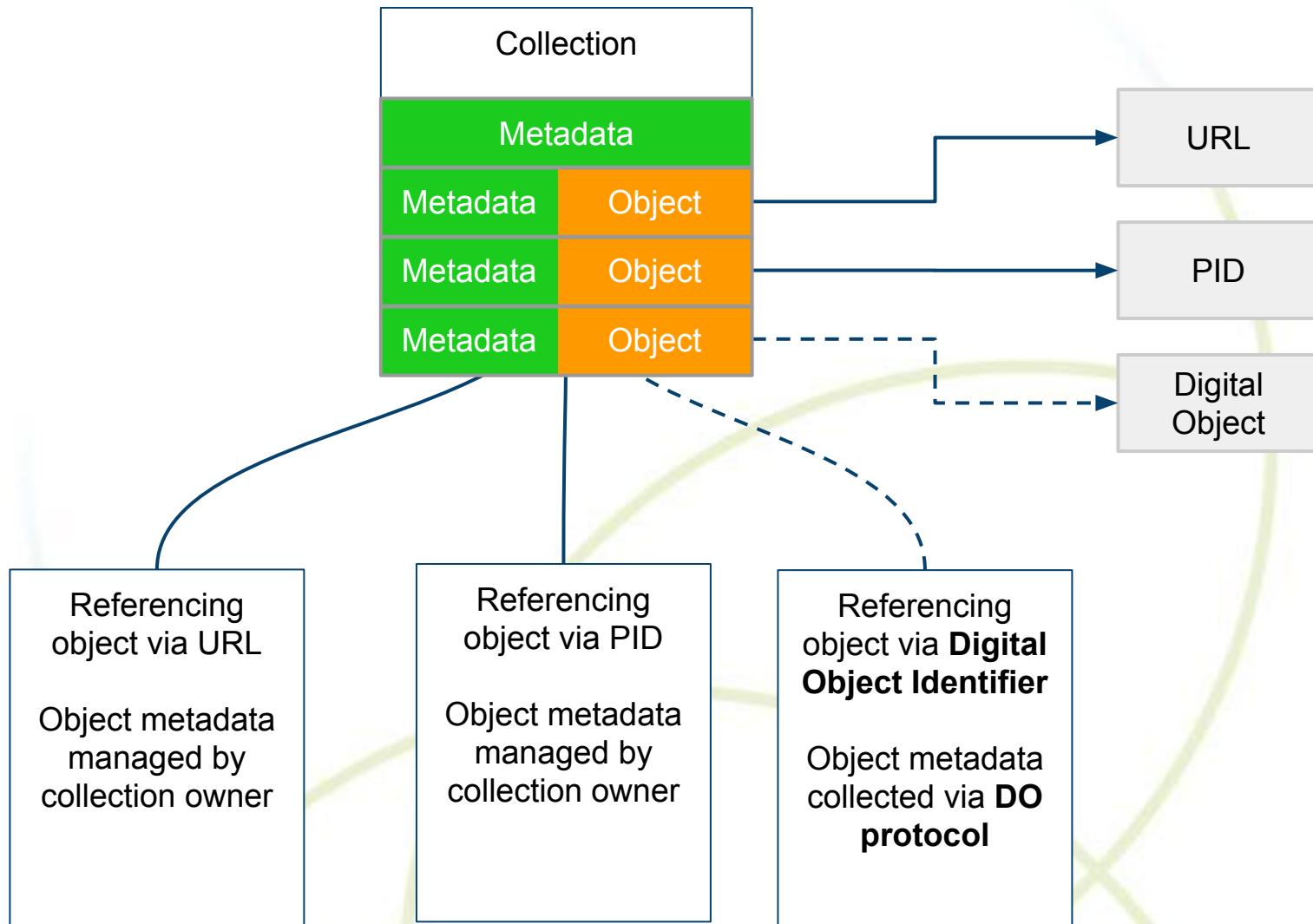




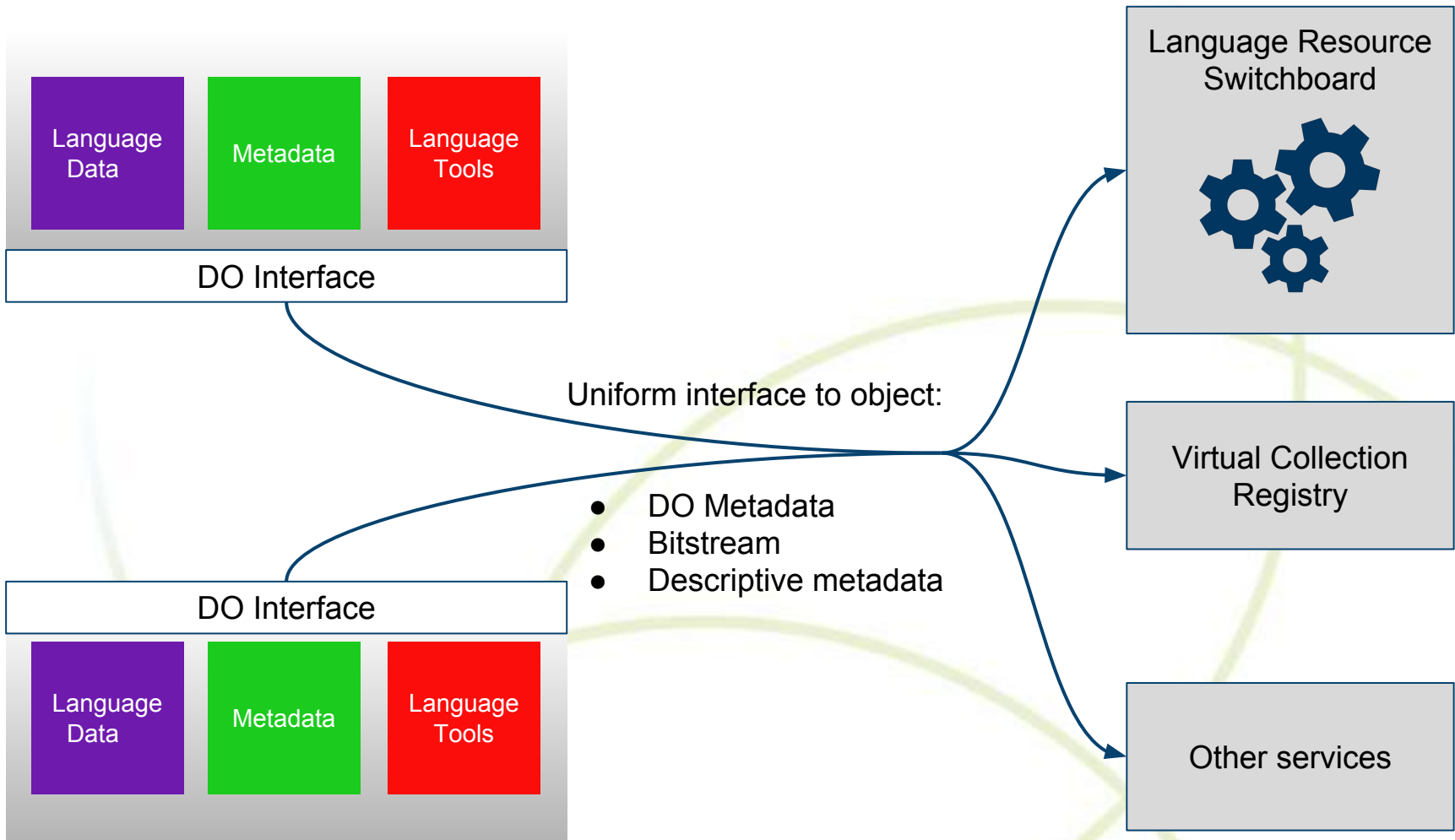
# CLARIN Digital Object requirements



# Virtual Collections



# Digital Object Protocol



# Summary

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