CLARIN and the adoptation of the RDA's
Data Foundation and Terminology

Dieter Van Uytvanck
Technical Director CLARIN ERIC
dieter@clarin.eu
2015-03-08

**RDA Adoption Day**
San Diego

CLARIN?

- **C**ommon **La**nguage **R**esources and Technology **In**frastructure

- European (ESFRI) Research Infrastructure – ERIC since February 2012

- aims at providing easy and sustainable access for scholars in the **humanities and social sciences**

  - to **digital language data** (in written, spoken, video or multimodal form)

  - to **advanced tools** to discover, explore, exploit, annotate, analyse or combine them

CLARIN centres

- A **distributed architecture**: (http-accessible) files, web applications and web services spread all over Europe
- Nodes in the network: **centres** (http://clarin.eu/centres)
- Some of them password-protected (licenses, privacy, …)
- User base: also spread over Europe (and rest of the world)

services to researcher
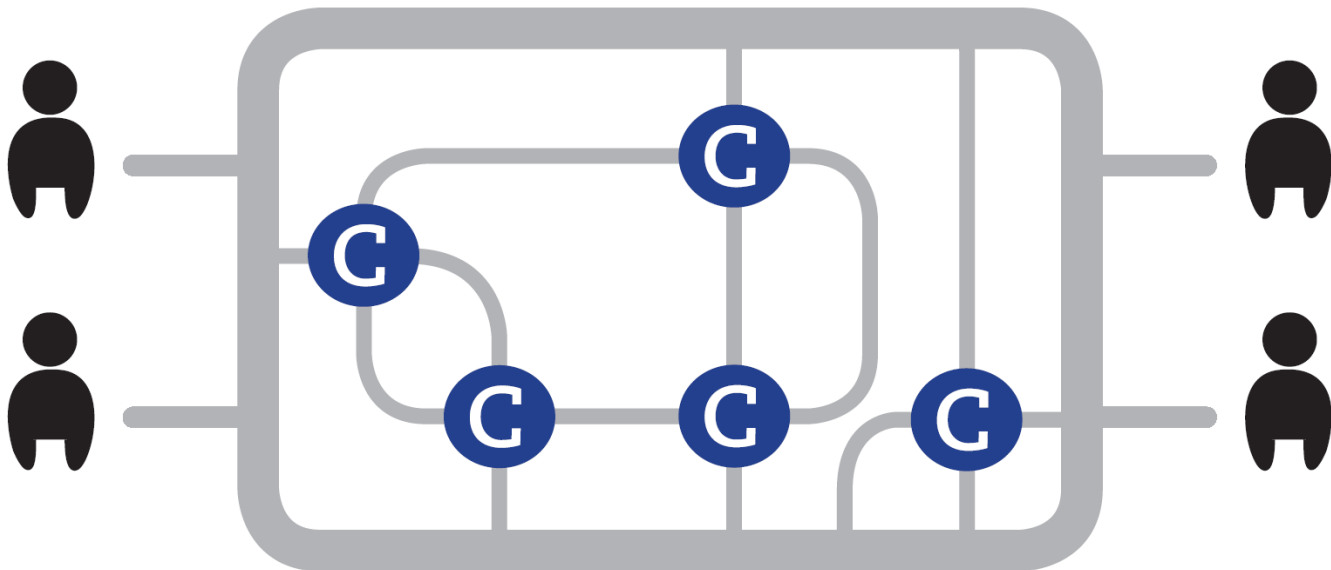
Organisation CLARIN

- **Members**: countries (AT, DK, CZ, DE, EE, GR, LT, NL, NO, PL, PT, SE) and international organisations (Dutch Language Union)

- **Resources & Services provided** (http://clarin.eu/services):
  - Access to language resources (including federated login when needed)
  - Access to language resource processing applications/services
  - Depositing services
  - Metadata catalogue: Virtual Language Observatory
  - "Glue components" like the Virtual Collection Registry (http://clarin.eu/vcr)
  - Consulting services
  - (+ a whole set of technical services behind the scenes)

The 28 CLARIN centres

Core of DFT

(Source: DFT 2: Analysis & Synthesis)

In the core is the concept of a **digital object** which:

- is represented by a **bitstream** (instantiations of which can be stored in several repositories)
- is referenced by a **persistent ID**
- is described by **metadata** and
- can be **aggregated to digital collections**.

DFT in CLARIN's architecture

digital object

Language **Data**

Metadata

Language **Tools**

bitstream

describes

DFT in CLARIN's architecture

| Language **Data** | Metadata | Language **Tools** |

describes

single text or recording
!
corpus
!
lexicon
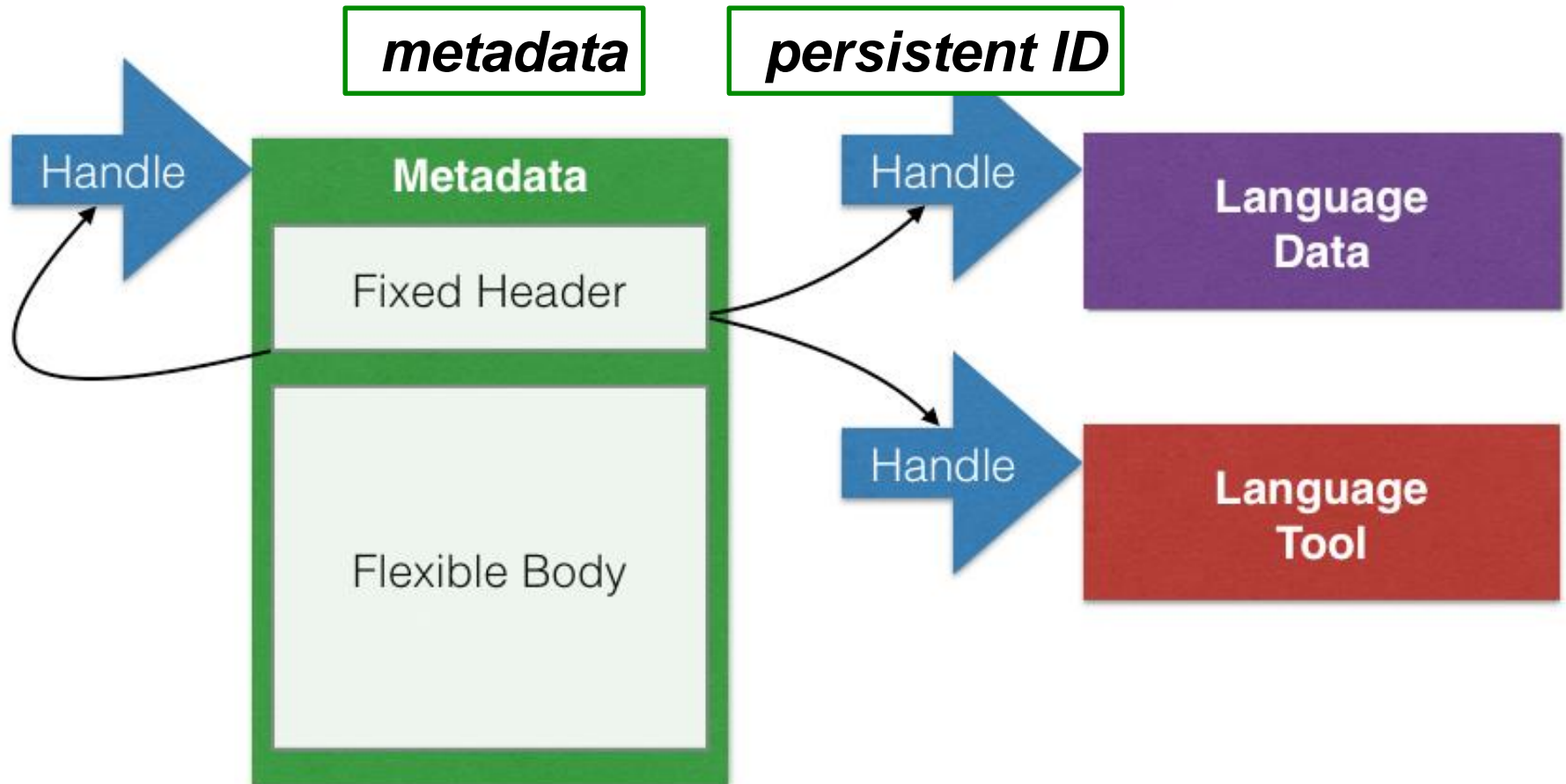!
wordnet
!
grammar
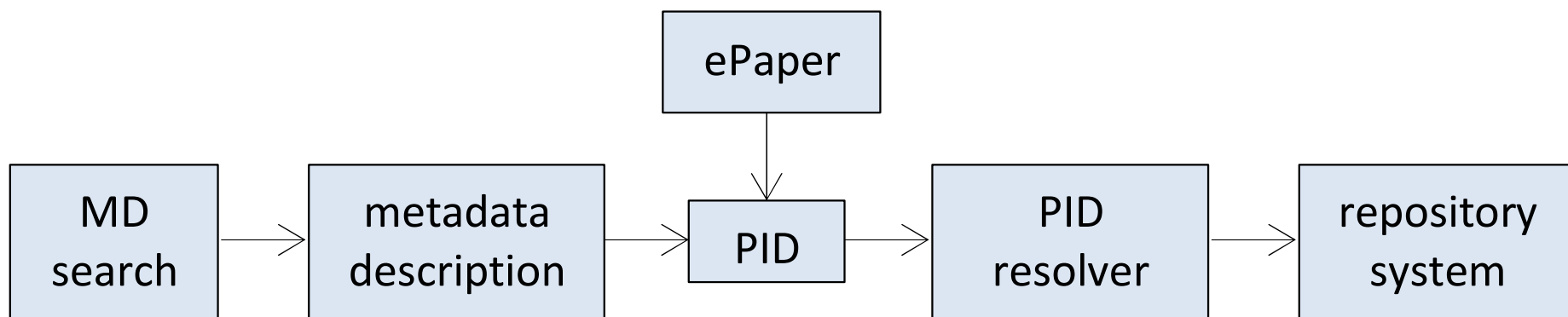!
…

web application
!
web service
!
web service pipeline
!
stand-alone application
!
…

DFT in CLARIN's architecture

DFT: Access Methods

DFT in CLARIN's architecture: Virtual Collections

Registry

my collection

digital collections
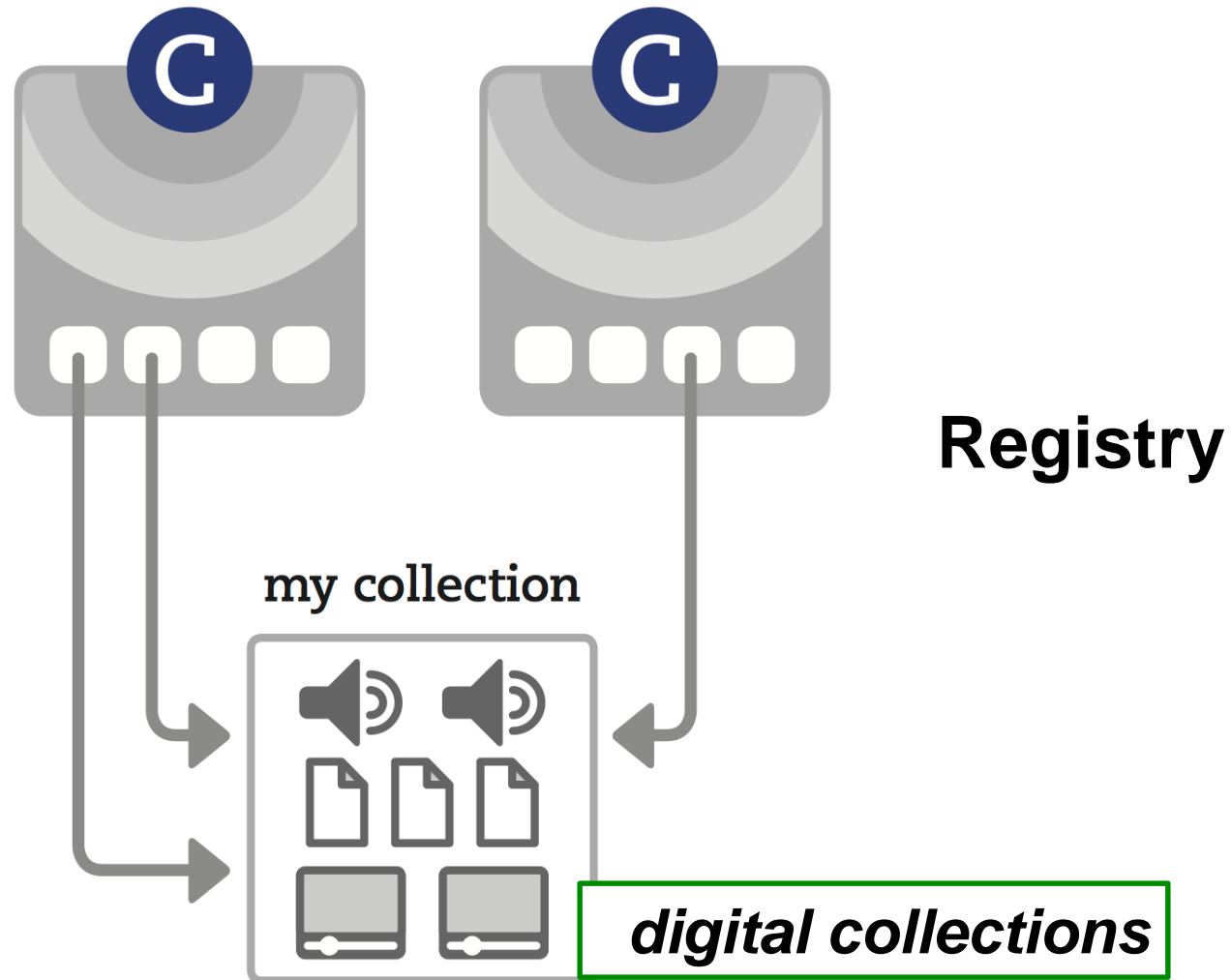
DFT in CLARIN's architecture: Virtual Collections

- VC registry available
- http://clarin.eu/content/virtual-collections

Absolute spatial deixis and proto-toponyms in Kata Kolok

General

Name:	Absolute spatial deixis and proto-toponyms in Kata Kolok
Type:	extensional
Creation Date:	2014-09-26
Description:	Digital references for De Vos, C. (2014). Absolute spatial deixis and proto-toponyms in Kata Kolok. NUSA: Linguistic studies of languages in and around Indonesia, 56, 3-26.
Purpose:	research
Reproducibility:	intended
Persistent identifier:	hdl:11372/VC-1001
Keywords:	■ sign language ■ Kata Kolok

Creators

Person:	Connie de Vos
Organisation:	Max Planck Institute for Psycholinguistics
Website:	http://www.mpi.nl/people/vos-connie-de
Role:	Researcher

Resources

Reference	Type
Journal Article (fulltext) This paper presents an overview of spatial deictic structures in Kata Kolok, a sign language which is indigenous to a Balinese village community.	Resource
Footnote 3 - video Absolute versus absolute transpositional pointing signs	Resource
Footnote 4 - video COME-HERE-FROM-A and GO-FROM-HERE-TO-B	Resource

Feedback as early adopter

- Very handy to have a lingua franca when discussing research infrastructure architectures

- It was good to be involved as adopting community from the start of the work

- Would be good to have a single and concise summarizing document / webpage (with a PID :)

Thank you for your attention!

For more details:
http://clarin.eu > Services / Centres / Technology