CLARIN and the adoption 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?

- **Common Language Resources and Technology Infrastructure**
- 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](http://clarin.eu/centres))
- Some of them password-protected (licenses, privacy, …)
- User base: also spread over Europe (and rest of the world)
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](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](http://clarin.eu/vcr))
  - Consulting services
  - (+ a whole set of technical services behind the scenes)
The 28 CLARIN centres
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

bitstream

Language Data

Metadata

Language Tools

describes
DFT in CLARIN’s architecture

Language Data

Language Tools

Metadata

describes

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

web application
!
web service
!
web service pipeline
!
stand-alone application
!
...
DFT in CLARIN’s architecture

**metadata**

**persistent ID**

- **Handle**
  - **Metadata**
    - Fixed Header
    - Flexible Body
  - **Language Data**
  - **Language Tool**
1.2.2 Access Methods

We can basically identify 2 methods to access a bitstream that encodes the information we wish to retrieve. These methods can be identified as

1. Direct access
2. Indirect access

Direct access can be realized in different ways:

A. MD search
B. Metadata description

Indirect access is realized through:

A. PID
B. PID resolver
C. Repository system
DFT in CLARIN’s architecture: Virtual Collections
DFT in CLARIN’s architecture: Virtual Collections

- VC registry available
- [http://clarin.eu/content/virtual-collections](http://clarin.eu/content/virtual-collections)

---

### Absolute spatial deixis and proto-toponymy in Kata Kolok

**General**

- **Name:** Absolute spatial deixis and proto-toponymy in Kata Kolok
- **Type:** extensional
- **Creation Date:** 2014-09-26
- **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](http://www.mpi.nl/people/vos-connie-de)
- **Role:** Researcher

**Resources**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Type</th>
</tr>
</thead>
</table>
| 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