CLARIN as European Adopter of RDA Outputs

Erhard Hinrichs
CLARIN ERIC Co-Director
erhard.hinrichs@uni-tuebingen.de
2015-09-24

RDA 6th Plenary, Paris
CLARIN?

- **Common Language Resources and Technology Infrastructure**
- European (ESFRI) Research Infrastructure – ERIC since February 2012 with currently 15 European member countries.
- 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
The 28 CLARIN centres
RDA output on DFT and CLARIN’s data architecture

metadata  persistent ID
Additional Areas of Adopting RDA Output

- Certification of CLARIN Centers according to the Data Seal of Approval.

- Adoption of a component-based metadata schema for language data and tools with a registered semantics.

- Building bridges to RDA: participation in RDA Working Groups and Conferences, profiting from RDA White Papers and commenting on them.
Thank you for your attention!

For more details:

http://clarin.eu > Services / Centres / Technology
DFT in CLARIN’s architecture

digital object

bitstream

describes
DFT in CLARIN’s architecture

- Single text or recording
- Corpus
- Lexicon
- Wordnet
- Grammar
- ...

- Language Data
- Metadata
- Language Tools

- Web application
- Web service
- Web service pipeline
- Stand-alone application
- ...

describes
1.2.2 Access Methods

We can basically identify 2 methods to access a bitstream that encodes the information we want to access.

1. **MD search**
2. **Metadata description**
3. **PID**
4. **PID resolver**
5. **Repository system**

These methods involve using a PID resolver to access the metadata description of the ePaper, which is then used to access the repository system.
DFT in CLARIN’s architecture: Virtual Collections

Registry

my collection

digital collections