Sharing of research data in the domain(s) served by CLARIN

Franciska de Jong
CLARIN ERIC
f.m.g.dejong@uu.nl

Workshop Open Science and Sharing Research Data
Brussels, 30 January 2018
CLARIN in seven bullets

• **CLARIN** is the Common Language Resources and Technology Infrastructure
• **ERIC** status since 2012, ESFRI 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** environment
• and that serves as an ecosystem for **knowledge sharing**.
CLARIN ERIC in members and centres

A consortium of:
- 19 members: AT, BG, CZ, DE, DK, DLU, EE, FI, GR, HU, IT, LT, LV, NL, NO, PL, PT, SE, SI
- 2 observers: FR, UK;
- >40 centres
CLARIN centres

• **Distributed architecture:** (http-accessible) files, web applications and web services spread all over Europe

• Nodes in the network: **centres**

• Currently:
  – 20 certified **B-centres**
  – over 40 registered centres in total

• Tools and data from different CLARIN centres are **interoperable**, so that data collections can be combined and tools from different sources can be chained to perform complex operations to support researchers in their work.
CLARIN in resource types

- Parliamentary records
- Literary texts
- Social Media data
- Historical letters
- Oral History data
- Disciplinary libraries
- Institutional archival data
- Broadcast archives
- Newspaper archives
- …
Harvesting of metadata for Findability

Diagram showing the flow of data and tools related to language data and metadata.
Certification CLARIN B-centres

- Data Seal of Approval -> CoreTrustSeal
- B-centre requirements
  (checked by Assessment Committee)
  - CMDI metadata
  - handle-based PIDs
  - federated Identity Management

- NWO accepts DMPs referring to our B-centres; How about Science Europe?
CLARIN feedback on the Framework Document and the Protocols approach

• **What is your opinion on the protocols’ approach?**
  – Important: ’owned’ by the domain
  – Room for diversity is key for adoption, Protocols Framework can help to give support / stimulate the long tail
  – Comply or explain: idem

• **Does the Framework give useful support in setting up domain-specific protocols?**
  – “template to be (partly) filled in”?

• **Are there elements missing in the Framework?**
  – emphasis on training, dissemination
  – diversity of scale-related issues: individual researcher vs. team vs. institute (“data officer in project of two people”?)
CLARIN/domain-specific aspect

• State of play in your domain: are there protocols in place or are there intentions to develop protocols?
  – interoperability of data and services; implementation coupled to assessment, certification, and automated (technical) checks
  – Support for DMP part of value proposition

• Which aspects of the Framework are especially useful for your domain?
  – could help balance national requirements and CLARIN approach

• Which aspects are negligible in your domain?
  – none
see you @

www.clarin.eu
or
f.m.g.dejong@uu.nl
DMPs in CLARIN

• CLARIN centres are stable repositories for project data
• Centres data policies start from national requirements
• Centre assessments help to
  – make policy choices explicit
  – ensure interoperability