CLARIN-D-Ressourcen:

Was kann ich nutzen und wie geht das?

Axel Herold

17. / 18. November 2014,
Textkorpora in Infrastrukturen für die Geistes- und Sozialwissenschaften, CLARIN-D-Workshop

CLARIN-D

CLARIN — Common Language Resources and Technology Infrastructure

- Sprachressourcen und -werkzeuge für Geistes- und Sozialwissenschaftlerinnen erschließen
- Integration vieler verschiedenartiger linguistischer Ressourcen und Werkzeuge
- **CLARIN-D**: BAS, BBAW, IDS Mannheim, MPI Nijmegen, Uni Hamburg, Uni Leipzig, Uni Saarland, Uni Stuttgart, Uni Tübingen (RZ Garching, FZ Jülich)
- **CLARIN ERIC** (europäischer Rahmen): CLARIN-AT, LINDAT-CLARIN (CZ), CLARIN-D, DLU, CLARIN-DK, CELR (EE), CLARIN-NL, CLARIN-PL, SWE-CLARIN, CLARIN-BG, CLARIN-LT (SWE-CLARIN, CLARINO)

CLARIN-D

Vision: (europaweit) vernetzte Forschungsinfrastruktur

- Auffindbarkeit von Ressourcen
- Erschließung für verschiedene Nutzergruppen
  (mehr dazu: Boehlke)
- Persistenz
- Interoperabilität
- offene Schnittstellen
- Zugriffskontrolle
- Hosting-Services
- Dokumentation, Schulung, Hilfe

CLARIN-D

CLARIN-D: Zentren im Verbund

- Repositorien
  (Primär- und Metadaten, Versionierung, Archivierung,
  mehr dazu: Zimmer)
- Endpunkte für Korpusrecherchen
- sprachtechnologische Webservices
  (mehr dazu: Trippel)
- Identity-Provider
- ...

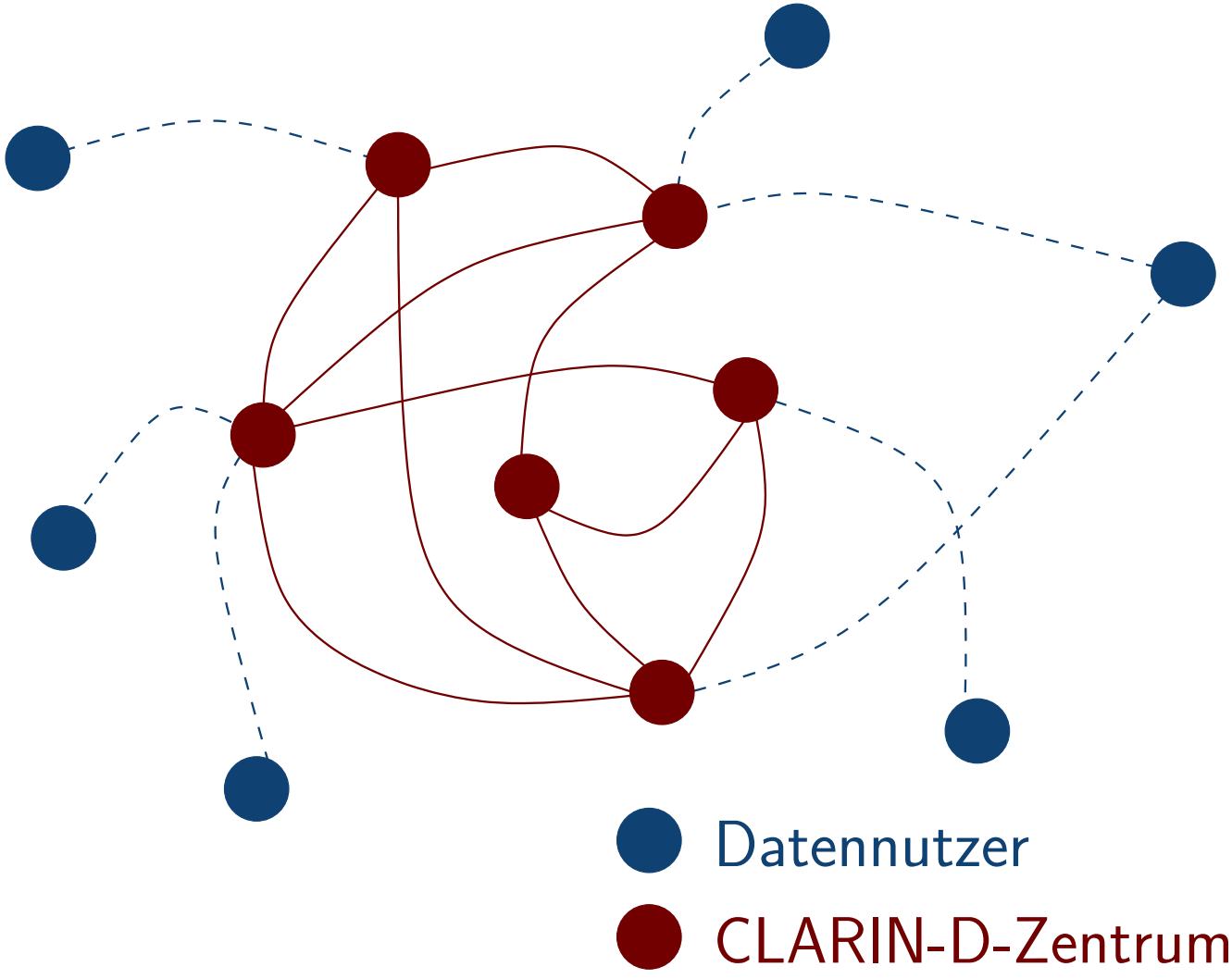
CLARIN-D

Zentrale Dienste

- Virtual Language Observatory (VLO)
- Virtual Collection Registry (VCR)
- Federated Content Search (FCS)
- WebLicht Webservice-GUI
- ...

CLARIN-D

CLARIN-D: Zentren im Verbund



CLARIN-D

Einige mögliche Szenarios

- Metadatensuche, Ressource(n) lokal nutzen (VLO $\longrightarrow$ konkretes Repositorium)

CLARIN-D

Einige mögliche Szenarios

- Metadatensuche, Ressource(n) lokal nutzen (VLO $\longrightarrow$ konkretes Repositorium)
- linguistische Analyse/Annotation eigener Daten (WebLicht)

CLARIN-D

Einige mögliche Szenarios

- Metadatensuche, Ressource(n) lokal nutzen (VLO $\longrightarrow$ konkretes Repositorium)
- linguistische Analyse/Annotation eigener Daten (WebLicht)
- Archivierung eigener Forschungs(primär)daten (Repositorium mit entsprechendem Profil, VCR)

CLARIN-D

Einige mögliche Szenarios

- Metadatensuche, Ressource(n) lokal nutzen (VLO $\longrightarrow$ konkretes Repositorium)
- linguistische Analyse/Annotation eigener Daten (WebLicht)
- Archivierung eigener Forschungs(primär)daten (Repositorium mit entsprechendem Profil, VCR)
- Bereitstellen eigener Ressourcen und Dienste (CLARIN-Standards)
- ...

Generell: Nutzung der Infrastruktur

- über Benutzer-Schnittstellen (GUIs)
- über standardisierte technische Schnittstellen

CLARIN-D

CLARIN-Ressourcen

- Überblick: Virtual Language Observatory
- facettierte Suche
- ca. 600 000 Ressourcen
- **Achtung**: Granularität
- stets aktuelle Daten durch Repositorien bereitgestellt
- (keine prinzipiellen Formatbeschränkungen)
- (mehr dazu: van Uytvanck)

CLARIN-D

Metadateninfrastruktur: CMDI

- Component Metadata Infrastrukture
- keine prinzipiellen Einschränkungen der Beschreibungsdomäne
- modular, adaptierbar (Profile, Komponenten)
- Interoperabilität durch semantische Annotation
  (CLARIN Concept Registry, CCR)
- momentan: ISO-Standardisierung

CLARIN-D

Datentypen

- Texte
- Wörterbücher
- Tonaufnahmen
- Videoaufnahmen
- Annotationen
- Datensätze
- Webservices, Anwendungen
- (keine prinzipiellen Typbeschränkungen)

CLARIN-D

Unterstützte Primärdatenformate

- Unterstützungsgrad abhängig von Infrastrukturkomponenten
- Plain-Text Formate (viele Webservices)
- nativ: TCF (Text Corpus Format, auch für Wörterbücher)
- ausgewählte TEI-Dialekte (z. B. DTABf)
- anwendungsspezifische Audio- / Videoformate
- vereinzelt Unterstützung für spezialisierte Datenformate (z. B. Wörterbücher)
- (mehr zu einzelnen Themen: Haaf / Jurish)

Strategie: Bereitstellen von Konvertern

CLARIN-D

CLARIN-D-Benutzerhandbuch

- „CLARIN-D in a nutshell“ – gute Einstiegsmöglichkeit und Hintergrundinformationen
  - grundlegende Eigenschaften und Modelierungsprinzipien für Ressourcen und Werkzeuge
  - spezifische Eigenschaften und Anforderungen
- `http://de.clarin.eu/de/sprachressourcen/benutzerhandbuch.html`

Weitere Hilfe

- Helpdesk
- Beratung durch Zentren
- (mehr dazu: Lehmberg)

CLARIN-D

Offene Punkte

- Weiterentwicklung des Datenangebots
  (in Abstimmung mit den Facharbeitsgruppen)
- nativer Umgang mit verbreiteten Primär(text)formaten
  (bspw. TEI-Integrator, TEI-Dokumente in WebLicht nutzen)
- breitere Erschließung von nicht textbasierten Daten
  (Ton-, Filmaufnahmen)
- Harmonisierung der Metadaten

CLARIN-D

Danke! Fragen?

http://de.clarin.eu/

http://www.clarin.eu/