

ZUGANG ZU VERTEILTEN KOLLEKTIONEN

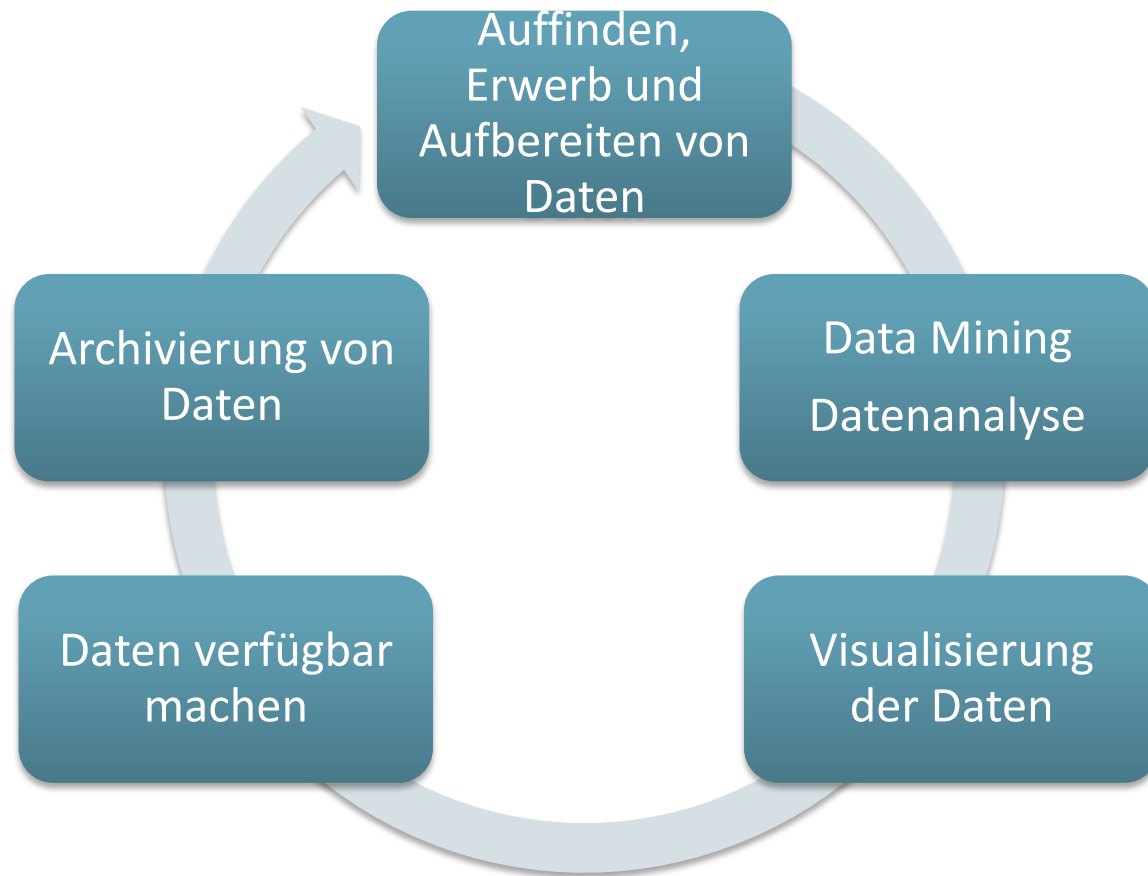
ERHARD HINRICHS & THORSTEN TRIPPEL

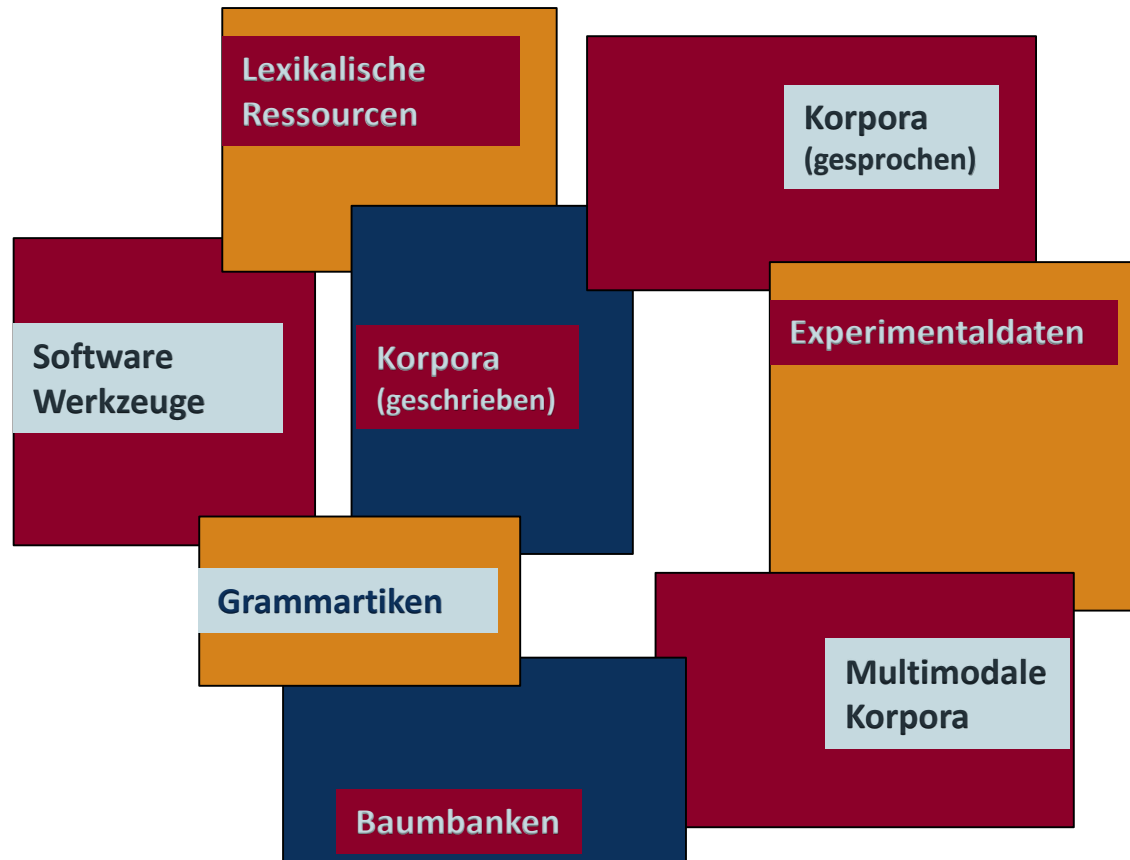
EBERHARD KARLS UNIVERSITÄT TÜBINGEN



- **Tübingen (Koord.)** – (Computerlinguistik (CL)): Annotierte Korpora, ling. Wissenskomponenten, Web-Services
- Leipzig – (Informatik): Web-Services und Textkorpora
- BBAW Berlin – (Zentrum Sprache): Deutsche Sprache: Wörterbücher, (hist.) Textkorpora
- Stuttgart – (CL): Web-Services und Textkorpora
- IDS Mannheim – (Germanistik/CL): Deutsche Sprache, Textkorpora
- LMU München – (BAS/CL) Deutsche Sprachdaten, multilinguale Daten, Web Services
- MPI Nijmegen – (Psycholinguistik): Bedrohte Sprachen, multimodale Sprachressourcen
- Hamburg – (CL): Multilinguale Sprachdaten, Transkriptionswerkzeuge
- Saarland – (CL): Multilinguale Textkorpora und Web-Services

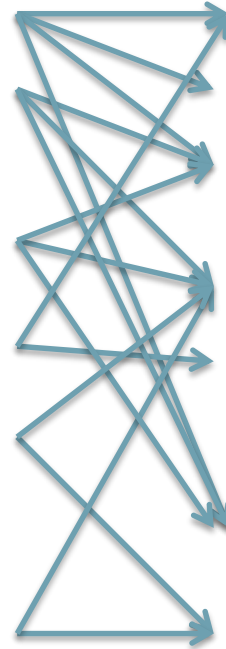
- Einfacher Zugang zu Sprachdaten mit einheitlichem Login (Single Sign-On)
- Verteilte Dienste für die Forschung
- Archivierung und nachhaltiger Zugang zu Sprachdaten





Datentypen

- Sprachsignale
- Textkorpora (basieren auf Druckerzeugnissen oder Webseiten)
- Lexikalische Ressourcen (Beispiele aus Textkorpora)
- Experimente/Umfragen
- Zusammenfassung von Daten aus verschiedenen Institutionen (Querschnittsstudien, Linked Data)
- Disziplinübergreifende Datennutzung

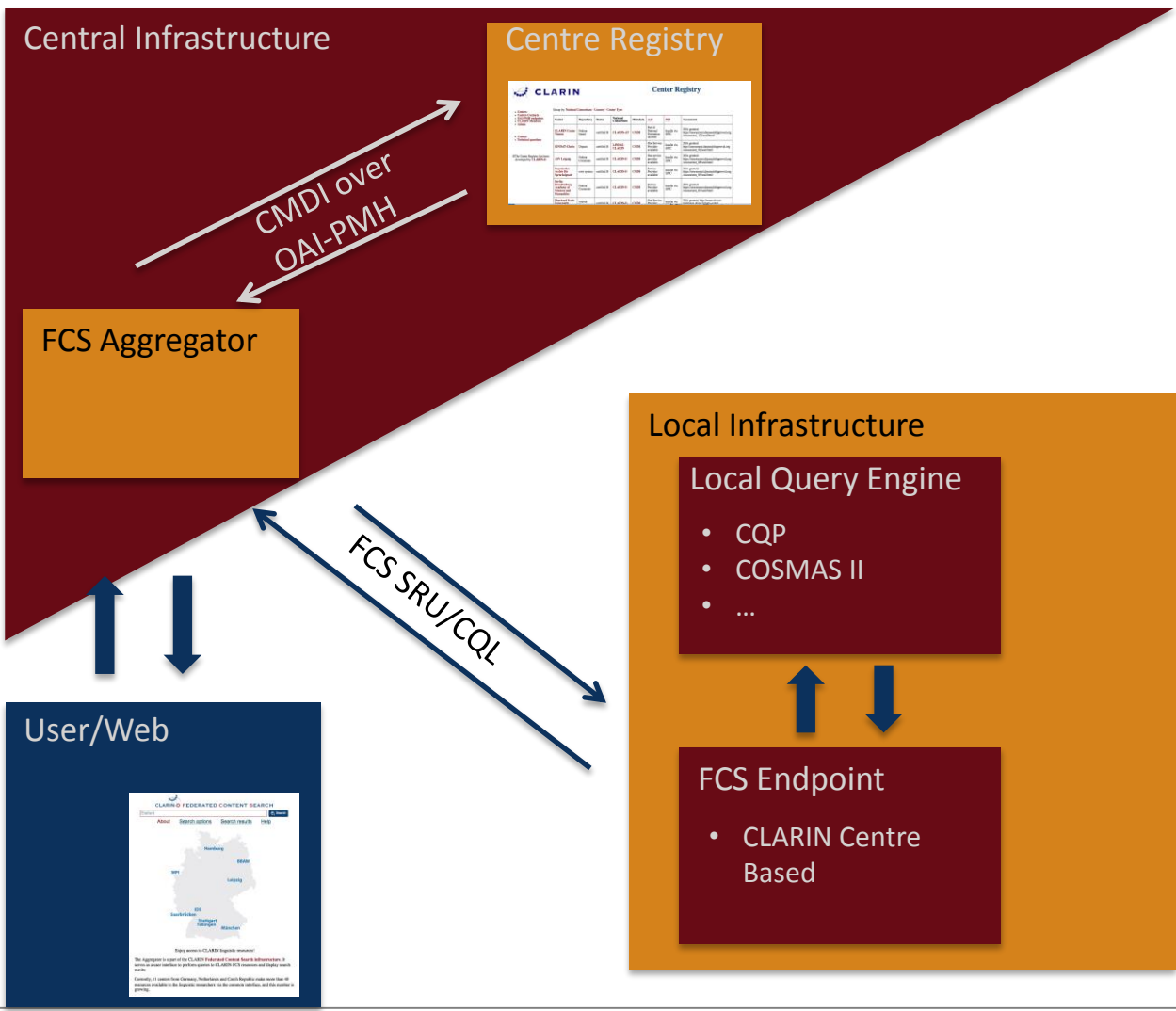


Interoperabilitätshürden

- Persönlichkeitsrechte
- Ethik
- Copyright
- Attribution
- "Experimentdiebstahl" vor Veröffentlichung
- Zugangsberechtigungen
- Schnittstellen

- Wissenschaftsprivileg für Medienerzeugnisse im Copyright → hilft nur bei Korpora, etc.
- "open access" → hilft nicht bei
 - Persönlichkeitsrechten
 - Ethik
 - Copyright
 - "Experimentdiebstahl"
- Technische Lösung → Authorisierung von Einzelpersonen und Gruppen

- Verteilte Suche → ohne zentralen Index
- Offene Schnittstellen → OAI-PMH, SRU/CQL
- Authentifizierung → AAI über Shibboleth
- Authorisierung →
 - in der Regel für die akademische Welt offen
 - weitere Einschränkungen auf bestimmte Benutzer unklar



- Lösungen mit Zugangsbeschränkungen einbeziehen
 - Keine zentrale Indizierung möglich
 - Abfragespezifizierung von Teilquellen
- Attribution bei vernetzten Forschungsdaten
- Anfragemodelle für Cross-Disciplinary-Data und Schnittstellen (Umgang mit Zugangsbeschränkung)
- "Forschungsdaten Rückversicherung": Übernahme von fachbezogenen Infrastrukturdienstleistungen