

Digitale Forschungsinfrastrukturen in den Geisteswissenschaften: CLARIN-D und DARIAH-DE

Prof. Dr. Erhard Hinrichs, CLARIN-D

Prof. Dr. Wolfram Horstmann, DARIAH-DE



Zwei Initiativen: CLARIN-D und DARIAH-DE

- Aus den Bedarfen der Geisteswissenschaften entstanden
- In die europäische Forschungsinfrastrukturlandschaft eingebunden
- Vom BMBF gefördert
- Auf technischer Ebene kooperierend
- Engere Verzahnung geplant



Datenressourcen und Rechercheumgebung

Digitale Textsammlungen

Lexikalische und terminologische Ressourcen

Annotierte Forschungsdaten

Multimodale- und Sprachaufnahmen

Experimentelle Daten

Werkzeuge

Virtuelle Forschungs-umgebungen

Annotations-werkzeuge

Speech-Tools

Datenmanagement-Werkzeuge

Wissensvermittlung

Benutzerhandbuch

Legal Information Platform

Helpdesk

Beispielverwendungen

Kurse & Tutorien

Individuelle Beratung

Schwerpunktdisziplinen

Philologien

Kognitionswissenschaften

Sprachwissenschaften

Geschichtswissenschaften

qualitative Sozialwissenschaften
und Politikwissenschaft

Datenressourcen und Rechercheumgebung

Werkzeuge

Digitale Textsammlungen

Deutsches Textarchiv

GermaParl

Deutsches Referenzkorpus (DeReKo)

Lexikalische und terminologische Ressourcen

DWDS

OWID

GermaNet

Wortschatz

Annotierte Forschungsdaten

Tübinger Baumbank Collection

TIGER

Multimodale- und Sprachaufnahmen

Sprachkorpora des BAS

DOBES + Sprachdokumentationsdaten

Experimentelle Daten

Mind Research Repository

Wissensvermittlung

Benutzerhandbuch

Helpdesk

Kurse & Tutorien

Legal Information Platform

Beispielverwendungen

Individuelle Beratung

Schwerpunktdisziplinen

Philologien

Sprachwissenschaften

qualitative Sozialwissenschaften
und Politikwissenschaft

Kognitionswissenschaften

Geschichtswissenschaften

Datenressourcen und Rechercheumgebung

Werkzeuge

TüNDRA

WebLicht

GermaParl

WebAnno

ELAN

EXMARaLDA

WebMaus

OCTRA

EMU

Repositorienbetrieb

Datenmanagementplan-Werkzeuge

Datenübernahmeservices

**Virtuelle
Forschungsumgebungen**

Annotationswerkzeuge

Speech-Tools

Datenmanagement-Werkzeuge

Wissensvermittlung

Benutzerhandbuch

Helpdesk

Kurse & Tutorien

Legal Information Platform

Beispielverwendungen

Individuelle Beratung

Schwerpunktdisziplinen

Philologien

Sprachwissenschaften

qualitative Sozialwissenschaften
und Politikwissenschaft

Kognitionswissenschaften

Geschichtswissenschaften

Digitale Textsammlungen

Lexikalische und terminologische Ressourcen

Annotierte Forschungsdaten

Multimodale- und Sprachaufnahmen

Experimentelle Daten

Wissensvermittlung

Dokumentation

- Benutzerhandbuch
- Informationsplattform zu rechtlichen und ethischen Fragestellungen
- Dokumentation und Beispiele für die Verwendung von Standards

Support von Nutzenden

- Helpdesk
- Individuelle Beratung
- Beispielverwendungen

Kurse und Tutorien

- Sammlung von Lehreinheiten
- Kurse und Tutorien bei Sommerschulen und Konferenzen

Schwerpunktdisziplinen

Philologien

Sprachwissenschaften

qualitative Sozialwissenschaften
und Politikwissenschaft

Kognitionswissenschaften

Geschichtswissenschaften

Datenressourcen und

Digitale Textsammlungen

Lexikalische und terminologische
Ressourcen

Annotierte
Forschungsdaten

Multimodale- und
Sprachaufnahmen

Experimentelle Daten

Wissensvermittlung

Schwerpunktdisziplinen

Facharbeitsgruppen für unterschiedliche Disziplinen

> 200 Forschende in ganz Deutschland in
8 Facharbeitsgruppen

- Philologien (Amerikanistik, Anglistik, Germanistik, Klassische Philologie, Romanistik, Slavistik)
- Sprachwissenschaften (inkl. Computerlinguistik, Sprachtechnologie, Phonetik und allgemeine Linguistik)
- Kognitionswissenschaften
- Geschichtswissenschaften
- qualitative Sozialwissenschaft und Politikwissenschaft
- Kulturwissenschaften (inkl. Anthropologie und Sprachdokumentation)

Datenressourcen und Rechercheumgebung

Digitale Textsammlungen

Deutsches Textarchiv
DeReKo
GermaParl

Lexikalische und terminologische Ressourcen

DWDS
GermaNet
OWID
Wortschatz

Annotierte Forschungsdaten

Tübinger Baumbank Collection
TIGER

Multimodale- und Sprachaufnahmen

Sprachkorpora des BAS
DOBES + Sprachdokumentationsdaten

Experimentelle Daten

Mind Research Repository

Werkzeuge

Virtuelle Forschungsumgebungen

TüNDRA: Visualisierungs- und Analysewerkzeug
WebLicht: Verkettung von Webservices zur automatischen Analyse von Sprachdaten
PoIMine

Annotationswerkzeuge

WebAnno
ELAN
EXMARaLDA

Speech-Tools

WebMAUS zur Alignierung von Transkription und Sprachsignal
Transkriptionseditor OCTRA
EMU zur Auswertung empirischer Sprachdaten

Datenmanagement-Werkzeuge

Repositorienbetrieb
Datenübernahmeservices
Datenmanagementplan-Werkzeuge

Wissensvermittlung

Benutzerhandbuch
Legal Information Platform
Helpdesk
Beispielverwendungen
Kurse & Tutorien
Individuelle Beratung

Schwerpunktdisziplinen

Philologien
Kognitionswissenschaften
Sprachwissenschaften
Geschichtswissenschaften
qualitative Sozialwissenschaften und Politikwissenschaft

Datenressourcen und Rechercheumgebung

Digitale Textsammlungen

Deutsches Textarchiv

Lexikalische und terminologische Ressourcen

Annotierte Forschungsdaten

Multimodale- und Sprachaufnahmen

Experimentielle Daten

Werkzeuge

Virtuelle Forschungs-umgebungen

TüNDRA: Visualisierungs- und Analysewerkzeug

WebLicht: Verkettung von Webservices zur automatischen Analyse von Sprachdaten

Annotationswerkzeuge

WebAnno

Speech-Tools

Datenmanagement-Werkzeuge

Wissensvermittlung

Benutzerhandbuch

Legal Information Platform

Helpdesk

Beispielverwendungen

Kurse & Tutorien

Individuelle Beratung

Schwerpunktdisziplinen

Philologien

Kognitionswissenschaften

Sprachwissenschaften

Geschichtswissenschaften

qualitative Sozialwissenschaften und Politikwissenschaft

Datenressourcen und Rechercheumgebung

Digitale Textsammlungen

Deutsches Textarchiv

Lexikalische und terminologische Ressourcen

Annotierte Forschungsdaten

Multimodale- und Sprachaufnahmen

Experimentielle Daten

Werkzeuge

Virtuelle Forschungsumgebungen

TüNDRA: Visualisierungs- und Analysewerkzeug

WebLicht: Verkettung von Webservices zur automatischen Analyse von Sprachdaten

Annotationswerkzeuge

WebAnno

Speech-Tools

Datenmanagement-Werkzeuge

Wissensvermittlung

Benutzerhandbuch

Legal Information Platform

Helpdesk

Beispielverwendungen

Kurse & Tutorien

Individuelle Beratung

Schwerpunktdisziplinen

Philologien

Kognitionswissenschaften

Sprachwissenschaften

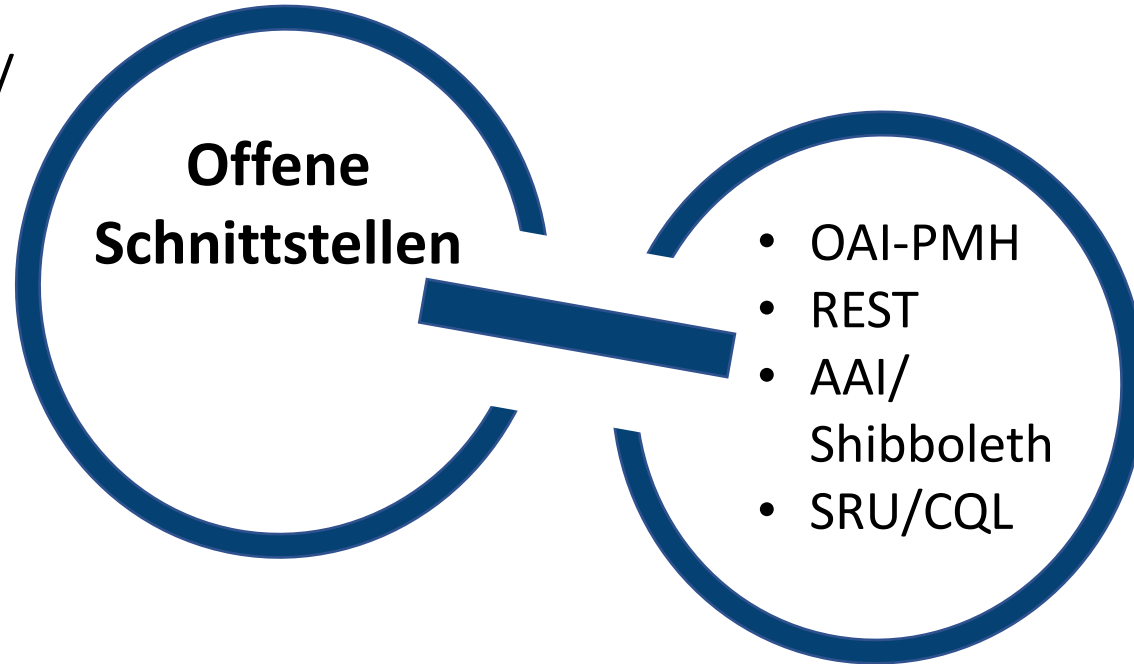
Geschichtswissenschaften

qualitative Sozialwissenschaften und Politikwissenschaft

CLARIN-D Architektur

Zentrenangebote

- Datenmanagement/
Repositorien
- Webservices
- Webanwendung
- Beratung
- Hilfe



Offene Schnittstellen

- OAI-PMH
- REST
- AAI/
Shibboleth
- SRU/CQL

Föderierte Werkzeuge

- Federated Content
Search (FCS)
- WebLicht Chainer
- Virtual Language
Observatory (VLO)
- Virtual Collection
Registry

Normen, Standards und andere Best Practices

CLARIN-D: offener Zentrenverbund

- zertifizierte Datenzentren
 - gemeinsame technischen Infrastruktur und Standards
 - Datenmanagementangebote an Dritte
- angebunden an europäische Infrastruktur
- offene Verbundsarchitektur



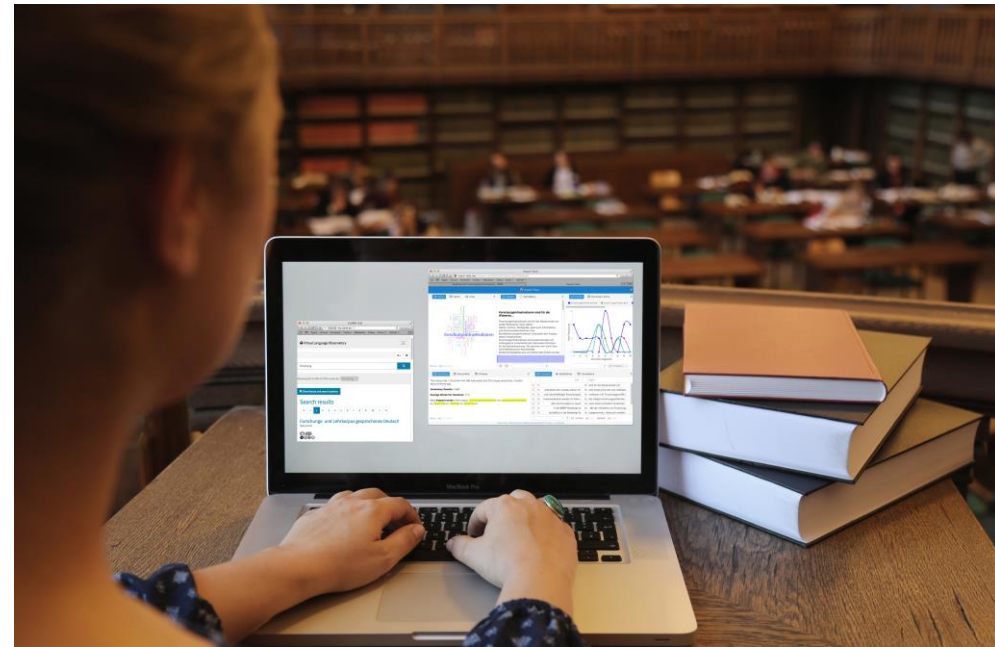
- Zentren im Konsortium
- Assoziierte Zentren
- Kooperierende Rechenzentren



Kurzer Überblick in Zahlen

- **1.677.412** Forschungsdatensätze (über europäischen Verbund)
- **> 170.000** Downloads aus deutschen Zentren jährlich
- **> 400** Webservices in WebLicht integriert
- **>5.800.000** Service Calls
- **>100** Annotationsvorhaben mit WebAnno
- **> 33.000** registrierte Endbenutzer
- **> 800** Lizenznehmer
- **462** kooperierende Projekte
- **2.421** Institutionen
- **35** beteiligte Fachdisziplinen

Stand Mai 2018






DARIAH-DE

DARIAH-DE – ein offenes Konsortium aus 19 Partnern

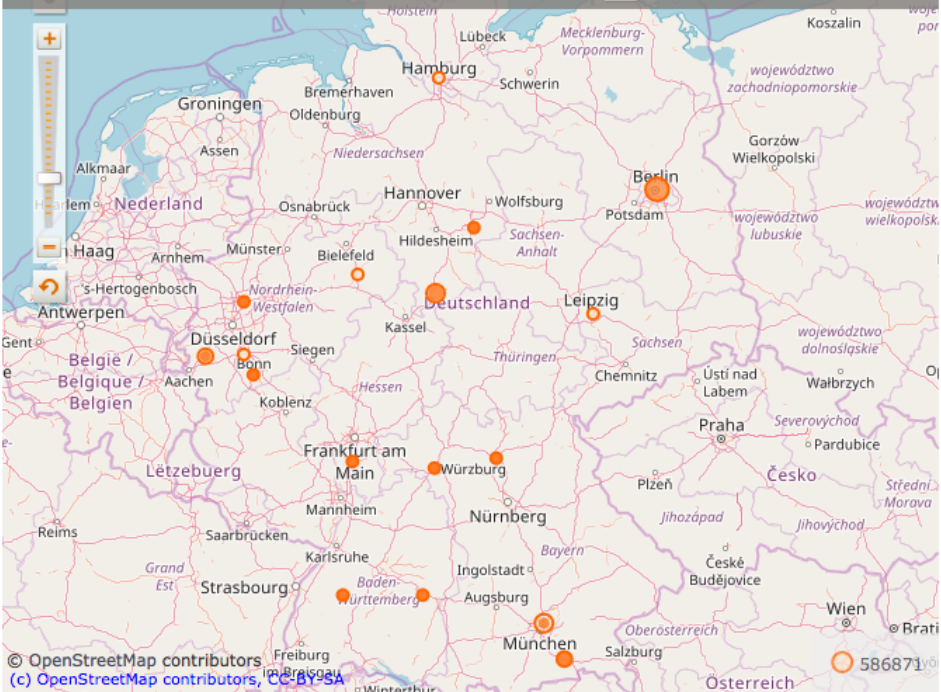
- 3 Universitäten,
- 4 Forschungsinstitute,
- 4 Rechenzentren,
- 2 Bibliotheken,
- 2 Forschungsverbünde,
- 2 Akademien der Wissenschaften,
- 1 NGO,
- 1 KMU,
- + neue Partner

Load Data **Load Overlay** **Dataset Information**

Static Data KML File URL Magnetic Link
Please choose file... 586871  [load KML](#)

Background map **Map selector tools** **Filter** **Publication Place**

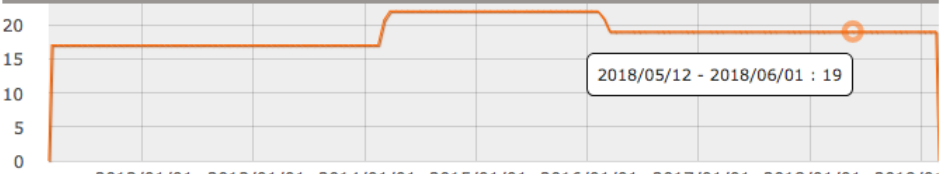
Open Street Map 27 results Slupsk



© OpenStreetMap contributors
(c) OpenStreetMap contributors, CC-BY-SA

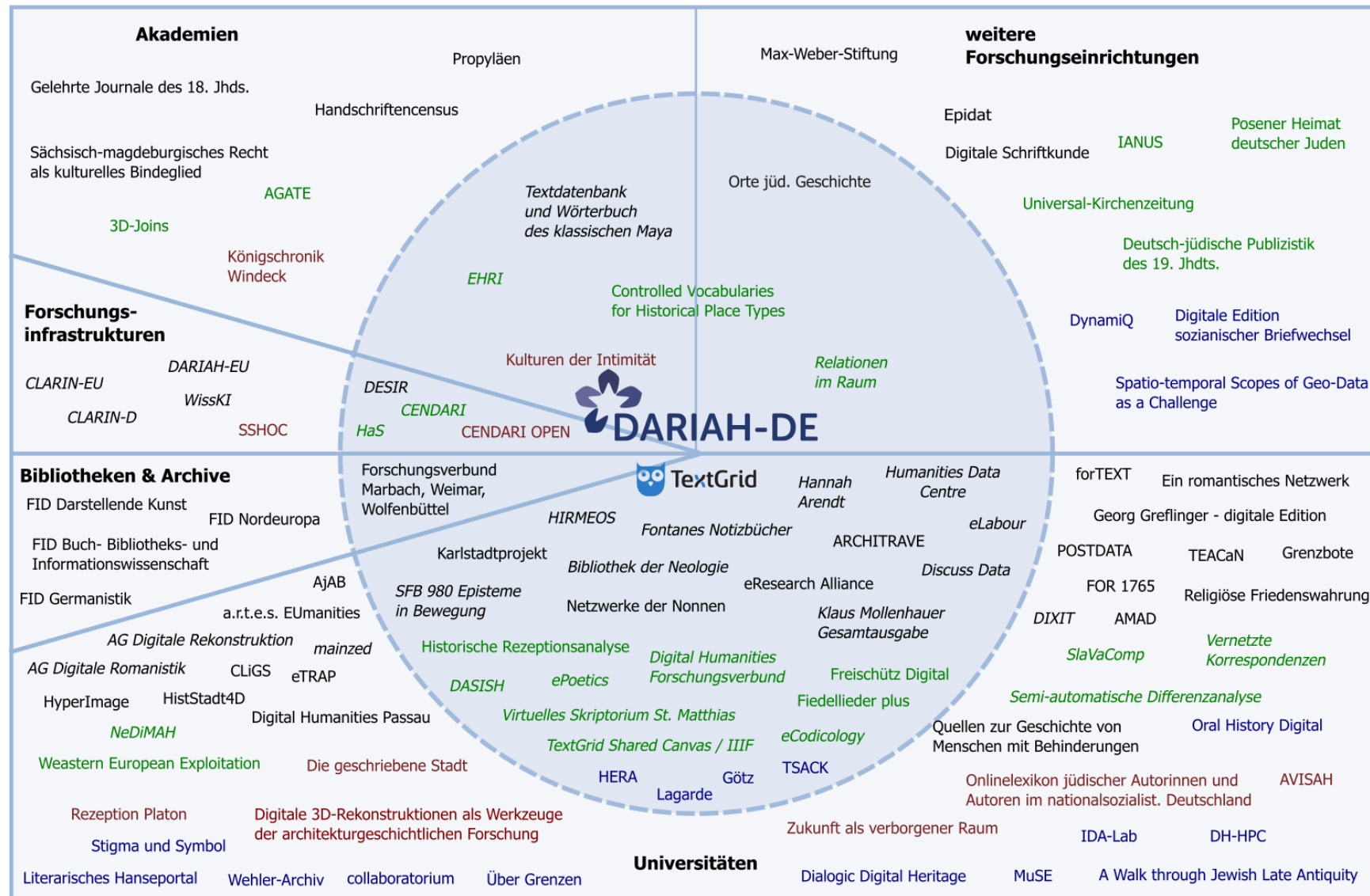
Time start **Time unit** **Scaling**

2011 continuous normal



2012/01/01 2013/01/01 2014/01/01 2015/01/01 2016/01/01 2017/01/01 2018/01/01 2019/01/01

Kooperationen im Kontext von DARIAH-DE und TextGrid



DARIAH-DE in Zahlen

Nutzerzahlen

4559 Accounts insgesamt
89 Confluence Spaces
312 Gruppen

Forschungsdaten

Generische Suche: 2.471.415 Dokumente
Collection Registry: 204 Sammlungsbeschreibungen; 258 Akteure

TextGrid: 182.335 Images; 107.362 XML-Texte

DHd-Blog: 1.147 Beiträge; 172 Autorinnen und Autoren; 73 Institutionen aus dem deutschsprachigen Raum

DHd-Kanal: 30.359 Aufrufe; 253 Abonnenten; 15 Playlists

DARIAH-DE Portal: 51.200 Seitenansichten in 2017

DARIAH-DE stellt mehrere Tools (z.B. Geobrowser) als Webanwendung bereit, die ohne Anmeldung genutzt werden können. Darüber werden keine Statistiken erhoben.

Wissenschaftsgeleitete Cluster-Struktur

- Wissenschaftliche Sammlungen
- Quantitative Datenanalyse
- Annotieren, analysieren, visualisieren
- Wissenschaftliche Begleitforschung
- eInfrastruktur
- Institutionalisierung von DARIAH-DE
- + Planung neuer Cluster (z.B. Objekt-Cluster)

Schwerpunktdisziplinen

- Archäologie
- Epigraphik
- Geschichtswissenschaft
- Jüdische Studien
- Kulturwissenschaften
- Kunstgeschichte
- Literatur- und Sprachwissenschaft
- Mediävistik
- Musikwissenschaften
- Philologie
- Philosophie
- Theologie / Kirchengeschichte
- Wissenschaftsgeschichte
- Sozialwissenschaften

DARIAH-DE Angebote und Services

Fachspezifische Tools und Angebote



CosmoTool

Ein Werkzeug, welches biographische Informationen aus unterschiedlichen Quellen zu inter- und nationalen Bewegungsprofilen historischer Persönlichkeiten zusammenfasst.



Geo-Browser

Der Geo-Browser ist ein webbasiertes Tool, das Raum-Zeit-Relationen von Daten- und Quellensammlungen analysieren und visuell darstellen kann.



ConedaKOR

ConedaKOR dient der Verwaltung und Präsentation akademischer Objektsammlungen aus den bildbasierten Kultur- und Geisteswissenschaften.



Digivoy

Die Grundidee der Voyant Tools ist es, web-basierte, vielfältige explorative und analytische Zugänge zu beliebigen Texten oder Textsammlungen zu ermöglichen.



TextGrid

Virtuelle Forschungsumgebung für die Geisteswissenschaften



MEI Score Editor

Desktop-Tool zum Anzeigen und Bearbeiten von MEI-kodierten Musikdateien. Es bedient Forschungsfragen in der Musikologie.



DKProWrapper

DKPro Wrapper ist ein in Java geschriebenes NLP-Tool, das linguistische Informationen aus Texten extrahiert. Seine Benutzung wird ausführlich beschrieben im Tutorial **NLP Based Analysis of Literary Texts**.



Topics

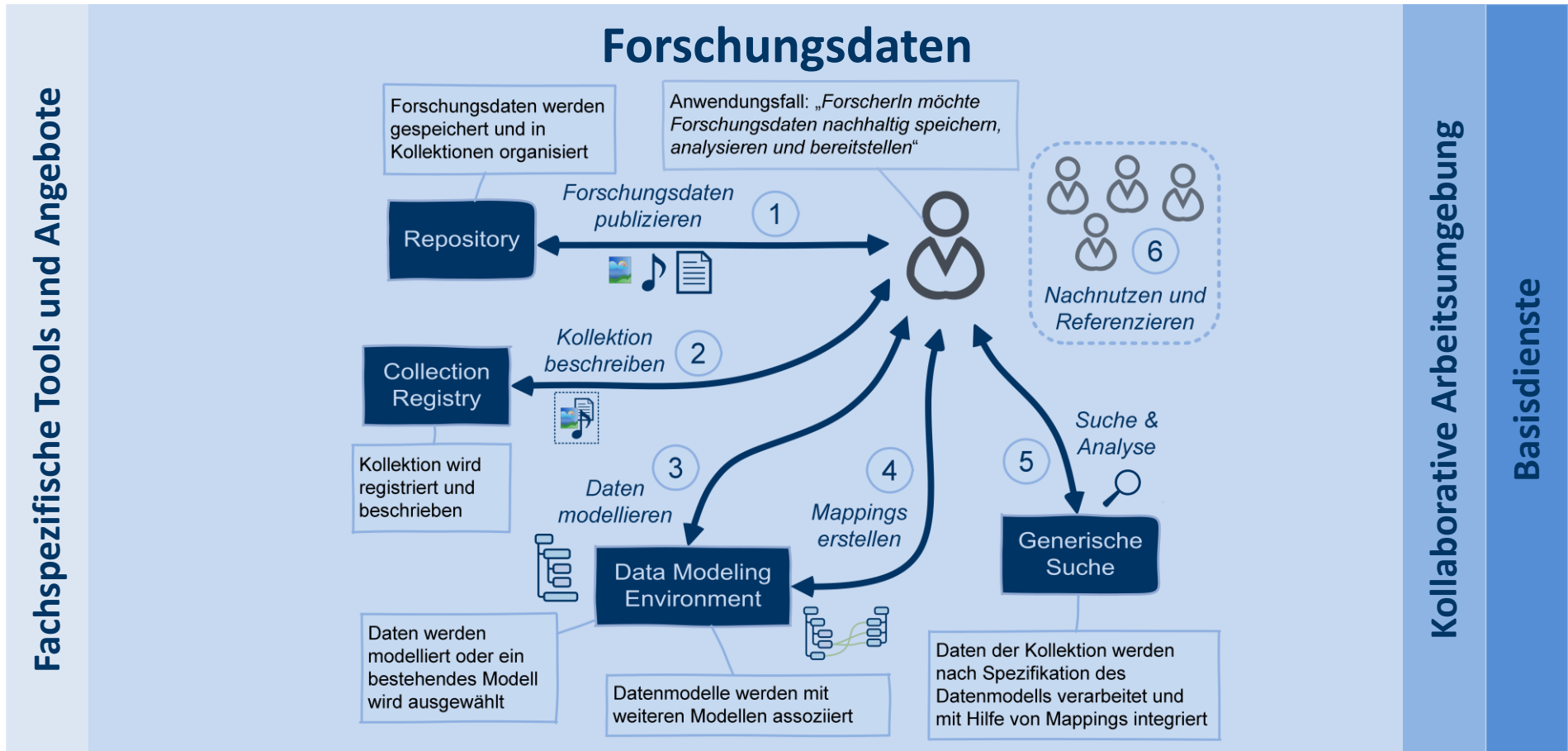
Topics ist eine Python-Bibliothek für Topic Modeling, die auf die Bedürfnisse der Literaturwissenschaft zugeschnitten ist.

Forschungsdaten

Lehre und Community

Basisdienste

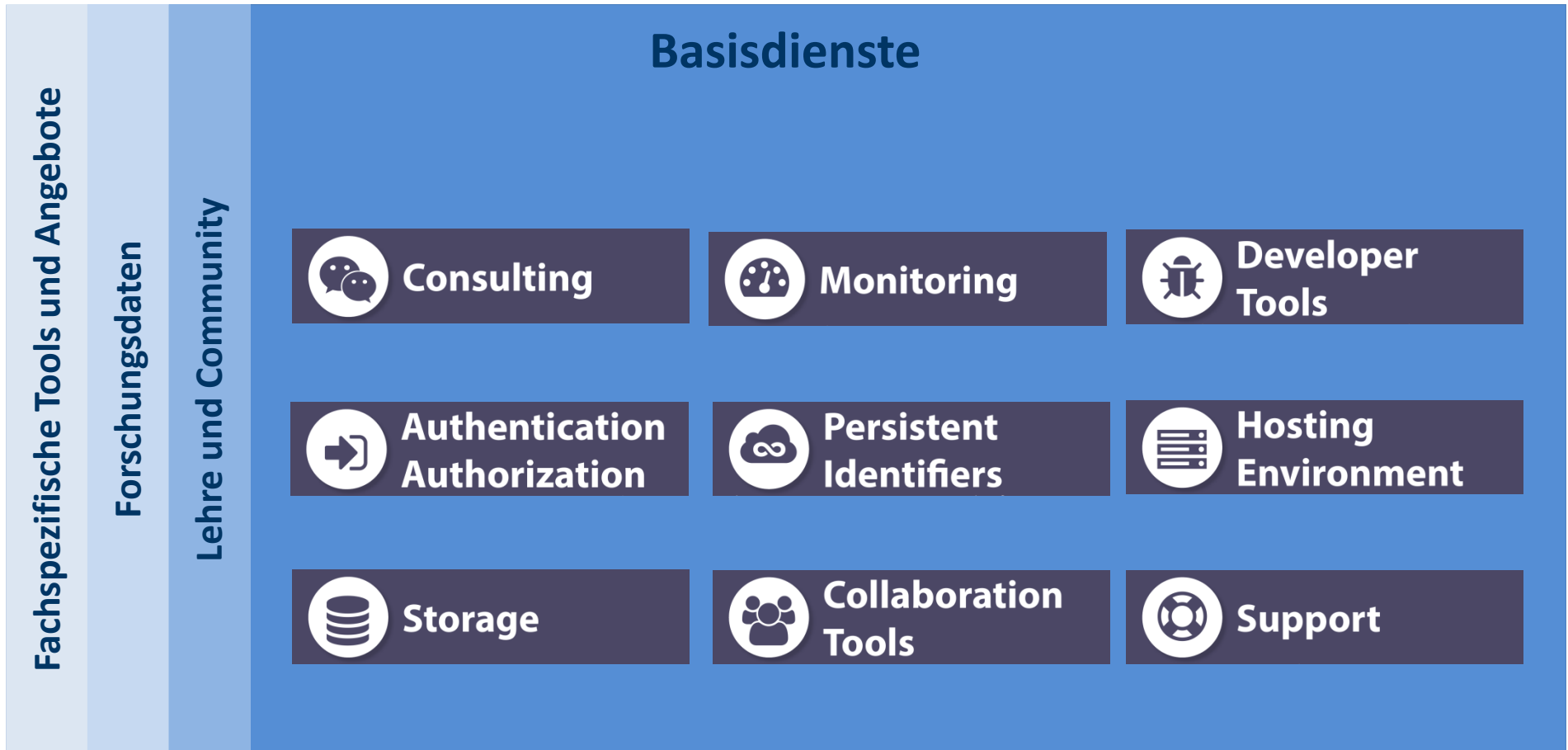
DARIAH-DE Angebote und Services



DARIAH-DE Angebote und Services

Fachspezifische Tools und Angebote	Forschungsdaten	Lehre und Community		Basisdienste
		Veröffentlichungen <ul style="list-style-type: none">– Working Papers– DHd Blog– DHd Kanal– Schulungs- und Lehrmaterialien– Tutorials– Bibliographie Doing Digital Humanities– DH-Handbuch	Veranstaltungen <ul style="list-style-type: none">– Experten- und Methodenworkshops– Sommerschools Empfehlungen <ul style="list-style-type: none">– Best Practices– Impactomatrix Nachwuchsförderung <ul style="list-style-type: none">– DARIAH-DE DH-Award– Reisestipendien	

DARIAH-DE Angebote und Services



Stakeholdergremium „Wissenschaftliche Sammlungen“

[Memorandum: Digitalisierung von Kulturgut](#)

<https://dhd-blog.org/?p=9849>

veröffentlicht am 17. Mai 2018

bisher unterschrieben von 21 Institutionen und
50 Einzelunterzeichnenden

DARIAH-DE 2019–21

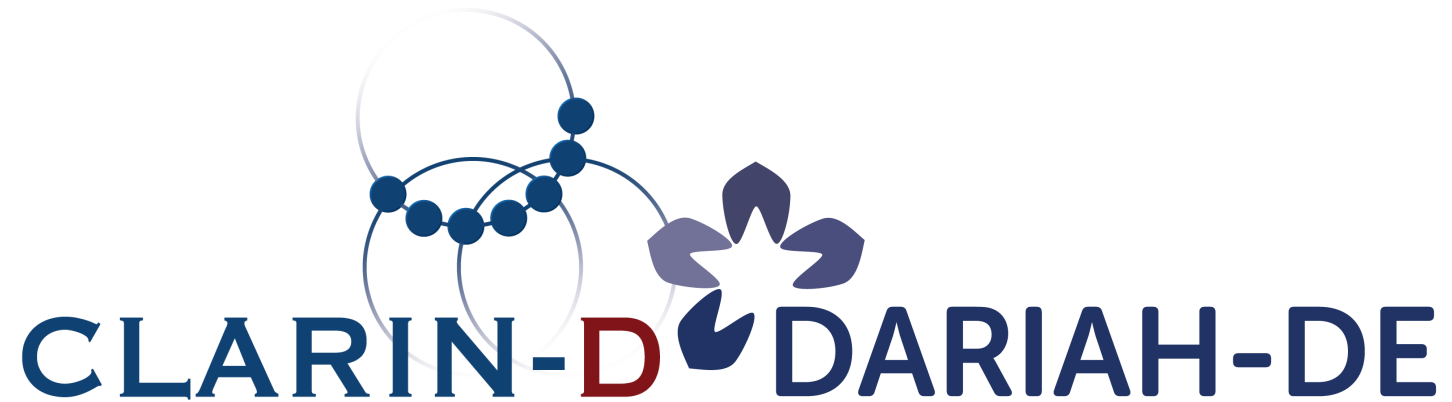
- Betriebskooperationsvereinbarung
Durchführung des Basisbetriebs nach Ablauf des regulären Förderzeitraums durch das BMBF für zwei Jahre
- Wissenschaftsgeleitete Aufgaben-Matrix
Dienste, Daten, Expertise, Dissemination, Management
- Servicekatalog
Jeder Partner hat die Verantwortung für ein oder mehrere Aufgabengebiete des Servicekatalogs und wird diese durch Eigenmittel 2019–21 finanzieren.

Zusammenarbeit CLARIN-D und DARIAH-DE

...



Weiterentwicklung der bisherigen Zusammenarbeit



Forschungsdaten

- Weiterentwicklung und Vereinheitlichung von Datenbeständen und Repositorien
- Beratung zum Datenmanagement
- Empfehlungen zum Umgang mit ethischen und rechtlichen Fragen
- Publikationen zu Best-Practice-Empfehlungen
- Interoperabilität zu bestehenden Werkzeugen
- Interoperabilität föderierter Suchfunktionalitäten

Werkzeuge und virtuelle Forschungsumgebungen

- Interoperabilität der virtuellen
Forschungsumgebungen Textgrid, WebLicht+Tundra
- Erweiterung der Funktionalität und Ergänzung
weiterer Methoden
- Integration von Annotationswerkzeugen für die
automatische und manuelle Verarbeitung
- Spezifikation und Entwicklung von Schnittstellen
zwischen großen Werkzeugen

Standards und Verfahren

- Entwicklung und Vereinheitlichung von Empfehlungen für die Verwendung von Datenformaten und Standards
- Erstellung von Handreichungen für Datenersteller
- Kooperation mit Standardisierungsgremien bei der Entwicklung und Überarbeitung von Standards

Digitale Fertigkeiten in Lehre, Ausbildung & Weiterbildung

- Zusammenführung und Ausbau des Angebots von Lehr- und Lernmaterialien
- Nachwuchsförderung mittels Awards für NachwuchswissenschaftlerInnen und Reisestipendien zu nationalen / internationalen DH-Konferenzen
- Durchführung von Workshops, Expertenseminaren, Doktorandenveranstaltungen und Sommerschulen
- Durchführung von Konferenzen und Tagungen in Kooperation mit Fachverbänden sowie Tutorien im Rahmen dieser Veranstaltungen

Vernetzung und Koordination der technischen Entwicklungen

- Ortsverteilte Infrastruktur auf gemeinsamer technischer Grundlage
- Infrastrukturbasisdienste
- Betrieb und Weiterentwicklung von Diensten für die geistes- und kulturwissenschaftliche Forschung und Lehre