

Text+

**Plenary
2026**



@textplus_NFDI
#TextPlusPlenary

Collections
Lexical Resources
Editions
Infrastructure/
Operations

TEXT+

ABSCHLUSS

DER ERSTEN PROJEKTPHASE

text-plus.org



Text+

Plenary
2026



@textplus_NFDI
#TextPlusPlenary

5. TEXT+ PLENARY

PROGRAMM

Montag, 15. Juni 2026

12:00 – 13:30	<i>Lunch (Katakomben)</i>
13:30 – 14:00	Begrüßung, Eröffnung und Einführung in das Programm (Aula)
14:00 – 15:30	Ergebnisse aus Task Area: Infrastructure & Operations (Aula)
15:30 – 16:05	<i>Kaffeepause (Katakomben)</i>
16:05 – 17:30	Ergebnisse aus Task Area: Editions (Aula)
17:30 – 18:00	Posterslam (Aula)
18:00 – 19:00	Postersession (Fuchs-Petroluh-Saal)



Collections
Lexical
Resources
Editions
Infrastructure/
Operations

Text+ im Kontext der NFDI – Stand, Rolle und Perspektive

Analyse der aktuellen Position und
zukünftigen Entwicklungen

Funded by
DFG Deutsche
Forschungsgemeinschaft

German Research Foundation

Project number 460033370

Part
of
nfdi Nationale
Forschungsdaten
Infrastruktur



Willkommen beim 5. Text+ Plenary

2022: Projektfortschritt und die aktuellen
Entwicklungen, Mannheim

2023: Connecting People and Data, Göttingen

2024: Große Sprachmodelle (LLMs) und deren
Nutzung

2025: Connecting Infrastructures -
Connections between European Research
Infrastructures, Göttingen

2026: Abschluss der ersten Projektphase

Text+

Text+ Plenary 2022

Text+

Text+ Plenary 2023: Connecting People and Data

3. Text+ Plenary 2024

Text+

4. Text+ Community-Plenary 2025 (Göttingen)

Text+

5. Text+ Plenary 2026

Text+

15.–16. Juni 2026
Mannheim, Schloss
Europa/Berlin Zeitzone

Geben Sie Ihren Suchbegriff ein

Überblick

Tagungsprogramm

Anmeldung

Postersession: Text+
Plenary 2026

Hotels

Informationen zum
Datenschutz

Ansprechpersonen

Contact

✉ textplus-events@ids-ma...

Text+ ist ein Konsortium der Nationalen Forschungsdateninfrastruktur (NFDI) und hat zum Ziel, sprach- und textbasierte Forschungsdaten langfristig zu erhalten und ihre Nutzung in der Wissenschaft zu ermöglichen.



Text+

<https://www.>



Sprache, Daten und methodischer Wandel

- Veränderung in Geisteswissenschaften
- Methodischer Wandel durch KI
- Forschungsdaten: Qualität und Nachhaltigkeit
- Text+ als Brücke



Text+ auf einen Blick: Auftrag und Struktur

- Verteilte Forschungsinfrastruktur
- Text+ Data Space
- FAIR-Prinzipien
- Skalierbarkeit



Text+ im NFDI-Ökosystem

- Schlüsselrolle für sprach- und textbasierte Daten innerhalb der Nationalen Forschungsdateninfrastruktur.
- Vernetzung mit Humanities@NFDI und anderen Konsortien
- Integration in OneNFDI
- Kooperation und Synergien



Governance und Verantwortung innerhalb der NFDI

- Leitung der Konsortialversammlung
- Engagement in zentralen Strategiegruppen, Sektionen und Task Forces der NFDI
- Einbringung von Forschungsperspektiven und Anpassung an strukturelle Entwicklungen.
- Langfristige Partnerschaft: Mitgestaltung des NFDI-Systems



Sektionen, Metadaten, Arbeitsgruppen und Standardisierung

- NFDI-Sektionen
 - Common Infrastructures
 - Ethical, Legal and Social Aspects
 - Industry Engagement
 - International Engagement
 - (Meta)daten, Terminologien, Provenienz
 - Training & Education
- Normungsprozesse: nationalen und internationalen Normung
- Beispiel: Geschichteter Metadatenansatz
 - Kern-, domänenspezifische und zentrumsspezifische Ebenen für ausgewogene Metadatenstandards
- Standardisierung als Enabler



Governance-Struktur innerhalb von Text+

- Rotationsprinzip der Antragstellenden Einrichtung
 - Erste 5 Jahre: Leibniz-Institut für Deutsche Sprache
 - Zweite 5 Jahre: Universität Göttingen
- Funktion:
 - Kontinuität
 - geteilte Verantwortung
 - langfristige Stabilität



Kontinuität und personelle Entwicklung im Konsortium

- Langfristige Zusammenarbeit im Konsortium
- Übergänge als Teil der Governance-Kultur
 - Plenary Mannheim (vor 2 Jahren): Verabschiedung von Erhard Hinrichs als Sprecher
 - Wolfram Horstmann (Konzeptionsphase)
 - Regine Stein - in neuer institutioneller Rolle
 - Neue mitantragstellende Einrichtungen
 - Neue institutionelle Partner



Europäische Einbindung: CLARIN, DARIAH und nationale Knoten

- Enge Zusammenarbeit mit
 - CLARIN
 - DARIAH
- Rolle von Text+ als nationaler Beitrag zu europäischen Forschungsdateninfrastrukturen
- Abstimmung mit dem Ministerium:
 - Verein als juristische Person: Verein Geistes- und kulturwissenschaftliche Forschungsinfrastrukturen e.V.
 - Nationale Knoten über den Verein
- Inhaltliche Beiträge durch Text+: Nationale Beiträge in europäische Infrastrukturen



Zentrale Dienste von Text+

- Umfangreiche Dienstinfrastruktur: Data Space, Registry, Federated Content Search
- Zentrale Anlaufstelle für Forschende: Webportal und Helpdesk
- Service-Katalog für Transparente Darstellung der Angebote
- Verzahnte Dienstleistung im Forschungs-Daten-Lebenszyklus



Die verteilte Infrastruktur als Stärke

- Dezentrale Text+ Zentren: operative Rückgrat der Infrastruktur
- Kohärente gemeinsame Architektur
- Nachhaltigkeit und geteilte Verantwortung
- kurze Wege, fachnaher Beratung und Zugang für Forschende



Integration von Base4NFDI-Diensten

- Prinzip der Integration von NFDI-Basisdiensten
- Verbessertes Nutzungserlebnis
- Fokussierung auf Community



Wo stehen wir heute?

- Leistungsfähige Infrastruktur für die geisteswissenschaftliche Forschung mit vielfältigen Diensten.
- Breites Dienstportfolio gut in die nationale Forschungsdateninfrastruktur (NFDI)
- Nutzungslücke in Forschung
- Zentrale Herausforderung: Diskrepanz zwischen Infrastruktur und Nutzung



Zwei Jahre als Entscheidungsphase

- Strategische Bedeutung: Text+ als unverzichtbaren Bestandteil der Forschungslandschaft.
- Konsequente Ausrichtung auf Nutzung in der Forschung.
- Stärkere Sichtbarkeit in der Community
- Integration in Forschungsprozesse



Forschung im Zentrum: Nutzung statt Angebot

- Nutzung im Forschungsprozess: für die Fragestellungen der Forschenden
- Forschungsdatenmanagement als Enabler: Erkenntnisgewinn und Werkzeug für Forschung.
- Bedürfnissen der Forschenden
- Integration in tatsächliche Forschungsprozesse



Text als Ressource: KI und LLMs

- Neue Forschungsmöglichkeiten durch den Einsatz von LLMs in den Geisteswissenschaften
- Hochwertige, kuratierte und rechtlich einwandfreie Sprach- und Textdaten Rolle von Text+
- Zentrale Infrastruktur und Expertise für KI-gestützte geisteswissenschaftliche Forschung
- Etablierung von Text+ als Referenzinfrastruktur



Programm der nächsten zwei Tage

- Infrastructure/Operations
- Editions
- Lexical Resources
- Collections
- Ausblick





<https://events.gwdg.de/event/1274>

15.–16. Juni 2026
Mannheim, Schloss
Europe/Berlin Zeitzone

Geben Sie Ihren Suchbegriff ein

Überblick

Tagungsprogramm

Anmeldung

Postersession: Text+
Plenary 2026

Hotels

Informationen zum
Datenschutz

Ansprechpersonen

Contact

✉ textplus-events@ids-ma...

Text+ ist ein Konsortium der Nationalen Forschungsdateninfrastruktur (NFDI) und hat zum Ziel, sprach- und textbasierte Forschungsdaten langfristig zu erhalten und ihre Nutzung in der Wissenschaft zu ermöglichen.



Collections
Lexical
Resources
Editions
Infrastruktur
e/Operations

Funded by

DFG Deutsche
Forschungsgemeinschaft

German Research Foundation

Project number 460033370

Part
of

nfdi Nationale
Forschungsdaten
Infrastruktur

Diese Präsentation entstand im Rahmen der Arbeit des Vereins „Nationale Forschungsdateninfrastruktur (NFDI) e.V.“. Die NFDI wird von der Bundesrepublik Deutschland und den 16 Bundesländern finanziert, und das Konsortium Text+ wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert – Projektnummer 460033370. Die Autoren bedanken sich für die finanzielle Unterstützung und Förderung. Darüber hinaus gilt unser Dank allen Institutionen und Akteuren, die sich für den Verein und seine Ziele engagieren.



<https://www.text-plus.org>



5. TEXT+ PLENARY

PROGRAMM

Montag, 15. Juni 2026

12:00 – 13:30 *Lunch (Katakomben)*

13:30 – 14:00 **Begrüßung, Eröffnung und Einführung in das Programm (Aula)**

14:00 – 15:30 **Ergebnisse aus Task Area: Infrastructure & Operations (Aula)**

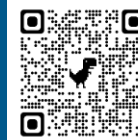
15:30 – 16:05 Kaffeepause (Katakomben)

16:05 – 17:30 **Ergebnisse aus Task Area: Editions (Aula)**

17:30 – 18:00 **Posterslam (Aula)**

18:00 – 19:00 **Postersession (Fuchs-Petroluh-Saal)**

<https://www.text-plus.org>





Ergebnisse aus den Task Areas

Gefördert durch

DFG Deutsche
Forschungsgemeinschaft

Förderungsnummer 460033370

Teil der

nfdi Nationale
Forschungsdaten
Infrastruktur



Übersicht

15.6.2026, 14:00 – 15:30

Ergebnisse aus Task Area Infrastructure & Operations (IO)

15.6.2026, 16:00 – 17:30

Ergebnisse aus Task Area Editions

16.6.2026, 9:00 – 10:30

Ergebnisse aus Task Area Lexical Resources

16.6.2026, 11:00 – 12:30

Ergebnisse aus Task Area Collections



Infrastructure / Operations

Highlights of the 1st funding phase
Text+ Plenary 15-16.6.2025

Funded by

DFG Deutsche
Forschungsgemeinschaft
German Research Foundation

In cooperation with

nfdi Nationale
Forschungsdaten
Infrastruktur

Thematische Blöcke

- 01** Big Picture
- 02** Technische Basisinfrastruktur
- 03** Search & Retrieval
- 04** Software Services, Data Processing Pipelines & LZA
- 05** Interoperability/Re-Usability
- 06** Außendarstellung & Community Outreach
- 07** Diskussion
- 08** IO in der NFDI, Standortbestimmung

01

Big Picture

Stefan Buddenbohm → Alexander Steckel

Text+ Vision

The text- and language-oriented humanities and social sciences extensively leverage the possibilities of digitisation in their research, teaching, and transfer, establishing a common data culture.

Covering the **entire humanities research landscape.**

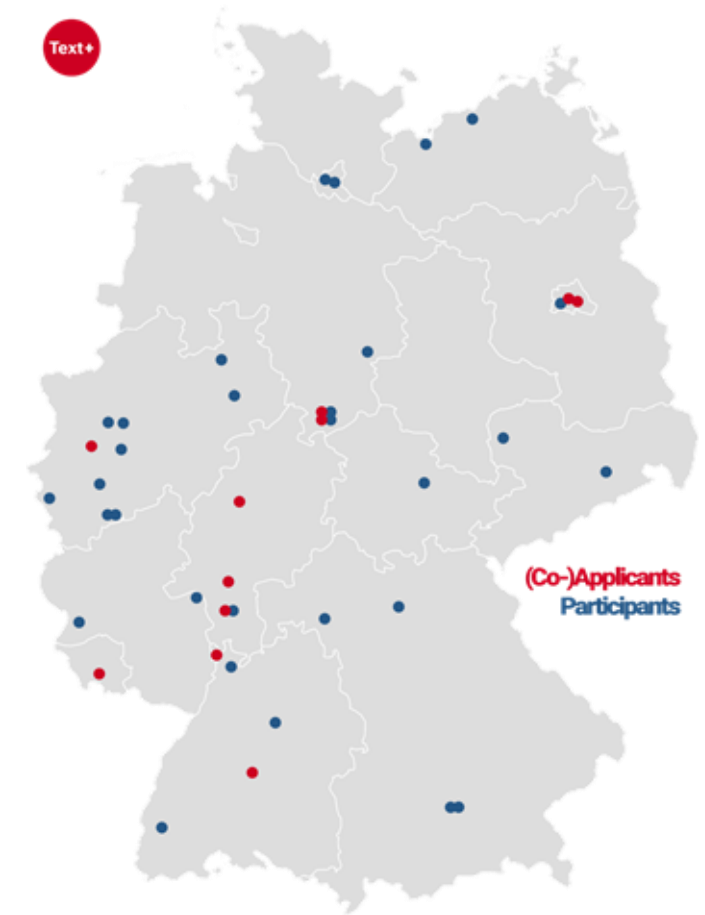
Services for the **entire spectrum of text-based research.**

Strong combination of **university and non-university institutions.**

21 universities, 8 academies of sciences and humanities, 12 computing centres, libraries, and non-university research institutions

5 task areas with co-spokespersons from universities and non-university research institutions:

- **Collections:** Philippe Genêt (DNB)
- **Lexical Resources:** Alexander Geyken (BBAW) and Axel Herold (BBAW)
- **Editions:** Andreas Speer (Academy of Sciences North-Rhine Westphalia) and Kilian Erasmus Hensen (CCeH)
- **Infrastructure Operations:** Philipp Wieder (GWDG) and Stefan Buddenbohm (SUB Göttingen)
- **Administration:** Philipp Wieder (GWDG) and Andreas Witt (IDS Mannheim)



Der rote Faden der Daten. Vom Standard zum System zum Nutzen: Konsistente Metadaten und PIDs sind das Rückgrat, das verstreute Forschungsdaten in eine navigierbare digitale Landkarte verwandelt.

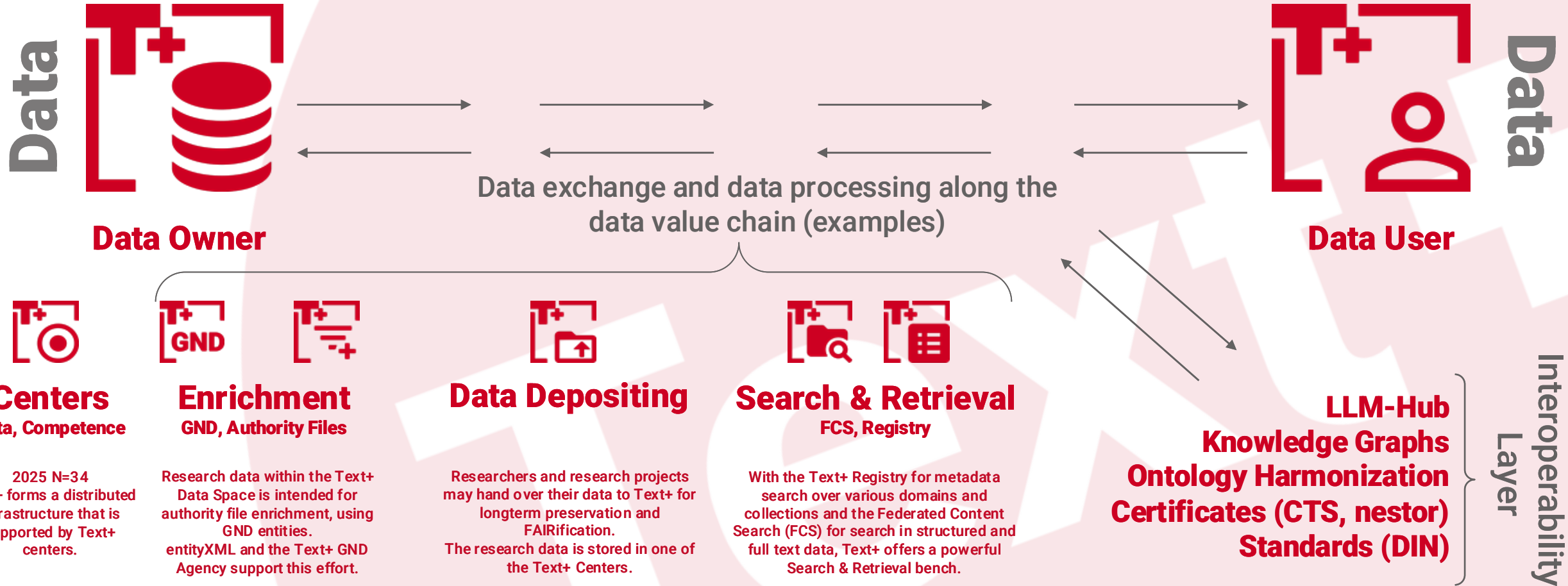
Vielfalt unter einem Dach. Föderiert, skalierbar, vernetzt: Die föderierte Text+ Dateninfrastruktur bündigt die Komplexität verteilter Datenbestände und macht sie zentral durchsuchbar.

Tiefer graben mit Hybrid Search. Jenseits der Oberfläche: Die Verknüpfung von Registry und Federated Content Search schlägt die Brücke von Metadaten direkt in die Tiefe komplexer Text- und Sprachinhalte.

Präzision durch Identität. Vom Namen zum Netzwerk: Dank Normdaten werden aus historischen Varianten eindeutige Entitäten – für eine Forschung, die versteckte Verbindungen in Wissensgraphen sichtbar macht.

IO enables data sharing

Text+ in the Data Space

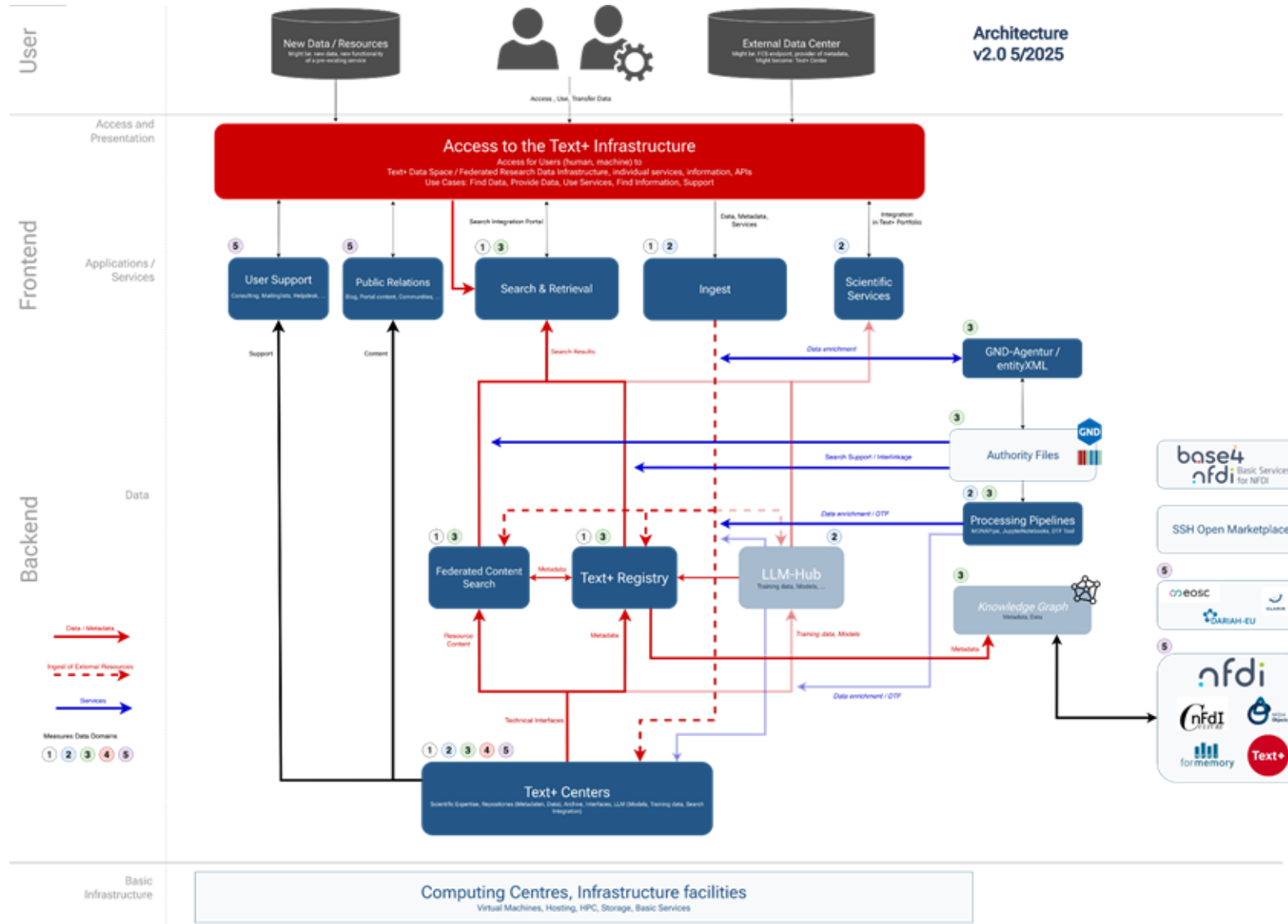


02

Basic Infrastructure & Ground Work

Stefan Buddenbohm → Alexander Steckel

Architecture



IO Enables Project Collaboration



Einmal anmelden, alles nutzen

Identity & Access

- SAML, Shibboleth und OpenID Connect
- Keine dienstspezifischen Logins

Compute & Storage

- OpenStack (Self-Service & API)
- Zugriff auf Speicher

AI & Advanced Compute

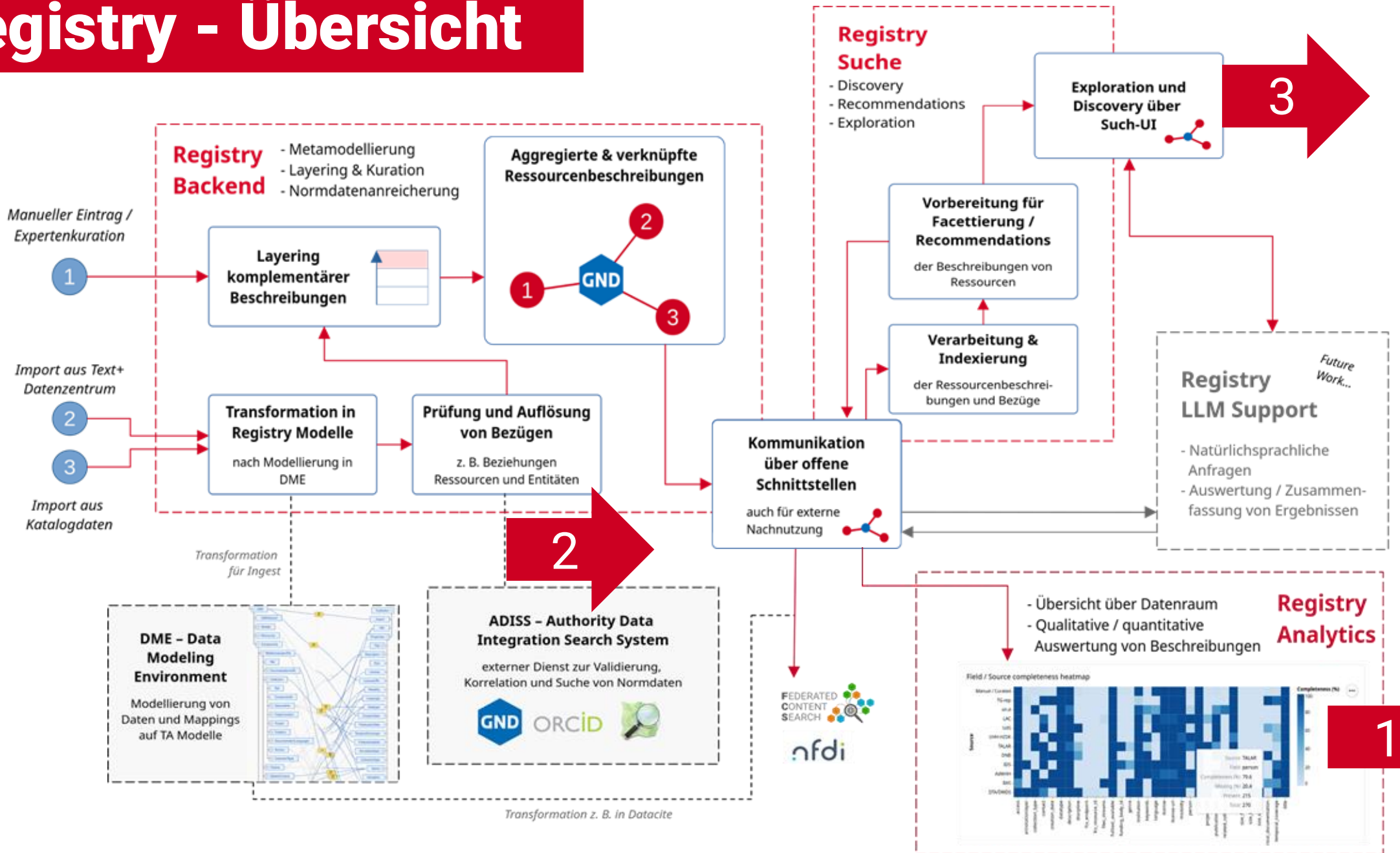
- Text+-spezifischer AI Assistant
- OpenAI-kompatible API
- Finetuning & RAG-fähig

03

Search & Retrieval

Thomas Eckart, Tobias Gradl

Registry - Übersicht



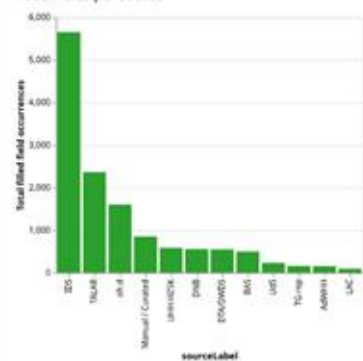
Registry - Analytics

Registry Analytics Overview **Graph** Textual collections Editions Lexical Resources

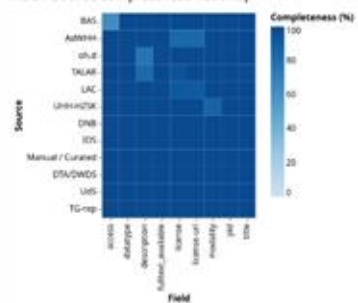
Textual collections dbl-click / Shift+click to maximize charts

Mandatory fields Toggle Data CSV

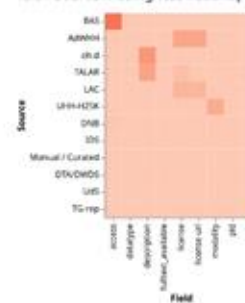
Model fields per source



Field / Source completeness heatmap

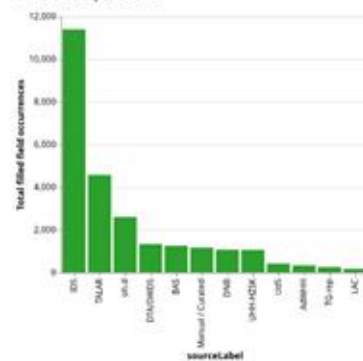


Field / Source missingness heatmap

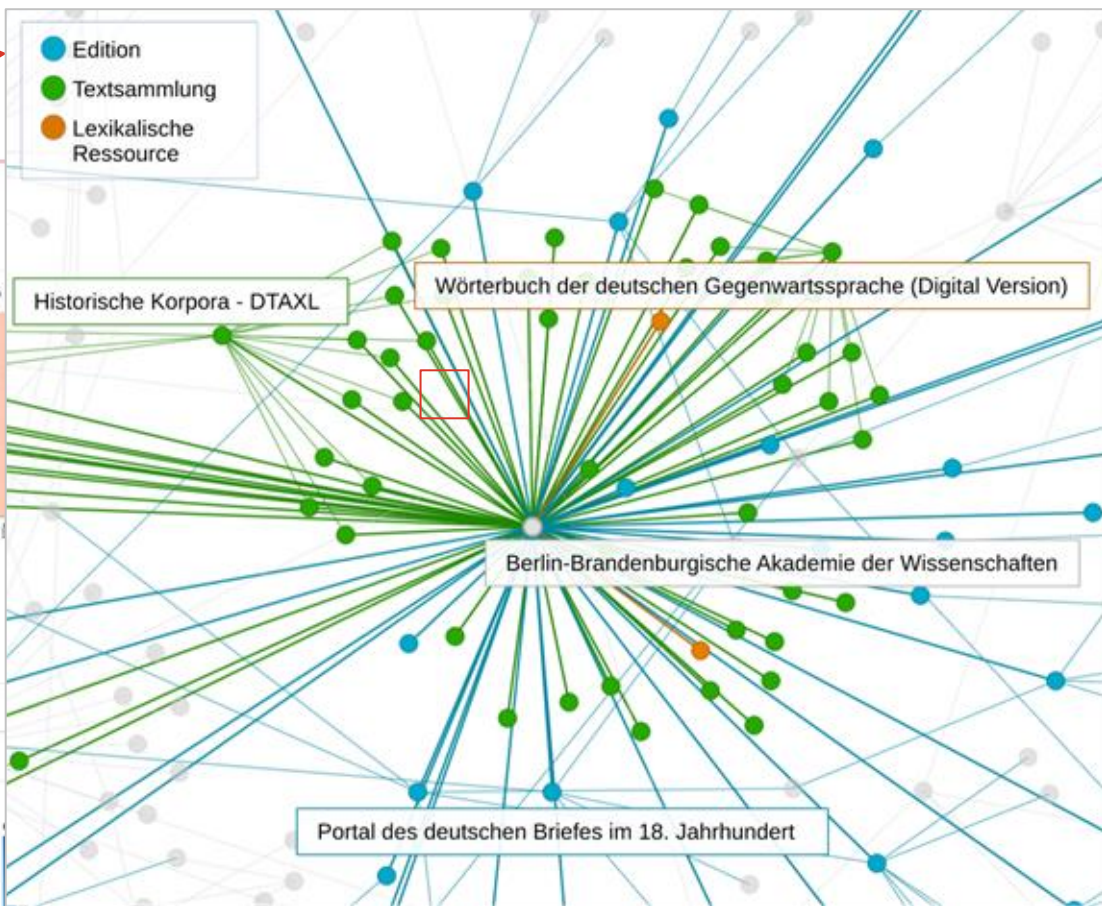
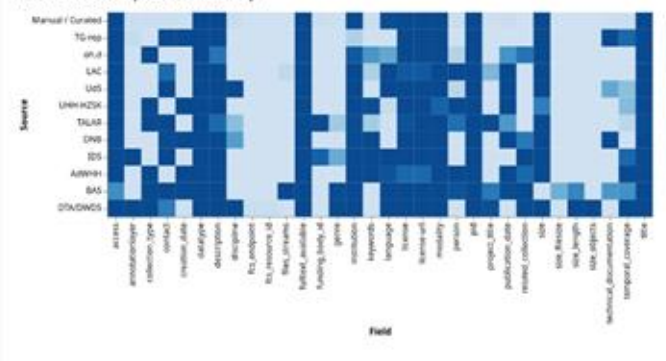


All fields Toggle Data CSV

Model fields per source



Field / Source completeness heatmap

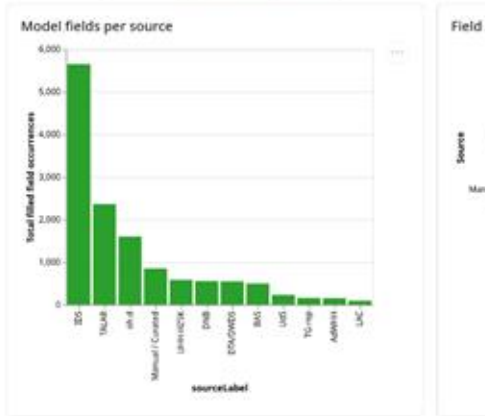


Registry - Analytics

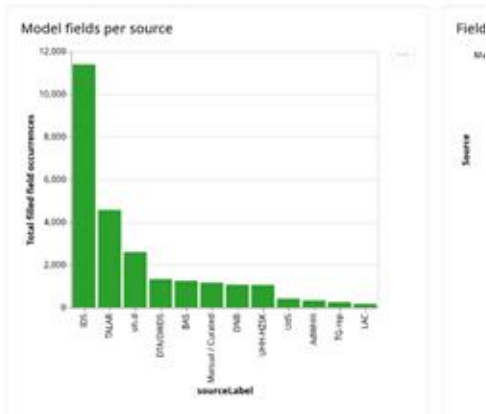
Registry Analytics Overview Graph Textual collections Edit

Textual collections dbl-click / Shift+click to maximize:

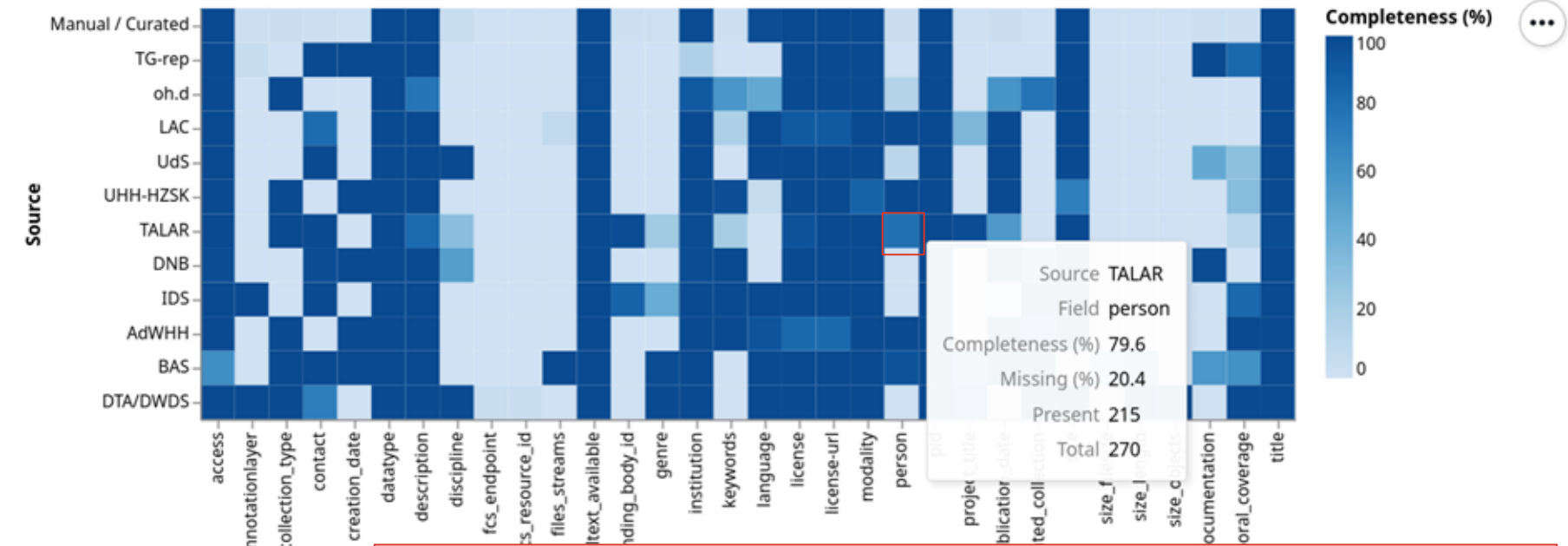
Mandatory fields



All fields



Field / Source completeness heatmap



aktuell: erste Kennzahlen zu Metadaten...

- hier Anteil Metadatenfeld / Einträge je Datenzentrum
- (noch) nicht: inhaltliche Qualität (z. B. Freitext vs. Identifier)

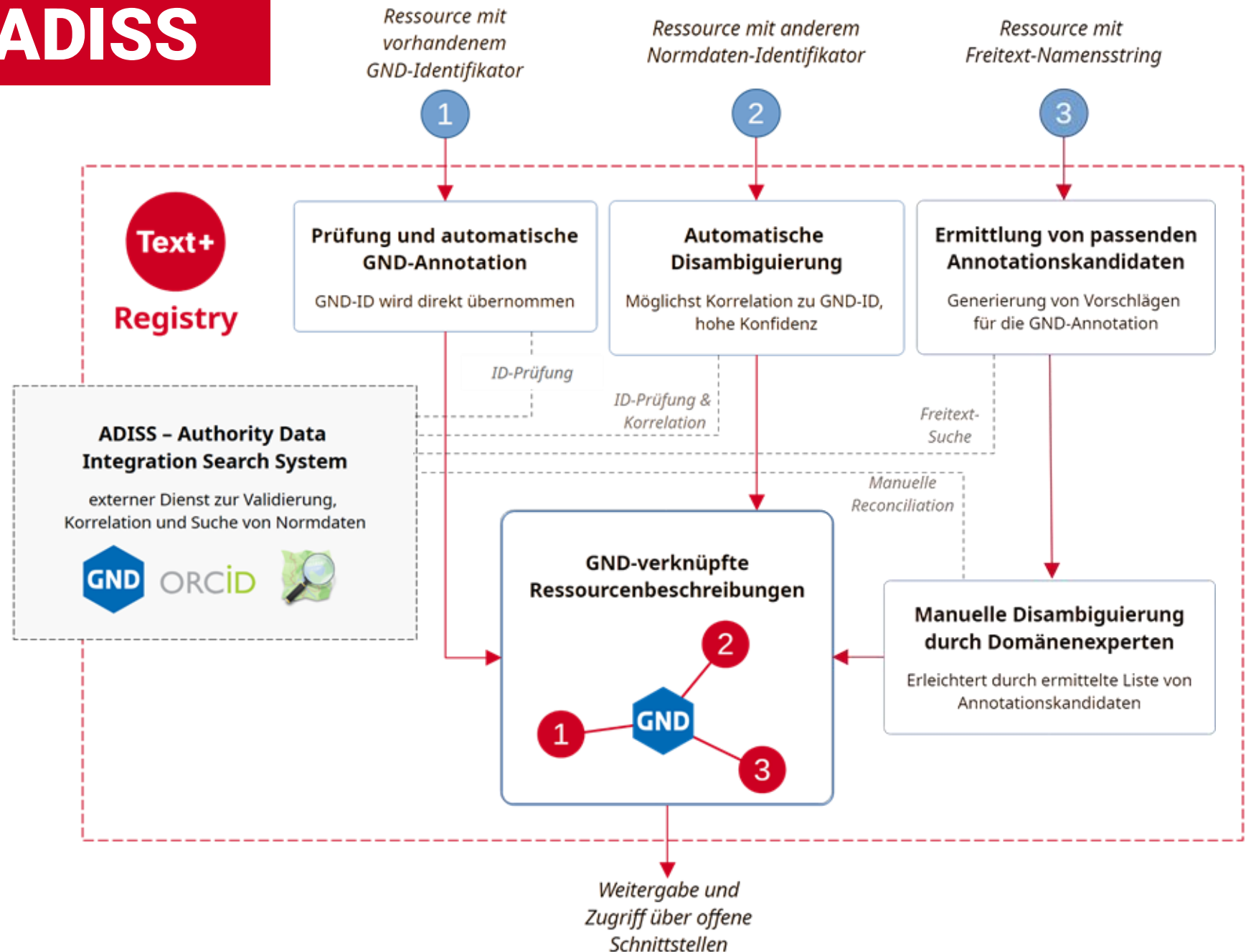
Gerne auf Darstellung des eigenen Datenzentrums achten: <https://registry.text-plus.org/analytics/>



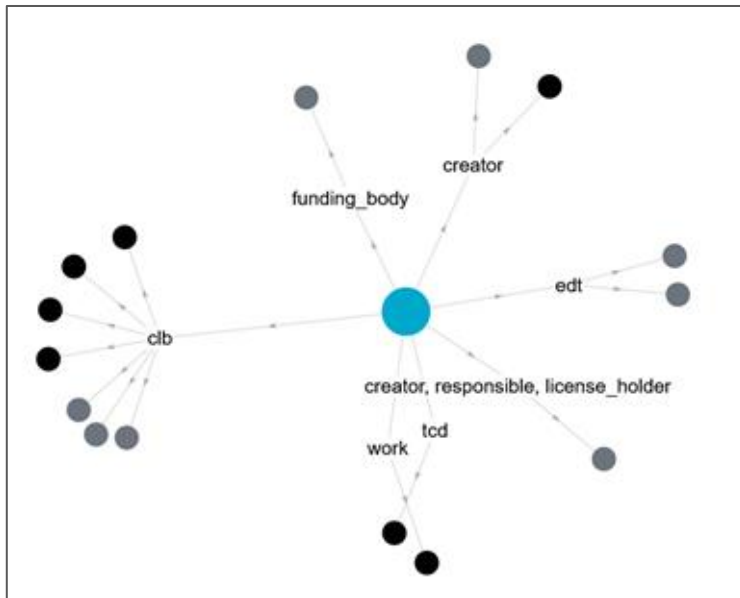
Registry - ADISS

ADISS

- Integrierte Suche über Normdatenbanken und Knowledge Bases
- **Ziel:** Möglichst umfassende GND-Verknüpfung der Ressource, selbst bei externem oder fehlendem Identifikator



Registry - Suchunterstützung



Titel

eng Historisch-kritische Ausgabe der Werke und Briefe Adalbert Stifters

deu Historisch-kritische Ausgabe der Werke und Briefe Adalbert Stifters

Abstract

The historical-critical edition of the works and letters of Adalbert Stifter (1805-1868) comprises a total of about 50 volumes. In addition to the poetic work, the edition will also contain all documents of his official activities as Imperial and Royal School Councillor in Linz and as Provincial Conservator of Upper Austria, as well as his extensive correspondence. All volumes contain a text-critical apparatus and a commentary with detailed explanations of individual passages, as well as information on the origins and reception of the texts by Stifter. Die Historisch-Kritische Ausgabe der Werke und Briefe von Adalbert Stifter (1805-1868) umfasst insgesamt rund 50 Bände. Neben dem dichterischen Werk wird die Edition auch alle Dokumente seiner

ADISS – Authority Data Integration Search System

externer Dienst zur Validierung, Korrelation und Suche von Normdaten

The ADISS box is a light grey rectangle containing the title, a description, and logos. An arrow points from the 'Adalbert Stifters' text in the previous block to the ADISS box.

Adalbert Stifter <https://d-nb.info/gnd/118618156>
Anfangs Privatlehrer in Wiener Adelsfamilien, u. a. im Haus des Fürsten Metternich. 1848-1851 Redakteur des „Wiener Boten“ und der „Linzer Zeitung“ ...

Namen und Beschreibungen verknüpfter Entitäten

Named Entity Recognition in Titel und Beschreibung

Auflösen mit ADISS

Nutzung verknüpfter Entitäten um Ressourcen zu kontextualisieren.

Personen und Institutionen werden getaggt.

Erkannte Entitäten werden mit ADISS automatisch aufgelöst und zur Anreicherung genutzt.

Registry - Suche (aktuelle Version)

- Kontinuierliche Weiterentwicklung mit den Datendomänen
 - Suchfilter
 - Erklär- und Hilfetexte
- Neues Datenmodell für lexikalische Ressourcen

Registry - Der Text+ Katalog

Übergreifender Katalog Editionen Korpora- und Textsammlungen Lexikalische Ressourcen Lexikalische Ressource (v2) ?

Filter

Importquelle

Feldtyp

Lexikalischer Typ

Zweisprachiges Wörterbuch (6)

Einsprachiges Wörterbuch (6)

Fachwörterbuch (2)

Modalität

Objektsprache

Sprachperiode

N/A (8)

Gegenwartssprache (4)

Deutsch in seiner Gesamtheit (2)

Sprachregion

14 von 14 Ressourcen

Lexikalische Ressource (v2)

[Lexical and Morphological Xhosa Dictionary](#)

A dictionary for Xhosa based on the Bantu Language Model (BLM). It contains lexical and morphological data and English translations. The data is based on extensive manual work by J.A. Louw, was digitized and enhanced by and under the supervision of Prof. Sonja Bosch (University of South Africa), and converted from CSV files to this TEI Lex-0 dataset by the Saxon Academy of Sciences and Humanities in Leipzig.

SAW Lemma Übersetzung Numerus Wortart Grundform Verwandter Begriff Zweisprachiges Wörterbuch Geschrieben

Lexikalische Ressource (v2)

[WAHRIG Deutsches Wörterbuch](#)

WAHRIG Deutsches Wörterbuch (WDW) wurde von dem Lexikografen Gerhard Wahrig erarbeitet und erschien 1966 in der 1. Auflage, damals noch unter dem Titel Das große deutsche Wörterbuch. Seit mehreren Jahrzehnten gilt dieses einsprachige Wörterbuch als bewährtes Standardwerk der deutschen Gegenwartssprache. Sowohl im In- als auch im Ausland dient es Muttersprachlerinnen und Muttersprachlern sowie Deutschlernenden gleichermaßen als wertvolle Orientierungshilfe im Umgang mit der deutschen Sprache. Nach dem Tod Gerhard Wahrigs im Jahre 1978 führte seine Tochter Renate Wahrig-Burfeind das lexikografische Werk des Vaters fort.

DTA/DWDS Antonym Zitat Definition Etimologie Genus Hyperonym Hyponym Lemma Numerus Phonetik Wortart Segmentierung Synonym

Einsprachiges Wörterbuch Geschrieben Gegenwartssprache Deutschland

Lexikalische Ressource (v2)

[SentimentWortschatz: a sentiment annotated wordlist \(v2.0\)](#)

Registry - Neue Benutzeroberfläche



<https://c102-142.cloud.gwdg.de/test-registry/>

Entwicklung der neuen Benutzeroberfläche mit Vue.js

- flexibleres und moderneres Framework
- leichtere Einbindung ins Webportal
- konfigurierbar für weitere Anwendungsfälle

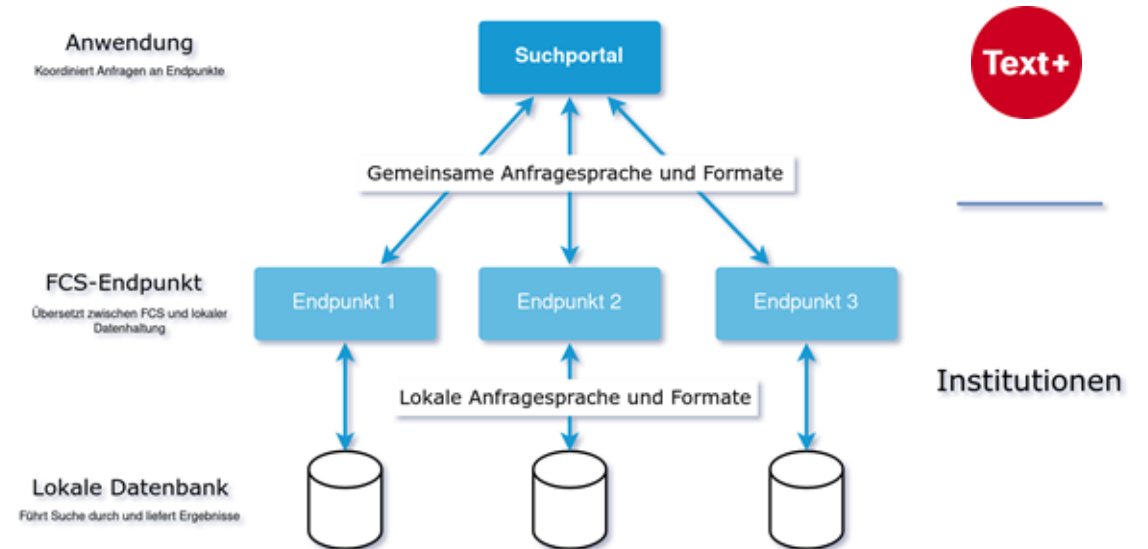
The screenshot displays the 'Registry Search' interface. On the left, there is a 'Filters' sidebar with categories like 'Allgemein', 'Zugang & Verfügbarkeit', 'Inhalte der Editionen', and 'Aufbau der Editionen'. Below these are 'Disziplinäre Zuordnung' options, including 'Disziplinäre Zuordnung (DFG)' with a search filter and a tree view of disciplines such as 'Geistes- und Sozialwissenschaften (1710)', 'Sozial- und Verhaltenswissenschaften (27)', 'Geisteswissenschaften (1493)', 'Alte Kulturen (59)', 'Geschichtswissenschaften (765)', 'Kunst, Musik, Theater- und Medienwissenschaften (93)', 'Sprachwissenschaften (36)', 'Literaturwissenschaft, Kulturwissenschaft (441)', 'Sozial- und Kulturanthropologie, Außereuropäische Kul...' (76), 'Theologie (76)', 'Philosophie (139)', 'Lebenswissenschaften (6)', and 'Naturwissenschaften (16)'. The main area features a search bar with 'Suchen...' and a magnifying glass icon. Below the search bar are navigation tabs: 'ÜBERGREIFENDE SUCHE', 'KOLLEKTIONEN', 'EDITIONEN' (selected), and 'LEXIKALE RESSOURCEN'. There are also buttons for 'VOLLTEXTSUCHE' and 'SEMANTISCHE SUCHE'. The search results section shows '1728 Ergebnisse'. The first result is 'Edition der evangelischen Kirchenordnungen des 16. Jahrhunderts', with a description: 'Church orders are a new type of source emerging in the 16th century and an invaluable treasure trove for research. They reflect the position of the respective sovereign or magistrat determination of the state of confession and the resulting religious norms. The church ordinances of the various territories and imperial cities by no means only contain internal ch... (worship regulations such as prayer forms, liturgies and information on the frequency, times and contents of the sermon), but also numerous organisational regulations which exte...'. Below the description is a table of metadata:

Angabe der Zugänglichkeit	Frei
Medium	Text
Disziplinäre Zuordnung (DFG)	Geisteswissenschaften, Geistes- und Sozialwissenschaften, Theologie, Evangelische Theologie
Typ	Hybrid
Datenquelle / Provenienz	AGATE
Sprache	Deutsch

The second result is 'Prosopographia Imperii Romani', with a description: 'The Prosopographia Imperii Romani is a biographical dictionary, which essentially covers the ruling class of the Roman Empire in the early and high imperial period. The investigat... Aktium 31 BC, which sealed the monarchical future of Rome. It ends with the reign of Diocletian (284-305), as the state and social structure of the empire had changed considerably material, which is completely recorded, is wide-ranging. It encompasses the entire literature of the imperial era, i.e. not only the historiography and the works of the poets, but also c...'. Below the description is a table of metadata:

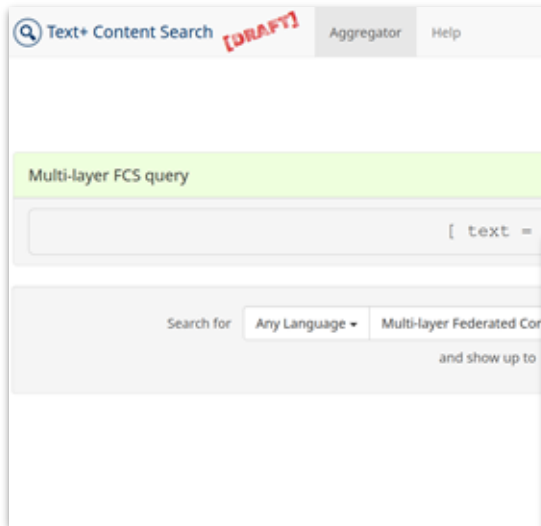
Angabe der Zugänglichkeit	Frei
Medium	Text
Disziplinäre Zuordnung (DFG)	Geisteswissenschaften, Geistes- und Sozialwissenschaften, Theologie, Evangelische Theologie
Typ	Hybrid
Datenquelle / Provenienz	AGATE
Sprache	Deutsch

- **Zusammenarbeit mit allen Datendomänen**
 - Ausrichtung von Workshops und Hackathons
 - Mehr beteiligte Datendomänen (1 → 3)
 - Mehr beteiligte Institutionen (+100%)
 - Mehr Endpunkte (+200%)
 - Mehr verfügbare Ressourcen (+400%)

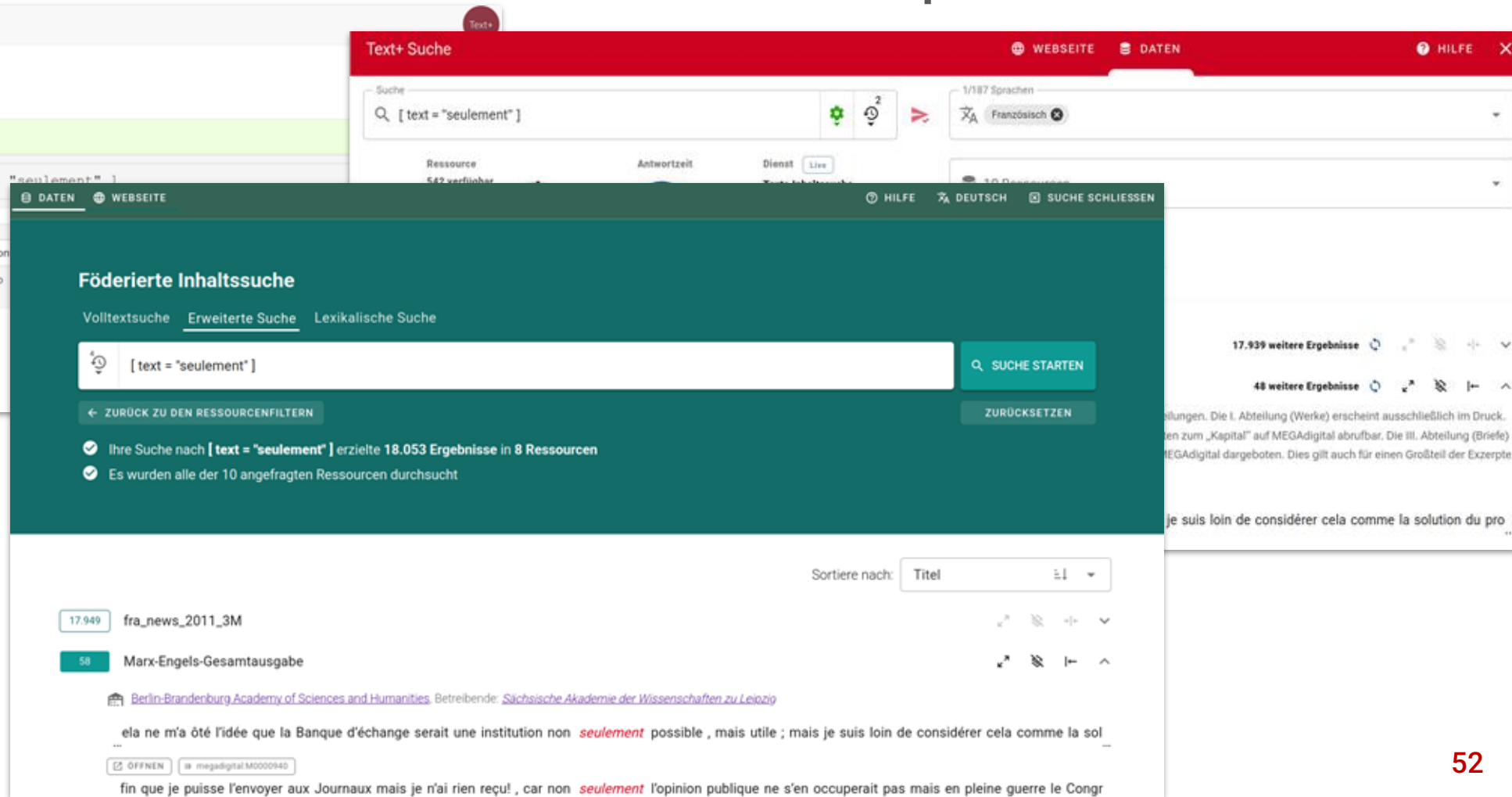


FCS – Benutzeroberflächen

1. ab August 2022



2. ab September 2023



3. seit Juni 2026

Zunehmende Annotation von Textinhalten mit Normdaten-Entitäten, z.B.

- Annotation von Personen, Organisationen, Ereignissen und Orten in Textkorpora und Texteditionen
- Bedeutungsdefinitionen in Wörterbüchern und Enzyklopädien

Autor (Metadaten)



Friedrich III.

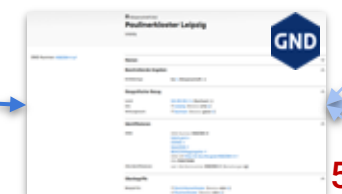
Publikationsort
(Metadaten)



Torgau



Giovanni Nanni



Paulinerkloster Leipzig

Nr. 125 Kf. Friedrich *an*
Prior [Johann Oertel] des Dominikanerklosters Leipzig
6. April 1514 (Donnerstag nach Judica) · Torgau · Brief · Konzept · deutsch

A:
LATH – HStA Weimar, EGA, Reg. Kk 747, fol. 1rv (Konzept).

[1] Kf. Friedrich bedankt sich bei dem Prior [Johann Oertel] des Dominikanerklosters St. Paul zu Leipzig für das übersendete Verzeichnis der Klosterbibliothek, um das er gebeten hatte. [2] Darin hat er die „Expositio super apocalypsim“ des Annius von Viterbo entdeckt, die er für die Erarbeitung der gerade entstehenden Chronik benötigt. Um daraus Abschriften anzufertigen, bittet er, dass das Buch einem Terminierer mit einem Begleitschreiben in der kommenden Woche mitgegeben wird. Zur nächsten Leipziger Ostermesse soll das Buch den Mönchen unversehrt zurückgegeben werden. [3] Darüber hinaus sucht der Kf. noch einige Bücher, die in einem beiliegenden Verzeichnis genannt sind.

Quelle: Briefe und Akten zur Kirchenpolitik Friedrichs des Weisen und Johans des Beständigen 1513 bis 1532

FCS – Live Demo der EntityFCS



DATEN WEBSITE HILFE DEUTSCH SUCHE SCHLIESSEN

Föderierte Inhaltssuche

Volltextsuche Erweiterte Suche Lexikalische Suche

Bsp.: [pos = "ADV"](2,4) SUCHE STARTEN

1 Die erweiterte Suche über Annotationslayer erlaubt die Abfrage von annotierten Textkollektionen, -korpora und vergleichbarer Ressourcen über die Anfragesprache FCS-QL, die sich an der populären Korpusanfrage-Sprache CQP orientiert. Dabei können für Wörter im Text komplexe Suchkriterien angegeben und diese für längere Textstellen miteinander kombiniert werden.

445 verfügbare Ressourcen

Zur Beschleunigung der Suche können Sie die gewünschten Ressourcen vorab auswählen.

445 Ressourcen 187 Sprachen

Sortiere nach: Titel

Aachener Zeitung afr_wikipedia_2018_3 00K afr-za_web_2018_1M Aloys Hirt: Briefwechsel und Amtliche Schriften amh_wikipedia_2018_ 30K Ancient Greek Prose (Vanessa Gorman) ara_newscrawl_2013_ 1M

Schnelleinstieg

Machen Sie sich mit der Benutzeroberfläche vertraut.

1 Suchmodus wählen

Wählen Sie den passenden Suchmodus.
Die ECS unterstützt für verschiedene Suchinteressen die Volltextsuche, die erweiterte Suche und die lexikalische Suche. Diese wirken sich sowohl auf die Art der formulierbaren Anfragen als auch auf die Ergebnispräsentation aus. Weitere Informationen dazu finden Sie im Reiter 'Suchtechnologie und Ressourcen'.

WEITER BEREICH ZEIGEN

2 Ressourcen filtern optional

3 Suchbegriff eingeben

4 Ergebnisse ansehen

Suchtechnologie und Ressourcen

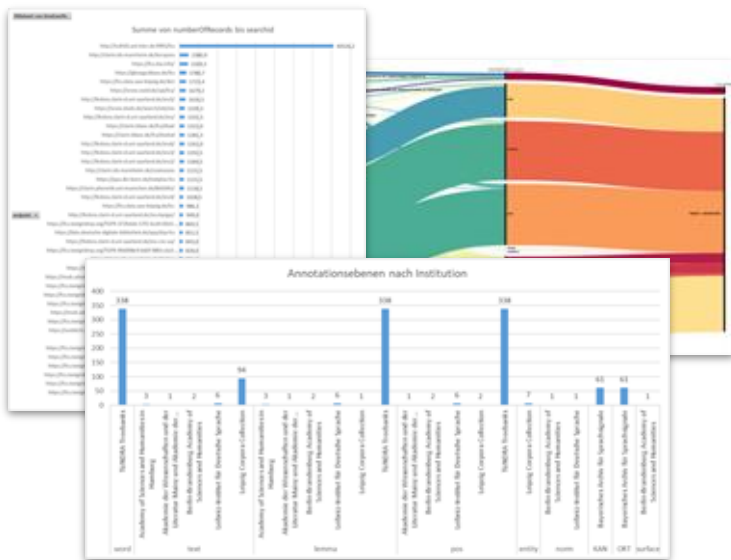
Verwenden Sie die erweiterten Möglichkeiten zu Abfrage und Darstellung.

FAQ

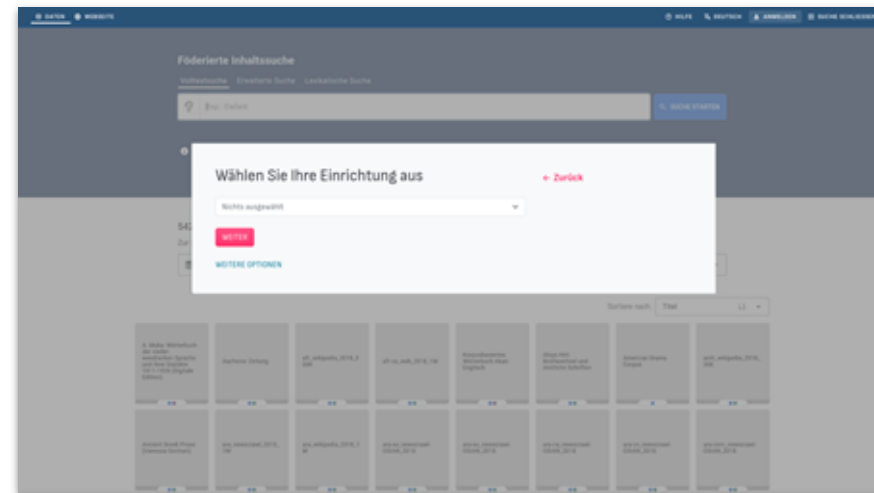
Häufig gestellte Fragen

→ <https://text-plus.org/fcs>

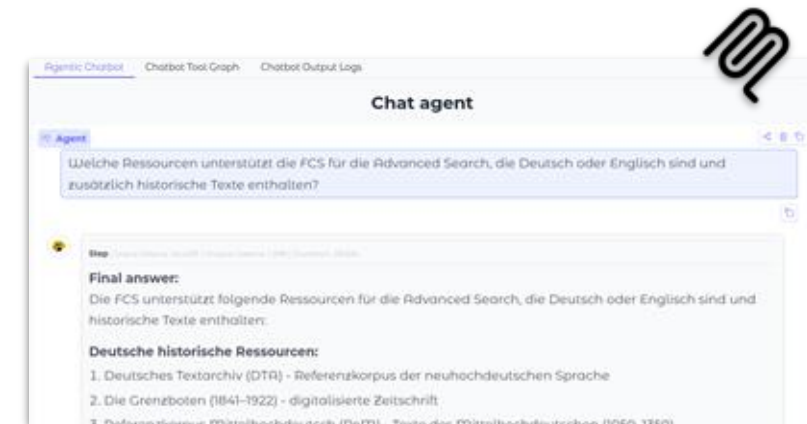
FCS – Aktuelle Arbeiten / Vorarbeiten



Kontinuierliches Qualitätsmonitoring
des FCS-Ökosystems



AAI-Integration zur Unterstützung
zugriffsbeschränkter Ressourcen



Ertüchtigung der FCS für
agentenbasierte Arbeitsprozesse

Hybrides Suchsystem



Muka Zur FCS Suche

5 von 5 Ressourcen

Lexikalische Ressource (v2) Zur FCS Ressourcen Links

[A. Muka: Dictionary of the Lower Sorbian Language and Its Dialects 1911-1928](#)

Das „Wörterbuch der nieder-wendischen Sprache und ihrer Dialekte“ von Arnošt Muka wurde 1911 und 1928 ... Muka: Dictionary of the Lower Sorbian Language and Its Dialects 1911-1928 ... Muka: Wörterbuch der nieder-wendischen Sprache und ihrer Dialekte 1911-1928 ... Arnošt Muka's "Dictionary of the Lower Sorbian Language and Its Dialects" was published in two volumes

SAW

A. Muka: Wörterbuch der nieder-wendischen Sprache und ihrer Dialekte 1911-1928 (Digitale Edition)

Korpusbasiertes Wörterbuch Akan-Englisch

B. Šwjela: Niedersorbisch-deutsches Wörterbuch 1961 (Digitale Edition)

Bayerisches Wörterbuch (BWB)

Kleines Wörterbuch der Verlaufsformen im Deutschen

Deutsche Wortfeldetymologie in europäischem Kontext

Deutsches Sprichwörter-Lexikon von Karl Friedrich Wilhelm Wander

Serbski Institut / Sorbisches Institut

A. Muka: Wörterbuch der nieder-wendischen Sprache und ihrer Dialekte 1911-1928 (Digitale Edition)

Das "Wörterbuch der nieder-wendischen Sprache und ihrer Dialekte" von Arnošt Muka wurde 1911 und 1928 als zweibändiges Werk veröffentlicht. Eine umfassende Sammlung niedersorbischer Lexik, die weiterhin auch Eigennamen, Phraseologismen und Sprichwörter umfasst, sowie eine Vielzahl detaillierte Beschreibungen mit teilweise enzyklopädischem

2 Suchvorschläge

translation = "Schlepp"

Alle Einträge die eine Übersetzung enthalten die mit "Schlepp" beginnt.

> ZEIGE MEHR

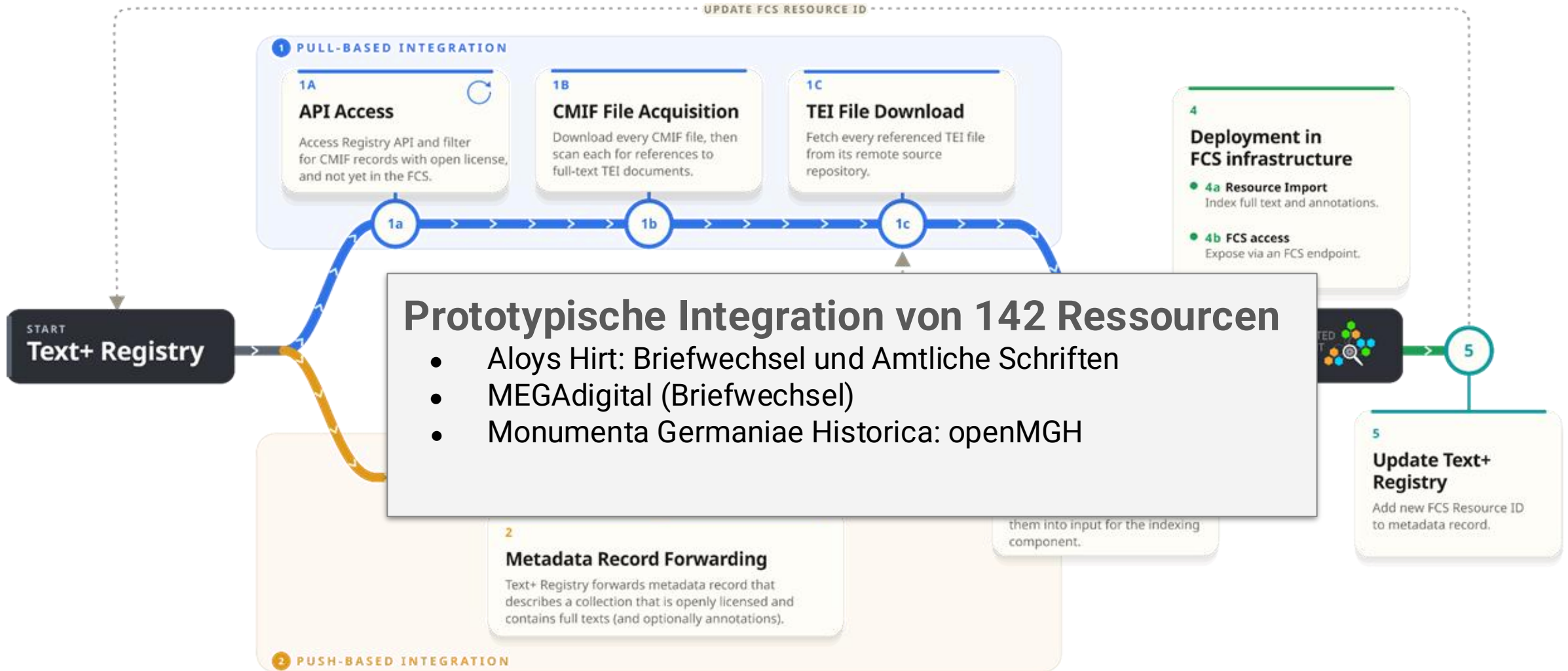
ÖFFNE REGISTRY EINTRAG

ÖFFNE RESSOURCEN-WEBSEITE

ÖFFNE REGISTRY EINTRAG



Hybrides Suchsystem – Integration Service Prototyp



Prototypische Integration von 142 Ressourcen

- Aloys Hirt: Briefwechsel und Amtliche Schriften
- MEGAdigital (Briefwechsel)
- Monumenta Germaniae Historica: openMGH

04

Software Services, Data Processing Pipelines & LZA

George Dogaru

Jupyter Notebooks und Jupyter4NFDI

- Mehrere Tools basierend auf Jupyter Notebooks und Jupyter4NFDI
 - Fluffy Publication Workflow für TGRep (Web-Anwendung)
 - Notebook Actions (Notebook mit einfacher UI)
- Integration weiterer Tools in Jupyter4NFDI
 - MONAPipe
- Aktive Mitgestaltung des Jupyter4NFDI-Dienstes

Tools zur sofortigen Verwendung in Jupyter4NFDI

hub.nfdi-jupyter.de/user/

File Edit View Run Kernel Git Tabs Settings Help

start-tgrep.ipynb

Markdown git Notebook

Jupyter4NFDI-Integration: Zentrales Notebook ermöglicht Auswahl weiterer Notebooks/Tools

Notebooks for text processing, OCR, PDF generation and more

General purpose notebook with a forms-based UI:

- Notebook Actions **Fluffy Publication (TGRep), OCR, XSLT etc.**

Process texts from TGRep using one of the available notebooks:

nlp_notebooks **MONAPipe-Notebooks**

- _basics.ipynb
- named_entity_recognition.ipynb

textgrid_rep **Weitere Notebooks für Texte in TGRep**

- explore_text.ipynb

- gitlab.gwdg.de/textplus/collections/tgrep-jupyter-notebooks
- gitlab.gwdg.de/textplus/textplus-io/nb-actions

TGRep-Jupyter4NFDI-Integration (in Entwicklung)

dev.textgridrep.org/browse/textgrid:mgq9.0

TextGrid Repository

Search term Content ▾ Documentation ▾

Advanced Search

TextGrid Digitale Bibliothek > Dickens, Charles > Romane > Oliver Twist oder Der Weg eines Fürsorgezöglings > Oliver Twist oder Der Weg eines Fürsorgezöglings

Metadata

File Type
text/xml

PID
hdl:11858/00-1734-0000-0002-7ECF-E
Citation Suggestion

Download

Object (TEI)
Metadata (XML)
Tech. Metadata (XML)
Plain Text (txt)
E-Book (epub)
HTML
ZIP

Tools

Voyant ⓘ
Annotate ⓘ
Switchboard ⓘ
Jupyter Notebook
TEI V

Charles Dickens
Oliver Twist
oder
Der Weg eines Fürsorgezöglings
(Oliver Twist, or, The Parish Boy's Progress)

[3] Erstes Kapitel
Schildert den Ort, wo Oliver auf die Welt kam, sowie die seine Geburt begleitenden Umstände.

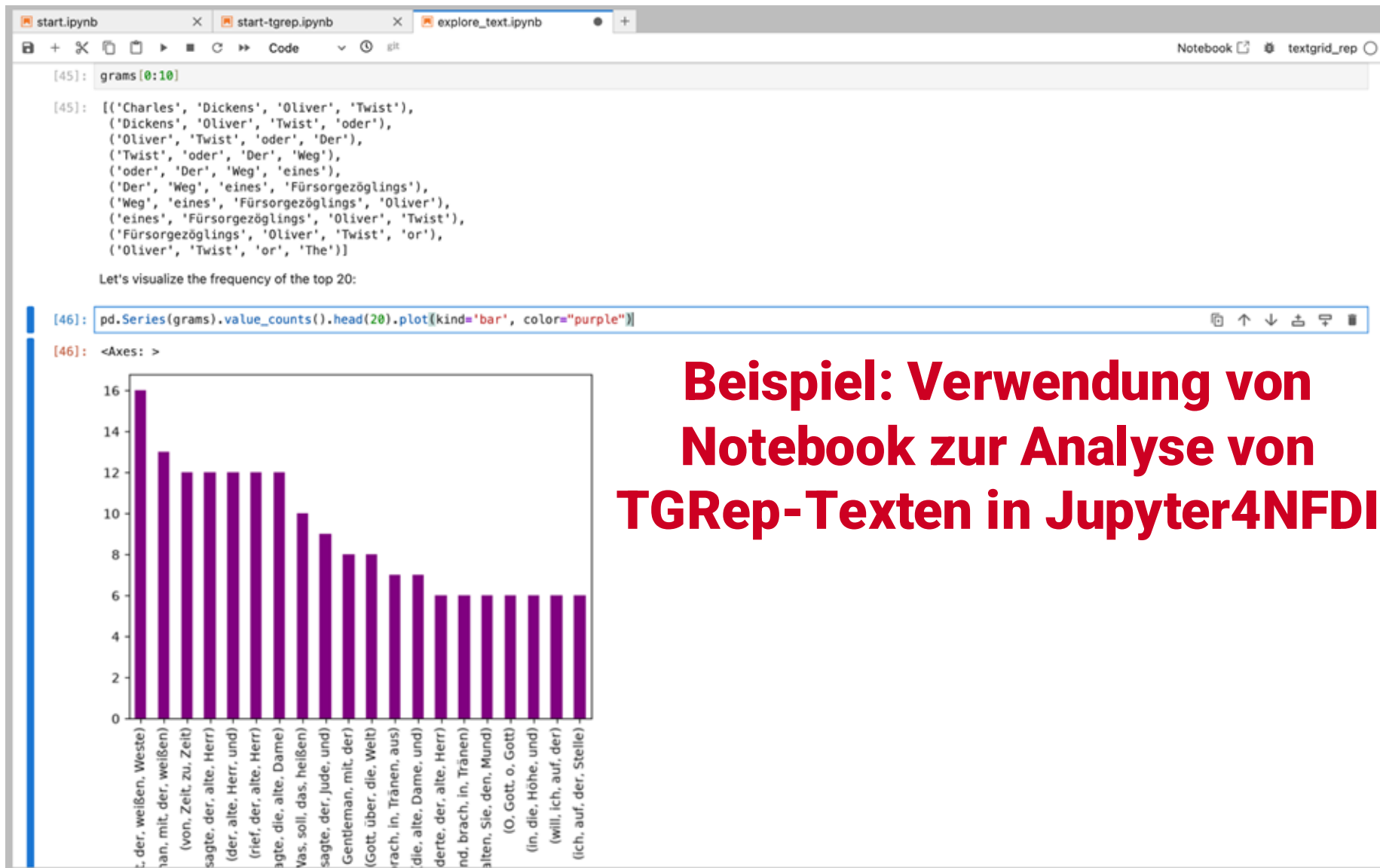
Link öffnet Übersichtsnotebook in Jupyter4NFDI und übermittelt die URL des Texts an den Jupyter Server

Unter andern öffentlichen Gebäuden in einer gewissen Stadt, die ich nicht nennen, der ich aber auch andererseits keinen erdichteten Namen beilegen möchte, befand sich eines, wie es wohl die meisten Städte, ob groß oder klein, besitzen, nämlich ein Arbeitshaus; und in diesem wurde eines Tages der kleine Weltbürger geboren, dessen Name dieses Buch trägt.

Lange Zeit, nachdem der Arzt des Kirchspiels ihm zum Eintritt in diese Welt der Mühen und Sorgen geholfen, schien es recht zweifelhaft, ob er lange genug würde am Leben bleiben, um überhaupt einen Namen nötig zu haben.

<https://jupyter.textgrid.d.sub.uni-goettingen.de/open/textgrid:mgq9.0>

TGRep-Jupyter4NFDI-Integration (in Entwicklung)



Beispiel: Verwendung von Notebook zur Analyse von TGRep-Texten in Jupyter4NFDI

TGRep-Jupyter4NFDI-Integration (in Entwicklung)

```
[5]: # Show first 20 detected entities
print("Standard spaCy NER - First 20 entities:")
print("=" * 60)
for ent in doc.ents[0:20]:
    print(f"{ent.text:30} {ent.label_:10}")

Standard spaCy NER - First 20 entities:
=====
Charles Dickens          PER
Oliver Twist             PER
Oliver Twist             PER
The Parish Boy's Progress MISC
Schildert                PER
Lange Zeit               MISC
Oliver Twist             PER
Oliver                   PER
Oliver                   PER
Oliver                   PER
Der Arzt                 MISC
g'habt                   PER
Gott o Gott             PER
Mrs                      PER
Thingummy                PER
niedergebeugt           PER
Das arme Kleine          MISC
nacht hergeschafft       ORG
Herrn Vorstands          ORG

[6]: # Render the visualization for entities
displacy.render(doc, style="ent", jupyter=True)

Charles Dickens PER
Oliver Twist PER

oder

Der Weg eines Fürsorgezöglings

( Oliver Twist PER , or, The Parish Boy's Progress MISC )

Erstes Kapitel

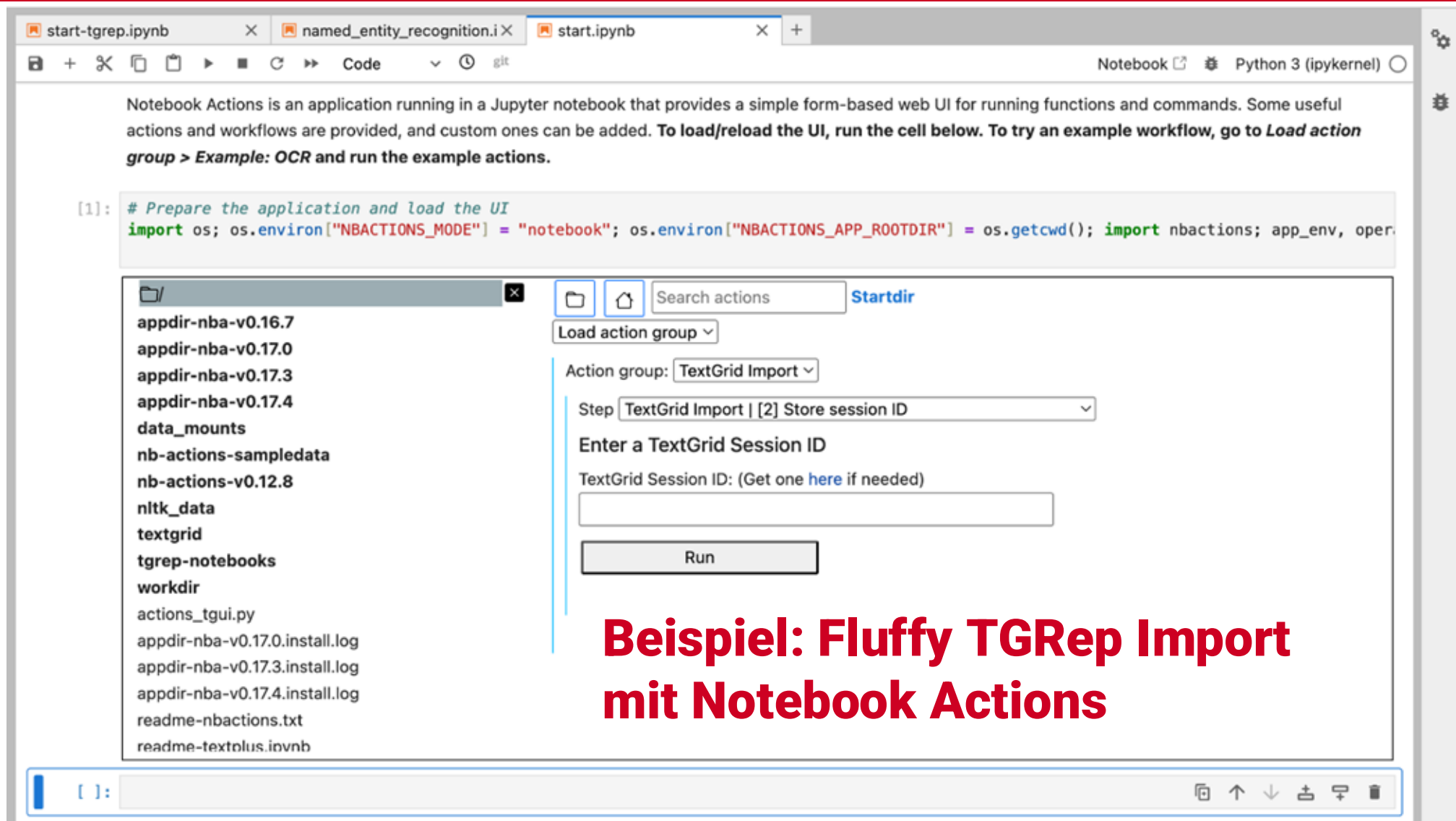
Erstes Kapitel

Schildert PER den Ort, wo Oliver auf die Welt kam, sowie die seine Geburt begleitenden Umstände.

Unter andern öffentlichen Gebäuden in einer gewissen Stadt, die ich nicht nennen, der ich aber auch andererseits keinen erdichteten Namen beilegen möchte, befand sich eines, wie es wohl die meisten Städte,
```

Beispiel: NER für Text aus TGRep mit MONAPipe in Jupyter4NFDI

Tools zur sofortigen Verwendung in Jupyter4NFDI



The screenshot shows a Jupyter Notebook with three tabs: 'start-tgrep.ipynb', 'named_entity_recognition.i...', and 'start.ipynb'. The active tab is 'start.ipynb', which contains a code cell with the following text:

```
[1]: # Prepare the application and load the UI
import os; os.environ["NB_ACTIONS_MODE"] = "notebook"; os.environ["NB_ACTIONS_APP_ROOTDIR"] = os.getcwd(); import nbactions; app_env, oper...
```

Below the code cell is a web-based UI for running actions. It features a file browser on the left with a list of folders and files, including 'appdir-nba-v0.16.7', 'appdir-nba-v0.17.0', 'appdir-nba-v0.17.3', 'appdir-nba-v0.17.4', 'data_mounts', 'nb-actions-sampleddata', 'nb-actions-v0.12.8', 'nltk_data', 'textgrid', 'tgrep-notebooks', 'workdir', 'actions_tgui.py', and several install logs. The main area of the UI has a search bar, a 'Startdir' button, a 'Load action group' dropdown, and an 'Action group' dropdown set to 'TextGrid Import'. Below this is a 'Step' dropdown set to 'TextGrid Import | [2] Store session ID', a text input field for 'Enter a TextGrid Session ID', and a 'Run' button.

Beispiel: Fluffy TGRep Import mit Notebook Actions

Tools zur sofortigen Verwendung in Jupyter4NFDI

The screenshot displays a Jupyter Notebook environment. On the left, a file explorer shows the directory structure of a project named 'appdir-nba-v0.17.4'. The main area contains a code cell with the following output:

```
[Example: OCR | OCR With Tesseract] Processing source files ...  
100%|██████████| 135/135 [02:14<00:00, 1.00it/s]  
/workdir/examples/example_OCR  
main-000.hocr  
main-000.jpg  
main-001.hocr  
main-001.jpg  
main-002.hocr  
main-002.jpg  
main-003.hocr  
main-003.jpg  
main-004.hocr  
main-004.jpg  
main-005.hocr  
main-005.jpg  
main-006.hocr  
main-006.jpg  
main-007.hocr  
main-007.jpg  
main-008.hocr
```

Below the code cell, a success message is displayed: "[Example: OCR | OCR With Tesseract] Success/Errors: 67/68 files. Result folder: workdir/examples/example_OCR".

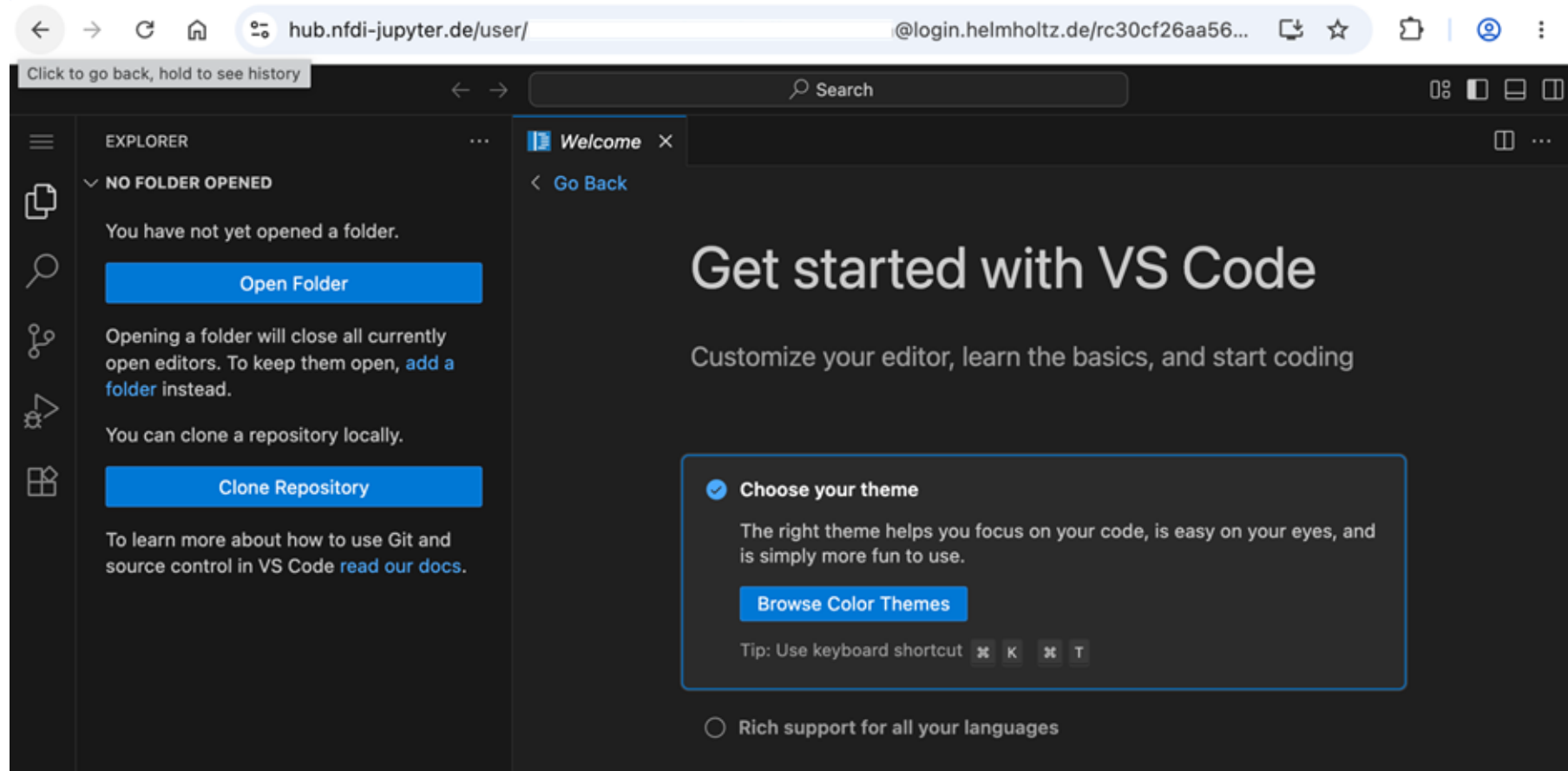
On the right side, an Action panel is visible, showing the configuration for the 'Example: OCR' action group. The configuration includes:

- Action group: Example: OCR
- Item: Example: OCR | OCR With Tesseract
- Source file or folder (multiple values are possible): {root_dir}/workdir/examples/example_OCR
- Destination folder: (empty)
- format: HOCR
- Language (multiple values are possible): eng
- First remove the contents of destination folder:
- For PDF and text: merge result files into one output file:

Buttons for 'Run' and 'Generate command' are located at the bottom of the Action panel.

**Beispiel: OCR (Tesseract)
mit Notebook Actions**

Jupyter Notebooks und Jupyter4NFDI



Beispiel: Visual Studio Code in Jupyter4NFDI

Jupyter4NFDI-Link:

<https://hub.nfdi-jupyter.de/share/jDMq5oVJu0E>

Repo: [kreuzert/jupyter-vscode-example](#)

Jupyter Notebooks und Jupyter4NFDI

- Authentifizierung und Autorisierung schon dabei
- Editoren, Viewer, Upload/Download bereits vorhanden
- Maintenance des Dienstes sichergestellt
- Nicht nur Notebooks sind möglich, sondern auch:
 - (fast) beliebige Software Stacks dank Repo2Docker- & Docker-Image-Feature
 - Web-Anwendungen können auf einem Jupyter-Server laufen

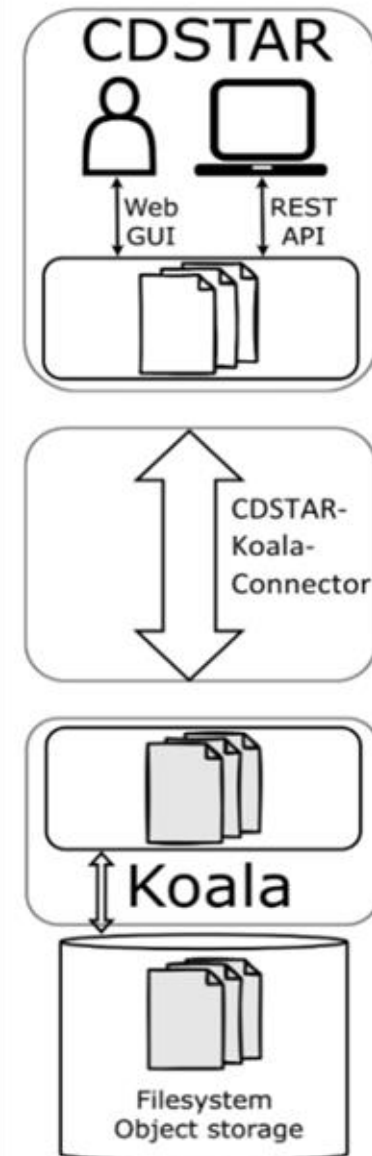
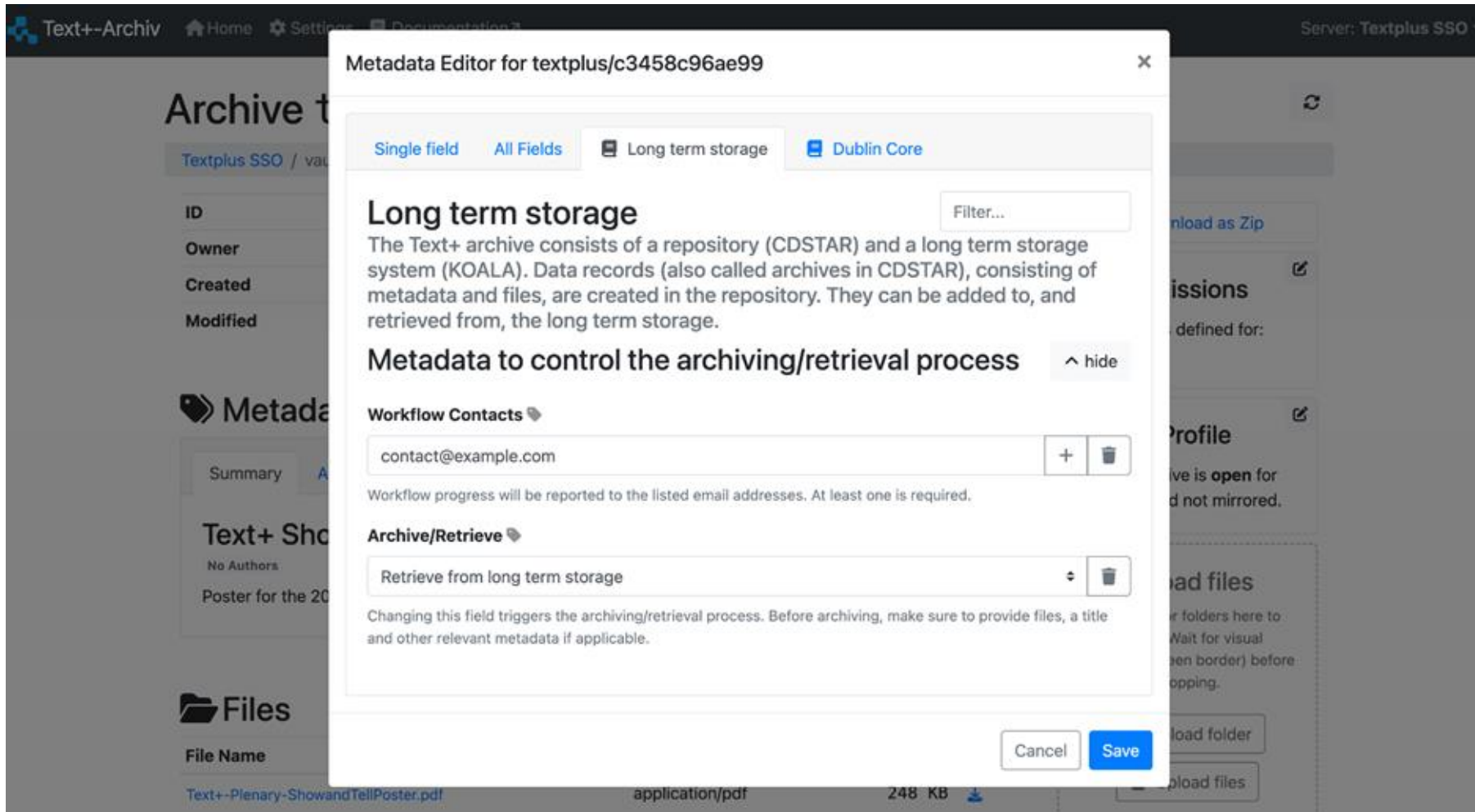


Langzeitarchivierung: Das Text+-Archiv

Langzeitarchivierung in Text+

- An den zahlreichen Datenzentren
 - auf Thema oder/und Formate spezialisierte Lösungen
- Im **Text+-Archiv** (demnächst)
 - Generische Fallback-Lösung ohne Einschränkung des Datenformats
 - Bietet Bitstream Preservation
 - Kreis der Nutzenden flexibel konfigurierbar über die AcademicCloud-AAI (zunächst Text+-Zugehörige)
 - Basiert auf der Langzeitarchivierungslösung KOALA
 - und auf der Repository-Lösung CDSTAR
 - Data Upload/Download, konfigurierbare Metadateneingabe, PIDs, flexible Rechteverwaltung, Suche
 - Datensätze können frei oder nur für bestimmte Nutzende/Gruppen zugänglich sein
 - So können auch Daten ein Zuhause bekommen, die nicht ohne Weiteres veröffentlicht werden können (Projekt abgelaufen, legal issues etc.)

Langzeitarchivierung: Das Text+-Archiv



Text+-Archiv: Metadaten-Editor (links) und Architektur (rechts)

Adressierung der Bedarfe in den User Stories

- Tools und Infrastruktur für Datenverarbeitung und -publikation
- Für TextGrid Rep und allgemein
- **Generische Lösung für Langzeitarchivierung für offene sowie geschützte Daten**
- Bedeutender Teil der Bedarfe in den Text+ User Stories wird direkt oder indirekt adressiert, s. auch gitlab.gwdg.de/textplus/process-textplus-userstories/-/tree/main/analysis-2026?ref_type=heads)

Adressierung der Bedarfe in den User Stories

Very Short Coverage Table

Measure set	Substantial	Partial	Indirect/minor	None	Substantial + partial	Including minor
Long-term preservation / Textplus Archive	33	52	11	17	85 / 113	96 / 113
Text/data processing tools	13	41	31	28	54 / 113	85 / 113
Either measure, best fit	42	56	8	7	98 / 113	106 / 113

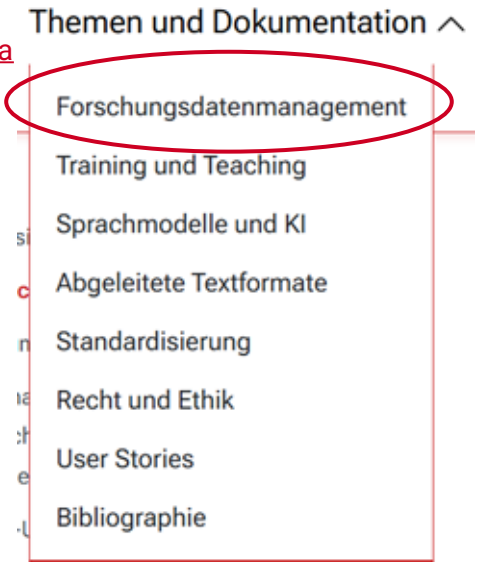
05

Interoperability & Re- Usability

Melina Jander, Barbara Fischer

Research Data Management

<https://text-plus.org/themen-dokumentation/forschungsdatenmanagement/>



Information Page on Research Data Management

- Research data in Text+
- General information on RDM & additional links
- RDM support provided by Text+, e.g.:
 - Research Data Management Organiser

RDMO: Text+-Specific Questionnaire with questions tailored to our community and concrete application examples

- Provided by the eRA Göttingen
- Access via Academic ID
- Closely linked to the consulting services and Helpdesk



Quelle: <https://rdmorganiser.github.io/>

Research Data Management – Incubator Project with DMP4NFDI

Objectives:

- Adaptations to the Text+ own RDMO catalogue so that interoperability with the NFDI-wide DMP template is established
- Co-development of a consortium-wide service offering
- Promotion of the service through information events and blog post
- Survey on needs related to DMPs and RDMO (running until 30 June 2026)



The NFDI DMP Template Framework

Version: 1.0.0



Research Data Management – Inkubatorprojekt mit DMP4NFDI

Survey about DMPs and RDMO (you're invited to partake until 30.06.26)



The poster features a red circular logo with the text "Text+" on the left. To its right is a red icon of four hands holding gears. Further right is a large black and white QR code. Below these elements, the text reads: "Das Leben ist kein Wunschkonzert? FDM in Text+ schon! Macht mit bei unserer Umfrage zu DMPs und RDMO in den Geisteswissenschaften!". A red arrow curves upwards from the bottom right towards the QR code. The background is light blue with a hexagonal grid pattern.

Text+

Das Leben ist kein Wunschkonzert?
FDM in Text+ schon!

Macht mit bei unserer Umfrage zu DMPs
und RDMO in den
Geisteswissenschaften!

The Text+ Agency as enabler of our data space

01

Service Portfolio of the Agency

- project consultation services
- authority data training and orientation
- data analysis of data deliveries
- GND reconciliation
- data preparation and editing
- data conversion
- data import into the GND

02

entityXML & Workflow for GND-compliant Enrichment of Research Data

- XML-based data format for entity descriptions
- exchange an storage format for data providers and agency
- allows project-specific applications through custom namespaces



03

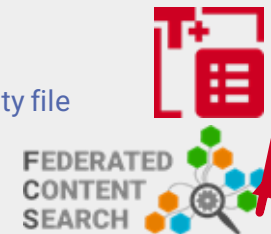
User Groups

- beginning in 2022 with 3 digital editions from the University of Göttingen
- 5 Text+ internal projects
- 4 research projects & GND users at large

04

Search & Retrieval Solutions

- Text+ Registry (able to contribute to graphs)
- Federated Content Search (full texts)
- both services with enhanced user experience due to authority file integration
- both with open interfaces to other infrastructures and services beyond Text+



The Text+ Data Space...

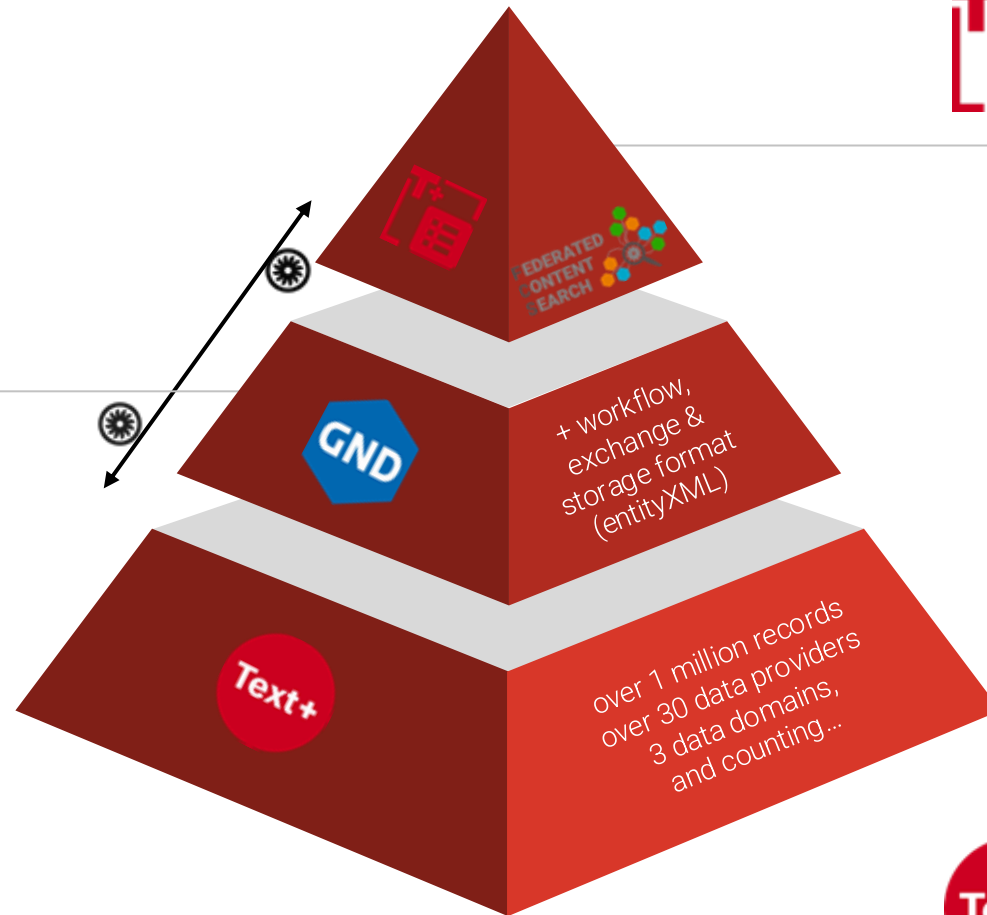


GND - Gemeinsame Normdatei / GND-Agentur Text+ (service)

GND as connecting, enriching infrastructure layer for the Text+ Data Space. Recurring element in the UIs of our search applications and an (upcoming) requirement for research data joining the data space.

The Integrated Authority File aka GND provides 10 million entities spanning people, organisations, geographic locations, events, and subjects. Increasing daily with about 700 new items.

Plus providing a transnational integrating governance infrastructure, enhancing community building through events, meetups and concise workshops to enable the Text+ community to become an active member in the GND network with its committees and working groups.



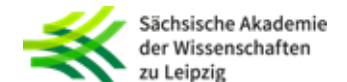
Text+ Registry & FCS - Federated Content Search (applications)

1

The Text+ Registry is a central cross-domain discovery tool for structuring, linking and presenting resources such as editions, collections, lexical resources, services, and entities (Text+ Centres). This is a continuation of DARIAH activities reaching back to 2011.

The FCS is a discovery solution for linguistic resources in a federated research environment with currently over 500 resources (440 for annotated full texts) and nearly 40 endpoints from 16 institutions. FCS has a CLARIN background.

Both applications are easily accessible and documented through/in our portal.



3

Text+ Data Space (research data)

Growing linked and enriched collection of research data accessible via the search & retrieval tools FCS and Registry. Data providers fill in metadata schema to register their resources using the GND as controlled vocabulary. Phase I of Text+ focussed on the infrastructural set up of this pyramid.

Phase II will focus on upscaling the data space itself.

... and its Search & Retrieval Services

06

Communication, Dissemination, Outreach

Alexander Steckel, Melina Jander

Data, Services & Information for the Community

Text+

DE EN

Daten und Dienste ▾ Themen und Dokumentation ▾ Vernetzung ▾ Über uns ▾ Aktuelles ▾

Suche

Helpdesk

Infrastruktur für text- und sprachbasierte Forschungsdaten

Text+ stellt im Rahmen der NFDI einen umfangreichen Bestand an Textsammlungen und Korpora, Editionen und lexikalischen Ressourcen zur Nachnutzung bereit. Zum Angebotsportfolio gehört neben eigenen Daten und Diensten die Möglichkeit, Forschungsdaten der Community aufzunehmen, findbar zu machen und langfristig zu bewahren.

Daten suchen

Daten geben

Dienste

Beratung

Schaufenster

The screenshot shows the 'Registry - Der Text+ Katalog' website. It features a navigation bar with 'Registry' and several 'Federated ...' links. Below the navigation, there's a search bar and a list of resources. The first resource is 'Fachwissenschaftskorpus Collection' with a description in German. The second resource is 'OrbiNews Release 2.0' with a description in English. The interface includes a 'Filter' sidebar on the left and a 'Suche' button on the right.

Text+ Blog

Edit empfiehlt #4: Corpus Musicae Ottomanicae (Ressourcen-Reigen Spezial)

Sven Gronemeyer

2. April 2026 11:18



Als Hauptstadt des Osmanischen Reiches war Istanbul nicht nur das Machtzentrum der Hohen Pforte, sondern auch der kulturelle Schmelztiegel eines Vielvölkerstaats. Mit seiner Ausdehnung war das Osmanische Reich seit jeher ein Transitraum.

zum Blog

aktuelle Veranstaltungen

Di.
09.
Juni
2026

09:00–10:00
Text+ Research
Rendezvous
Virtuell

Data, Services & Information for the Community

Text+

EN Support | Helpdesk

Über uns - Forschungsdaten - Vernetzung - Links/Downloads - Aktuelles/Veranstaltungen -

PLENARY 2022 #Textplusplenary

MONTAG, 12. SEPTEMBER 11 bis 18 Uhr + Abendempfang

DIENSTAG, 13. SEPTEMBER 9 bis 16 Uhr, Internat. Arbeitsschichten

SCHLOSS MANNHEIM

Am 12. und 13. September 2022 findet das Text+ Plenary in Mannheim statt. [Weitere Informationen](#)

Der Verbund Text+ wird text- und sprachbasierte Forschungsdaten langfristig erhalten und ihre breite Nutzung in der Wissenschaft ermöglichen.

Die Gemeinsame Wissenschaftskonferenz hat Text+ als Konsortium der bundesweiten Initiative zum Aufbau einer nationalen Forschungsdateninfrastruktur (NFDI) bewilligt. Text+ startete offiziell im Herbst 2021 nach mehrjähriger Vorbereitungszeit und wird zunächst für fünf Jahre durch die Deutsche Forschungsgemeinschaft (DFG) gefördert.

Die Text+ Infrastruktur ist auf Sprach- und Textdaten ausgerichtet und konzentriert sich zunächst auf digitale Sammlungen, lexikalische Ressourcen und Editionen. Diese sind von hoher Relevanz für alle sprach- und textbasierten Disziplinen, speziell für Sprachwissenschaftler, Literaturwissenschaftler, Philologen, Klassische Philologen, Archäologen, außerräumliche Kulturen und Sprachen sowie sprach- und textbasierte Forschung der Sozial-, Wirtschafts-, Politik- und Geschichtswissenschaften.



Text+

Daten und Dienste - Themen und Dokumentation - Vernetzung - Über uns - Aktuelles -

Suchen - Hilfe/Link

Infrastruktur für text- und sprachbasierte Forschungsdaten

Text+ stellt im Rahmen der NFDI einen umfangreichen Bestand an Textsammlungen und Korpora, Editionen und lexikalischen Ressourcen zur Nachnutzung bereit. Zum Angebotsportfolio gehört neben eigenen Daten und Diensten die Möglichkeit, Forschungsdaten der Community aufzunehmen, findbar zu machen und langfristig zu bewahren.

Daten suchen - Daten geben - Dienste - Beratung

Registry

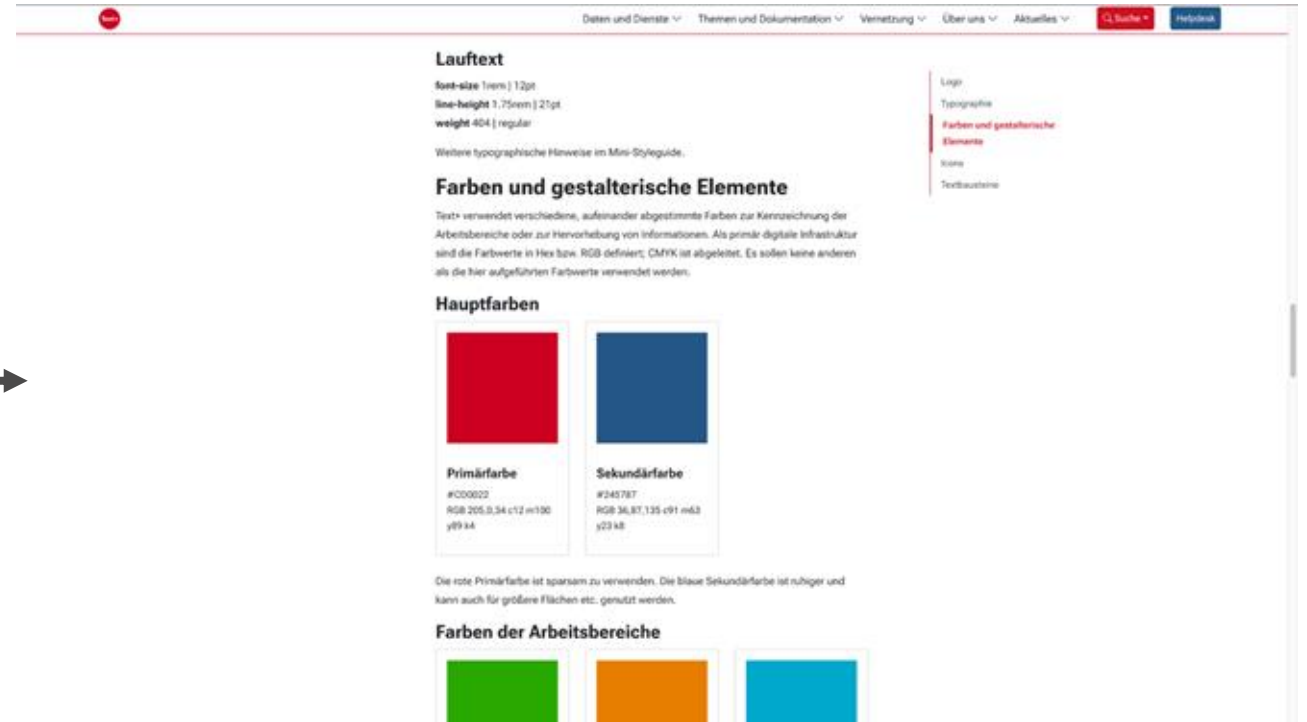
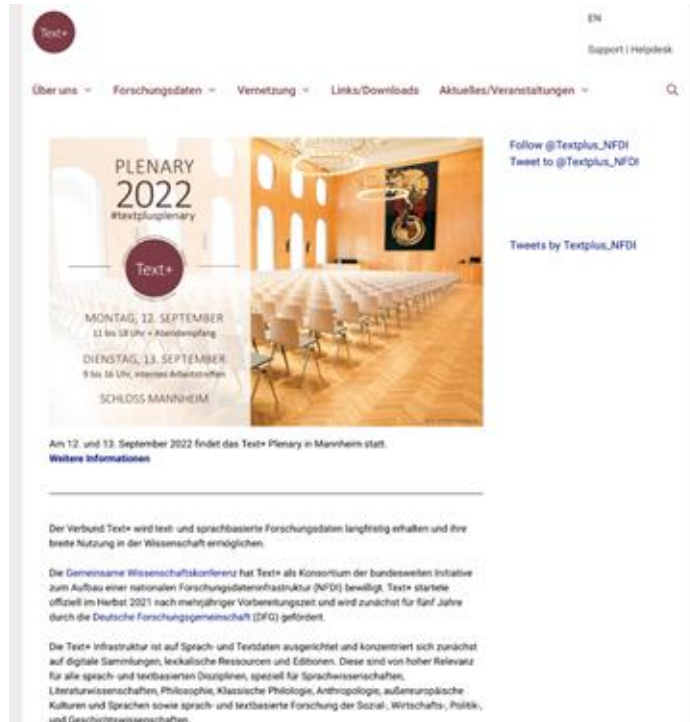
Registry - Der Text+ Katalog

aktuelle Veranstaltungen

09. Juni 2026 Text+ Research Rendezvous

- Seite über Konsortium (Antragsphase) umgewandelt in **Seite über Inhalte**
- aus Eigendarstellung wurde ein **Portal für Forschende**

Branding



optisch: Re-branding der Gesamtmarke Text+

- schlankes, modernes Auftreten
- **textlastig**, aber keine Bleiwüste
- klarer Fokus auf Inhalte
- **barrierearm**

Content

- Portalstruktur ausgerichtet nach User-Sicht (~~nicht: Konsortialsicht~~)
- Was kann ich hier tun?
- Wofür kann ich das gebrauchen?
- viele **Calls-to-action**
- Übersichtsseiten wie T+-Architektur
- komplett **zweisprachig** DE / EN

Daten und Dienste ▾ Themen und Dokumentation ▾ Vernetzung ▾ Über uns ▾ Aktuelles ▾

Suche Helpdesk

Infrastruktur für text- und sprachbasierte Forschungsdaten

Text+ stellt im Rahmen der NFDI einen umfangreichen Bestand an Textsammlungen und Korpora, Editionen und lexikalischen Ressourcen zur Nachnutzung bereit. Zum Angebotsportfolio gehört neben eigenen Daten und Diensten die Möglichkeit, Forschungsdaten der Community aufzunehmen, findbar zu machen und langfristig zu bewahren.

Daten suchen Daten geben Dienste Beratung

Schaufenster

Registry

Registry - Der Text+ Katalog

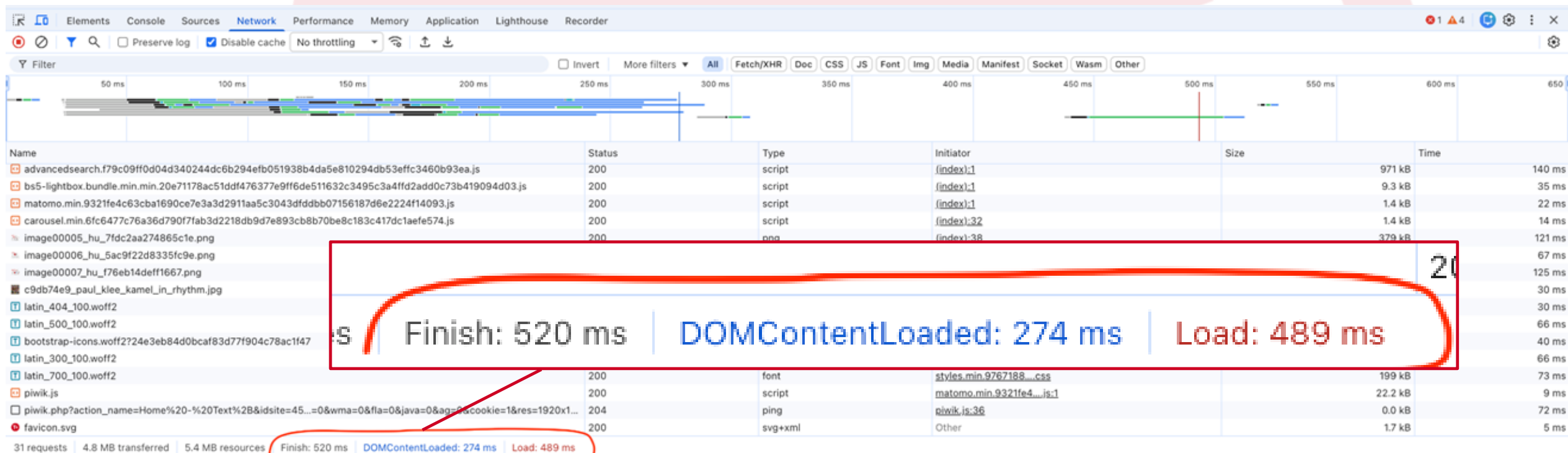
Filter: Importquelle, Sprache(n)

aktuelle Veranstaltungen

Di.	09:00-10:00
09.	Text+ Research
Juni	Rendezvous
2026	Virtuell

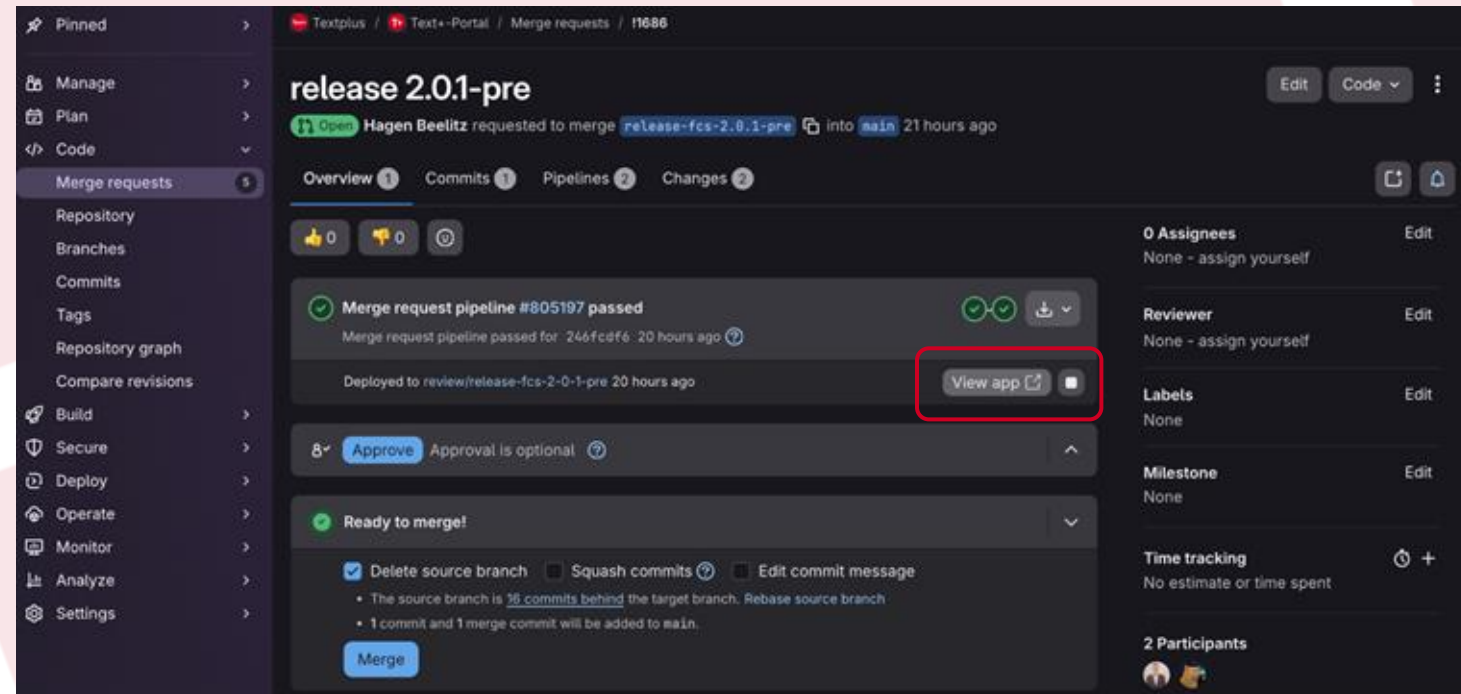
Best Practices I

- Hugo: **Static-Site-Generator**. Macht aus dem Content in Kombi mit dem Theme eine funktionierende Webseite
- Python: aktualisiert die externen Daten, die die Webseite einbindet = **keine Datenbank @runtime**
- **performant**
- plattformunabhängig (desktop first, mobile second)
- **zuverlässig**



Best Practices II

- Versionierung und Sicherheit (**GitLab**)
- **ortsverteilter Zugriff** im föderalen Konsortium
- multiple Deployments von Branches in je eigener **Vorschau-Umgebung** (10+x Branches mit eigenen ausgelieferten Seiten) zur Entwicklung



Best Practices III

- Inhalte
 - [CC-Lizenz](#)
- Quellcode:
 - [Public Repo](#)
 - [EUPL v1.2](#)



Fr.
12.
Juni
2026

11:00 – 12:00
The CORAL Platform: LLMs in Service of Oral History
Virtuell

Mo.–Di.
15.–16.
Juni
2026

5. Text+ Community Plenary 2026 in Mannheim
Schloss Mannheim

weitere Veranstaltungen

[Impressum](#) [Datenschutzerklärung](#) [Erklärung zur Barrierefreiheit](#) [Kontakt und Beratung](#)

Alle nicht anders gekennzeichneten Inhalte [CC BY-SA 4.0](#) [Der Quellcode dieser Seite](#) wird bereitgestellt unter [EUPL v1.2](#)

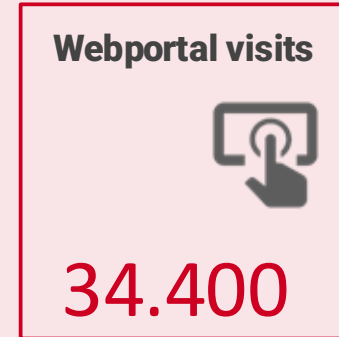
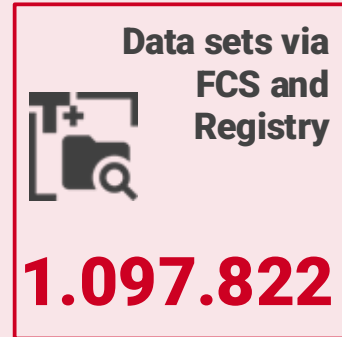
Text+ wird gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – 460033370 | Text+ ist Teil der Nationalen Forschungsdateninfrastruktur (NFDI)

Community Outreach – Events

Services for the community
interner Outreach: Plenary 25 mit DAE als Beispiel



10 contributes to Text+ by




Community activities

IO contributes to Text+ by



Text+ Centres
... and counting

33



Data sets via
FCS and
Registry

1.097.822



758

Helpdesk &
consultation
requests solved



Events
... and counting

286

Spotlight: Text+ IO Lectures



Die Text+ Task Area Infrastructure/Operations bietet Schulungen/Vorträge zu verschiedenen Themen an. Allen Text+ IO-Lectures ist gemein, dass sie entweder relevant für die Arbeit in den Text+ Datendomänen sind oder von Interesse für die Text+ Infrastruktur insgesamt.

Die Zielgruppe der IO-Lectures sind in erster Linie die Text+ Datendomänen, also Editionen, Lexikalische Ressourcen sowie Sammlungen. Darüber hinaus dürfen sich aber auch Kolleginnen und Kollegen aus anderen Kontexten, insbesondere aber anderen NFDI-Konsortien eingeladen fühlen.

Die Teilnahme an den Text+ IO-Lectures ist kostenlos, es ist lediglich eine Anmeldung mit wenigen persönlichen Informationen notwendig.

Für Fragen, Hinweise oder Wünsche für eine bestimmte IO-Lecture bitte an das Text+ Operations Office schreiben: textplus-operations-office@lists.gwdg.de

the project
bibliography

Started in March 2023
12 lectures until Feb 2025

community activities


Community Outreach – Spotlight

Event Series “Show and Tell: Tools, Services, Projects” – A Place for Exchange on the Diversity of Topics within the NFDI

- Started in January 2025
- More than 20 events with different guests
- Topics:
 - Tools, Standards & Technologies
 - Updates from Text+
 - NFDI Core Services
 - Insights into Other Consortia



Work Results Associated with IO



Artwork

4


Blog Posts



72

14

Book & Book Section




Computer Program

4




14

Conference Paper



Dataset

3



Document

2


Journal Article



18


2

Preprint



Presentation

63



12

Report



Standard

3

How IO enters Phase 2



T+ DATA IO-M1

Federated Research Data Infrastructure

- Federated Content Search
- Registry
- Data integration with data providers
- Data enrichment in domains / curation with data providers

T+ PROCESSES IO-M2

Service Integration w/ Data Domains

- (users) UX/RE of services offered by T+
- (infrastructure) of services comprising the T+ service portfolio

T+ PLATFORM IO-M3

Platform and Core Services

- Portal (maintenance, deployment pipeline)
- base4nfdi Base4NFDI (adaption of basic services in T+)
- LLM as basic service
- Provision of data centres resources for data domains
- IT service management (ITIL, monitoring)
- Project cloud services (collaboration, IAM)

T+ ECOSYSTEM IO-M4

Coherence

- Policy framework (quality assurance, software development, inbound)
- Interfacing (EOSC, Humanities@NFDI, ERICs, NFDI, outbound)
- Indicators / impact alignment with data domains and admin
- Community activities (helpdesk, workshops, dissemination)

Outlook

- Text+ aims for a further integrated, navigable and growing data space.
- Text+ Registry (metadata) and Text+ FCS (full texts and structured data) enable users to discover our data space.
- The data space currently consists of more than 30 institutions (Text+ Centers) with well over 1 million data sets.
- We promote the use of and enrichment with GND identifiers as well as other authority files.
- Upscaling of the Text+ Data Space as major effort until the end of funding phase 1 and ongoing in phase 2 (start 10/2026).

07

Fragen & Diskussion

Philipp Wieder

Text

Text+

Plenary
2026



@textplus_NFDI
#TextPlusPlenary

5. TEXT+ PLENARY

PROGRAMM

Montag, 15. Juni 2026

12:00 – 13:30	<i>Lunch (Katakomben)</i>
13:30 – 14:00	Begrüßung, Eröffnung und Einführung in das Programm (Aula)
14:00 – 15:30	Ergebnisse aus Task Area: Infrastructure & Operations (Aula)
15:30 – 16:00	<i>Kaffeepause (Katakomben)</i>
16:00 – 17:30	Ergebnisse aus Task Area: Editions (Aula)
17:30 – 18:00	Posterslam (Aula)
18:00 – 19:00	Postersession (Fuchs-Petroluh-Saal)



Editionen

User Demands & Community Activities

Die Community auf der Couch:

FDM-Beratung im Bereich Editionen

Editionen finden & vernetzen:

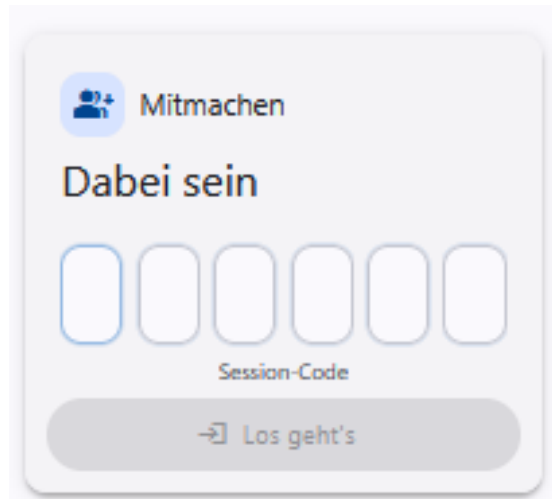
Die Text+ Registry als Brücke im Informationsökosystem

Diskussionspanel:

Was braucht die Editionscommunity? - Perspektiven in Text+ und NFDI 2.0



<https://arsnova.eu>



Mitmachen

Dabei sein

Session-Code

Los geht's

Code: V8JP94





Editionen

User Demands & Community Activities

Philipp Hegel (AdWL | Mainz) und Martin Fechner (BBAW)

Die Community auf der Couch:
FDM-Beratung im Bereich Editionen

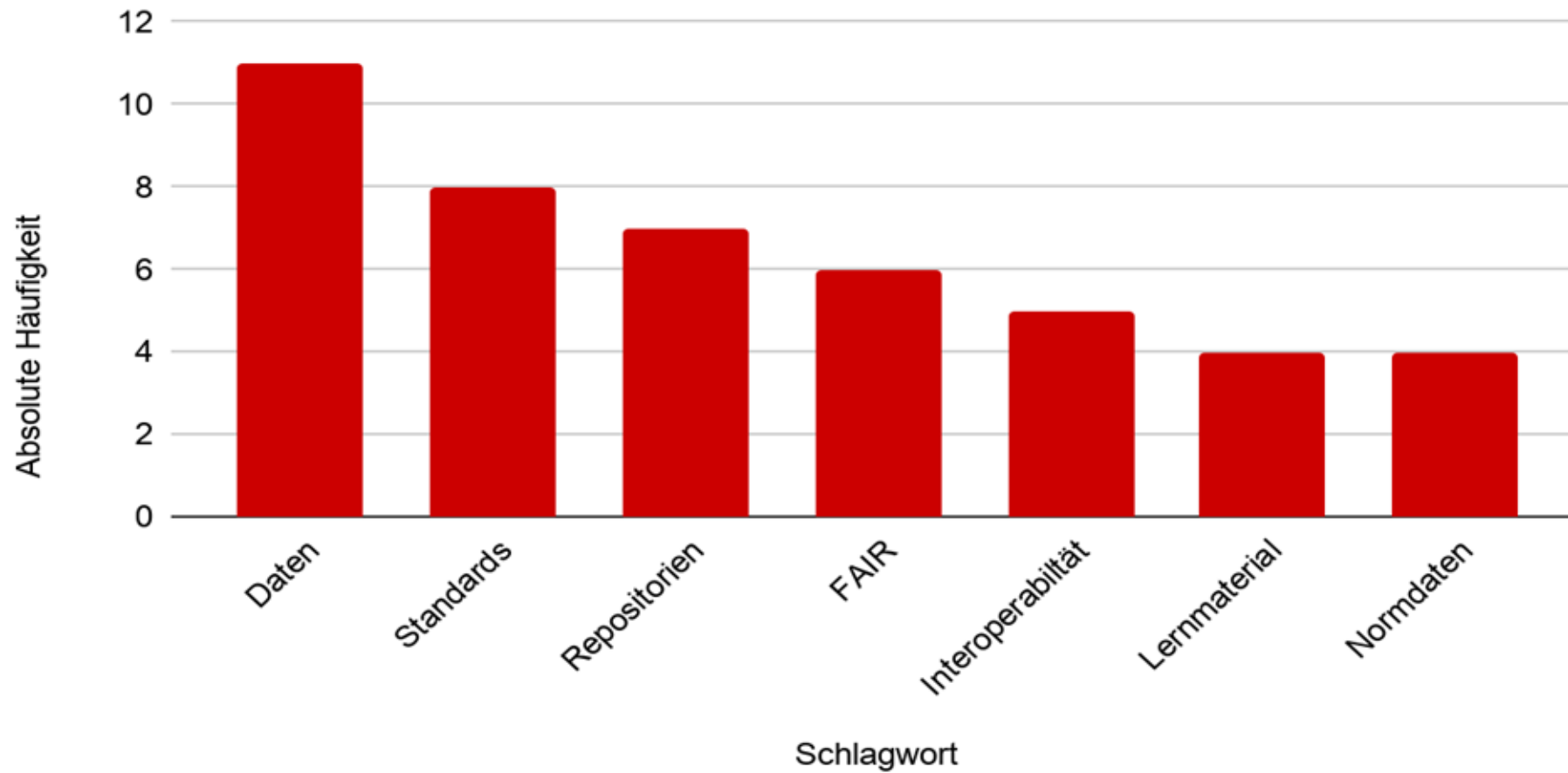
Editionen finden & vernetzen:
**Die Text+ Registry als Brücke
im Informationsökosystem**

Diskussionspanel:
**“Was braucht die Editionscommunity?”
Perspektiven in Text+ und NFDI 2.0**



Community Demands

>3 bei n=24



Datenpublikation und -qualität

I consider an **evaluation and quality assurance** of my data desirable (e.g. by scientific advisory boards) and the transfer to **repositories** to make the data, which is generated outside of large research networks, more visible and secure in the long term.



- Registry
- Repositories



Standardisierung

Due to the heterogeneous initial situation, the requirements are manifold, different workflows must be addressed in order to obtain a coherent result. **Standard workflows or tools for a conversion of Word data into TEI/XML data** would be enormously helpful. The tools should be freely usable.



- Standardisierung
- Software Stacks
- Tool Registry



Community Activities

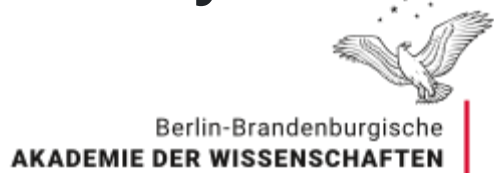
It would be helpful and motivating if I could also **receive advice and technical support** from the NFDI Text+ consortium as an independent researcher, e.g. **referral to local or regional experts, institutions, mentors, training courses.**



- Consulting, Expertise
- Training, Workshops
- Tutorialsammlung
- Curriculare Empfehlungen



Community – Workshops, Veranstaltungen & Kooperationen



DH-Kolloquium



Text+

ediarum

.MEETUP



InFoDiText+

FAIR-February



NFDI4Memory



NFDI4Objects

ediorom Summer Schools



DHd Workshop



Community – Inhalte, Formate & Curriculare Arbeit

- FAIR-Prinzipien
- Forschungsdatenmanagement (FDM)
- Datenqualität
- Metadaten
- Einbindung externer Daten
- Transkription
- Interoperabilität
- APIs, XML, XSLT
- LLMs und KI



FAIR February 2026: Veranstaltungsreihe



- Findbarmachung
- Interoperabilität
- offene Lizenzen
- Dokumentation für Nachnutzung
- Usability
- Barrierefreiheit



Community – Tutorials & Veröffentlichungen

Curriculare Empfehlungen

- Wolff, F. (2025, December 4). Prompt Engineering und Digitale Editionen. Zenodo. <https://doi.org/10.5281/zenodo.17820796> ↗
- Hegel, Ph. (2026, May 22). Transkribieren: Mehr als nur abschreiben? Zenodo. <https://doi.org/10.5281/zenodo.20341522> ↗

Tutorialsammlung

- Die Task Area Editions in Text+ bietet mit dem Tutorial-Verzeichnis eine Sammlung von Lernmaterialien und Tutorials, die relevant sind für die Arbeit mit digitalen Editionen.
- [Link zur Tutorialsammlung](#) ↗



Beiträge

- Im [Text+ Blog](#) ↗ dokumentieren wir Highlights aus unserem Veranstaltungsprogramm.
- [Handreichung zu digitalen Editionen](#) ↗
- [Veröffentlichungen, Ankündigungen und mehr](#) ↗





Editionen

User Demands & Community Activities

Die Community auf der Couch:
**FDM-Beratung im Bereich
Editionen**

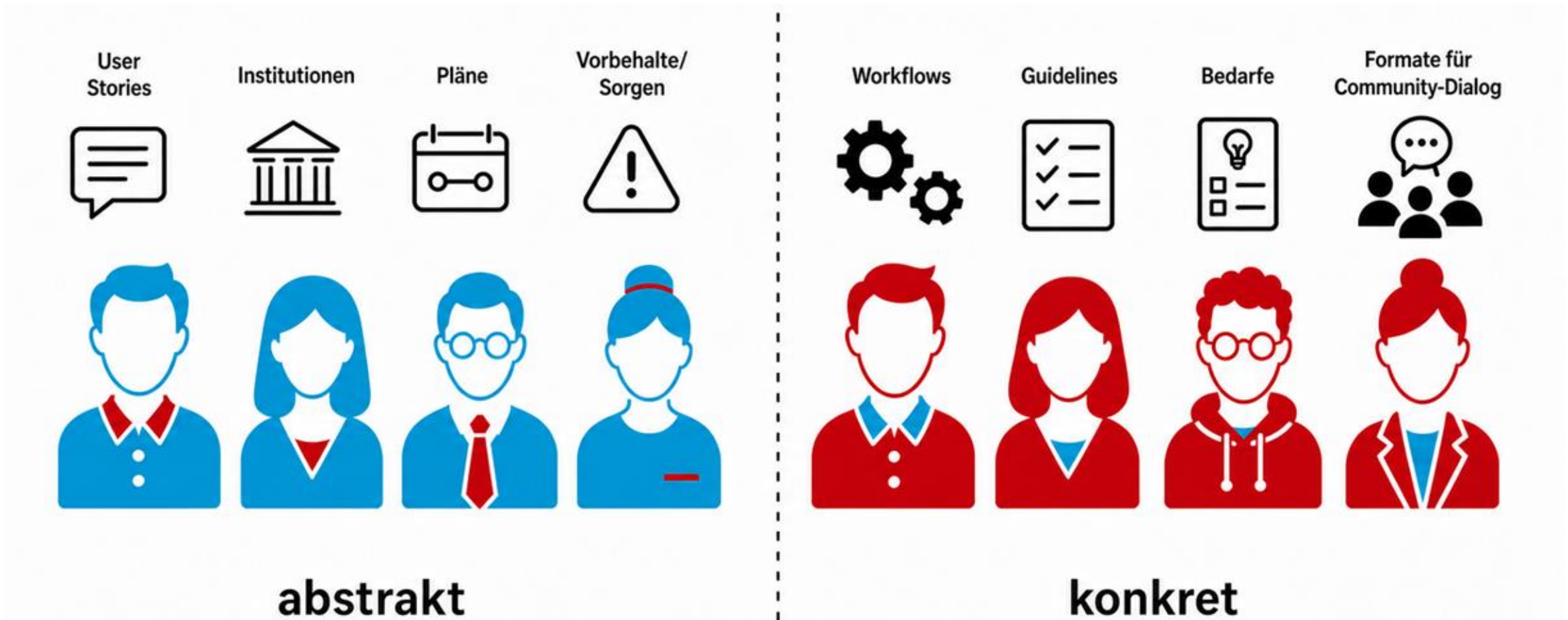
Jonathan Blumtritt (AWK NRW) und
Martin Sievers (AdWL | Mainz)

Editionen finden & vernetzen:
**Die Text+ Registry als Brücke
im Informationsökosystem**

Diskussionspanel:
“Was braucht die Editionscommunity?”
Perspektiven in Text+ und NFDI 2.0



Der Gesamtprozess



Eigene Darstellung (mit ChatGPT erstellt)



Die Entwicklungsphasen des Consulting in der TA Editions

Aufbau

- Begriffsdefinition „Consulting“
- Entwicklung von Workflows: technisch, organisatorisch und sozial
- Protokollvorlage

Entwicklung

- Research Rendezvous
- Erwartungsmanagement
- Agent:innen-Treffen (Text+ Helpdesk)
- Konsortienübergreifende Beratungen
- Mitarbeit in NFDI-weiten Strukturen
- Fortbildung der Agent:innen

Konsolidierung

- Erweiterung des Research Rendezvous zur FDM-Sprechstunde Humanities@NFDI
- Überarbeitung der Protokollvorlage (Abgleich mit Beratungsrealität und TA-übergreifenden Erkenntnissen)



Wann ist eine Beratung eine Beratung?

Ein Beratungsvorgang im Verständnis der TA Editions setzt sich zusammen aus:

- Initiale Anfrage mit editionsbezogenem Anliegen über Helpdesk, Research Rendezvous, direkte Ansprache, ...
- Persönlicher Austausch (Vieraugenprinzip)
- einmaliger Kontakt, mehrfacher Austausch bis hin zu Gesprächsreihen
- Kernprinzipien wie Protokollierung der Gespräche

Seit August 2022 wurden **92 Beratungsvorgänge protokolliert** (im Kontext Editions)

- Markdownvorlage
- Abgestimmtes Kategorienschema
- Ablage in der Nextcloud
- Über die Helpdesk-Agent:innen Konsolidierung und Nutzung in Text+ insgesamt



Das Consulting in der TA Editions in Zahlen

Gegenstand / Thema

26	×	Antragsberatung
30	×	Community
Activities		
13	×	CroCo
27	×	Data Depositing
7	×	DMP
39	×	Datenqualität
31	×	Kooperation
28	×	LZV
Präsentationsschicht		
2	×	LLM / KI / ML
22	×	LOD / Normdaten
5	×	Rechtl. / eth.
Fragestell.		

Text+ Dienste

2	×	T+ CSP
22	×	T+ Community
16	×	T+ Data Depositing
3	×	T+ FCS
8	×	T+ Flexfund
2	×	T+ GND-Agentur /
entityXML		
12	×	T+ Registry

27 × Software

Text+ (Benutzung)

4 × Software



Das Consulting in Zahlen

Beispielfrage
n

Gegenstand / Thema

26	×	Antragsberatung
30	×	Community
Activities		
13	×	CroCo
27	×	Data Depositing
7	×	DMP
39	×	Datenqualität
31	×	Kooperation
28	×	LZV
Präsentationsschicht		
2	×	LLM / KI / ML
22	×	LOD / Normdaten
5	×	Rechtl. / eth.
Fragestell.		

Was muss ich bei **Antragstellung** beachten? Was sollte in Kapitel 2.4 meines DFG-Antrags stehen?

Welche **Kooperationsmöglichkeiten** gibt es?

Kann jemand zu unserem **Workshop** kommen und eine Einführung zu gängigen Editions-Frameworks geben?

27 × Software

(Benutzung)

4 × Software



Das Consulting in Zahlen

Beispielfrage
n

Gegenstand / Thema

26	×	Antragsberatung
30	×	Community
Activities		
13	×	CroCo
27	×	Data Depositing
7	×	DMP
39	×	Datenqualität
31	×	Kooperation
28	×	LZV
Präsentationsschicht		
2	×	LLM / KI / ML
22	×	LOD / Normdaten
5	×	Rechtl. / eth.
Fragestell.		

Welche **Best Practices** gibt es für die **Nachhaltigkeit von Editions-Websites**, die auf TEI Publisher, Docker oder ähnlichen Technologien basieren, und wie kann sichergestellt werden, dass die **Website auch nach Projektende weiterbetrieben** wird?

Wie kann die **Langzeitverfügbarkeit meiner Briefedition** sichergestellt werden, wenn die Website aktuell von einer Hilfskraft erstellt wurde und keine institutionelle Verantwortung dafür besteht?

Wie kann die Website nach Projektende **langfristig** erhalten und **finanziert** werden?

Welches Repository eignet sich für die Langzeitarchivierung?

Welche **Kooperationsmöglichkeiten** bestehen?

27 × Software

(Benutzung)

4 × Software



Das Consulting in Zahlen

Beispielfrage
n

Gegenstand / Thema

26	×	Antragsberatung
30	×	Community Activities
13	×	CroCo
27	×	Data Depositing
7	×	DMP
39	×	Datenqualität
31	×	Kooperation
28	×	LZV
Präsentationsschicht		
2	×	LLM / KI / ML
22	×	LOD / Normdaten
5	×	Rechtl. / eth.
Fragestell.		

Ich möchte OCR/HTR-Texterkennung einsetzen, Entity Recognition und teil-automatisierten Normdatenabgleich betreiben. Welche **Tools** sollte ich dafür einsetzen?

Wie füge ich verschiedene Tools zu einem **Workflow** zusammen?

27 ×
Text+ (Benutzung)

Software



Das Consulting in Zahlen

Beispielfrage
n

Gegenstand / Thema

26	×	Antragsberatung
30	×	Community
Activities		
13	×	CroCo
27	×	Data Depositing
7	×	DMP
39	×	Datenqualität
31	×	Kooperation
28	×	LZV
Präsentationsschicht		
2	×	LLM / KI / ML
22	×	LOD / Normdaten
5	×	Rechtl. / eth.
Fragestell.		

Wie kann die Datenqualität sichergestellt werden, wenn die **Quellen aus dem Mittelalter** oder früher stammen und **keine einheitliche Namenskonvention** existiert?

Wie kann die **Verknüpfung mit externen Normdaten** umgesetzt werden, wenn die Personen oder Orte nicht in der **GND** erfasst sind?

Wie kann die Datenqualität und **Nachnutzbarkeit** sichergestellt werden, wenn ein eigenes Vokabular verwendet wird?

Wie kann die Datenqualität der **Transkriptionen von Handschriften** im 14. Jh. sichergestellt werden, wenn die **Kurrentschrift schwer lesbar** ist?

Wie kann die Datenqualität der **Personen- und Ortsangaben** in den Briefen sichergestellt werden, wenn **Tausende Personen ohne GND-ID** existieren?

27 × Software

(Benutzung)

4 × Software



Fazit und Ausblick

Aus einem **abstrakten** TA - und Measure-getriebenen **Plan** mit vielen **Unbekannten** ist auf Basis von **User Stories** und **institutionellem** Commitment im Rahmen eines fortlaufenden Prozesses ein durch **Guidelines** und **Workflows** entlang **konkreter Bedarfe aus der Community** geleitetes Consultingangebot in Text+ entstanden, bei dem **engagierte Agent:innen** über **niedrigschwellige Kontaktmöglichkeiten** und Austauschformate mit der Community **interagieren** und im **Dialog** sind.

Die in Text+ aufgebauten Strukturen helfen auch bei der **Integration in FDM-Strukturen** außerhalb der NFDI sowie **konsortienübergreifende Beratungsstrukturen** und werden ein wichtiger Bestandteil der zweiten Förderphase sein.

Weiterentwicklung und Zusammenarbeit





Editionen

User Demands & Community Activities

Die Community auf der Couch:
FDM-Beratung im Bereich Editionen

Editionen finden & vernetzen:
**Die Text+ Registry als Brücke
im Informationsökosystem**

Nils Geißler (AWK NRW) und
Daniela Schulz (HAB Wolfenbüttel)

Diskussionspanel:
“Was braucht die Editionscommunity?”
Perspektiven in Text+ und NFDI 2.0



Warum eine Editionsregistry?

Editionen besser finden, sichtbar machen und verknüpfen.

Die Ausgangslage: verteilt und schwer durchschaubar

Informationen zu Editionen liegen verstreut in verschiedenen Systemen und Formaten.

-  Bibliothekskataloge und Discovery-Systeme
-  Fachportale und Verzeichnisse
-  Projektwebseiten und Forschungsdatenbanken
-  Institutionelle Repositorien und Publikationsdienste
-  Forschungsinformationssysteme und weitere Quellen

Für Forschende bedeutet das:

-  Viele Suchorte
-  Unterschiedliche Metadatenstandards
-  Kaum sichtbare Beziehungen
-  Hoher Zeitaufwand für Recherche

Die Editionsregistry als Brücke

Ein zentraler Einstiegspunkt, der Informationen sammelt, zusammenführt, anreicht und wieder zur Verfügung stellt – offen, vernetzt und forschungsunterstützend.



Warum Verknüpfungen so wichtig sind

-  Welche Editionen zum gleichen Werk existieren?
Editionen, Übersetzungen, Fassungen und Versionen im Überblick.
-  Welche Personen, Institutionen und Projekte sind beteiligt?
Wer arbeitet zusammen? Wer baut aufeinander auf oder zitiert einander?
-  Welche Quellen und Ressourcen sind verbunden?
Gemeinsam genutzte Quellen, Daten, Standards oder Software.
-  Welche Zusammenhänge werden sichtbar?
Kontexte erkennen, Forschung erleichtern, neue Perspektiven eröffnen.
-  FAIRer Zugang zu Forschungsdaten
Auffindbar, zugänglich, interoperabel und nachnutzbar.

Informationen aus unterschiedlichen Quellen werden ...



Für Forschende, Lehrende und Lernende

Einfacher finden. Besser verstehen. Gemeinsam weiterforschen.



Datenmodell und Minimaleintrag

Niedrigschwellig starten. Anschlussfähig bleiben.

Heterogene Editionslandschaft

Viele Formen. Viele Medien.
Ein gemeinsames Modell.

-  Gedruckte Editionen
-  Digitale Editionen
-  Audio-Editionen
-  Film-Editionen
-  Hybride Publikationsformen
-  Projekte, Personen, Institutionen ...



Text+ Registry (Editions)

Gemeinsame Basis – offen und vernetzt

Minimaleintrag – Pflichtfelder

- | | |
|---|---|
|  Titel der Edition / des Projekts |  Zeitliche Einordnung der Editionsobjekte |
|  Ressourcentyp
(Edition, Collection, Dienst ...) |  Beziehungstyp zu anderen Versionen (z. B. Vorgängerversion) |
|  Publikationsjahr(e) |  Datenquelle / Provenienz
(z. B. händisch, Import) |
|  Disziplinäre Zuordnung |  Text+ ID (automatisch vergeben) |
|  Medium der Editionsobjekte
(z. B. Text, Audio, Film) |  Timestamp (automatisch generiert) |



Mit wenige Angaben sichtbar und recherchierbar.
Darüber hinaus: beliebig anreicherbar.

Technische Basis & Anschlussfähigkeit



DataCite als Grundlage

Kompatibel, standardisiert
und international anerkannt.



Domänenübergreifend

Gemeinsame Basis innerhalb
von Text+ für Recherchen
über Domänengrenzen hinweg.



Anschluss an die NFDI

Vorbereitet für Wissensgraphen
und weitere übergreifende
Strukturen.

Anschluss an externe Dienste & Infrastrukturen



RSD Instanz

Austausch und Aggregation über die
Research Software Directory Instanz
von Text+.



SSH Open Marketplace

Sichtbarkeit und Nachnutzung im
gesamten SSH Open Marketplace –
für mehr Reichweite und Impact.



Auffindbar

Bessere Sichtbarkeit
durch einheitliche,
strukturierte Daten.



Verknüpfbar

Beziehungen zwischen
Editionen, Projekten,
Personen und Institutionen.



Nachnutzbar

Über offene Schnittstellen
(API / OAI-PMH)
und maschinenlesbar.



Interoperabel

Domänenübergreifend
recherchierbar und
kompatibel.



Zukunftsfähig

Bereit für Wissensgraphen,
Linked Open Data und
KI-Anwendungen.



Registry - Der Text+ Katalog

Die [Text+ Registry](#) ist ein Recherche- und Informationssystem, das text- und sprachbasierte Forschungsdaten verzeichnet und sie in einer zentralen Übersicht zusammenführt.

[Übergreifender Katalog](#)
[Editionen](#)
[Korpora- und Textsammlungen](#)
[Lexikalische Ressourcen](#)


Suchbeispiele:

- [Suche nach digitalen Editionsdaten eines portugiesischen Dichters, dessen Name mir gerade entfallen ist.](#)
- [Für meine Dissertation suche ich eine digitalisierte Ausgabe von deutschsprachigen Reisetagebüchern.](#)

Vorgestellte Editionen



[Buber-Korrespondenzen Digital. Das Dialogische Prinzip in Martin Bubers Gelehrten- und Intellektuellennetzwerken im 20. Jahrhundert](#)

Der Religionsphilosoph Martin Buber (1878–1965) ist der wohl bedeutendste und bis heute



[Der Sturm – Digitale Quellenedition zur Geschichte der internationalen Avantgarde](#)

DER STURM. Digitale Quellenedition zur Geschichte der internationalen Avantgarde, erarbeitet und herausgegeben von Marjam



[Kalila and Dimna - Anonym Classic](#)

 W
R
S




Filter

- Zugang & Verfügbarkeit
- Inhalte der Editionen
- Aufbau der Editionen
- Disziplinäre Zuordnung
- Forschungsdaten
- Allgemein

reise Suche

20 von 25 Ressourcen

Edition Ressourcen Links

Philipp Hainhofer: Reiseberichte & Sammlungsbeschreibungen 1594–1636

"Die Editionswebsite gibt Ihnen Zugriff auf die digitale Edition der **Reiseberichte** des Augsburger ... Die **Reiserelationen** bilden den Kern von Hainhofers schriftlichem Nachlass und sind eine außerordentliche ... Gesichtspunkten wie der Erforschung von Residenzen und Sammlungen, Hof-, Adels- und Zeremonialkultur, **Reise** ... Philipp Hainhofer: **Reiseberichte & Sammlungsbeschreibungen 1594–1636**

Digitale Edition (Portal, Webpräsentation) **Text** **Deutsch** 103-01 SDE

Edition Ressourcen Links

Online-Edition der fünf Reisetagebücher Hans Posses

Mit der Auswertung der **Reisetagebücher** nimmt sich das Projekt eines Desiderats der Forschung zur NS-Kunstpoltik ... Online-Edition der fünf **Reisetagebücher** Hans Posses

Digitale Edition (Portal, Webpräsentation) **Text** **Deutsch** 103-01 SDE

Edition Ressourcen Links

Online-Edition der fünf Reisetagebücher Hans Posses

Mit der Auswertung der **Reisetagebücher** nimmt sich das Projekt eines Desiderats der Forschung zur NS-Kunstpoltik ... Online-Edition der fünf **Reisetagebücher** Hans Posses

Digitale Edition (Portal, Webpräsentation) **Text** **Deutsch** 103-01 SDE

Filter

Zugang & Verfügbarkeit

Typ

- Digitale Edition (Portal, Webpräsentation) (25)

Angabe Der Zugänglichkeit

Inhalte der Editionen

Medium

- Text (25)
- Bild (3)

Sprache(N)

- Deutsch (25)

Aufbau der Editionen

Editionstyp (Editorisches Modell)

Bestandteile Der Edition

Disziplinäre Zuordnung

Disziplinäre Zuordnung (DFG)

Einordnung Basisklassifikation

Forschungsdaten

Allgemein



Text+ Startseite

Edition

Grand Tour digital. Digitale Edition frühneuzeitlicher Selbstzeugnisse von Bildungsreisen

Ressourcen Links

Metadaten Empfehlungen Graph

Titel
Grand Tour digital. Digitale Edition frühneuzeitlicher Selbstzeugnisse von Bildungsreisen

Abstract
Im Fokus dieses Erschließungsvorhabens stehen 24 größtenteils deutschsprachige Reisetagebücher (insgesamt ca. 10.300 Seiten), die zwischen 1550 und 1770 von jungen Männern verfasst wurden und sich in der Herzog August Bibliothek (HAB)...

Zitation
Grand Tour Digital - Digitale Edition frühneuzeitlicher Selbstzeugnisse von Bildungsreisen Herausgegeben und bearbeitet von: Angela Göbel, Maximilian Görmar. Technische Umsetzung: Maximilian Görmar. Wolfenbüttel 2022-2025.

Sprache
Deutsch

Sprache(n)
Deutsch
Französisch
Latein

Angabe der Lizenz
Creative Commons Attribution Share Alike 4.0 International

entitlyd
c59bcc06-44f1-429b-84be-2ceb97d6c094

Personen
Angela Göbel

Links
<http://proj.beschreibung.hab.de/?link=114>
<https://brandtourdig.hypotheses.org/>
http://selbstzeugnisse.hab.de/edition_grand_tour
<https://brandtourdig.hypotheses.org/>

Datenformate (in Erfassung, Präsentation und Bereitstellung)
TEI/XML
HTML

Kurztitel
Grand Tour digital

Editionstyp (Editorisches Modell)
Diplomatische Edition
Quellenedition

Datenquelle / Provenienz
<https://www.hab.de/digitale-editionen/>

Metadaten



Graphansicht



Metadaten Empfehlungen Graph

Institution
Deutsche Forschungsg...
Herzog August Bibliothek

Person
Maximilian Görmar
Thomas Mandl

Nicht verknüpft
Angela Göbel
Selbstzeugnisse der Frü...

Empfehlungen



Metadaten Empfehlungen Graph

Edition

Selbstzeugnisse der Frühen Neuzeit in der Herzog August Bibliothek, Digitale Edition des Diariums von Herzog August dem Jüngeren.
Hrg. von Inga Hanna Ralle. Techn. Konzeption und Begleitung David Maus. Metadaten, Recherche und Korrektur Jaqueline Krone, Projektbegleitung Prof. Dr. Ulrike Gleixner. Gefördert durch das Niedersächsische Ministerium für Wissenschaft und Kultur. Wolfenbüttel: Herzog August Bibliothek, 2015–2017.
"Die Edition präsentiert das Tagebuch mit dem Originaltitel „Ephemerides. Sive Diarium“,...

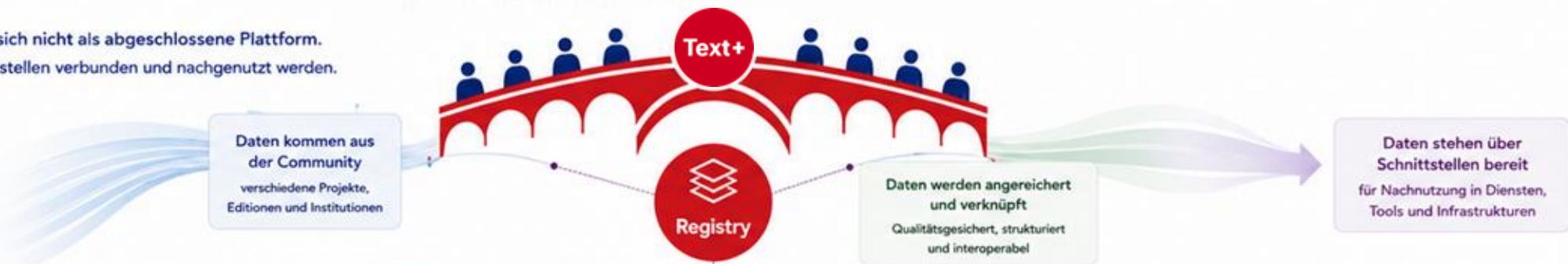
Edition

Mittelfränkische Urkunden, 1300 bis 1330 – Edition
Hg. von Dr. Andrea Rapp unter Mitwirkung von Sabine Bender und Ruth Rosenberger, Trier: Universität Trier, 2003.
"Die Edition führt das von Friedrich Wilhelm begründete Corpus der altdeutschen Originalurkunden bis 1300 durch ein regionalbezogenes Korpus bis in das Jahr 1330 fort; vereinzelt sind Urkunden bis in das Jahr 1350 erfasst. Die Edition erweitert die Grundlagen für...



Interoperabilität & Anschlussfähigkeit

Die Editionsregistry versteht sich nicht als abgeschlossene Plattform.
Ihre Daten können über Schnittstellen verbunden und nachgenutzt werden.



Vielfältige Quellen

- Bibliotheken & Kataloge
- Fachportale & Verzeichnisse
- Projektdaten & Repositories
- Wissenschaftliche Communities
- Forschungsinformationssysteme
- ... und viele mehr

Beispiel in der Praxis: PhilFinder (FID Philosophie)

The screenshot shows the PhilFinder website. At the top, it says 'FACHINFORMATIONSDIENST PHILOSOPHIE' and 'Zugriff auf 32.374 Wissensobjekte'. There is a search bar labeled 'Wissensobjekte suchen' with a 'Suchtext...' input field. Below the search bar are four featured sections: 'Neu hier?', 'Person des Jahres' (highlighting Immanuel Kant), 'Random Fun Fact' (about philosopher life expectancy), and 'Interesse an der Karatierung von Daten?'. A star icon at the bottom left indicates a note: 'Der PhilFinder nutzt die Editionsdaten der Registry über eine API. So werden Editionen auffindbar, verknüpft und in neue Kontexte gebracht.'

Über API: Daten maschinenlesbar abfragen

Beispiel-Query

```
POST https://registry.text-plus.org/api/search/edition
{
  "query": "https://d-nb.info/gnd/118559796",
  "filter": {"properties.languages.@reference": ["deu"]}
}
```

- Offene, dokumentierte Schnittstellen (JSON / REST)
- Gezielte Abfrage und Filterung
- Vereinfachte Integration in eigene Dienste

Heute für den PhilFinder.
Morgen für noch mehr.

- Weitere Fachportale und Discovery-Systeme
- Wissensgraphen & NFDI
- Weitere NFDI-Dienste und Infrastrukturen
- ... und darüber hinaus



Editionen

Weiterführende Links

- Editionen online eintragen (Anmeldung via AcademicID, ORCID oder GitHub): <https://registry.text-plus.org>
- Dokumentation der Registry: <https://registry.text-plus.org/docs>
- Spezifikation des Datenmodells: <https://zenodo.org/records/13379995> (TA Editions)
- Mehr zum Arbeitsbereich Editionen: <https://text-plus.org/editionen>
- Kontakt: <https://text-plus.org/helpdesk>

Was bietet die Editionenregistry?

Die Registry der Datendomäne Editions in Text+ soll Editionen nachweisen, um deren Auffindbarkeit und Sichtbarkeit signifikant zu erhöhen und insbesondere auch den Zugriff auf deren Forschungsdaten – im Sinne der FAIR Prinzipien – zu befördern. Dabei steht sie vor der Herausforderung, eine in mehrfacher Hinsicht heterogene Editionslandschaft in einem einheitlichen Nachweissystem abzubilden. Ein Ausdruck dieser Heterogenität sind die zahlreichen verschiedenen Manifestationen (gedruckt, E-Book, digital, hybrid etc.) in denen Editionen vorkommen und die Diversität der Plattformen, die Zugang zu diesen gewähren sowie deren zahlreiche Bestandteile. Anspruch der Registry ist hierbei, Editionen möglichst inklusiv und vollumfänglich nachzuweisen. Dies bedeutet konkret die Erfassung von einer Vielzahl verschiedener Informationen, wie z.B. die disziplinäre/n Zugehörigkeit(en) oder die Sprache(n) der Edenda, also des zugrundeliegenden Quellenmaterials. In ihrem Umfang beschränkt sich die Registry nicht auf bereits abgeschlossene Editionen, sondern bezieht auch laufende Editionsprojekte mit ein, sofern zu diesen bereits ausreichend Informationen vorliegen, um diese sinnvoll zu erfassen.

Hinsichtlich ihres Skopus ist die Registry grundsätzlich nicht beschränkt. Kern bilden die Editionen der Text+ Partnerinstitutionen, die initial und sämtlich verzeichnet werden. In zweiter Instanz möchte die Registry die deutsche Editionslandschaft möglichst umfangreich abbilden, erhebt jedoch hier weder einen Anspruch auf Vollständigkeit, noch beschränkt sie sich grundsätzlich auf Editionen, die in Deutschland oder unter Beteiligung von deutschen Institutionen entstanden sind. Ziel ist, Informationen zu möglichst vielen Editionen zu erfassen, um damit eine Vielzahl von Auswertungsmöglichkeiten zu bieten und möglichst viele Bedarfe von Nutzenden abzudecken. Auch sollen Schiefenzen (z.B. hinsichtlich Disziplinen oder bestimmter Genres etc.) vermieden werden.

Wer kann die Editionenregistry nutzen und wie?

Die Registry soll grundsätzlich allen interessierten Personen als Rechercheinstrument zur Verfügung stehen. Wonach und zu welchem

Weiterführende Links

- Was bietet die Editionenregistry?
- Wer kann die Editionenregistry nutzen und wie?
- Was ist die „Erfassungseinheit“?
- Welche Voraussetzungen müssen für eine Aufnahme erfüllt sein?
- Was sind die Kategorien, nach denen die Ressourcen beschrieben werden?
- In welchem Verhältnis stehen Text+ Portfolio und Editionenregistry zueinander?
- Wie kommen Editions-Metadaten in die Registry?
- Was ist ein „minimaler Eintrag“?
- Beispiel: Capitularia
- Beispiel: Pessoa Digital
- Beispiel: Bibliothek der Neologie
- Welche Referenzvokabulare oder Normdatenschemata werden verwendet?
- Wie verhalten sich die Elemente im Datenmodell zum Referenz-Vokabular DataCite?
- Wie funktionieren Workflow und Redaktionsmodell?
- Meine Edition ist noch nicht in der Registry auffindbar. Wie kommt sie in die Registry?
- Ich habe einen Fehler in den Angaben eines Datensatzes entdeckt bzw. möchte Informationen ergänzen. Wie gehe ich vor?
- Ich würde gerne mehr über die Editionenregistry erfahren. Wo finde ich Materialien?

Die Editionsregistry (ERY) und das zugrundeliegende Datenmodell – Dokumentation und Spezifikation

Deliverable E1.1 Release of the registry; metadata and API specifications

Das vorliegende Dokument wurde im Rahmen des Konsortiums Text+ im Kontext der Arbeit des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e.V. verfasst. NFDI wird von der Bundesrepublik Deutschland und den 16 Bundesländern finanziert, und das Konsortium Text+ wird gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 460033370. Die Autor:innen bedanken sich für die Förderung sowie Unterstützung. Ein Dank geht außerdem an alle Einrichtungen und Akteur:innen, die sich für den Verein und dessen Ziele engagieren.

This document was created in the context of the work of the association German National Research Data Infrastructure (NFDI) e.V. NFDI is financed by the Federal Republic of Germany and the 16 federal states, and the consortium Text+ is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 460033370. The authors would like to thank for the funding and support. Furthermore, thanks also include all institutions and actors who are committed to the association and its goals.

Version	1.1
Versionsdatum	2024-08
Redaktion	M1 der Task Area Editions

Editionen finden:
registry.text-plus.org
Nicht gefunden? Editionen eintragen:
text-plus.org/editionen/eintragen
Dokumentation der Registry:
registry.text-plus.org/docs
Spezifikation des Datenmodells:
zenodo.org/records/13379995
Mehr erfahren:
text-plus.org/editionen/
Kontakt:
text-plus.org/helpdesk



Text+
 Wo findet man
 meine Edition?

Mindeleintrag in Postkartenformat:
 Die Text-Registry ist das zentrale Verzeichnis aller Informationen über Editionen. Sie enthält eine möglichst vollständige Beschreibung der einzelnen Editionenbestandteile in einer rich-text auf Basis der deutschen Zeichensätze in Form von Text, Bild, Audio, Video, etc. Editionen werden als Einträge in der Registry erfasst und sind auffindbar. Die Registrierung der Editionen erfolgt über das Online-Formular. Bitte beachten Sie, dass die Registrierung der Editionen ein freiwilliges Engagement ist. Die Registrierung der Editionen erfolgt über das Online-Formular. Bitte beachten Sie, dass die Registrierung der Editionen ein freiwilliges Engagement ist.

Kontaktperson:
 E-Mail:

Titel:
Publ.-jahr:
Ausgabeform: Original Edition Print / E-Book Hybrid Forschungsdaten
Disziplin:
URL:
Erzierte Quelle(n):
Zeitraum:
Medium: Text Bild Audio Multimediale Medien





Die Editionenregistry in Zahlen



 **>1.700**

Umfassende Abdeckung der Editionslandschaft – international und interdisziplinär.



Kuratierte Ressourcen

 **knapp 1/3**
sind kuratiert

Qualitätsgesichert durch die Community – mit Sorgfalt und fachlicher Expertise.

Wissenschaftliche Sichtbarkeit

 **5** wissenschaftliche Publikationen

 **>15** Präsentationen & Workshops





Ausblick

Potenziell 2. Förderphase: Nächste Schritte



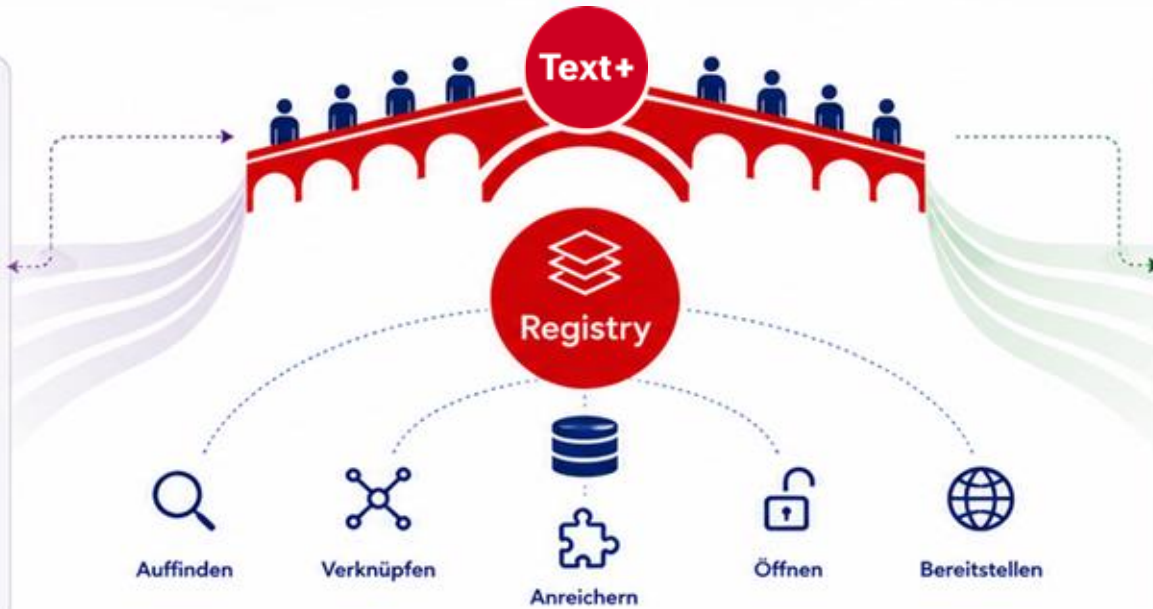
Anbindung an bibliothekarische Systeme
Integration in Suchräume und Workflows
bibliothekarischer Infrastrukturen.



Einbindung von Daten aus bibliothekarischem Kontext
Kataloge, Discovery-Systeme, FID-
Ressourcen und weitere Quellen.



Mapping & Interoperabilität
z. B. Mapping auf MARC21, Nutzung
bibliothekarischer Datenstandards
und Identifier.



Mehr Daten. Mehr Verknüpfungen. Mehr Nutzen.



Ressourcen besser auffindbar machen
Bibliotheksdaten und Editionen verbinden –
für umfassendere Recherche.



Kontexte sichtbar machen
Verknüpfungen zwischen Werken,
Personen, Institutionen und Projekten
noch besser erschließen.



Wissensgraphen & KI-Anwendungen ermöglichen
Anreicherung und Standardisierung als
Grundlage für innovative Dienste.



Nachhaltig vernetzte Forschungsinfrastruktur
Offen, interoperabel und bereit für
die Zukunft.



Community-Arbeit im Mittelpunkt
Der Schlüssel zum Erfolg – heute und morgen.



Austausch fördern
Dialog mit Forschenden, Bibliotheken,
Projekten und Infrastrukturen.



Gemeinsam gestalten
Bedarfe verstehen, Lösungen
ko-kreieren, Standards entwickeln.



Vertrauen aufbauen
Offenheit, Transparenz und
langfristige Zusammenarbeit.



Die mit ✨ gekennzeichneten Folien wurden mithilfe von ChatGPT erstellt.





Editionen

User Demands & Community Activities

Die Community auf der Couch:
FDM-Beratung im Bereich Editionen

Editionen finden & vernetzen:
**Die Text+ Registry als Brücke
im Informationsökosystem**

Diskussionspanel:
**„Was braucht die Editionscommunity?“
Perspektiven in Text+ und NFDI 2.0**



Diskussionspanel:

Was braucht die Editionscommunity? Perspektiven in Text+ und NFDI 2.0

Panelist:innen

- Andrea Rapp (Co-Lead TA Editions, Text+)
- Hannah Busch (SCC Editions)
- Swantje Dogunke (SCC Editions)
- Philipp Hegel (TA Editions)

Moderation

- Kilian Hensen (TA Editions)

Publikumsinteraktion

- Nils Geißler (FID Philosophie, TA Editions)



Perspektiven in Text+ und NFDI 2.0



Bild: Katrin Bimmer

Was zeichnet gute editorische
Infrastruktur aus,

Andrea Rapp?

- Co-TA Lead TA Editions (gemeinsam mit Andreas Speer) ab der zweiten Förderphase
 - Präsidentin der Akademie der Wissenschaften und Literatur | Mainz
 - Vizepräsidentin der Union der deutschen Akademien der Wissenschaften
 - Professorin für Germanistik [Computerphilologie/Mediävistik] an der TU Darmstadt
 - *Historische Fremdsprachenlehrwerke digital. Sprachgeschichte, Sprachvorstellungen und Alltagskommunikation im Kontext der Mehrsprachigkeit im Europa der Frühen Neuzeit (FSL digital)*
-
- **Engagements (Auswahl):**
 - Sachverständige für den Wissenschaftsrat
 - Beirat/Beisitz: Blogportal *Hypothèse*, *HAB*, *CCeH*
 - Boards: *Punch4NFDI*, *DARIAH-EU ERIC*, *MPG*
 - Vorsitz: *TextGrid*, *GKFI*
 - Ständiger Ausschuss der *Leopoldina*



Perspektiven in Text+ und NFDI 2.0



Was zeichnet gute editorische
Infrastruktur aus,

Hannah Busch?

- Wissenschaftliche Mitarbeiterin im Akademieprojekt *Die Formierung Europas durch Überwindung der Spaltung im 12. Jahrhundert* (BAdW & AWK NRW) am Cologne Center for eHumanities (CCeH) der Uni Köln
- **Studium & Stationen**
 - Deutsch-Italienische Studien und Editionswissenschaft in Bonn, Florenz und Berlin (FU)
 - Mitarbeiterin am Trier Center for Digital Humanities (2013–2018)
 - laufendes Promotionsvorhaben zur Anwendung von Deep Machine Learning Verfahren zur Datierung und Lokalisierung mittelalterlicher lateinischer Handschriften (Huygens Instituut (KNAW) Amsterdam, Universität Leiden)
- **Engagements (Auswahl):**
 - SCC Editions in *Text+*
 - Redaktion Wissenschaftsblog *Mittelalter - interdisziplinäre Forschung und Rezeptionsgeschichte* [Vertretung im wissenschaftlichen Beirat des *Handschriftenportals*]
 - Executive Board des *Digital Medievalist*
 - Mitgründerin des Podcasts *Coding Codices*



Perspektiven in Text+ und NFDI 2.0



Was zeichnet gute editorische
Infrastruktur aus,

Swantje Dogunke?

- Fachreferentin an der Thüringer Universitäts- und Landesbibliothek Jena für Geschichte, Alte Geschichte, Kulturanthropologie
- **Studium & Stationen**
 - Museologie sowie Bibliotheks- und Informationswissenschaft in Leipzig (HTWK) und Berlin (HU)
 - Museologin an der Klassik Stiftung Weimar sowie der Overbeck-Gesellschaft Kunstverein Lübeck
 - Mitarbeiterin im Forschungsverbund Marbach Weimar Wolfenbüttel und an der Bauhaus-Universität Weimar
- **Engagements (Auswahl):**
 - Vorstandsmitglied der digiCULT-Verbund eG
 - Arbeitsgruppe *DH in Mitteldeutschland* an der Sächsischen Akademie der Wissenschaften zu Leipzig (SAW)
 - Forschungsgruppe *Netzwerk digitale Geisteswissenschaften und Citizen Science* der Universität Erfurt
 - Fachredaktionsmitglied der *ZfdG* (bis 2019)
 - SCC Editions in *Text+*



Perspektiven in Text+ und NFDI 2.0



Bild: Christina Strvali

Was zeichnet gute editorische
Infrastruktur aus,

Philipp Hegel?

- Leitung der Hans-Kelsen-Forschungsstelle, Arbeitsstelle Frankfurt, an der Akademie der Wissenschaften und Literatur | Mainz, Projekt *Hans Kelsen Werke*
- **Studium & Stationen**
 - Studium Allgemeine und Vergleichende Literaturwissenschaft, Geschichte und Philosophie (Konstanz, Bielefeld), Editionswissenschaft FU Berlin; Promotion TU Darmstadt
 - Institut für Sprach- und Literaturwissenschaft (TU Darmstadt) und SFB *Episteme in Bewegung* (FU Berlin)
 - Vertreter der TA Editions im Bereich *Community Activities* (Curriculare Empfehlungen, Workshops und Consulting)
- **Engagements (Auswahl)**
 - AG *Referenzcurriculum* im DHd-Verband
 - Arbeitsgemeinschaft für germanistische Edition

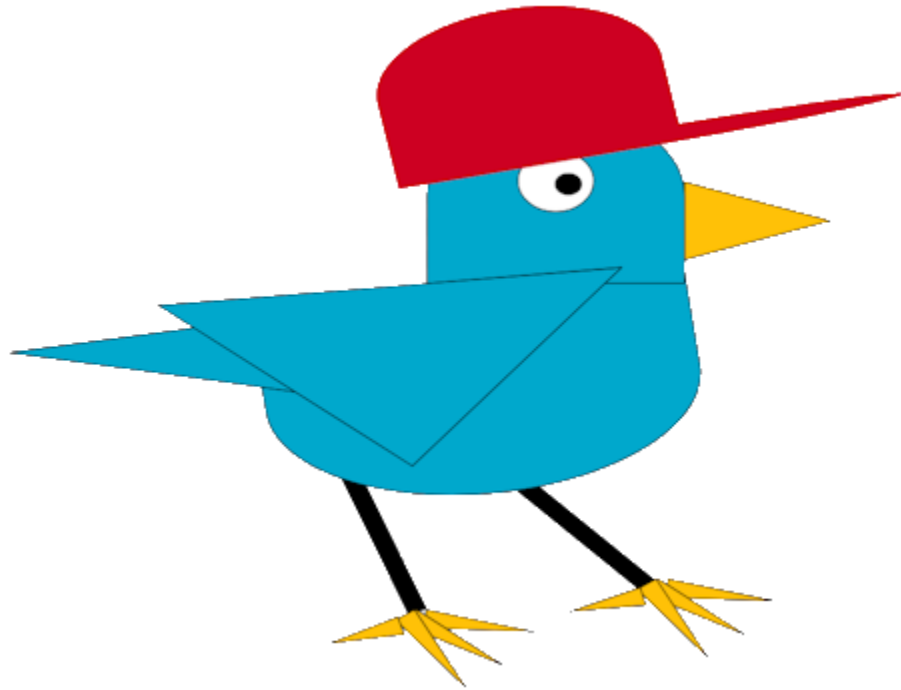


Diskussionspanel:

***Was braucht die Editionscommunity?
Perspektiven in Text+ und NFDI 2.0***



Ergebnisse aus Task Area Editions



Vielen Dank!



Text+

Plenary
2026



@textplus_NFDI
#TextPlusPlenary

5. TEXT+ PLENARY

PROGRAMM

Montag, 15. Juni 2026

12:00 – 13:30	<i>Lunch (Katakomben)</i>
13:30 – 14:00	Begrüßung, Eröffnung und Einführung in das Programm (Aula)
14:00 – 15:30	Ergebnisse aus Task Area: Infrastructure & Operations (Aula)
<i>15:30 – 16:00</i>	<i>Kaffeepause (Katakomben)</i>
16:05 – 17:35	Ergebnisse aus Task Area: Editions (Aula)
17:35 – 18:05	Posterslam (Aula)
18:05 – 19:00	Postersession (Fuchs-Petrolub-Saal)
<i>19:00 – 22:00</i>	<i>Abendempfang (Rektoratshof)</i>

Text+

Plenary
2026



@textplus_NFDI
#TextPlusPlenary

5. TEXT+ PLENARY

PROGRAMM

Dienstag, 16. Juni 2026

09:00 – 10:30	Ergebnisse aus Task Area: Lexical Resources (Aula)
<i>10:30 – 11:00</i>	<i>Kaffeepause (Katakomben)</i>
11:00 – 12:30	Ergebnisse aus Task Area: Collections (Aula)
12:30 – 13:00	Abschluss und Ausblick auf die zweite Phase (Aula)
<i>13:00 – 14:00</i>	<i>Lunch (Katakomben)</i>
14:00 – 15:30	<i>Parallele Sitzungen (intern, nur auf Einladung)</i>
	Gemeinsames Treffen des Scientific Coordination Committees Lexical Resources und der Task Area Lexical Resources (Aula)
	Treffen des Scientific Coordination Committees Collections (Dozentenzimmer, O 126)
	FID Koop Treffen (Musikerzimmer, EO 154)



Lexical Resources

Text+ Plenary
16. Juni 2026, Mannheim

Gefördert durch

DFG Deutsche
Forschungsgemeinschaft

Förderungsnummer 460033370

Teil der

nfdi Nationale
Forschungsdaten
Infrastruktur



Die Task Area Lexical Resources in Text+

Collections
Lexical Resources
Editions
Infrastructure/
Operations

- Allgemeiner Überblick
- Kooperationsprojekte und SCC
- Gemeinsame Wörterbuchplattform
- Die TA im Gesamtprojekt
- KI als neues Themenfeld
- Ausblick



Die Task Area Lexical Resources in Text+

Collections
Lexical Resources
Editions
Infrastructure/
Operations

- **Allgemeiner Überblick**
- Kooperationsprojekte und SCC
- Gemeinsame Wörterbuchplattform
- Die TA im Gesamtprojekt
- KI als neues Themenfeld
- Ausblick



Überblick über die TA Lexical Resources

Collections
Lexical Resources
Editions
Infrastructure/
Operations

“The task area Lexical Resources deals with reference works and databases of word usage. Their conception and compilation constitute a research activity that produces important results in specific disciplines, e.g. lexicography or semantics. The resulting lexical resources are used for research purposes in many areas.”

Erstantrag, S. 72



Überblick über die TA Lexical Resources

Collections
Lexical Resources
Editions
Infrastructure/
Operations

beteiligt sind die zentralen Akteure im Bereich der digitalen Lexikographie

- 7 Institutionen
- ca. 20–25 Mitarbeitende
- 7 Kooperationsprojekte



Ressourcen

Collections
Lexical Resources
Editions
Infrastructure/
Operations

94 lexikalische Ressourcen und mehr als 10 verschiedene Ressourcentypen



Ressourcen

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Sprachen

Über 30 verschiedene Sprachen, Sprachstufen und Varietäten werden durch die Ressourcen abgedeckt.

Englisch
Arabisch
Ägyptisch
Latein
Altgriechisch
...

Sprachstufen des Deutschen:
Gegenwartsdeutsch
Mittelhochdeutsch
Althochdeutsch

Rund 20 deutschsprachige
Dialekte:

- Schwäbisch
- Bayerisch
- Fränkisch
- Rheinisch
- Sudetendeutsch
- Luxemburgisch
- Pfälzisch
- Elsässisch
- Lothringisch
- Westfälisch
- ...

Herausforderungen durch nicht-lateinische
Schriftsysteme!



Einsatzgebiete/Community

Collections
Lexical Resources
Editions
Infrastructure/
Operations

- Lexikographie (Nachnutzung von Wörterbüchern, Vernetzung, ...)
- Literaturwissenschaft (semantische Erschließung, ...)
- Editionswissenschaft (Kommentierung, Annotation, ...)
- Sprachwissenschaft (Bedeutungswandel, Wortbildungsforschung, Dialektologie, ...)
- Digital Humanities (Wissensgraphen, Linked Open Data, ...)
- Computerlinguistik und KI (NLP-Verfahren, Sprachmodelle, Information Extraction, ...)
- Lehre
-



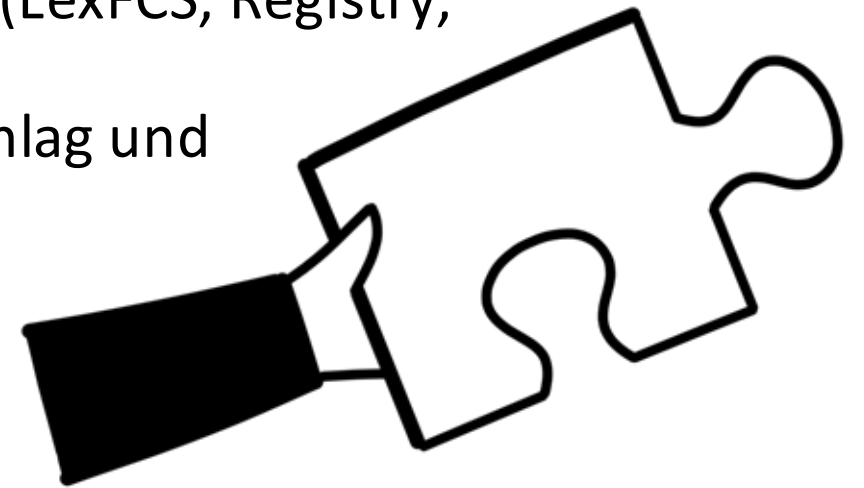
Die Task Area Lexical Resources in Text+

- Allgemeiner Überblick
- **Kooperationsprojekte und SCC**
- Gemeinsame Wörterbuchplattform
- Die TA im Gesamtprojekt
- KI als neues Themenfeld
- Ausblick



Kooperationsprojekte und SCC

- sehr spezifische, sehr heterogene Daten
- viele Nutzergruppen
- unser SCC ist
 - eine unverzichtbare Schnittstelle zu den Fachcommunities, die Dienste von Text+ nutzen
 - ein wichtiges fachliches Beratungsgremium (LexFCS, Registry, Repositories usw.)
 - ein strategischer Einflussfaktor durch Vorschlag und Begutachtung von **Kooperationsprojekten**



Kooperationsprojekte

- sieben Projekte
- 325.000 Euro
- kurze Laufzeit
- infrastrukturelle Einbindung in Text+ gemäß FAIR-Prinzipien

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Kooperationsprojekte

Collections
Lexical Resources
Editions
Infrastructure/
Operations

2023: **Text+-Schnittstelle für das
Mittelhochdeutsche Wörterbuch
(MWB-APIplus)**



Niedersächsische Akademie
der Wissenschaften
zu Göttingen

2023: **Integration NiederSorbisch-deutscher
Wörterbücher als lexikalische
Ressourcen in Text+ (INSERT)**



Serbski Sorbisches
institut Institut

2023: **Corpus Glossariorum Latinorum
Online Lemmatization (CGLO)**



BAYERISCHE
AKADEMIE
DER
WISSENSCHAFTEN



Kooperationsprojekte

Collections
Lexical Resources
Editions
Infrastructure/
Operations

2024: **Thesaurus Linguae Aegyptiae –
More Fair with APIs (TeLApi)**



2025: **Glossarium Graeco-Arabicum –
Open Data (GlossGA-OD)**



Kooperationsprojekte

Collections
Lexical Resources
Editions
Infrastructure/
Operations

**2026: Langzeitverfügbarkeit und Integration
der Altägyptischen Wörterbücher
im Verbund (AWV 2.0)**

kompetenzwerkD
Sächsisches Forschungszentrum und Kompetenznetzwerk
für Digitale Geisteswissenschaften und Kulturelles Erbe

**2026: Entwicklung eines sprachübergreifenden
Standard-Repräsentationsmodells
für framesemantische Daten (MuLingFrames)**

 **Institut**



Kooperationsprojekte

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Breite Sprachenabdeckung

- Alte Sprachen
 - Latein
 - klassisches Griechisch
 - klassisches Ägyptisch
- Historische Sprachstufen
 - Arabisch (8. bis 10. Jahrhundert)
 - Mittelhochdeutsch
- Moderne Sprachen
 - Deutsch, Französisch, Spanisch, Englisch
 - Niedersorbisch



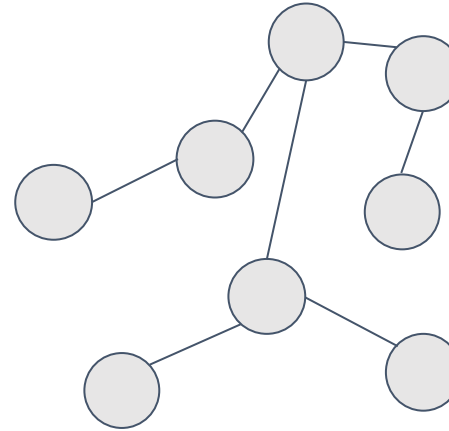
raben-swarz *Adj.* [Lexer](#) [Lexer-Nachtr.](#) [BMZ](#) (1)
raben-var *Adj.* [Lexer](#) [BMZ](#) [Findeb.](#) (3)
raben-vètache *swfm* [Findeb.](#)
rabi *M.* [Lexer-Nachtr.](#) [Findeb.](#)
rabîne *stF.* [Lexer](#) [BMZ](#) [Findeb.](#) (20)
Râbînîs *stM.*
rab-sâme *swM.* [Lexer](#)
rabusch *stM.* [Lexer](#)
rac *Adj.* [Lexer](#) [BMZ](#) → [règen](#) *stV.*
rac *stM.* [Lexer](#) [Findeb.](#)



Kooperationsprojekte

Breite thematische/konzeptionelle Abdeckung

- Sprachstufenwörterbücher
- Thesauri
- Glossarien
- bilinguale Wörterbücher
- Framenets



Kooperationsprojekte

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Technische Fragestellungen

- Umgang mit ressourcen- bzw. sprachspezifischen Fonts ← Ägyptisch
- Import von Daten aus TEI Lex-0 ← Niedersorbisch
- Implementierungsaufwand für zentrale Software ← Altgriechisch/Arabisch
- Einheitlicher Zugang zu verschiedenen Wörterbuchressourcen ← Niedersorbisch/
Lateinische Glossare
- Formatdefinition für FrameNet-Ressourcen ← Mehrsprachigkeit



Kooperationsprojekte

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Technisch-organisatorische Fragestellungen

- Verzahnung zwischen Kooperationsprojekten
- Kombination von Wörterbuch- und Korpuszugriff mittels APIs
- Zugriff auf beschränkte Ressourcen

- ← Altägyptisch
- ← Mittelhochdeutsch
- ← Niedersorbisch

→ auf Ebene der TA besonders das „I“ in „FAIR“

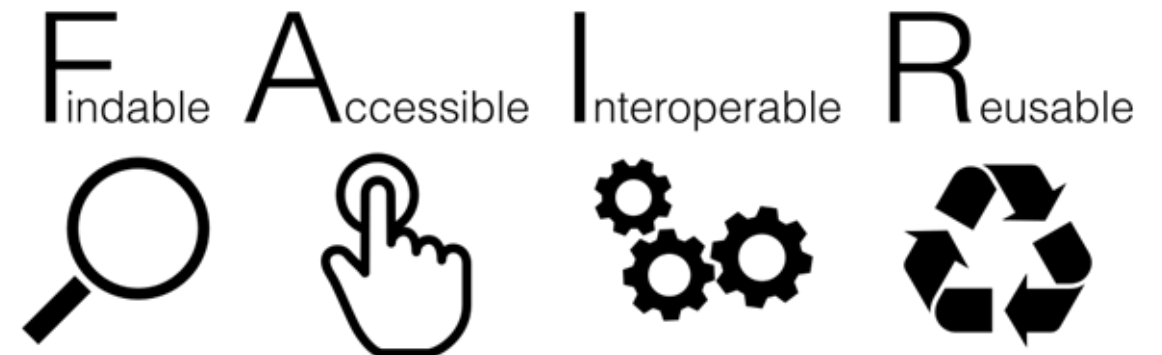


Abbildung: SangyaPundir, CC BY-SA 4.0 Intl.



Kooperationsprojekte

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Umgang mit diesen technischen/organisatorischen Fragestellungen durch

- Konverter für die Datenaufnahme
 - Einbau von neuen Features in die LexFCS
 - Ergänzungen der Registry – Verbreiterung des Ressourcenangebots
- Große Bereicherung für die Arbeit der TA LexRes
- Weitere Details bei der Beschreibung der **gemeinsamen Wörterbuchplattform**



Die Task Area Lexical Resources in Text+

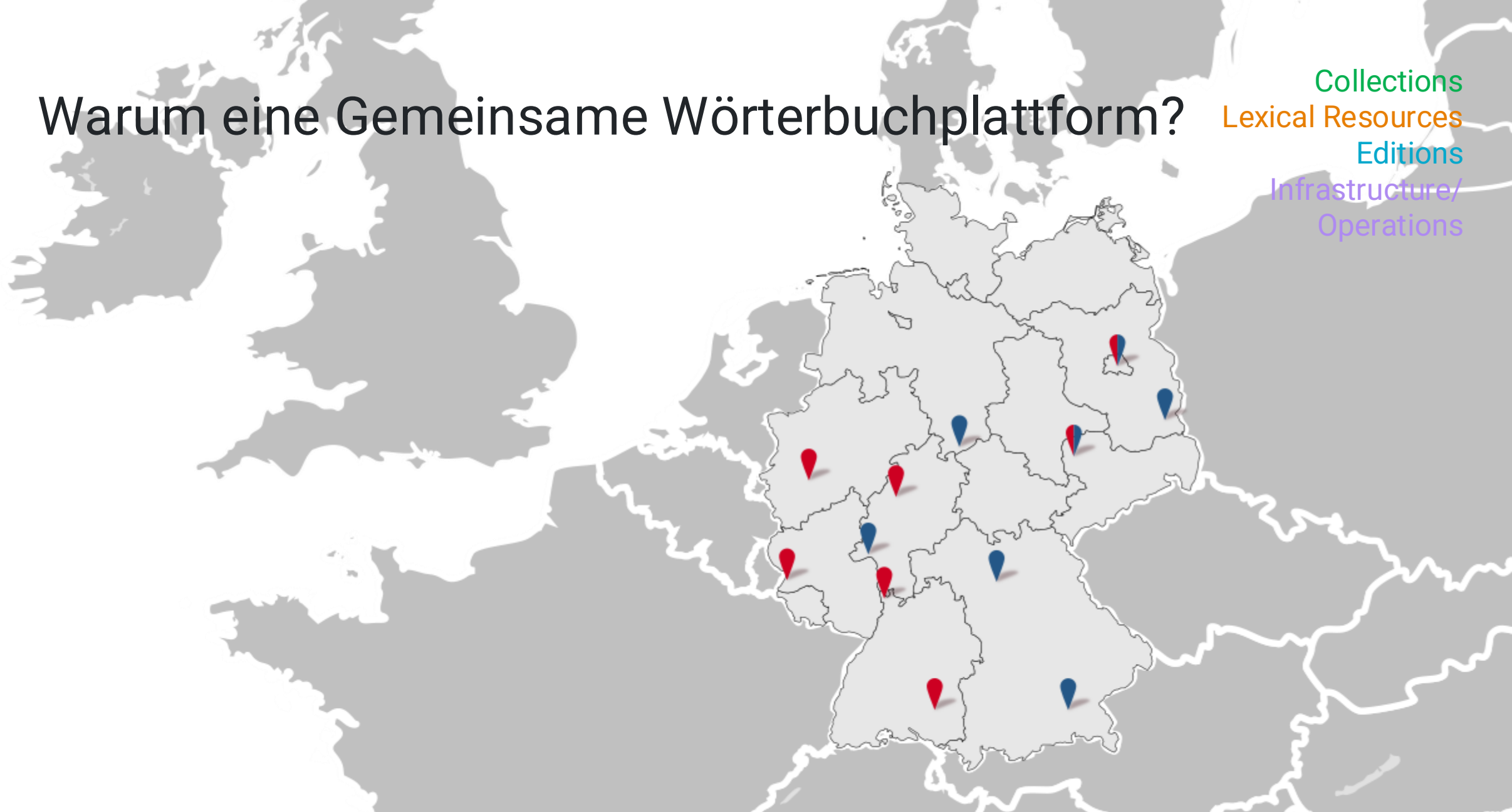
Collections
Lexical Resources
Editions
Infrastructure/
Operations

- Allgemeiner Überblick
- Kooperationsprojekte und SCC
- **Gemeinsame Wörterbuchplattform**
- Die TA im Gesamtprojekt
- KI als neues Themenfeld
- Ausblick



Warum eine Gemeinsame Wörterbuchplattform?

- Collections
- Lexical Resources
- Editions
- Infrastructure/
Operations



Anfragesprache

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Erweiterte Suche im Neologismenwörterbuch

Stichwort

Suche mit Platzhalterzeichen (*,?)

Neologismtyp

Aufkommen und Herkunft

Aufkommen

Herkunft (Sprache)

Herkunft (Typ)

Wortbildung

Wortbildung (Typ)

Wortbildung (Konstituenten)

Keine Merkmale ausgewählt.

Suchen Sie nach Artikeln mithilfe von über 150 verschiedenen Eigenschaften.

- Heben Sie die Details einer Kategorie farblich hervor, um gleiche oder ähnliche Werte schneller erkennen zu können.
- Blenden Sie die Details einer Kategorie ein oder aus. Sie können sich die Werte einer Kategorie auch anzeigen lassen, ohne bestimmte Kriterien auszuwählen.
- Setzen Sie ihre Auswahl für die Kategorie zurück.

Fahrrad|

- Fahrrad - Substantiv
- Fahrradabstellanlage - Substantiv
- Fahrradabstellplatz - Substantiv
- Fahrradabteil - Substantiv
- Fahrradanhänger - Substantiv
- Fahrradautobahn - Substantiv
- Fahrradbauer - Substantiv
- Fahrradbeauftragte - Substantiv
- Fahrradbereifung - Substantiv
- Fahrradbesitzer - Substantiv

ausklappen

fahrrad

Version 01/25

Wörterbuchnetz

... je weiter ich in diesem Studium fortgehe, desto klarer wird mir der Grundsatz: daß kein einziges Wort oder Wörtchen bloß eine Ableitung haben, im Gegenteil jedes hat eine unendliche und unerschöpfliche. Alle Wörter scheinen mir große und sich spaltende Strahlen eines wunderbaren Ursprungs, daher die Etymologie nichts tun kann, als einzelne Leitungen, Richtungen und Ketten aufzufinden und nachzuweisen, soviel sie vermag. Fertig wird das Wort nicht damit.“

Jacob Grimm an Friedrich Carl von Savigny, 20. April 1815

Fahrrad

14 Treffer

Fahrrad - Fahrrad - Fahrrad - FAHRRAD - FAHRRAD - Fahrrad - Fahrrad - Fahrrad - Drahtisches - Fahrrad - Fahrrad - Fahrrad - Fahrrad - Fahrrad

Fahrrad

Search term interpretation

- Ignore case
- Enable regular expressions

Edit distance

Enter an integer greater than 0

Word categories

Empty selection searches all categories.

- Adjectives
- Nouns
- Verbs

Word classes

Empty selection searches all classes.

- Allgemein (Adjectives, Verbs)
- Artefakt (Nouns)
- Attribut (Nouns)
- Besitz (Nouns, Verbs)
- Bewegung (Adjectives)
- Form (Nouns)
- Gefühl (Nouns, Adjectives, Verbs)

Orthographic variants

Empty selection searches all variants.

- Current form
- Current form, variant spelling
- Old form
- Old form, variant spelling



Darstellung

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Rover Synset Search Semantic Relatedness

Synset Search

Fahrrad [Show search options](#)

1 result

Bike, Drahtesel, Fahrrad, Rad, Stahlross, Stahlroß, Velo Artikel
n.

Component meronyms 3
Fahrradpedal, Pedal, Sattelstütze, Fahrradreifen, Radreifen ...

Hypernyms 1
Zweirad

Hyponyms 20
Citybike, Tourenrad, Niederrad, Alltagsrad, Lastenfahrrad, Lastenrad, Zweifaltrad, Klappfahrrad, Klapprad, BMX-Rad, E-Bike, Elektrofahrrad, Elektrorad, Damenfahrrad, Damenrad, Sportrad, Leihfahrrad, Leibrad, Dienstfahrrad, Di...

Related to 3
Fahrradhänger, Fahrradfahren, Fahrradfahrt, Radfahren, Radfahrt

Conceptual Relations

Relations

Hypernyms

Hyponyms

Component meronyms

Fahrrad, das

Wortaufgibt

Geografische Verteilung

Bedeutung

Etymologie

Fahrrad

Wort: **Fahrrad** Anzahl: 19.152 Rang: 2.893 Hauptklasse: 18

Artikelliste

Wortgraph

Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm / Neubearbeitung (A-F)

FAHRRAD, n. *Bd. 9, Sp. 57*

FAHRRAD, n. *Bd. 9, Sp. 58*

FAHRRAD, n. *zus. mit fahren vb. puristisches ersatzwort für velociped n. zweirädriges fortbewegungsmittel mit pedalantrieb: 1894 wenn wir ... von dauermärschen, distanceritten, feldtelegraph, fahrrad, luftschiffahrt, brieftauben nichts wüßten MASSOW ref. 231. 1924 sie hüteten, wie jemand absprang, das fahrrad gegen die mauer lehnte FRANK bürger 144. 1995 die voraussetzungen für eine umweltfreundliche fortbewegung - mit dem öffentlichen nahverkehr, dem fahrrad oder den eigenenbeinen - sind somit günstig frankf. allg. ztg. 21.NZ*

FAHRRAD-, ca. 60 nominale zus. mit *fahrrad n. die meisten bildungen sind erst seit der 2. hälfte des 20. jhs. bezeugt:*

- fahrradbestandteil** n.: 1910 A. WEBER kampff 9.
- fahrradhändler** m.: 1910 ebd. 132. 1993 frankf. allg. ztg. 131.R12.
- fahrradindustrie** f.: 1901 HUBER Dtlid. 352. 1979 frankf. allg. ztg. 147.5.
- fahrradordnung** f.: 1899 TRIEFEL völkerrecht 55.

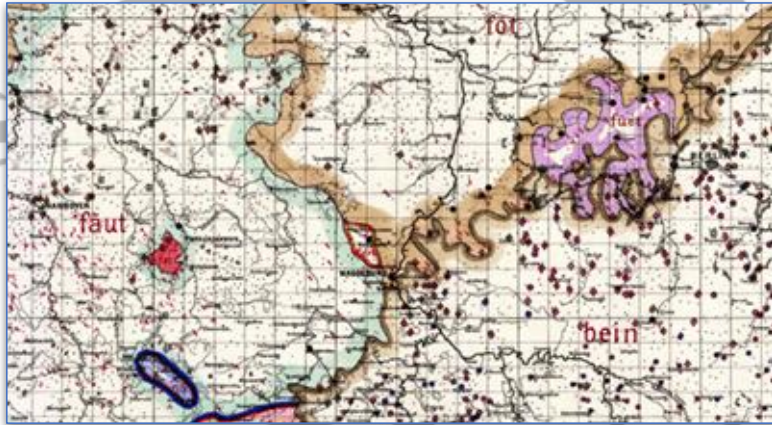


Kartensuche



Medientyp

Collections
Lexical Resources
Editions
Infrastructure/
Operations



PersonA::GlossType	▶ GEHÖREN1^
PersonA::Gloss	▶ SEIN-IHR1
PersonA::Mundbild	
PersonA::HamNoSys	▶ O...R...f

Video

Meyers Großes Konversationslexikon (6. Auflage, 1905-1909)

Dach die, s. 401 **Dach** die, s. 401

unterscheidet man folgende Arten: Das Sattel- od. Giebeldach (Fig. 2) hat zwei Dachflächen, einen First und zwei Traufen. Die offen bleibenden Seiten werden durch senkrechte, zu den Fronten gehörige Giebelwände geschlossen. Durchdringen sich zwei Satteldächer unter einem rechten Winkel, so entsteht das Kreuzdach (Fig. 7). Ein halbes Satteldach, d. h. ein D. mit nur einer Dachfläche, einem First und einer Traufe, nennt man ein Pultdach (Fig. 5), ein D. mit vier Dachflächen, vier in gleicher Höhe liegenden Traufen und nur einem First ein Walmdach (Fig. 3). Die dreieckigen Dachflächen sind die Walm. Liegen die Traufen der Walm höher als die der beiden andern Dachflächen, so entsteht das Krüppelwalmdach (Fig. 4). Ein aus vier und mehr gleichen Wälm gebildetes D. heißt ein Zeltdach (Fig. 6). Auf Türmen gestaltet sich dasselbe zur aufstrebenden schlanken Pyramide (Turmhelm). Das Kegeldach (Fig. 10) ist als Zell- oder Pyramidendach über einem Polygon von unendlich viel Seiten, d. h. über einem kreisförmigen Grundriß, aufzufassen; seine Dachfläche ist also ein Kegelmantel.



Bahnhofsapotheke, die

Grammatik Substantiv (Femininum) · Genitiv Singular: **Bahnhofsapotheke**

Aussprache

Worttrennung Bahn|hofs|apo|the|ke

Wortzerlegung ↗ Bahnhof ↗ Apotheke

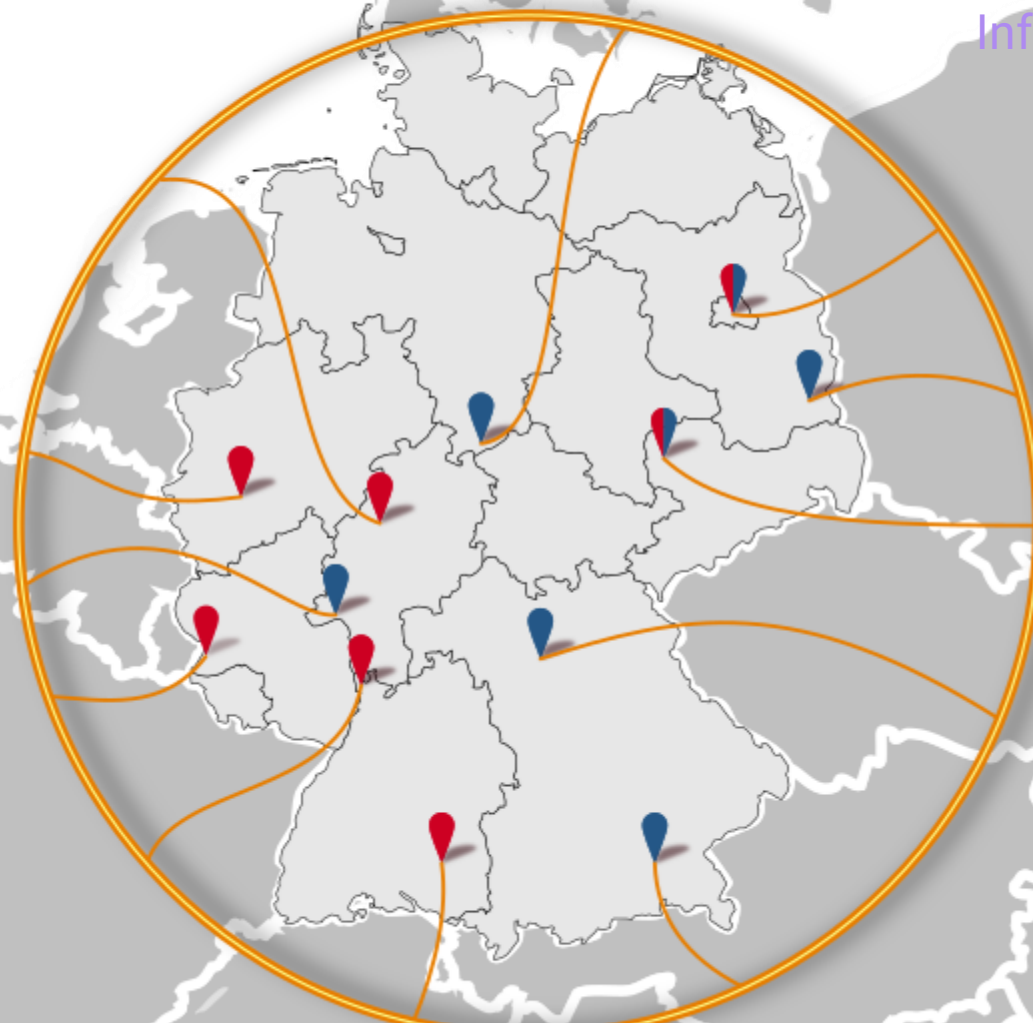


LexFCS

Collections
Lexical Resources
Editions
Infrastructure/
Operations

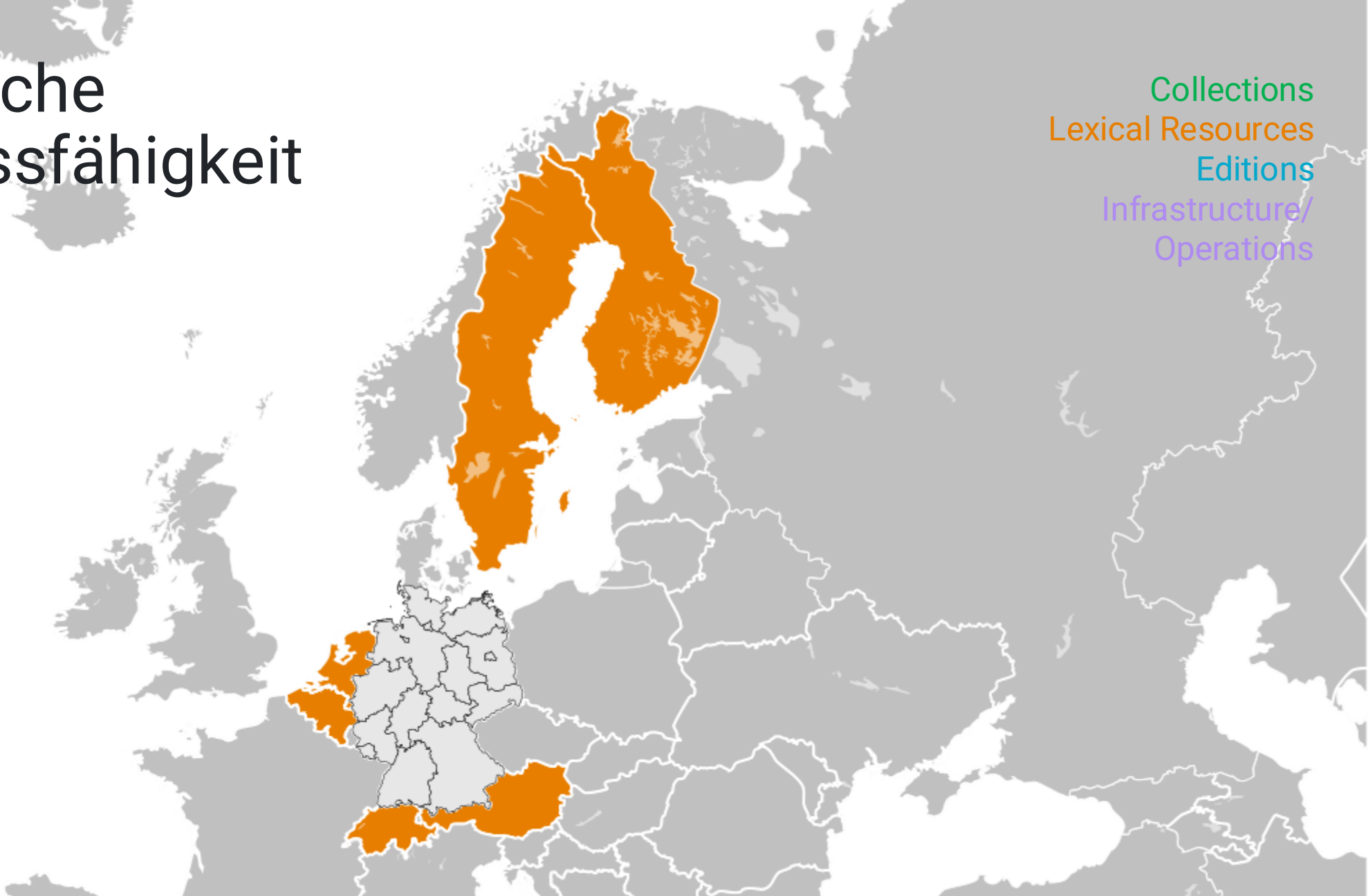
L
e
x

FEDERATED
CONTENT
SEARCH



Europäische Anschlussfähigkeit

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Wörterbuchplattform – Was?

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Suchsystem

Katalog

Portfolioerweiterung

Standards

User-Support

Entwickler-Support

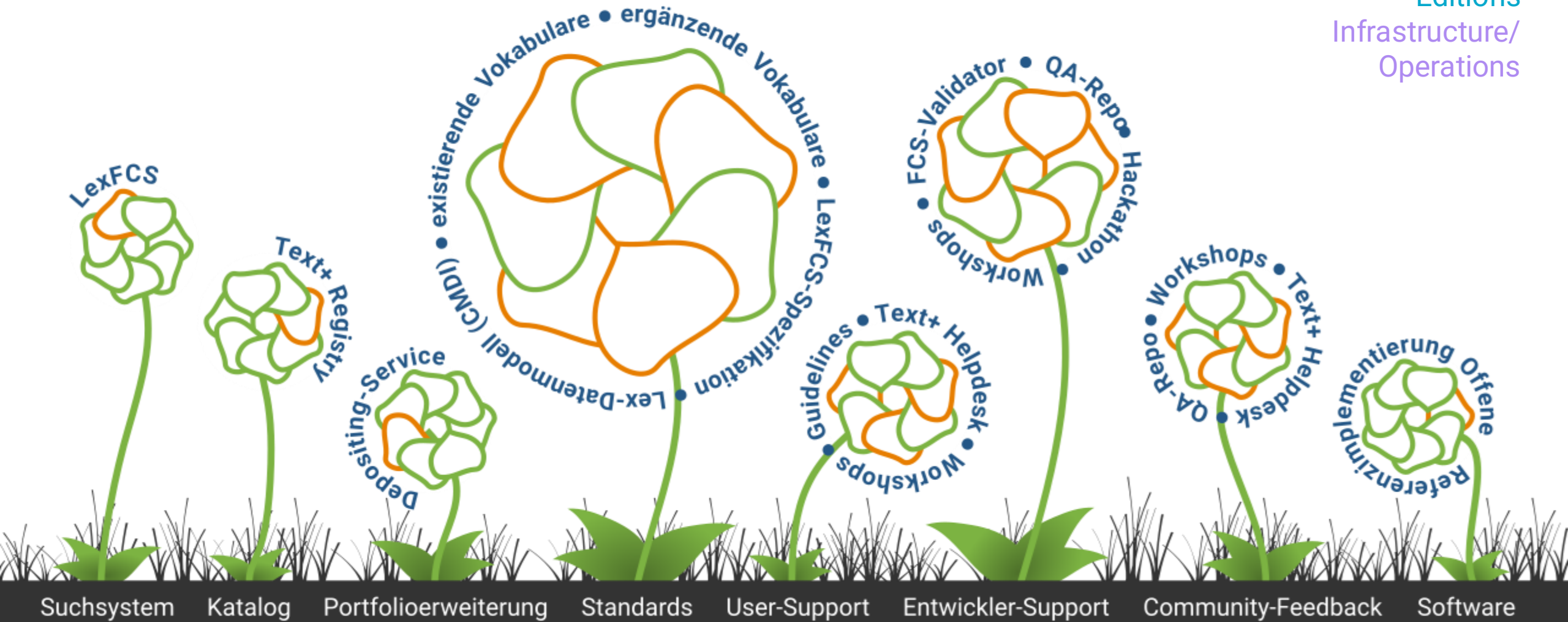
Community-Feedback

Software



Wörterbuchplattform – Was?

Collections
Lexical Resources
Editions
Infrastructure/
Operations



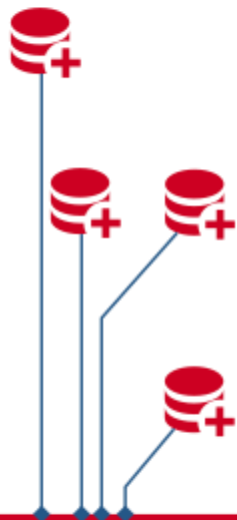
Wörterbuchplattform Entwicklung

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Wörterbuchplattform Entwicklung

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Text+ 01.10.2021–30.09.2026

Text+ II

CGLO

LR TeLApi

GlossGA - OD

AWV 2.0

INSERT

MuLingFrames

MWB-APIplus

2022

2023

2024

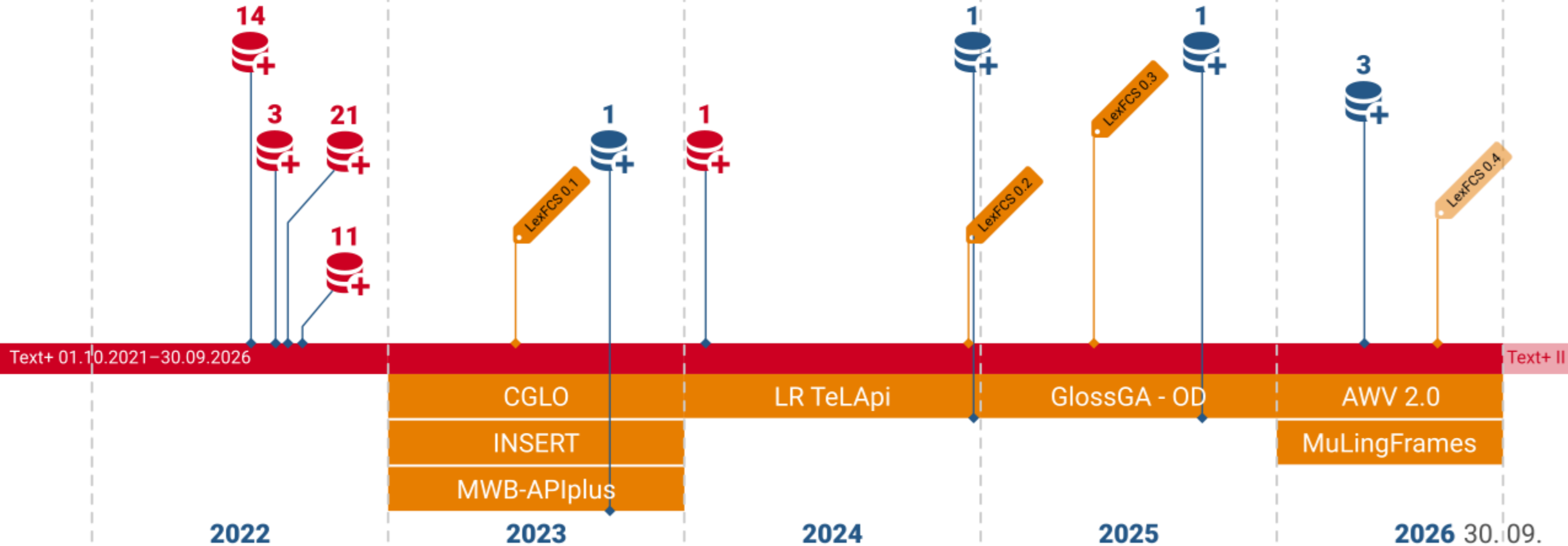
2025

2026 30.09.

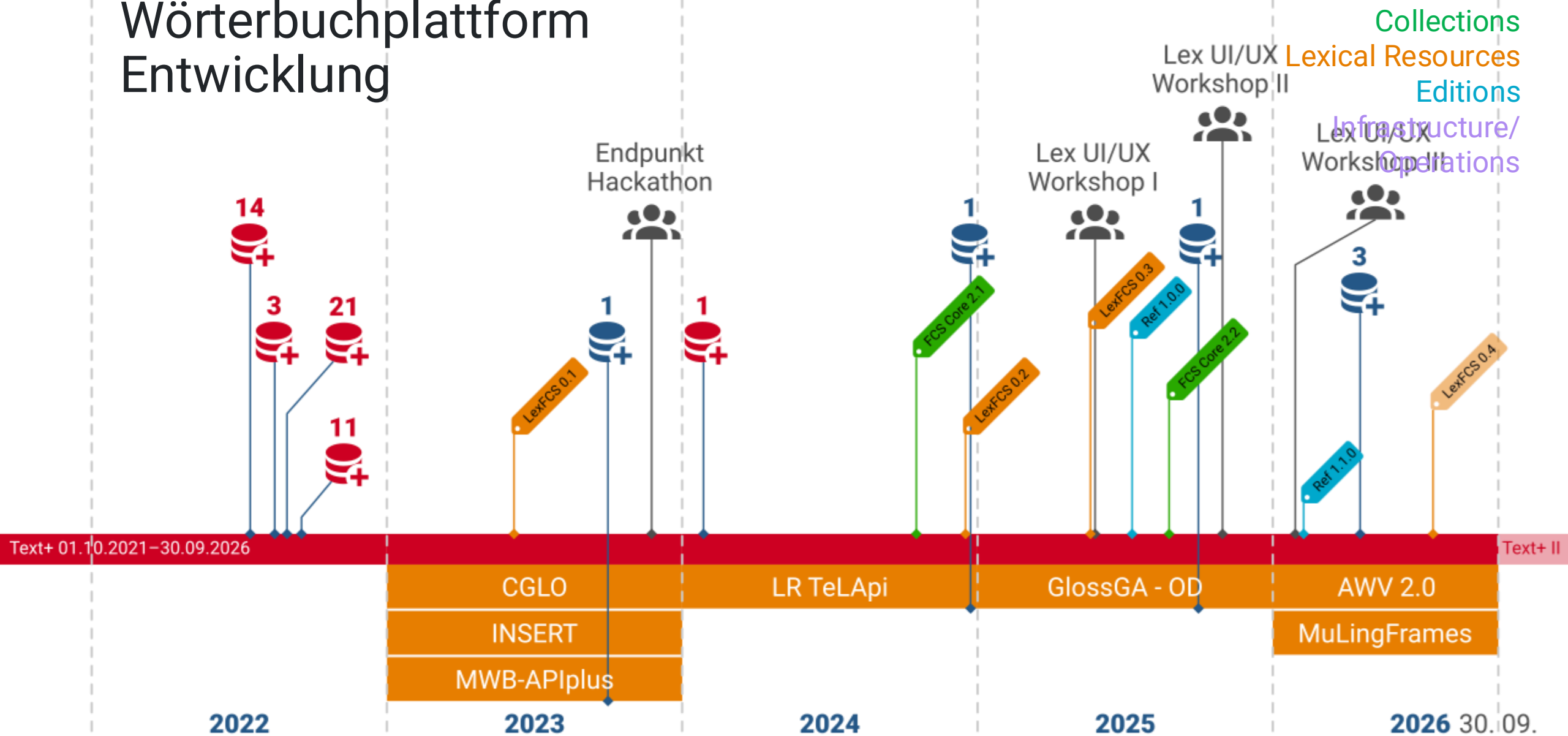


Wörterbuchplattform Entwicklung

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Wörterbuchplattform Entwicklung

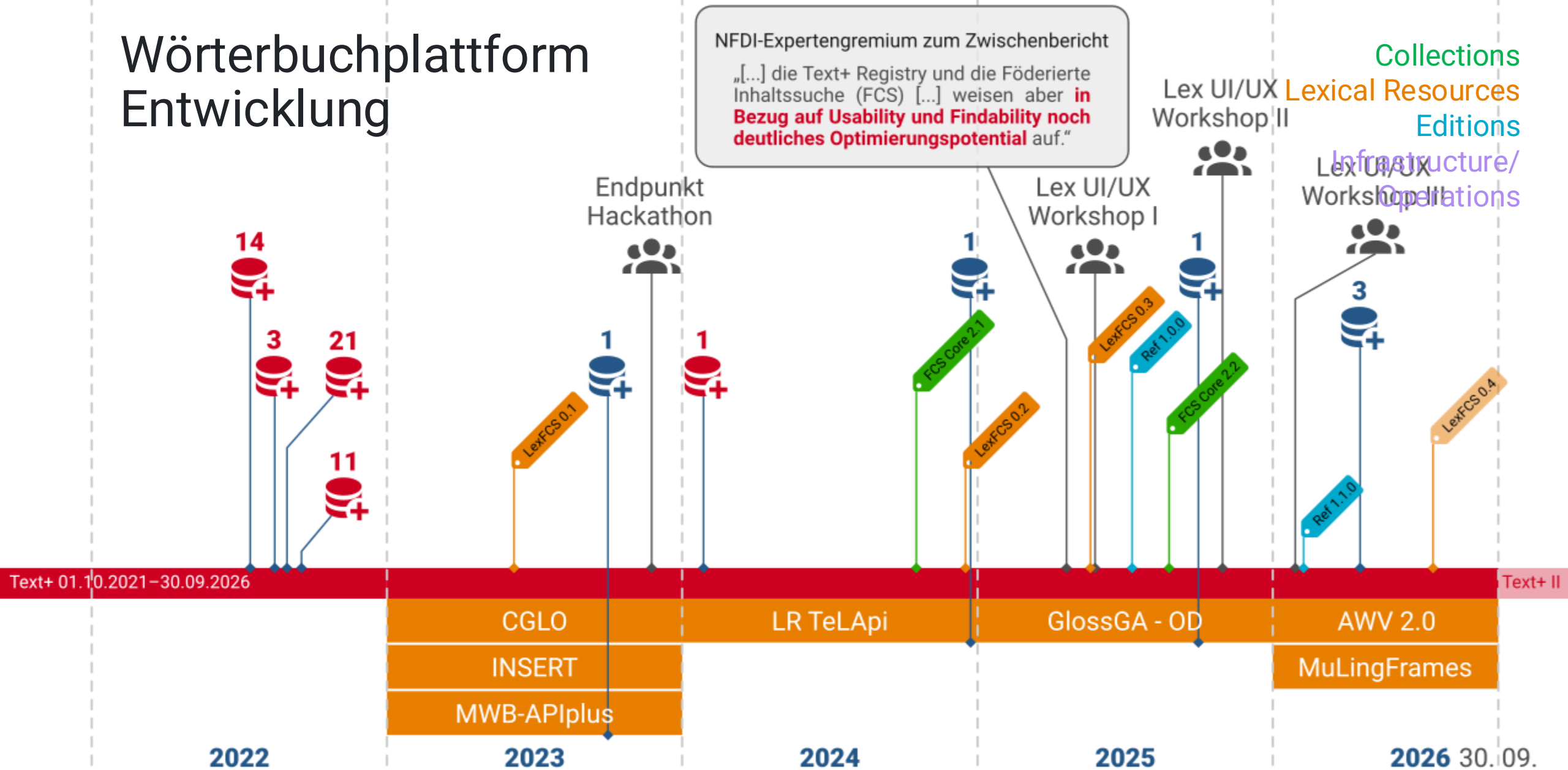


Text+ 01.10.2021-30.09.2026

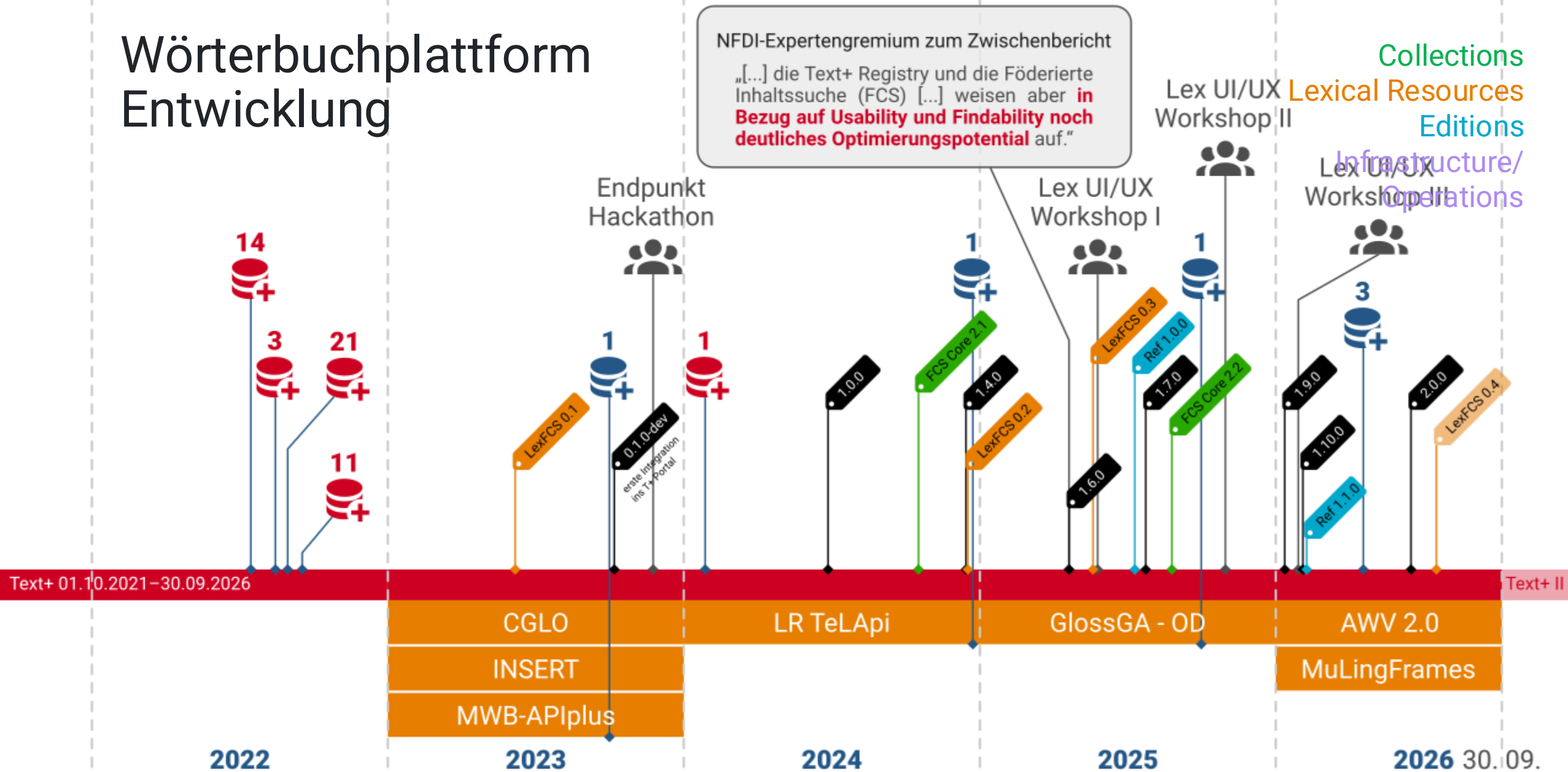
Text+ II



Wörterbuchplattform Entwicklung



Wörterbuchplattform Entwicklung



NFDI-Expertengremium zum Zwischenbericht
 „[...] die Text+ Registry und die Föderierte Inhaltssuche (FCS) [...] weisen aber **in Bezug auf Usability und Findability noch deutliches Optimierungspotential** auf.“

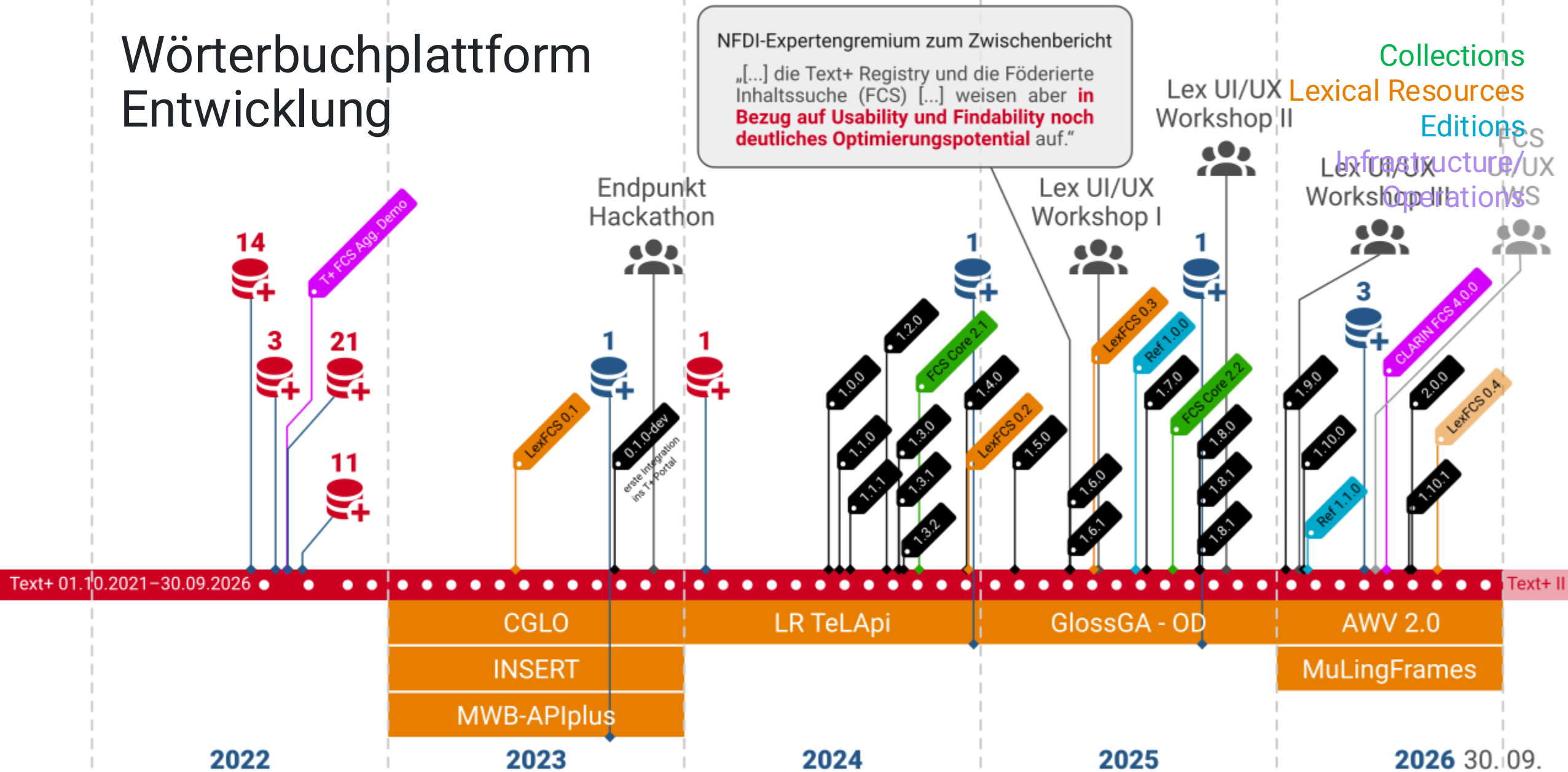
Collections
 Lexical Resources
 Editions
 Infrastructure/
 Operations

Text+ 01.10.2021-30.09.2026

Text+ II



Wörterbuchplattform Entwicklung



Live-Demo FCS-Frontend

<https://text-plus.org/fcs?qtype=lex>

([Testumgebung](#))



Live-Demo FCS-API

[Hier zum Jupyter Notebook](#)

([Testumgebung](#))



Die Task Area Lexical Resources in Text+

- Allgemeiner Überblick
- Kooperationsprojekte und SCC
- Gemeinsame Wörterbuchplattform
- **Die TA im Gesamtprojekt**
- KI als neues Themenfeld
- Ausblick



LexRes als Teil von Text+

- Datendomänen/Task Areas sind keine Silos
- Kooperation in AGs, durch Koordinations- und Steuerungsgruppe
- Ideen und Lösungen gemeinsam erarbeiten und teilen oder nachnutzen
- Task Areas haben Einfluss auf das Gesamtprojekt
- ... und auch darüber hinaus



Abbildung gemeinfrei



FCS: Anfrageunterstützung für Korpora

Föderierte Inhaltssuche

Volltextsuche Erweiterte Suche Lexikalische Suche

Suche Belegstellen für das Wort "Auto" gefolgt von einem Verb oder einem Adjektiv!



SUCHE STARTEN

ZURÜCK ZU DEN VORFILTERN

ZURÜCKSETZEN

Die Anfrage "Auto" [pos = "VERB|ADJ"] wurde automatisch generiert

Ihre Suche nach **Suche Belegstellen für das Wort "Auto" gefolgt von einem Verb oder einem Adjektiv!** erzielte **11 Ergebnisse** in **2 Ressourcen**

Suche unterbrochen: Es konnten 117 der 445 angefragten Ressourcen durchsucht werden

10

Europarl14



Mit den Aufrufen , als Alternative zu dem *Auto öffentliche* Verkehrsmittel zu nutzen , zu Fuß zu gehen oder das Fahrrad zu nehmen , müssen auch Maßnahmen zu der Sen ...

<https://hdl.handle.net/11022/0000-0007-FCD4-E>

Dies ist einer der wenigen Vorschläge , die jeden Bürger in der Europäischen Union ganz unmittelbar in seinem tagtäglichen Leben betreffen , und zwar in einer sehr empfi ...

<https://hdl.handle.net/11022/0000-0007-FCD4-E>

Ich musste auf dem Seeweg und mit der Bahn , mit Taxis und in dem *Auto reisen* , um hier nach Straßburg zu gelangen , was mehr als 24 Stunden dauerte , in denen ich ...



FCS: Ressourcenspezifische Beispielanfragen

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Föderierte Inhaltssuche

Volltextsuche Erweiterte Suche Lexikalische Suche

45 [SUCHE STARTEN](#)

[ZU DEN SUCHERGEBNISSEN →](#) [ZURÜCKSETZEN](#)

i Die erweiterte Suche über Annotationslayer erlaubt die Abfrage von annotierten Textkollektionen, -korpora und vergleichbarer Ressourcen über die Anfragesprache FCS-QL, die sich an der populären Korpusanfrage-Sprache CQP orientiert. Dabei können für Wörter im Text komplexe Suchkriterien angegeben und diese für längere Textstellen miteinander kombiniert werden.

Sächsische Akademie der Wissenschaften zu Leipzig

Briefe und Akten zur Kirchenpolitik Friedrichs des Weisen und Johannis des Beständigen 1513 bis 1532. Reformation im Kontext frühneuzeitlicher Staatswerdung

Das Akademievorhaben „Briefe und Akten zur Kirchenpolitik Friedrichs des Weisen und Johannis des Beständigen 1513 bis 1532. Reformation im Kontext frühneuzeitlicher Staatswerdung“ sammelt erstmals das Quellenmaterial zur Kirchenpolitik dieser beiden Reformationsfürsten. Dabei wird unter Kirchenpolitik das planvolle Handeln und Reagieren der Landesherren in kirchlichen Belangen verstanden, das sie unter Umständen auch an ihre Räte oder Verwaltungseliten übertragen konnten.

[Entität](#) [Textuelle Darstellung](#) [Deutsch](#)

[ZEIGE WENIGER](#) [ÖFFNE RESSOURCEN-WEBSEITE](#)

2 Suchvorschläge

- ["Melanchthon"](#)
Suche nach Belegen für das Wort "Melanchthon"
- [\[entity = "4035206-7" \]](#)
Suche nach Belegen für die Stadt Leipzig auf Basis der entsprechenden GND-Entitäten-ID

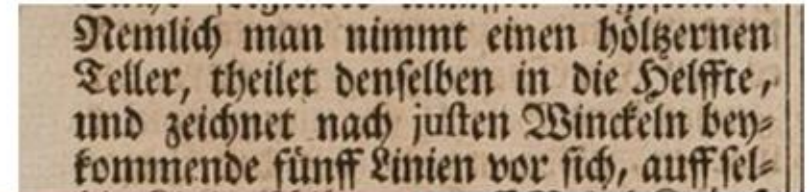


Seit 11/2025 Teil der FCS Core Spezifikation



Orthografische Normalisierung

- moderner Werkzeuge setzen meist moderne Orthografien voraus
- **Transnormer**, moderne Implementation zur orthografischen Normalisierung
- Idee: maschinelle Übersetzung
- Umsetzung: neuronale Modelle, Transformer-Architektur
- Training, Feintuning auf DTA-Daten
- Trainingsdaten und Modelle: huggingface.co/textplus-bbaw



Original

Nemlich man nimmt einen **höltzernen** Teller, **theilet** denselben in die **Helffte**, und zeichnet nach **juften Winckeln** **beykommende fünff** Linien vor sich [...].

Normalisierung

Nämlich man nimmt einen **hölzernen** Teller, **teilt** denselben in die **Hälfte**, und zeichnet nach **justen Winkeln** **beikommende fünf** Linien vor sich [...].



Anforderungsanalyse über Personas

Collections
Lexical Resources
Editions
Infrastructure/
Operations

- Workshop zur Entwicklung von Personas als repräsentative Nutzungsprofile
- Identifikation von Anforderungen, *Pain Points* und Potenzialen
- Wichtige Grundlage für eine nutzerorientierte Weiterentwicklung
- Erkenntnisse steuern die zukünftige Entwicklung der Plattform



Die Editionsphilologin
historische Textforschung

- arbeitet mit kritischen Editionen, Varianten, Metadaten
- braucht Registry Editions + FCS einfache Suche für Volltextstellen
- sucht oft nach Versionen, Editionstypen, editorischen Prinzipien



Heterogenität der Metadaten, unterschiedliche Editionsmodelle



Der Korpuslinguist
quantitative Analyse

- arbeitet mit großen Textsammlungen, Annotationen, Token-Informationen
- nutzt FCS erweiterte Suche (Annotationen!), ggf. Registry Collections
- braucht stabile APIs, Metadaten-harmonisierung, Exportformate



Unterschiede in Annotationstiefen, fehlende Filtermöglichkeiten



Die Lexikografin

- sucht nach Wörterbüchern, Lexika, Einträgen, Bedeutungsvarianten
- primär: Registry Lex + FCS lexikalische Suche
- interessiert an zeitlichen Entwicklungen, Varianten und Quellen



unklare Abgrenzung zwischen lexikalischer Suche und Volltextsuche

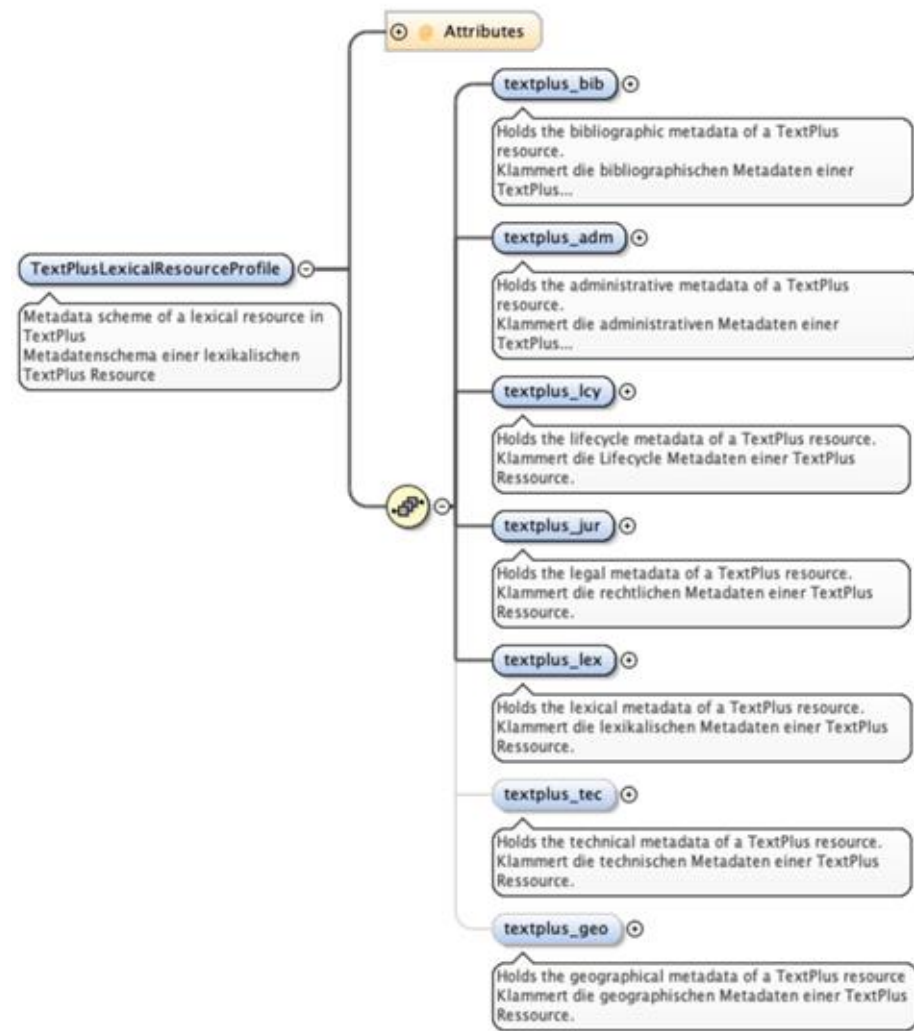


Workshop 26.01.2026, Berlin



Ein Metadatenmodell für Lexikalische Ressourcen

Collections
Lexical Resources
Editions
Infrastructure/
Operations



- 7 Institute liefern Metadaten an die Registry, in jeweils eigenen Formaten
 - Erschwert das Einspielen von Metadaten in die Registry (komplexer Mapping-Prozess).
 - Auch andere TAs liefern ihre Metadaten in diversen Formaten ab, was den Prozess weiter erschwert.
- ❖ Entwicklung eines **einheitlichen** Metadatenprofils für lexikalische Ressourcen
 - Verwendung **Komponenten**-basierter Metadaten (CMDI)
 - klare **Trennung** zwischen Komponenten, die unabhängig vom Ressourcentyp sind, und Komponenten, die Attribute für lexikalische Ressourcen enthalten.
 - **Wiederverwendbarkeit** für andere TAs
 - ❖ Sieben Hauptkomponenten
 - bibliografisch, administrativ, rechtlich, lifecycle, geografisch, technisch, **lexikalisch**
 - ❖ Umstellung bereits im [Gange](#)



Europäischer Einfluss

Collections
Lexical Resources
Editions
Infrastructure/
Operations

1. Verwendung von CMDI als Metadateninfrastruktur
 - a. definiert als ISO-Standard 24622-1 und 24622-2
 - b. große Verbreitung innerhalb von CLARIN-EU
 - c. basiert auf XML Technologie (Tool support; Validierung)

1. Weiterentwicklung von CLARIN Technologie:
 - a. FCS zur Inhaltssuche in Korpora
 - b. LexFCS zur Inhaltssuche innerhalb lexikalischer Ressourcen

1. Verwendung von TEI Lex-0
 - a. „Basisformat“ zum Austausch lexikalischer Daten
 - b. Beteiligung bei der Weiterentwicklung (organisiert durch DARIAH-EU)
 - c. breite Anwendung in vielen (Retro-)Digitalisierungsprojekten



TC 37 SC 4



Die Task Area Lexical Resources in Text+

Collections
Lexical Resources
Editions
Infrastructure/
Operations

- Allgemeiner Überblick
- Kooperationsprojekte und SCC
- Gemeinsame Wörterbuchplattform
- Die TA im Gesamtprojekt
- **KI als neues Themenfeld**
- Ausblick



Umgang mit KI / großen Sprachmodellen

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Text+ muss sich dem Thema stellen

- stand nicht im Arbeitsplan
- Text+ als NDFI Konsortium prädestiniert

Beispiele:

- LLM-Generierung von Beispielsätze für GermaNet
- LLM-basierte Erzeugung von Definitionen
- LLM-gestützte Verbesserung der Usability



Nutzung großer Sprachmodelle in GermaNet

Collections
Lexical Resources
Editions
Infrastructure/
Operations

- lexikalisch-semantisches Wortnetz
- v20: 231.500 Einträge
- Lexikonarbeit ist sehr anspruchsvoll

- Naheliegende Anwendung von LLMs: Generierung von Beispielsätzen
 - typisch, natürlich, informativ, verständlich
 - Atkins & Rundell, 2008
- in GermaNet bis jetzt aufwändig konstruiert

Respekt n. (Gefühl) 'respect n. (emotion)'

Paraphrase: *leichte Angst vor etwas*
'slight fear of something'

Example: *Er hatte vor Schlangen großen Respekt.*
'He had great respect for snakes.'

Hypernyms: *Angst, Bammel, Muffe, Schiss...*
'fear, apprehension, nervousness, dread...'

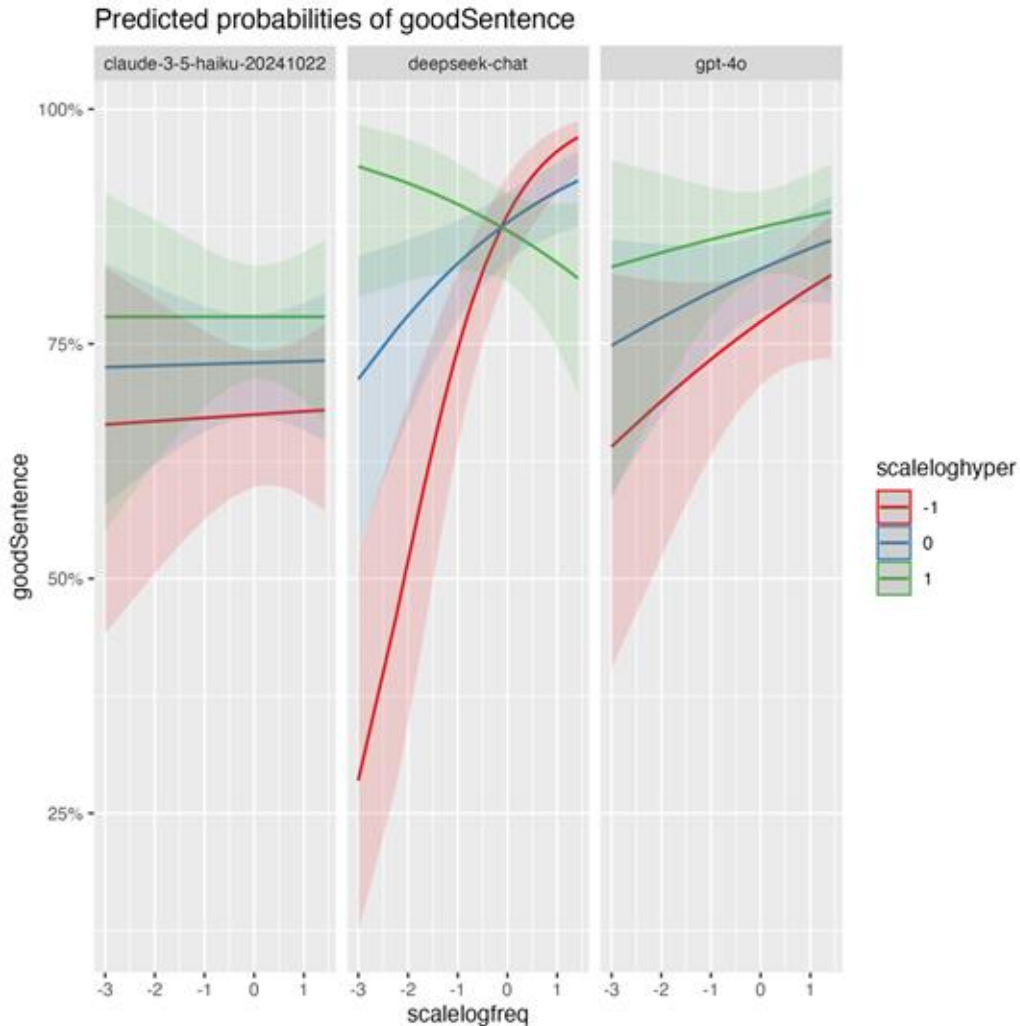
Hyponyms: *Heidenrespekt*
'tremendous respect'

Studie:

- Fokus auf Beispielsätze für **Nomen**
 - fast nicht in GermaNet vertreten
- Wahl eines repräsentativen Ausschnitts aus GermaNet
 - 100 monoseme Lemma
 - 300 Wortbedeutungen, diverse Polysemiegrade
- Evaluation mehrerer LLMs
- Generierung von jeweils 3 Beispielsätzen / LLM
- jeder Beispielsatz wird von 2 Annotatorinnen bewertet



Evaluation



- 900 Sätze für **monoseme** Lemmata, davon akzeptabel:
 - 84% (gpt-4o)
 - 88% (deepseek-V3-0324)
 - 75% (claude-haiku-3.5)
 - Lemma und Hyperonym Frequenzen haben Einfluss auf die Satzqualität
 - die 3 Sprachmodelle verhalten sich unterschiedlich
 - ignorant vs. additiv vs. trade-off
- 2700 Sätze für **polyseme** Lemmata, davon akzeptabel:
 - 72% (gpt-4o)
 - 74% (deepseek-V3-0324)
 - 61% (claude-haiku-3.5)



Konklusion

- Qualität der LLMs bereits mehr als ausreichend
 - 2 von 3 Beispielsätzen sind akzeptabel
 - aber **manueller Vetting-Prozess** ist nötig
 - grosse **Zeitersparnis**: statt (kreative) Satz-Erzeugung nur Satz-Selektion
 - Einbau von LLMs in **GernEdit** zur Unterstützung unserer Lexikographinnen
- Weitere Arbeiten zur Verwendung von LLM
 - Erkenne Wortbedeutung, die in einem gegebenen Beispielsatz verwendet wird
 - Gebe mir alle bekannten Wortbedeutungen für ein gegebenes Lemma



ZDL-Vergleichsstudie: Definitionen per KI?

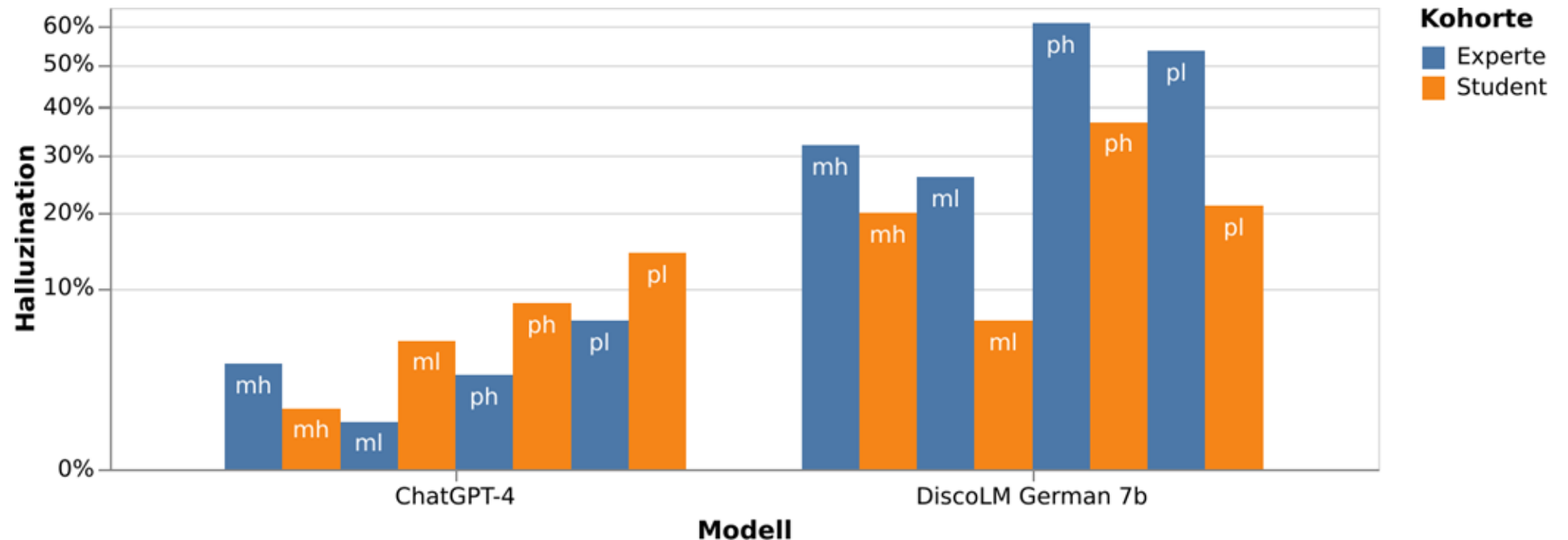
Können LLMs gute Definitionen auf der Basis guter Korpusbelege erzeugen?

- Setup
 - Menge von guten Belegen (GDEX)
 - Prompt: präzise Anforderungen an Definitionen
- Evaluation
 - Monosemie versus Polysemie
 - hoch- versus niedrigfrequente Wörter
 - ChatGPT- versus DWDS-Definitionen
 - Halluzinationen?
- Bewertung durch 10 Expert:innen und 15 CL-Studierende



ZDL-Vergleichsstudie: Definitionen per KI?

Beispiel: Halluzinationen



Fazit: sinnvoller Einsatz als »Copilot« möglich, aber nicht autonom



Erweiterter Einsatz von KI für Usability

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Föderierte Inhaltssuche

Volltextsuche Erweiterte Suche Lexikalische Suche

Ich suche Informationen zur Wortherkunft von "Auto!"

SUCHE STARTEN

ZURÜCKSETZEN

56 verfügbare Ressourcen

Zur Beschleunigung der Suche können Sie die gewünschten Ressourcen vorab auswählen.

3/56 Ressourcen

- Deutsche Wortfeldetymologie in europäischem Kontext
- Etymologisches Wörterbuch des Althochdeutschen
- Etymologisches Wörterbuch des Deutschen

3 Sprachen

Mockup



Die Task Area Lexical Resources in Text+

- Allgemeiner Überblick
- Kooperationsprojekte und SCC
- Gemeinsame Wörterbuchplattform
- Die TA im Gesamtprojekt
- KI als neues Themenfeld
- **Ausblick**



Ausblick auf die zweite Phase

- Fortführung und Erweiterung der Arbeiten
 - abgestimmter Zentrenbetrieb mit allen Text+Partnern
 - LexFCS, Registry weiterentwickeln; enger integrieren
 - Ziel: »Vom Katalog zur Gemeinsamen Forschungsplattform« (One-Stop-Shop):
- Datenertüchtigung und datendomänenübergreifende Vernetzung
- KI und Lexikalische Ressourcen
 - “AI-Readiness” der Daten zur Verbesserung der Nachnutzung (MCP-Protokolle)
 - semantischer Annotationservice: Normdaten- und Lesarten-sensitive Annotation in Volltexten
- noch stärkerer Fokus auf Nutzerbedürfnisse ...



Ausblick auf die zweite Phase

- noch stärkerer Fokus auf Nutzerbedürfnisse
 - Dissemination and outreach activities
 - Training and education
 - Consulting, helpdesk and support...
- noch stärkere datendomänenübergreifende Zusammenarbeit
 - engere Abstimmung in allen Measures
 - Anwendungsfall: nahtloser Übergang von Wörterbucheinträgen zu Korpus-Belegen
 - Vernetzung von WB-Informationen (Grammatik, Bedeutung, Belege) mit Editionen
- Kooperation mit weiteren NFDI-Konsortien intensivieren, besonders in den Bereichen
 - Metadaten
 - Knowledge-Graphen
 - Helpdesk und Consulting



Ausblick auf die zweite Phase

Vergrößerte TA “Lexical Resources” in Phase 2:

- alle bisher aktiven Partner (gefördert):
 - BBAW, DSA (Koord);
 - IDS, SAW, Uni(Köln|Trier|Tüb)
- weitere Partner (nicht gefördert):
 - Bayerische Akademie der Wissenschaften
 - Universität Rostock
 - Akademie der Wissenschaften in Hamburg
 - Niedersächsische Akademie der Wissenschaften zu Göttingen
 - Heidelberger Akademie der Wissenschaften



Vielen Dank!



Text+

Plenary
2026



@textplus_NFDI
#TextPlusPlenary

5. TEXT+ PLENARY

PROGRAMM

Dienstag, 16. Juni 2026

09:00 – 10:30	Ergebnisse aus Task Area: Lexical Resources (Aula)
<i>10:30 – 11:00</i>	<i>Kaffeepause (Katakomben)</i>
11:00 – 12:30	Ergebnisse aus Task Area: Collections (Aula)
12:30 – 13:00	Abschluss und Ausblick auf die zweite Phase (Aula)
<i>13:00 – 14:00</i>	<i>Lunch (Katakomben)</i>
14:00 – 15:30	<i>Parallele Sitzungen (intern, nur auf Einladung)</i>
	Gemeinsames Treffen des Scientific Coordination Committees Lexical Resources und der Task Area Lexical Resources (Aula)
	Treffen des Scientific Coordination Committees Collections (Dozentenzimmer, O 126)
	FID Koop Treffen (Musikerzimmer, EO 154)



Collections
Lexical Resources
Editions
Infrastructure/
Operations

The Collections collection

Gefördert durch

DFG Deutsche
Forschungsgemeinschaft

Förderungsnummer 460033370

Teil der

nfdi Nationale
Forschungsdaten
Infrastruktur

Führung durch die größten Meisterwerke
der TA Collections aus fünf Jahren Text+

16. Juni 2026, Mannheim, 11:00 Uhr





Collections
Lexical Resources
Editions
Infrastructure/
Operations

made with Canva AI



Eigene Daten und Daten Dritter: Collections Zentren

Collections
Lexical Resources
Editions
Infrastructure/
Operations



universität freiburg



BAYERISCHE
AKADEMIE
DER
WISSENSCHAFTEN



HEIDELBERGER AKADEMIE
DER WISSENSCHAFTEN
Akademie der Wissenschaften
des Landes Baden-Württemberg



IDS | LEIBNIZ-INSTITUT FÜR
DEUTSCHE SPRACHE



UNIVERSITÄT
ZU KÖLN

SUB | NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN



UNIVERSITÄT
DES
SAARLANDES



Eigene Daten und Daten Dritter: Registry

Text+ Startseite

Registry - Der Text+ Katalog

Die [Text+ Registry](#) ist ein Recherche- und Informationssystem, das text- und sprachbasierte Forschungsdaten verzeichnet und sie in einer zentralen Übersicht zusammenführt.

Übergreifender Katalog Editionen **Korpora- und Textsammlungen** Lexikalische Ressourcen

Suchbegriff eingeben oder alles durchsuchen Suche

Suchbeispiele:

- [Suche nach einem deutschsprachigen Briefkorpus, das im Volltext und unter einer offenen Lizenz bereitgestellt wird](#)
- [Im Rahmen eines neuen Forschungsprojekts zur Popularisierung von Wissen suche ich nach deutschsprachigen Zeitungen aus dem 19. Jahrhundert im Volltext.](#)
- [Für eine Studie zum Thema Bilingualität suche ich entsprechende nachnutzbare..gesprochensprachige Korpora inkl. Volltext](#)



Eigene Daten und Daten Dritter: Registry

Suchbeispiele

Collections
Lexical Resources
Editions
Infrastructure/
Operations

The screenshot shows the Text+ Registry interface. At the top left is the Text+ logo and 'Startseite'. On the right are icons for help, settings, and language (DE). The main heading is 'Registry - Der Text+ Katalog'. Below it are four tabs: 'Übergreifender Katalog', 'Editionen', 'Korpora- und Textsammlungen' (highlighted in green), and 'Lexikalische Ressourcen'. A search bar contains the word 'Brief' and a 'Suche' button. Below the search bar, it says '5 von 5 Ressourcen'. On the left is a 'Filter' sidebar with categories: 'Sprache' (Deutsch (5)), 'Datentyp', 'Lizenz' (public (5)), 'Volltext Verfügbar' (vorhanden (5)), 'Modalität', and 'Importquelle'. The main content area shows two search results. The first is 'Soldatenbriefe des 18. und 19. Jahrhunderts', described as a corpus of 170 letters from officers, sub-officers, and soldiers between 1745 and 1872. The second is 'Briefedition Jean Paul', described as a corpus of 5004 letters by Jean Paul from 1780, addressed to 627 recipients. Both results include tags for 'Deutsch', 'text', 'public', 'vorhanden', 'geschrieben', and 'DTA/DWDS'.



Eigene Daten und Daten Dritter: Registry

1497 Ressourcen

Sprache

deu · eng · ara · spa · fra · ukr · sel · ita · rus · lat · pol · kat · bul
· kir · ces · afr · fin · ben · jpn · ell

Datentyp

Text · Audio · Bild · Video · Tabelle

Lizenz

frei · frei für akademische Zwecke · beschränkt · unbestimmt

Volltext

vorhanden · teilweise vorhanden · nicht vorhanden

Modalität

geschrieben · gesprochen · multimodal · gebärdet

Importquelle (Text+ Zentrum)

IDS · TALAR · oh.d · UHH-HZSK · DNB · DTA/DWDS · BAS ·
UdS · TG-rep · AdWHH · LAC

Collections
Lexical Resources
Editions
Infrastructure/
Operations

☰ Filter

Sprache ▾

Datentyp ▾

Lizenz ▾

Volltext Verfügbar ▾

Modalität ▾

Importquelle ▾



Eigene Daten und Daten Dritter: Registry-Datenmodell


inhaltliche Metadaten

- Titel/Name der Sammlung
-  Lizenz, Lizenz-URL
-  Sprache
-  fachliche Zuordnung
- Schlagworte
- ...

technische Metadaten

- PID
- Zugangsinformation
- Hierarchie
- Bezug zu anderen Sammlungen
- ...

organisatorische Metadaten

- Datenverantwortliche Person/Institution
- Ansprechperson für Datenzugang
-  Förderer
- ...

Legende: = Pflicht, = optional,  = kontrolliertes Vokabular



Eigene Daten und Daten Dritter: FCS- Ressourcen



Collections
Lexical Resources
Editions
Infrastructure/
Operations

Map base: NordNordWest –
„Germany location map.svg”,
Wikimedia Commons, CC BY-SA 3.0.
Modified (added markers and
labels).



Eigene Daten und Daten Dritter: Data Depositing

Beispiele für neue Ressourcen

- User Stories
 - [CLiGS: Textbox](#)
 - [German IT-Blogs](#)
 - [NINNY](#)
- Kooperationsprojekte
- Andere
 - [Börsenblatt für den deutschen Buchhandel](#)
 - [European Literary Text Collection](#)
 - [Multilingual Parallel Bible Corpus](#)



Eigene Daten und Daten Dritter: Data Depositing

Beispiele für neue Ressourcen (**Registry**)

CLARIND-UdS:

- Sprachmodelle auf Silben- und Phonebene

The screenshot shows a web browser window with the URL `c102-142.cloud.gwdg.de/registry-search/doc/collection/5640feea-43e8-47e5-8b95-98fef09d5215`. The page header includes the Text+ logo and 'Startseite'. The main content area displays the title 'N-gram Language Models based on DeWaC German web corpus' under the category 'Korpus / Textsammlung'. There is a 'Ressourcen Links' button and three main action buttons: 'Metadaten', 'Empfehlungen', and 'Graph'. Below these, two panels show metadata details:

Titel eng N-gram Language Models based on DeWaC German web corpus	entityId 5640feea-43e8-47e5-8b95-98fef09d5215
Beschreibung eng The resource is a set of language models at syllable and phone level. The context length (n-gram length) of the models ranges from 1 to 4 for syllable level and 1 to 6 for phone level. For each n-gram length,...	sourceId c1c9b626-0a08-4962-9a02-04fd60f7cd5f
Sprache Deutsch	PID https://hdl.handle.net/hdl:21.11119/0000-0008-4141-2
Größe 15 Gigabyte	Zugang https://fedora.clarin-d.uni-saarland.de/sfb1102/#nlm-dewac
Lizenz public	Ansprechperson CLARIND-UdS: Repitorium für Sprachressourcen an der Universität des Saarlandes https://fedora.clarin-d.uni-saarland.de/
	Datentyp



Eigene Daten und Daten Dritter: Data Depositing

Beispiele für neue Ressourcen (FCS)

SUB/TG-rep:

- American Drama Corpus als ATF

Collections
Lexical Resources
Editions
Infrastructure/
Operations

The screenshot shows the Text+ search interface. At the top, there is a search bar with 'German' entered. Below the search bar, there are filters for '188 Sprachen' and '445 Ressourcen'. A progress bar indicates '445/445 Ressourcen', '0 ausstehend', and '57 passend'. The search results section shows 'Ihre Suche nach German erzielte 620.412 Ergebnisse (340 geladen) in 57 Ressourcen'. The results are displayed in a list format, with the first result being 'American Drama Corpus'. The list includes the following items:

- American Drama Corpus (10 items)
 - A collection of American drama texts focusing on the structural markup.
 - TextGrid Repository, Hosted: ["en": "TextGrid Repository"]
 - Ninette — Neat costume of a German peasant (ÖFFNEN | hdl:21.11113/0000-0014-A703-A)
 - In Fancy DANCING MASTER Drake and Ducks GERMAN (ÖFFNEN | hdl:21.11113/0000-0014-A1C0-A)
 - MYNHEER VON GOUTYBERGEN — German uniform. (ÖFFNEN | hdl:21.11113/0000-0014-A853-F)
 - Edna (with a broken German accent). (ÖFFNEN | hdl:21.11113/0000-0014-A460-4)
 - German . Mr. G. C. German. Mr. J. (ÖFFNEN | hdl:21.11113/0000-0014-A364-1)
 - A German Couple . A Guide . A Driver . (ÖFFNEN | hdl:21.11113/0000-0014-A667-B)



Eigene Daten und Daten Dritter: Data Depositing

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Kooperationsprojekte

Projekt	Registry	FCS	Text+ Blog
oh.d	oh.d		oh.d
KOLIMO+	KOLIMO+	kolimo+	
DigiMusTh	DigiMusTh	DigiMusTh (über DTAXL)	DigiMusTh
BeriaCollection	BeriaCollection		
DiPA+		DiPA+	
DLA Data+			DLA Data+
Diskmags			Diskmags



Integration von GRETIL in TextGrid Repository

1. TextGrid Repository
2. XML-TEI Dokumente und ihre Digitalisate
3. Seit 20 Jahren TextGrid
4. User Stories

Collections
Lexical Resources
Editions
Infrastructure/
Operations



TextGrid

Virtuelle Forschungsumgebung
für die Geisteswissenschaften



Integration von GRETIL in TextGrid Repository

Collections
Lexical Resources
Editions
Infrastructure/
Operations

- Fluffy Publication Workflow
- Zusammenarbeit zwischen Collections und IO
- Forschenden müssen keine neue Technologie oder Metadatenmodell lernen
- Ziele
 1. Daten veröffentlichen
 2. Metadaten nebenbei anreichern



Integration von GRETIL in TextGrid Repository

Collections
Lexical Resources
Editions
Infrastructure/
Operations

KOLIMO+



Distant [📖] Reading



F FOLGER
SHAKESPEARE
LIBRARY



MODERN
POETRY



GRETIL
Prākṛtam
Tamil
Samskr̥tam
Kawi
Pāli
Hindi
Marāṭhi
Bod̥skad̥
Malāyālam



Integration von GRETIL in TextGrid Repository

Collections
Lexical Resources
Editions
Infrastructure/
Operations

- GRETIL
- Göttingen Register of Electronic Texts in Indian Languages
- Seit 2001
- Wichtiges indologisches Korpus (Sanskrit)
- <https://gretil.sub.uni-goettingen.de>
- User Story dazu (Jonas Buchholz)



Integration von GRETIL in TextGrid Repository

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Probleme bis jetzt

- Keine langfristige Strategie
- Heterogenität (Kodierung, Format, Metadaten...)
- TEI, HTML, Plaintext, mit mehreren Dateien
- Ad-hoc-Lösungen
- Keine kontrollierte Metadaten
- Fehlende Funktionalitäten



Integration von GRETIL in TextGrid Repository

Collections
Lexical Resources
Editions
Infrastructure/
Operations

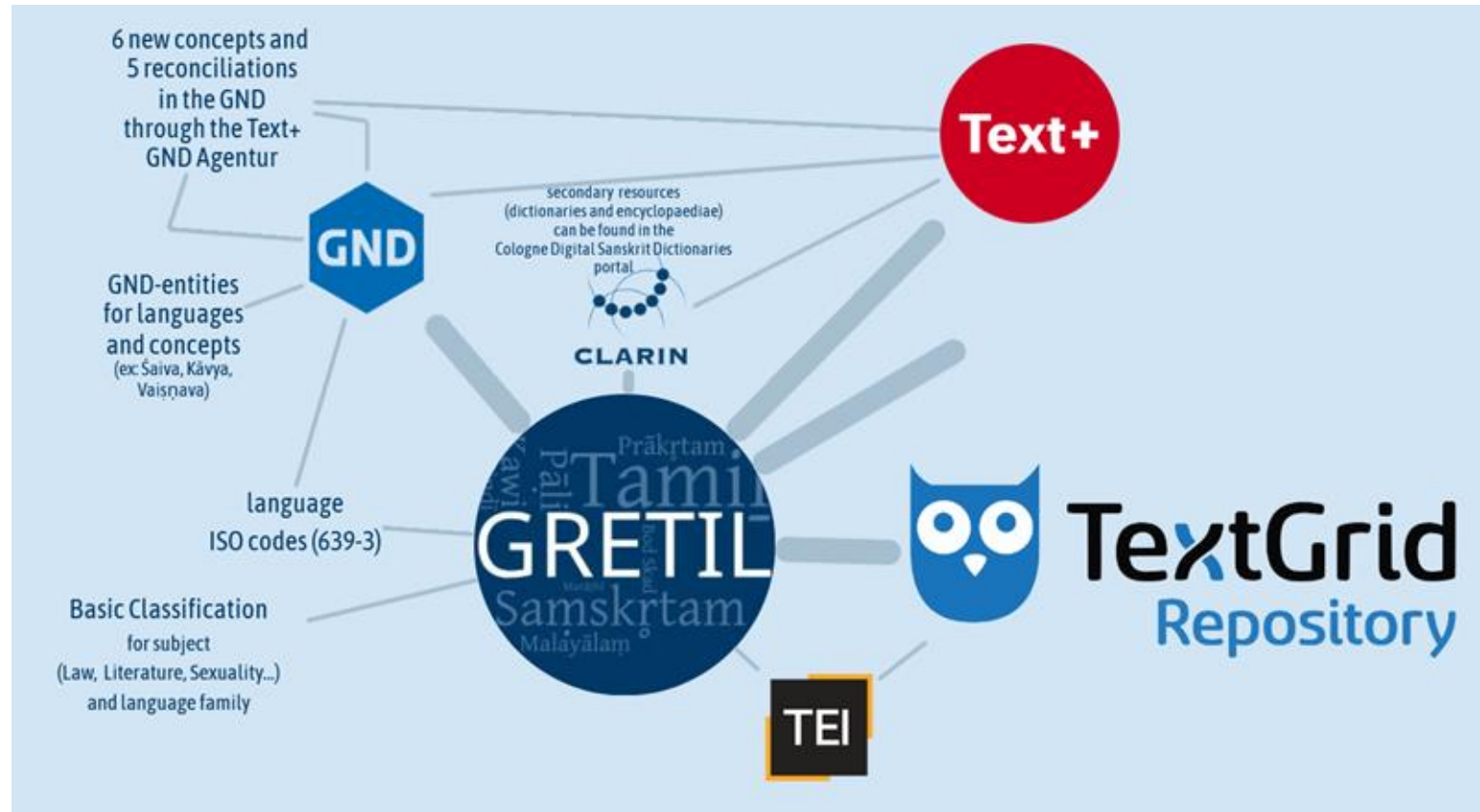


- FAIR-ification
- Anreicherung von Metadaten durch Standards
- Integration in Repositorien



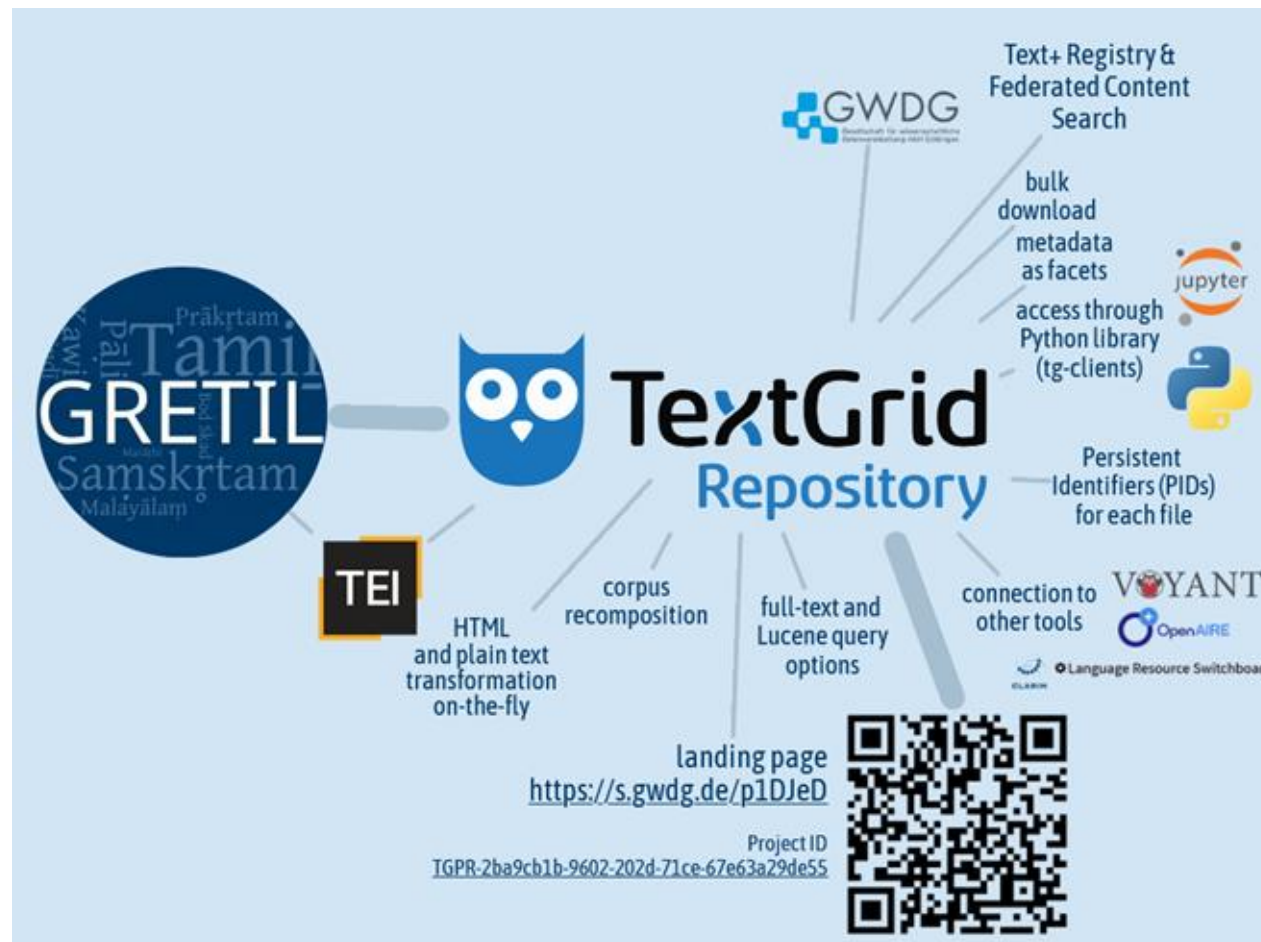
Integration von GRETIL in TextGrid Repository

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Integration von GRETIL in TextGrid Repository

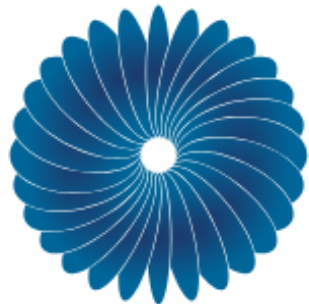
Collections
Lexical Resources
Editions
Infrastructure/
Operations



Integration von GRETIL in TextGrid Repository

- Integration von bereits vorhandenen Ressourcen
- Verbesserung der Daten
- TextGrid Repository ist vielfältiger
- Ressourcen sind leichter zu warten
- Data-Paper bei *Transformations*

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Transformations: A DARIAH Journal



Kooperationsprojekt „Text+oh.d“

- **Dauer:** Januar bis November 2025
- **Durchführung:** Universitätsbibliothek der FU Berlin, Abteilung „Forschungs- und Publikationsservices“, Team „Digitale Interviewsammlungen“
- **Projektteam:** Cord Pagenstecher (Projektleitung), Peter Kompiel (Projektmanagement), Thomas Schmidt und Christian Gregor (externe Entwickler)
- **Website:** https://www.fu-berlin.de/text_ohd

 Die Freie Universität Berlin ist nun Daten- und Kompetenzzentrum von Text+



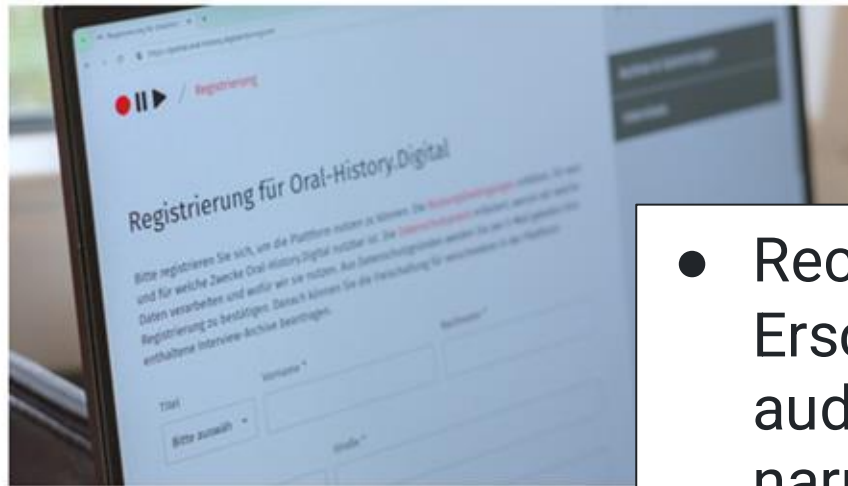


Oral-History.Digital

Erschließungs- und Recherche-Plattform für Audio- und Video-Interviews mit Zeitzeuginnen und Zeitzeugen

Institutionen

Archive und Sammlungen



- Recherche- und Erschließungsumgebung für audiovisuell aufgezeichnete narrative Interviews
- 51 Archive, 147 Sammlungen und 5.441 Interviews
- <https://portal.oral-history.digital>

Suche in den Interviews

Die Suche führt zu einzelnen Interviews aus verschiedenen Archiven. Filtern können Sie z. B. nach Thema, Sprache oder Geschlecht. Im Volltext können Sie nur in den für Sie freigeschalteten Archiven suchen.

[Interviews →](#)

Zugang zu den Interviews

Um die Persönlichkeitsrechte der Interviewten zu schützen, vor dem Zugriff auf die Archive eine Anmeldung. Bitte registrieren Sie sich im Portal und erhalten dann eine Freischaltung für einzelne Archive.

[Anmelden](#)

Metadatenachweis in der Registry

- Nachweis der Archive und Sammlungen in der Text+-Registry
- Entwicklung einer OAI-PMH-Schnittstelle
- Übermittlung der öffentlichen Metadaten im DataCite-XML-Format

The screenshot shows the Text+ Registry interface for the 'Archiv „Deutsches Gedächtnis“' collection. The page is titled 'Startseite' and includes a 'Text+' logo. The main content area displays the collection name and a 'Korpus / Textsammlung' label. Below this, there are buttons for 'Ressourcen Links', 'Metadaten', 'Recommendations', and 'Graph'. The 'Metadaten' section is expanded, showing the following information:

Titel	entityId
deu Archiv „Deutsches Gedächtnis“	b7708699-a440-4851-9ee7-725a8eb4fe9b
eng Archiv „Deutsches Gedächtnis“	
Beschreibung	sourceId
deu Im Archiv „Deutsches Gedächtnis“ werden subjektive Erinnerungszeugnisse wie lebensgeschichtliche Interviews, Autobiographien, Tagebücher und Briefsammlungen ganz...	42aa0fa9-5f3e-48b1-b287-d24c8f344313
eng The Archiv "Deutsches Gedächtnis" (Archive "German Memory") holds personal memories such as biographical interviews, autobiographies, diaries and collections of letters from different...	PID
	https://portal.oral-history.digital.de/catalog/archives/21894736
Größe	Zugang
1332 Interviews	https://portal.oral-history.digital.de/catalog/archives/21894736
Lizenz	Kollektionstyp
restricted for individual	Interviewsammlung
	Datentyp
	audio

Das Archiv „Deutsches Gedächtnis“ in der Text+-Registry,
<https://registry.text-plus.org/doc/collection/b7708699-a440-4851-9ee7-725a8eb4fe9b>



TEI-XML-Transformation von Transkripten



Dr. Günter Möstl

★ Interview merken Position kopieren

Transkript (ger) Inhaltsverzeichnis Suche im Interview Registereinträge

Und da wirzeiten wir nicht, sagte er: „Ah, das sind meine Grenzübergangsschuhe.“

<g(Gestikulieren) Das war ein Scherz alles.>

Und ähm, ja, wir sind, äh, locker auseinandergegangen: „Macht es gut. Also seht mal zu und_“

Nicht, dass das ein Abschied für immer ist oder irgendwas.

<g(Gestikulieren) <s(laut) Sie hatten uns, äh,> noch Gepäck, Wäsche und alles, mitgegeben.>

Sollten wir den Eltern abgeben.

Das war auf, äh, jeden Fall nützlich, <g(Gestikulieren) weil sie ja egal, ob das_ wenn es nicht geklappt hätte, sie hätten können umkehren,>

<g(Gestikulieren) wäre das <s(betont) ein_> sowieso eine Erleichterung gewesen.>

Und wenn es geklappt hätte auch. <i>

Ja, also wir_ wir fahren zurück nach Chemnitz.

Möstl, Dr. Günter, Interview ev002, 03.09.2019, Interview-Archiv
"Eiserner Vorhang", <https://archiv.eiserner-vorhang.de/de/interviews/ev002>

- Interviewmetadaten und Transkripte in einem TEI-XML-Dokument
- TEI-XML Standard „ISO 24624:2016 Language resource management – Transcription of spoken language“ ([Hedeland/Schmidt 2022](#))
- Zugang nach Freischaltung im Archiv

```
<!--annotationBlock xml:id="ab1372826" who="p1368424" start="T1_S1372826" end="T1_S1372827"-->
<-u xml:id="u1372826">
  <-<seg type="contribution" xml:id="s1372826" xml:lang="ger">
    <anchor synch="T1_S1372826"/>
    <w xml:id="s1372826_0">Das</w>
    <w xml:id="s1372826_1">war</w>
    <w xml:id="s1372826_2">ein</w>
    <w xml:id="s1372826_3">Scherz</w>
    <w xml:id="s1372826_4">alles</w>
    <pc xml:id="s1372826_5">.</pc>
    <anchor synch="T1_S1372827"/>
  </seg>
</u>
<-spanGrp type="kinesic">
  <span from="s1372826_0" to="s1372826_5">Gestikulieren</span>
</spanGrp>
<-spanGrp type="original">
  <span from="T1_S1372826" to="T1_S1372827"><g(Gestikulieren) Das war ein Scherz alles.></span>
</spanGrp>
</annotationBlock>
```



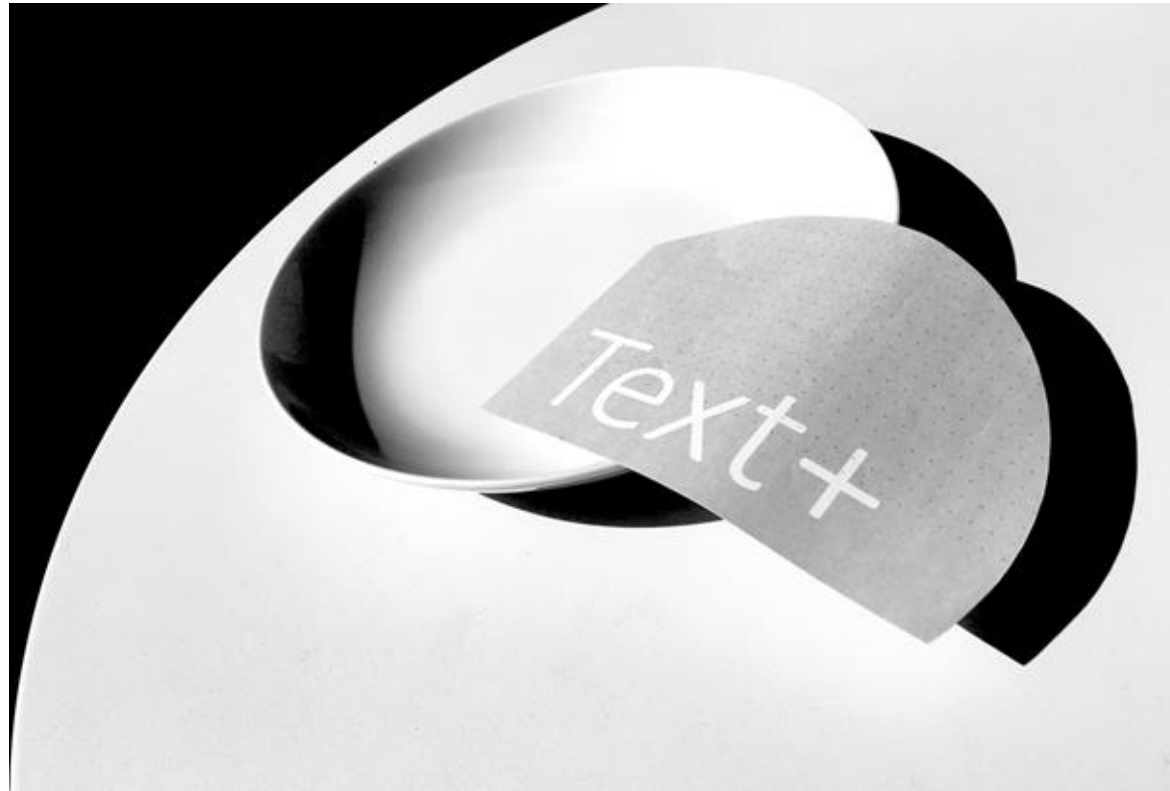


made with Canva AI



Von der Bubble bis zum Tellerrand und beyond...

Collections
Lexical Resources
Editions
Infrastructure/
Operations



AG Dissemination/Community:



NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

universität freiburg



LEIBNIZ-INSTITUT FÜR
DEUTSCHE SPRACHE



UNIVERSITÄT
DES
SAARLANDES



Berlin-Brandenburgische
AKADEMIE DER WISSENSCHAFTEN



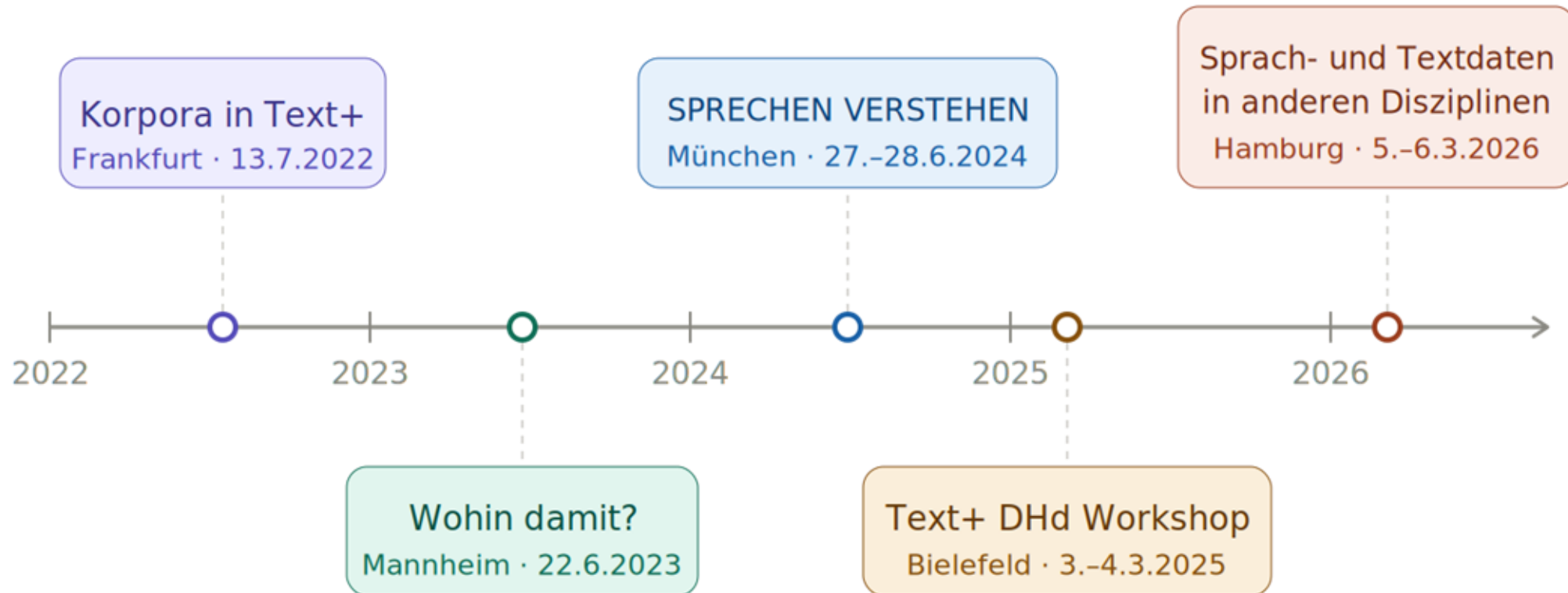
Zur Präsenz in der Box

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Workshops beyond the box

Collections
Lexical Resources
Editions
Infrastructure/
Operations



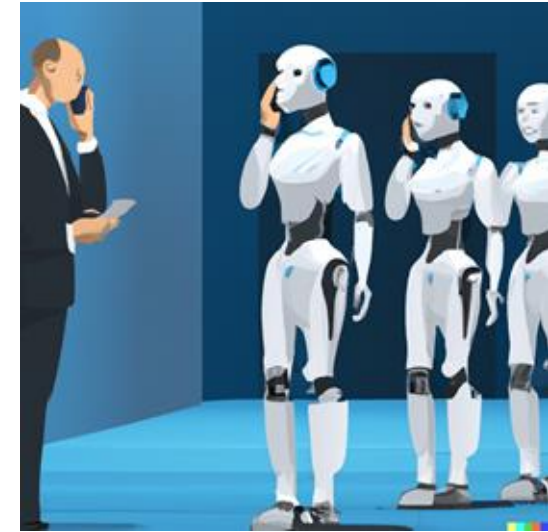
Beyond the Box I: Industrie



EP: ...der dann aus dieser angewandten Sicht erzählt hat, was mit Sprache im Auto passiert. Also das war für mich wirklich, wirklich erfrischend.

MH: Ich kann mich dem nur anschließen. Ich war ein bisschen erstaunt, dass ich das so überhaupt nicht auf dem Feld hatte, warum die gesprochene Sprache auch für die Wirtschaft so wahnsinnig spannend ist oder warum die so eine große Herausforderung ist. Ich muss sagen, dank des Vortrages von Herrn Prenninger spreche ich jetzt immer mit meinem Auto...

Das war für mich fast so ein Aha-Effekt. **Warum habe ich darüber noch nie nachgedacht?...**



generiert von Dall'e



<https://textplus.hypotheses.org/11995>



Beyond the Box II: Volkshochschulen

827 Volkshochschulen
2.659 Außenstellen in Deutschland

16

6 Mio. Teilnehmer*innen pro Jahr

vhs
in Zahlen

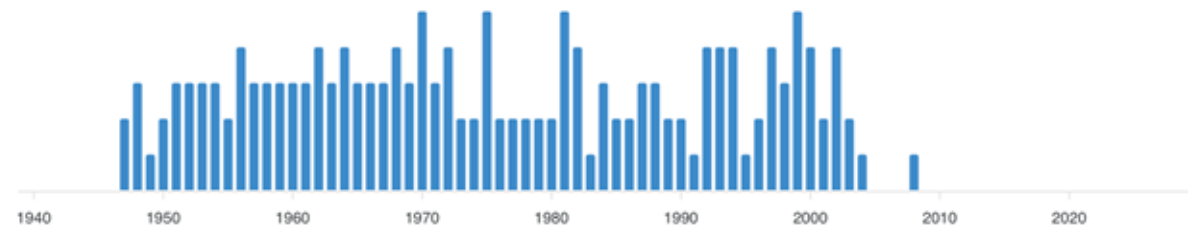
Landesverbände und ein Dachverband

184.000 Lehraufträge an Kursleitende

14,7 Mio. Unterrichtsstunden

Programme der Münchner Volkshochschule inklusive retrospektiver und zukünftiger Entwicklungen ausgehend von 2004

BESTANDS-STATISTIK
168 Elemente insgesamt verfügbar



Die Integration der 'Born Digitals', also seit 2004 in Form von PDF-Dateien gesammelter Hefte, steht derzeit noch aus. Für diese arbeiten wir an der ebenfalls **notwendigen Anonymisierung sowie Bereitstellung** in den für weiterführende Analysen notwendigen Formaten.



Beyond the Box III: Kriminaltechnik

Bedarfe an Datenkorpora für die Forschung in der forensischen Linguistik und Phonetik

- Newsletter-Rubrik „Neue Korpora“ (mit Filter)
- unedierte, authentische Texte und Sprache mit Zuordnung und Annotation, aus allen Gesellschaftsschichten, Mutter- und Fremdsprache
- KI-generierte Daten, Prompts und Ergebnisse
- u.v.a.m



Collections
Lexical Resources
Editions
Infrastructure/
Operations



made with Canva AI



Rechtliche und ethische Fragen

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Schwerpunkt der AG Legal: Text und Data Mining

Die DSM-RL und das Text und Data Mining

Das Training von KI und die Text und Data Mining Schranken

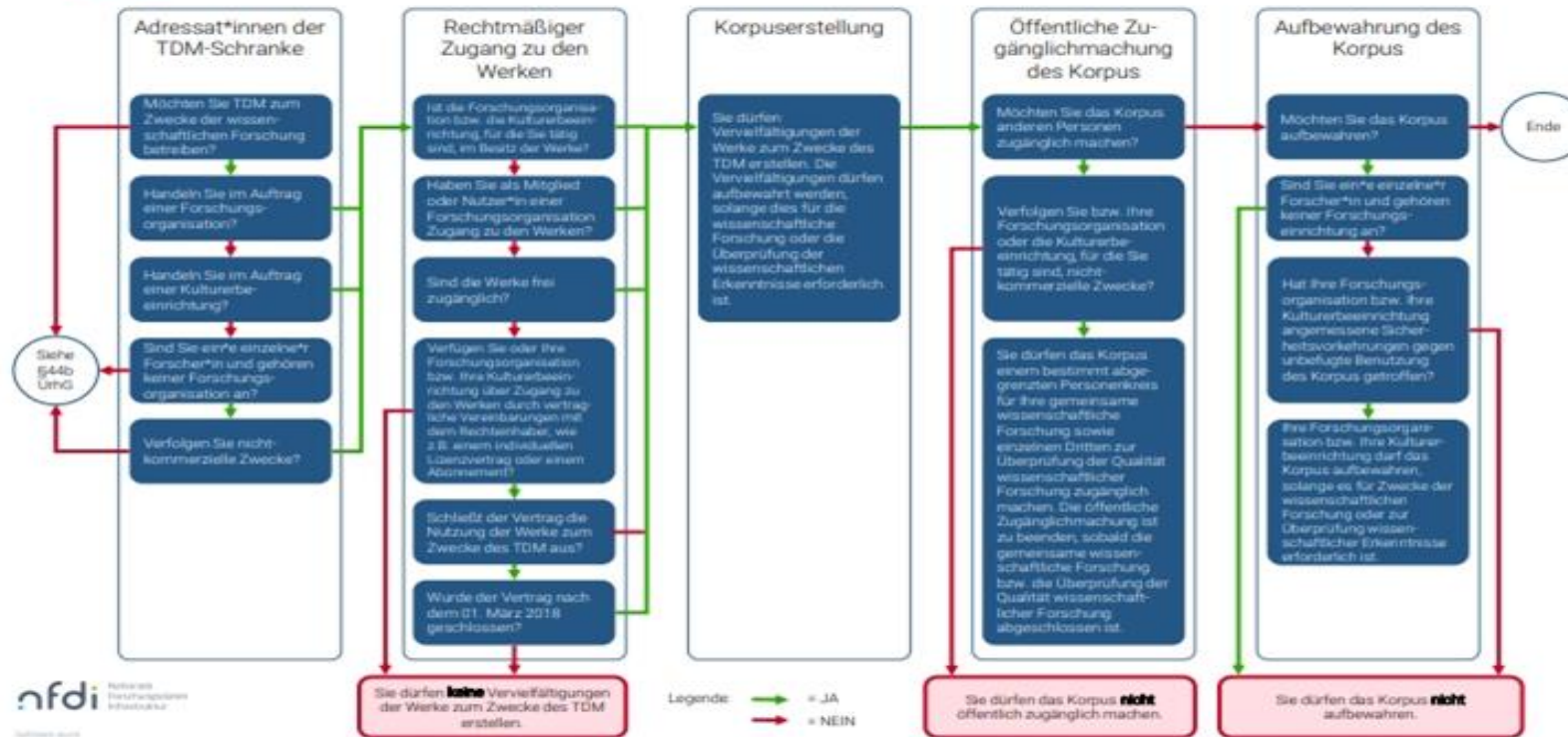
Legal Help Desk



Rechtliche und ethische Fragen



Entscheidungshilfe zur Text- und Datamining-Schranke im deutschen Urheberrecht §60d UrhG



Rechtliche und ethische Fragen

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Und falls ihr euch fragt, ob diese Ausnahme auch das Training von LLMs abdeckt, haben wir dazu ebenfalls eine **Stellungnahme** veröffentlicht (das tut sie! **Forschende brauchen keine Lizenz für das Training von LLMs!**).



Rechtliche und ethische Fragen

Collections
Lexical Resources
Editions
Infrastructure/
Operations



Collections
Lexical Resources
Editions
Infrastructure/
Operations

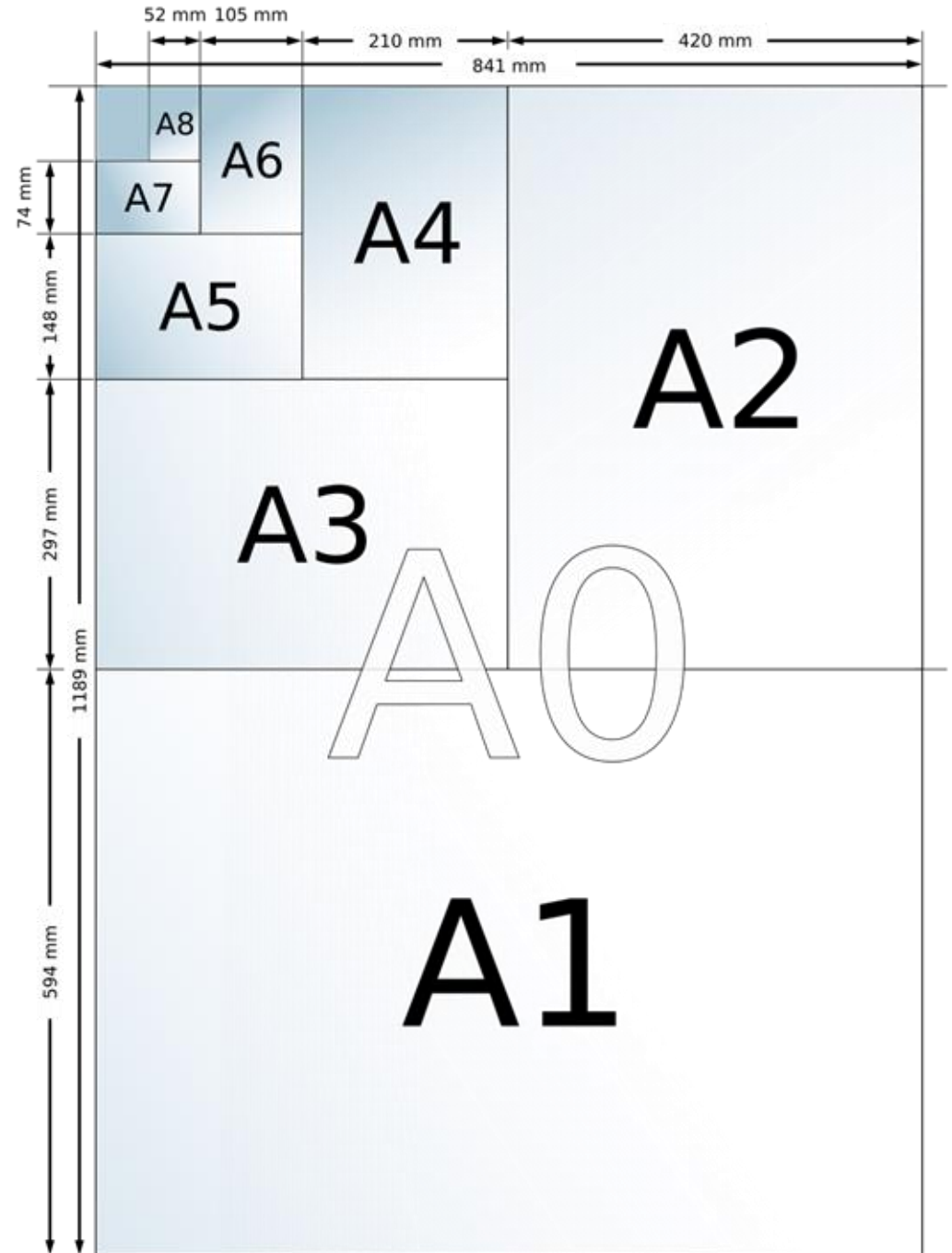


made with Canva AI



Institution

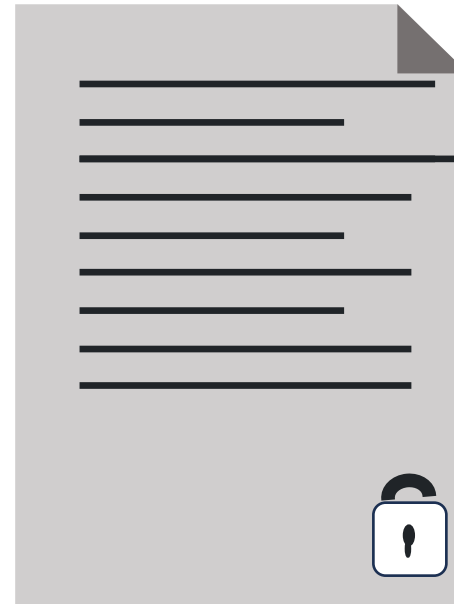
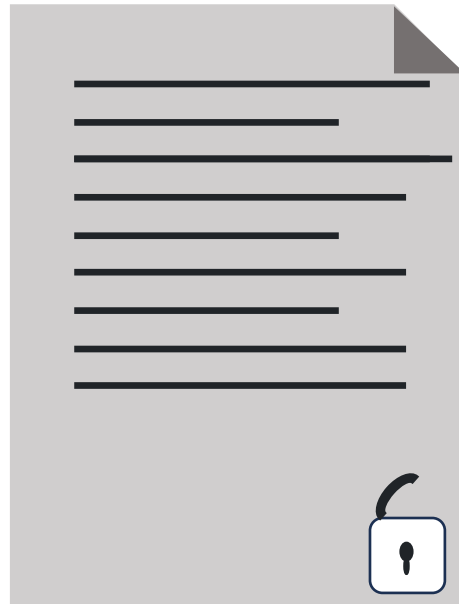
DIN
Unsichtbare Infrastruktur
Normung für jeden Tag



Konfliktfeld

- Offene Forschung

- Eigentum
- Urheberrecht
- Persönlichkeitsrecht



Transformation

DIN 19461

Original Text



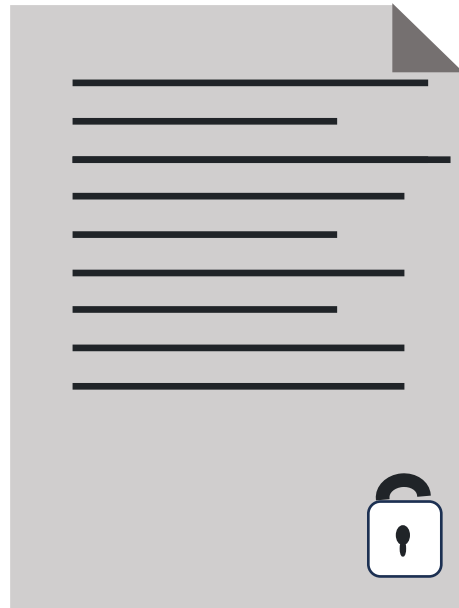
[Transformation]



Abgeleiteter Text

„Kontrollierte Transformation“

Komposition



Original Text



Anreicherung

e.g.

- POS
- NE
- Structure
- Keywords
- ...



Reduktion

- Beibehalten
- Löschen
- Ersetzen
- Randomisieren



Abgeleiteter
Text

Signatur

DIN 19461

Derived Text

Metadata (nach ISO 24622)

Werkzeuge

Parameter

Versionen

Provenienz

Kommentierungsmöglichkeit



Öffnung

DIN 19461



Richtlinien | Schulung

Folien | Verwendungsbeispiele



Forschungseinsatz

Interpretation

DIN 19461

Nicht nur eine Norm
sondern Infrastruktur
für
Offene Forschung

Vorherige Ausstellungen



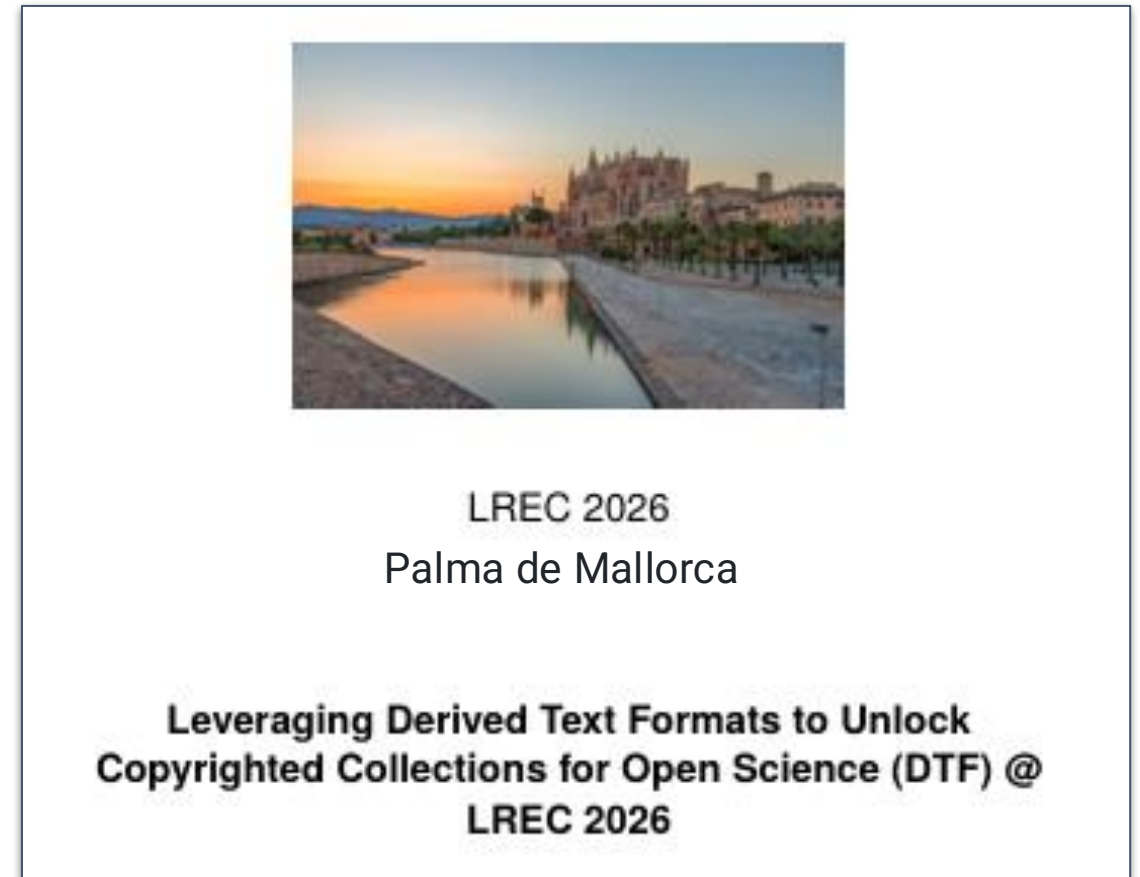
A poster for an exhibition titled "ATF: WTF?". The background is a light beige color with several overlapping squares in shades of tan, white, and black. At the top center, there is a red circle containing the text "Text+" in white. Below this, the title "ATF: WTF?" is written in a bold, black, sans-serif font. Underneath the title, the dates "12./13. November" are displayed in a smaller, black, sans-serif font. The main text of the poster, "Was Sie schon immer über ATFs wissen wollten*", is written in a large, bold, black, sans-serif font, with "ATFs" highlighted in red. The asterisk at the end of the sentence is also in red.

Text+


ATF: WTF?

12./13. November

Was Sie schon immer über ATFs wissen wollten*



A poster for an exhibition titled "LREC 2026". The top half of the poster features a photograph of a river flowing through a city at sunset. The sky is a mix of orange, yellow, and blue, and the buildings along the riverbank are silhouetted against the light. The water in the river reflects the colors of the sky. Below the photograph, the text "LREC 2026" and "Palma de Mallorca" is written in a black, sans-serif font. At the bottom of the poster, the full title of the exhibition, "Leveraging Derived Text Formats to Unlock Copyrighted Collections for Open Science (DTF) @ LREC 2026", is written in a smaller, black, sans-serif font.



LREC 2026
Palma de Mallorca

Leveraging Derived Text Formats to Unlock
Copyrighted Collections for Open Science (DTF) @
LREC 2026

Katalog zur Ausstellung

Handreichung

Foliensatz

Nutzungsszenarien

LREC-Proceedings



Rang	N-Gramm	Häufigkeit
1	gott sei dank	43
2	ja gnädigste frau	17
3	auch heute wieder	13
4	doch auch wieder	11
5	ist doch auch	11
6	ist immer so	10
7	gnädigste frau ist	10
8	war so war	10
9	nein gnädigste frau	9
10	wird ja wohl	9
11	ist doch recht	9
12	doch immer noch	9

*von_APPR_von Hohen-Cremmen_NN_Hohen-
Georg_NE_Georg zu_APPR_zu heller_AD
fiel_VVFIN_fallen schon_ADV_schon
bewohnten_ADJA_bewohnt In_APPR_i
Mittagsstille_ADJA_Mittagsstil
Gartenseite_NN_Gartenseite un
Park-TRUNC_Park- Dorfstraß
Seitenflügel_NN_Seitenflüg
die_ART_die hin_ADV_hin während_Ku
angebauter_ADJA_angebaut der_ART_die nach-
ein_ART_eine Schatten_NN_Schatten auf_APPR_auf
einen_ART_eine rechtwinklig_ADJD_rechtwinklig <SEG>
großes_ADJA_groß ,_PUN_, auf_APPR_auf mit_APPR_mit
in_APPR_in ein_ART_eine weiß_ADJD_weiß und_KON_und
über_APPR_über quadrierten_ADJA_quadrierten und_KON_und
diesen_PDAT_dies auf_APPR_auf Mitte_NN_Mitte
sein dann_ADV_dann
hinaus_ADV_hinaus*

MONAPipe Implementation: Derived Text Formatter

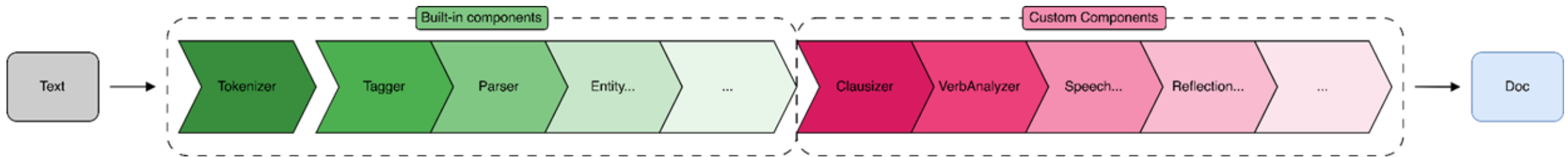
Plaintext Operationen

- Behalten/löschen (Keep)
 - *liebt den Salat. Ein gut gemachter auch ohne glänzen. Passend zu, Schnitzel oder einfach nur so macht dieser Krautsalat eine Figur.*
- Ersetzen (Replace)
 - *PRON liebt den ADJ Salat. Ein gut gemachter VERB auch ohne NOUN glänzen. Passend zu NOUN, Schnitzel oder einfach nur so macht dieser Krautsalat eine ADJ Figur.*
- Umtauschen (Randomize)
 - *nach und schnelle man Kochen zu Küche dem zum der groß aus. Denn Lust, Feierabend Bei den hat lange niemand in Gerichten sich einfache stehen. meisten sucht*



Modes of Narration and Attribution Pipeline (MONAPipe)

Collections
Lexical Resources
Editions
Infrastructure/
Operations



- Natural-language-processing pipeline implemented in Python/spaCy
- Integrates NLP developments in a shared framework



MONAPipe Installation and Usage

1. Installation via Python package manager

```
pip install monapipe
```

1. Simple usage in Python

```
# Build a pipeline
nlp = monapipe.model.load()
nlp.add_pipe("dependency_clausizer")
nlp.add_pipe("neural_reflection_tagger")
nlp.add_pipe("neural_attribution_tagger")
# Process a text
doc = nlp("All happy families are alike, ...")
```



Notebooks in Jupyter4NFDI

Collections
Lexical Resources
Editions
Infrastructure/
Operations

<https://hub.nfdi-jupyter.de/share/o9TJkcSoD0k>

File Edit View Run Kernel Tabs Settings Help Credits: 59 / 60

EntityRecognizer-ner+bert_x

Notebook Python 3 (ipykernel)

Component EntityRecognizer

This notebook shows the functionality of the named entity recognition in MONAPipe and spaCy for the available implementations.

```
[8]: # import the MONAPipe language model
import monapipe.model
```

```
[9]: # Check for Jupyter4NFDI Environment
import os
try:
    JUYPTER4NFDI = os.getenv("JUPYTERHUB_API_URL").startswith("https://hub.nfdi-jupyter.de/hub/api")
except:
    JUYPTER4NFDI = False
```

Implementation ner (spaCy standard)

The standard implementation of spaCy sets NER tags for persons/characters (PER), locations (LOC), organisations (ORG), and miscellaneous (MISC).

```
[10]: # create nlp object (without not required components)
nlp = monapipe.model.load(disable=['tok2vec', 'tagger', 'morphologizer', 'trainable_lemmatizer', 'parser'])
text = """
    Angela Merkel besuchte gestern New York und hielt eine Rede bei den Vereinten Nationen.
    """
doc = nlp(text)
for ent in doc.ents:
    print("ent_text: ", ent.text, "|", "ent_label: ", ent.label_)
```



Component: Entity Linker

WIKIDATA: Q308

Mercury is the smallest planet in the Solar System.

WIKIDATA: Q15869

Freddie Mercury , lead singer of Queen, was born in Zanzibar.

WIKIDATA: Q925

The element mercury was used in thermometers.



Entity Linker (SUB)

Collections
Lexical Resources
Editions
Infrastructure/
Operations

opentapioca entity linker (IDS)

- Links to Wikidata
- Re-trained and re-implemented algorithm

embedding entity linker (SAW)

- Links to GND
- Embedding-based candidate ranking

llm entity linker (GWDG)

- Links to Wikidata
- LLM-based candidate disambiguation



Implementation: LLM Entity Linker

Mercury is the smallest planet in the Solar System.

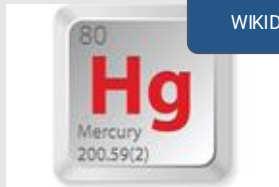
1. Candidates from Wikidata



WIKIDATA: Q15869



WIKIDATA: Q308



WIKIDATA: Q925



2. Disambiguation via LLM

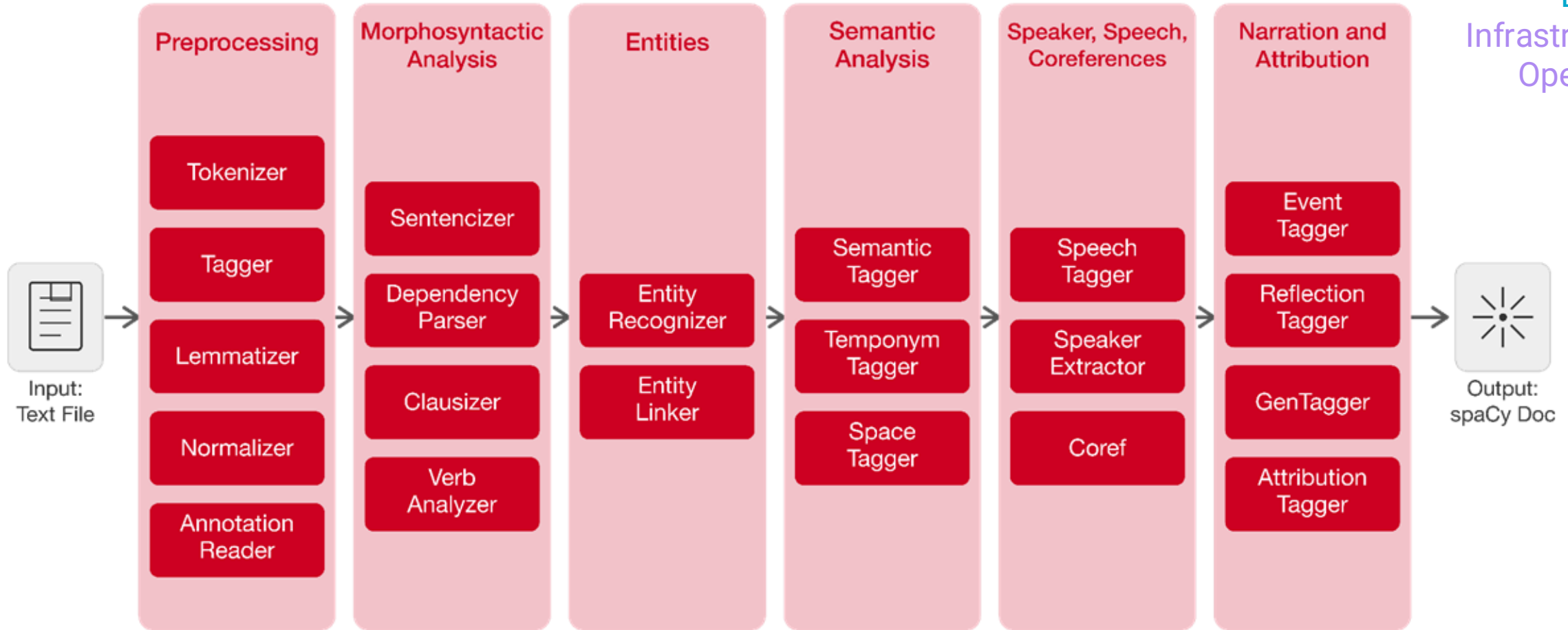
"You are an entity disambiguation assistant for Wikidata."
"FIRST write a short description ... who/what the entity is"
"THEN select the best Wikidata entry from the candidates."

WIKIDATA: Q308



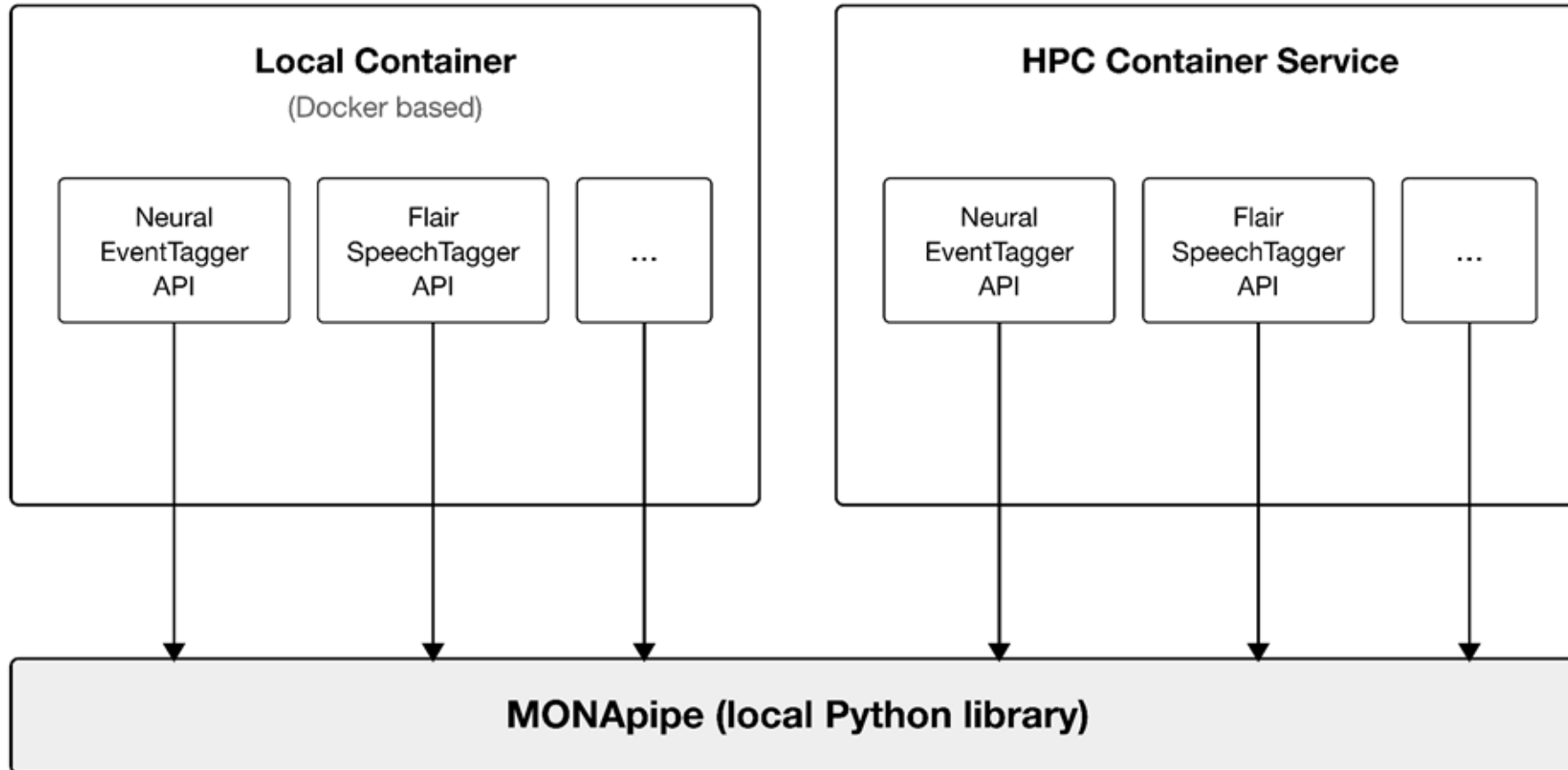
Custom Components

Collections
Lexical Resources
Editions
Infrastructure/
Operations



API components and Computation Service

Collections
Lexical Resources
Editions
Infrastructure/
Operations





Collections
Lexical Resources
Editions
Infrastructure/
Operations

made with Canva AI



6. Aufkommen von LLMs im Projektzeitraum: Task Force Große Sprachmodelle, Text+ Chatbot (*Umut*)

Text-Plus Chatbot (<https://text-plus.org/chatbot>)

- Entity Linking
- Named Entity Recognition
- Natural Language to LexCQL conversion
- Retrieval Augmented Generation

Hosted by GWDG (Usable via API key):

- Mona-Pipe components
- S2-Chunker
- NL -> LexCQL Converter



Entity Linking

Named Entity Recognition



Text+ AI Assistant
Powered by Academic Cloud

FCS-Search

Retrieval Augmented Generation



- CONVERSATIONS +
- LAST 7 DAYS
- Comparative Study Datasets 4d
 - YOLO Synthetic Training Data 4d
 - Octopus Blue Blood Secret 4d**
 - Basic Arithmetic Calculation 4d
 - Comparative Study Datasets 4d
 - Comparative Study Datasets 4d
 - Comparative Study Datasets 4d
 - Comparative Study Datasets 4d
 - Comparative Study Datasets 4d
 - Comparative Study Datasets 4d
 - Object Detection Dataset Comp... 4d
 - Octopuses Blue Blood Secret 4d
 - Math Basic Calculation 4d
 - Comparative Study Datasets 4d
 - Comparative Study Datasets Used 4d
 - Historical Document Layout Anal... 4d
- LAST MONTH
- Alan Turing Legacy 2w
 - You Only Look Once for History 2w
- LAST YEAR
- Alan Turing Legacy 3mo
 - New Conversation 3mo
 - Yo 3mo

T+ **AI Assistant**

Octopuses have **blue blood!**

Unlike humans, whose blood is red due to the iron in hemoglobin, octopuses use a copper-based protein called **hemocyanin** to transport oxygen. Copper turns blue when it binds with oxygen, which is why their blood is that color. This adaptation is particularly efficient for transporting oxygen in the cold, low-oxygen environments of the deep ocean where many octopuses live.

1 Source

BASE MODELS

- qwen3-coder-30b-a3b-instruct
- glm-4.7
- qwen3.6-35b-a3b

Show 14 more

CUSTOM MODELS

- Codiology** glm-4.7 ✓
- G glm-4.7
- ggt

Show 4 more

+ Add Custom Model

Model: Codiology ^

Undo Regenerate

- You **Collections** +
- Tell me a fun fact about octopuses.
 - Lexical Resources
 - Editions
 - Infrastructure/Operations

Available Tools

Collections
Lexical Resources
Editions
Infrastructure/
Operations

Add Custom LLM

Configuration
Model & prompt

Files
No files

Tools
2 enabled

Retrievers
0 enabled

External Tools

Enable tools for searching and retrieving information.

Wikipedia	<input checked="" type="checkbox"/>
Wikidata	<input checked="" type="checkbox"/>
Named Entity Recognition	<input type="checkbox"/>
FCS Resources	<input type="checkbox"/>
Google Scholar	Coming soon





CANCEL SAVE



Answers you can trust – straight from your documents.



Collections
Lexical Resources
Editions
Infrastructure/
Operations

Edit Custom LLM ✕

-  Configuration
Model & prompt
-  **Files**
2 uploaded
-  Tools
0 enabled
-  Retrievers
0 enabled

Upload Documents

Add files for your LLM's knowledge base. Supported: .txt, .pdf (max 50 MB)

-  2501.05485v1.pdf ✕
-  2023-06-02-Introduction_to_GPU_at_GWDG_Dorothea.pdf ✕

+ Add files 🗑 Delete all

CANCEL SAVE





Answers from **your document** — not the model's assumptions

OTHER CHATBOTS Overwrites the file with pretrained knowledge

TEXT+ CHATBOT Quotes the file & flags the contradiction

ASSISTANT

It appears that the uploaded document contains false information about Alan Turing.

Alan Turing was actually a British mathematician, computer scientist, logician, and philosopher who made significant contributions to the development of computer science, artificial intelligence, and cryptography. He is best known for his work on the theoretical foundations...

Highlighted text is the model's own knowledge — not in your file

AI ASSISTANT

✓ search_documents — 1 source retrieved

Looking through your uploaded documents for info about Alan Turing.

△ Note: this appears to be a fictional or incorrect account, as it contradicts historical facts.

The uploaded document "Alan-Turing-Incorrect.txt" states: Alan Turing was a "professional snail trainer" and held the "Guinness World Record for the fastest snail to cross the English Channel." He was a "skilled pastry chef" who invented the "Turing Tart," a dessert that became a staple in British bakeries.

Source · Alan-Turing-Incorrect.txt

Collections
 Lexical Resources
 Editions
 Infrastructure/
 Operations

WHY IT MATTERS Retrieval-first grounding. Text+ shows what's in your file and separates it from world knowledge — so you're never misled into thinking the model's answer came from your document.



● ✓ Evaluation passed — 95% confidence 4 steps ^

● ✓ Thinking — done

● ✓ search_wikipedia — done

● ✓ vector_store — 3 sources retrieved

● ✓ Evaluation passed — 95% confidence

Based on the provided documents, the **YOLO-gen** model is a generalist object detector used for **layout analysis in historical documents**.

Specifically, it is designed to address the challenge of analyzing historical manuscript images that come from different datasets with non-interoperable annotation schemes. It achieves this through a **codicologically-informed, hierarchical ontology** that harmonizes disparate datasets (e.g., e-NDP, CATMuS, and HORAE) into a unified framework.

Key characteristics of its use include:

- **Multi-Label Training:** It detects layout elements using a hierarchical structure (parent, intermediate, and subclass categories), allowing it to identify broad concepts like "Text," "Decoration," and "Initials" rather than just specific, localized tags.
- **Oriented Bounding Boxes (OBB):** It uses OBBs to faithfully represent the non-Cartesian and artistic nature of manuscripts, which is crucial for accurate layout analysis.
- **Foundational Model:** It serves as a strong "foundational" baseline for future research, proving that a single harmonized model can perform nearly on par with domain-specific specialist models.

📄 3 Sources ^

From Codicology to Code_ A Comparative Study of Transformer and YOLO-based Detectors for Layout Analysis in Historical Documents.pdf

Tool: s2_chunking

ross-Corpus Training: The YOLO-gen Foundational M

To move beyond single-corpus benchmarks, our final experiment addresses a core challenge in applying machine learning to historical documents: the lack of standardized annotation ontologies. Datasets like e-NDP, CATMuS, and HORAE, while invaluable, use particularized class schemes that are not directly interoperable. This experiment, therefore, tests a methodology for harmonizing these disparate datasets by creating a unified, hierarchical model.

The cornerstone of this approach is a *codicologically-informed reclassification* of the three source ontologies. While the specific tags in each corpus are distinct, they often refer to common, fundamental codicological elements. We created a unified ontology to capture this shared vocabulary, establishing high-level parent categories from scholarly concepts (Text, Paratext,

Collections
Lexical Resources
Editions
Infrastructure/
Operations



It reads your PDFs the way you do. (Layout aware) Modified version of S2 chunking by (Verma, 2025)¹

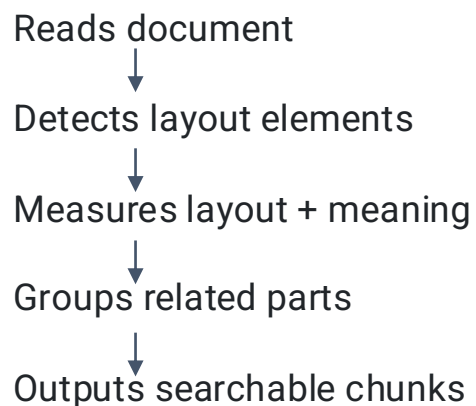



Figure 1: Three annotated pages from the HORAE (pray book), e-NDP (registers) and CATmUS (literature) datasets respectively.

Zone: Zone containing a running title (typically top of the page).

Zone: Zone containing a stamp (library, ownership, etc.).

Running Title Zone: Running title at the top of the page.

The raw category names (e.g., "DropCapitalZone") were mapped to descriptive phrases (e.g., "ornate drop capital letter zone") for use in VLM training configurations and output metadata, controlled by a script flag.

0.1.9 HORAE Dataset

This dataset was created as part of the HORAE project [2]. It is composed of pages from Books of Hours, best-selling personal prayer books in the late Middle Ages, known for their lavish illustrations. They represent a significant DLA challenge due to complex layouts featuring intricate borders, miniatures, and a wide variety of initials. We focus on 8 distinct layout classes that include decorative and structural elements.

- Order:** Margins filled with ornamental vine work or geometric patterns.
- Illustration:** Margins containing identifiable figures, scenes, or objects.
- Full-page or large-scale illustration:** Full-page or large-scale illustration depicting a narrative scene.
- Initial:** An enlarged initial letter with non-figurative ornamentation.
- Initial:** An enlarged initial letter containing an identifiable scene or figure.
- Decorative element:** decorative element filling the end of a line of text.
- Simple Initial:** A small or larger initial letter without significant decoration.
- Border Text:** Text written within the marginal space, often as captions or glosses.

0.2. Experimental Setups

For this comparative study the data sets were processed into standard COCO AABB format for models requiring axis-aligned bounding boxes and the YOLO OBB format (using minimum area rectangles derived from source polygons) for the oriented bounding box detector. A 90%/10% split was used to create trainval/test sets. The key characteristics are summarized in Table 1.

¹Verma, P. (2025). S2 chunking: A hybrid framework for document segmentation through integrated spatial and semantic analysis. arXiv. <https://arxiv.org/abs/2501.05485>



Nature Language to LexCQL

Finde Wörterbucheinträge, deren Definition sowohl
das Wort "car" als auch "clutch" enthält!



```
(definition =/lang=eng \"car\" OR lemma = \"car\") AND  
definition =/lang=eng \"clutch\"
```



From plain language to a database search

A PERSON ASKS

German word for house

● AI judgement ● Fixed rules ● Library / network

Collections
Lexical Resources
Editions
Infrastructure/
Operations

1 · UNDERSTAND

0 **Check what's in stock**
load_resources
Which databases & languages are available right now

1 **Understand the question**
Query Analyzer
Finds intent + key words; rates its own confidence

2 **Draft the request**
Query Planner
Which field to search, for what, how to combine

2 · BUILD THE FORMAL QUERY

3 **Sanity-check**
Query Guard
Blocks too-broad queries; fixes unsupported fields

4 **Write the official form**
IR Builder
Assembles exact wording — every bracket right

5 **Proofread**
Validate + Repair
Validator checks it; safe fallback if needed

LEXCQL READY ✓
translation = "house"

3 · SEARCH & ANSWER

6 **Send & gather**
run_tasks
Queries every database at once; notes any errors

7 **Tidy the results**
merge_evidence
Remove duplicates & junk; keep the best matches

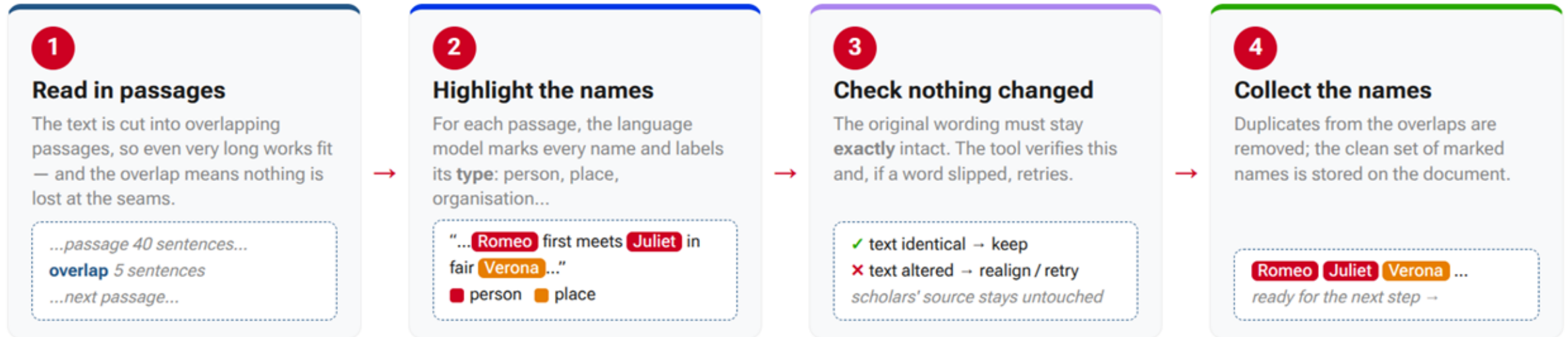
8 **Write the reply**
answer_synthesizer
Plain answer, with a source cited for every claim

9 **Grade its own homework**
answer_judge
Confirms every claim is backed by a real result



Teaching the machine to find every name in a text

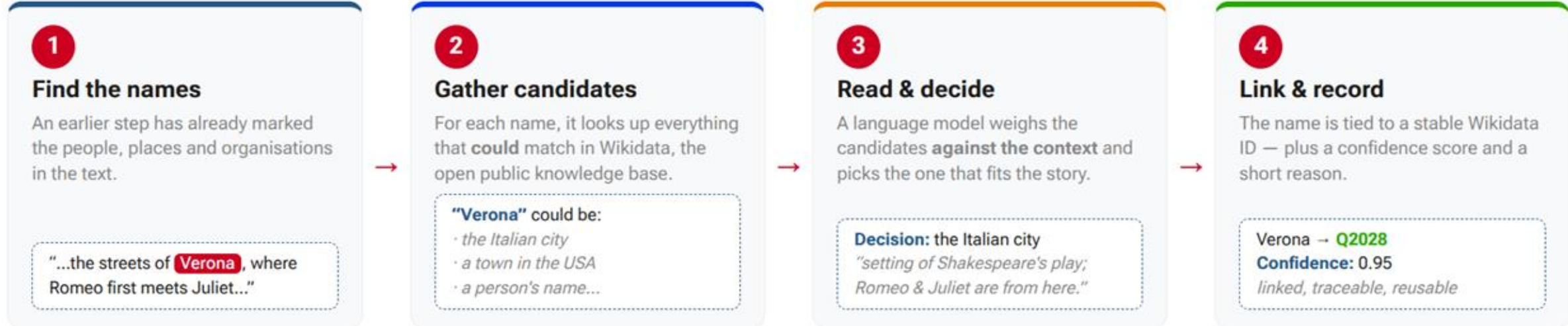
Before we can look anything up, we first have to *find* the people, places and organisations on the page — like a scholar going through a book with a highlighter, but automatically and across thousands of pages.



From a name on the page to a fact in a knowledge base

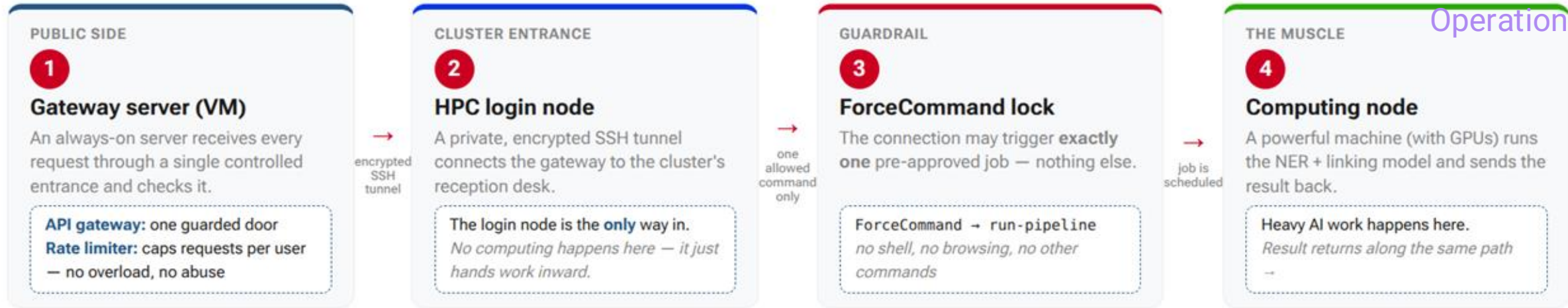
The same word can mean many things. This tool reads the surrounding text — like a careful human reader — to decide *which* real-world person or place a name refers to, and pins it to a permanent, shareable identifier.

Collections
Lexical Resources
Editions
Infrastructure/
Operations



From a user's request to a university supercomputer — safely

The language model is too heavy to run on a laptop, so it lives on the university's high-performance computing cluster (HPC). Every request travels a controlled, locked-down path to get there.



Why ForceCommand? — least privilege

The gateway needs a key to enter the cluster — but a key that opened *everything* would be dangerous if ever stolen. **ForceCommand** turns that key into a single-purpose one: no matter what arrives, the login node will *only* launch our pipeline — never an interactive shell or arbitrary commands. One door, one action, no surprises.



Text+

Plenary
2026



@textplus_NFDI
#TextPlusPlenary

5. TEXT+ PLENARY

PROGRAMM

Dienstag, 16. Juni 2026

09:00 – 10:30	Ergebnisse aus Task Area: Lexical Resources (Aula)
<i>10:30 – 11:00</i>	<i>Kaffeepause (Katakomben)</i>
11:00 – 12:30	Ergebnisse aus Task Area: Collections (Aula)
12:30 – 13:00	Abschluss und Ausblick auf die zweite Phase (Aula)
<i>13:00 – 14:00</i>	<i>Lunch (Katakomben)</i>
14:00 – 15:30	<i>Parallele Sitzungen (intern, nur auf Einladung)</i>
	Gemeinsames Treffen des Scientific Coordination Committees Lexical Resources und der Task Area Lexical Resources (Aula)
	Treffen des Scientific Coordination Committees Collections (Dozentenzimmer, O 126)
	FID Koop Treffen (Musikerzimmer, EO 154)



Abschluss und Ausblick auf die zweite Phase

Text+ Abschlussvortrag Phase I

Philipp Wieder, Andreas Witt



Positive Förderempfehlung

Der Förderantrag für Phase II wurde positiv bewertet – das Projekt wird als förderwürdig anerkannt.

Hohe Relevanz bestätigt

Text+ wird als relevante Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften anerkannt.

Kürzung ~35 %

Das bewilligte Budget liegt ~35% unter dem Antrag. Priorisierung und Anpassung des Arbeitsprogramms sind notwendig.

~35%

Budgetkürzung gegenüber
dem Antrag

Kein 'business as usual'

01 Arbeitsprogramm wird angepasst

Umfang und Detailtiefe einzelner Maßnahmen werden reduziert.

02 Priorisierung notwendig

Zentrale Dienste und Community-Mehrwert stehen im Vordergrund.

03 Strategischer Fokus

Text+ konzentriert sich auf die unverzichtbaren Kernleistungen.

Phase II

Was macht Text+ für die Forschung unverzichtbar?

Fokus auf Nutzung

Forschende aktiv einbinden – Text+ als unverzichtbares Werkzeug in der täglichen Forschungspraxis etablieren.

Integration in Forschung

Enge Verzahnung mit laufenden Forschungsprojekten, Disziplinen und der NFDI-Gemeinschaft.

Task Areas und Co-Spokespersons

TA1

Library & Archive Holdings



Elke Teich
(U Saarland)



Philippe Genêt
(DNB)

TA2

Language Corpora & Elicited Data



Angelika Zirker
(U Tübingen)



Andreas Witt
(IDS Mannheim)

TA3

Lexical Resources



Hanna Fischer
(U Marburg)



Alexander Geyken
(BBAW)

TA4

Editions



Andrea Rapp
(TU Darmstadt)



Andreas Speer
(AWK NRW)

TA5

Infrastructure /Operations



Vivien Petras
(HU Berlin)



Philipp Wieder
(U Göttingen/GWDG)



Regine Stein
(VZG)

TA6

Administration



Philipp Wieder
(U Göttingen/GWDG)



Andreas Witt
(IDS Mannheim)

M1

Federated Data Infrastructure

Föderierte Dateninfrastruktur für alle Text+-Zentren und Datenbereiche; FCS & Registry.

M2

Interfaces between Data Domains

Schnittstellen zwischen Sprachressourcen, Editionen, Lexika und Archivbeständen.

M3

Enabling Text & Language Data for AI

Hochqualitative, kuratierte Trainingsdaten für Sprachmodelle; automatische Annotation.

M4

Implicit & Explicit Structures

Erkennung latenter Strukturen in Sprachdaten; Wissensrepräsentation.

M5

Community Activities

Dissemination, Training, Helpdesk, Consulting, Outreach für die Forschungscommunity.

Infrastructure Operations (TA 5)

Text+

IO-M1

Federated Data Infrastructure

DATA: FCS & Registry, data integration with data centres, data enrichment.

IO-M2

Service Integration w/ Data Domains

PROCESSES: UX/UI, technology stack coherence across all services.

IO-M3

Platform and Core Services

PLATFORM: enabling role for LLM service, Portal, Base4NFDI, IT service management.

IO-M4

Coherence

ECOSYSTEM: alignment with NFDI, EOSC, standards, indicators and impact across data domains.

M1

Koordination & Governance

Koordination der Arbeitspakete, Governance-Strukturen, flexible Mittelvergabe, Zusammenarbeit mit der Förderinstitution.

M2

Communication & Outreach

Öffentlichkeitsarbeit, Community Engagement, Koordination der Kommunikationsaktivitäten im Konsortium.

M4

Monitoring & Sustainability

Fortschrittsüberwachung, Berichtswesen, Nachhaltigkeitsplanung für die Text+-Infrastruktur nach 2031.

M5

NFDI & Europa-Vernetzung

Einbindung in OneNFDI, Kooperation mit Base4NFDI, Engagement in europäischen Forschungsinfrastrukturen (EOSC).

2026 – 2028 (–2031)

1

Qualität statt Quantität

Weniger Maßnahmen, aber konsequent hochwertig und nachhaltig umgesetzt.

2

Nutzung in der Forschung

Text+-Dienste als Standard in geisteswissenschaftlichen Forschungsprojekten etablieren.

3

Sichtbarkeit in der Community

Aktive Präsenz, Events, Helpdesk und Outreach – Text+ muss sichtbar und bekannt sein.

KI als Chance

Die KI-Entwicklung schafft neuen Bedarf an qualitativ hochwertigen, kuratierten Textdaten – genau das, was Text+ liefert.

01 LLMs erhöhen Bedarf

Große Sprachmodelle brauchen hochwertige, kuratierte Trainingsdaten in Massen. Text+ ist zentral positioniert.

02 Qualitätsdaten entscheidend

Nur qualitätsgesicherte, FAIR-aufbereitete Daten ermöglichen verlässliche KI-Modelle für die Wissenschaft.

03 Text+ zentral positioniert

Als NFDI-Konsortium mit breiter Datenbasis und Community-Verankerung ist Text+ prädestiniert für diese Rolle.

Die nächsten Jahre entscheiden.

Text+ muss unverzichtbar werden.

Fokus auf Qualität · Community · Nutzung



Text+

Fragen, Anmerkungen und Diskussion

Prioritäten

Community

Nutzung