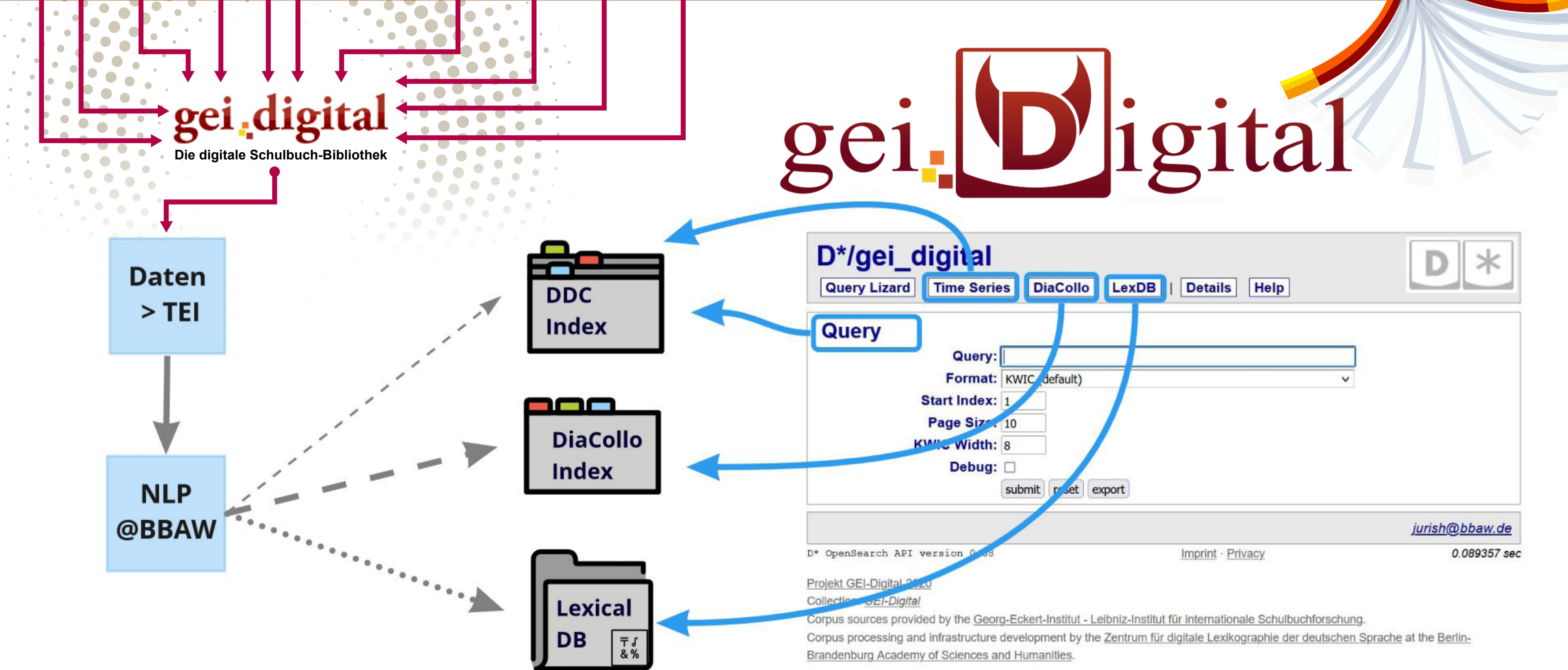


DiaCollo für GEI-Digital. Ein experimentelles Projekt zur weiteren Erschließung digitalisierter historischer Schulbuchbestände

Dr. Maret Nieländer, Christian Scheel (Leibniz-Institut für Bildungsmedien | Georg-Eckert-Institut, GEI),
Dr. Bryan Jurish (Berlin-Brandenburgische Akademie der Wissenschaften, BBAW)



Schulbücher sind relevante Quellen für

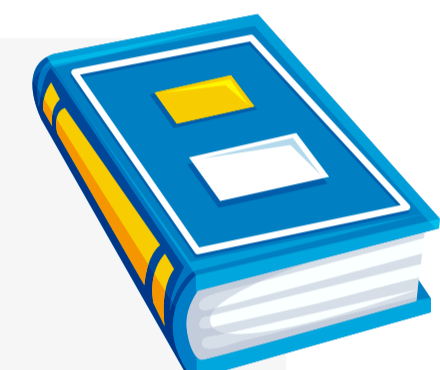
- ★ Historische Kulturwissenschaften,
- ★ Wissenschaftsgeschichte,
- ★ (Kinder-)Literaturwissenschaft,
- ★ Sozialgeschichte,
- ★ Ideen- und Begriffsgeschichte,
- ★ Soziolinguistik, Lexikographie, u.v.a.m.

Schulbücher vermitteln Wissensbestände, Wissensordnungen und Werte an Kinder und Jugendliche. Die Inhalte und Darbietung unterscheiden sich zum Beispiel nach:

- ★ Zeit und Ort/Region der Publikation
- ★ Schulfach, Schulform, Schulstufe
- ★ Alter, Konfession, Gender der Schüler:innen

1. Die Ausgangslage

Die digitale Schulbuchbibliothek "GEI-Digital" (<https://gei-digital.gei.de/>) stellt Digitalisate, automatisch generierte Volltexte und (schulbuchspezifische) Metadaten unter der Lizenz CC 0 zur Verfügung. Über die Benutzeroberfläche können Nutzer:innen nach Metadaten filtern und nach Zeichenfolgen in Metadaten und Volltexten suchen.



2. Das Projekt

Ende 2020 waren in "GEI-Digital" 5036 Schulbücher aus den Jahren 1648 bis 1921 mit OCR-generiertem Volltext verfügbar. Diese Werke wurden für die Analyse mit computerlinguistischen Werkzeugen aufbereitet:

- Konvertierung nach TEI
- Vorverarbeitung mit den an der BBAW für historische deutsche Texte genutzten NLP-Tools für Tokenisierung, Normalisierung, Wortartenerkennung und Lemmatisierung
- Indexierung
- Verknüpfung mit einer Instanz der an der BBAW genutzten D*-Korpusmanagement-Umgebung und den dort verfügbaren Werkzeugen
- Erstellung eines ausführlichen Tutorials für die digitalen Werkzeuge und das "GEI-Digital-2020" Korpus

3. Das Ergebnis

Nutzer:innen können das "GEI-Digital-2020" Korpus nun über die Weboberfläche von D* auf vielfältige Art durchsuchen und explorieren:

- Einfache und komplexe Suchen
- Sortieren und Filtern
- Frequenzanalysen
- Typische Wortverbindungen: Diachrone Kollokationsanalysen mit "DiaCollo" in frei wählbaren Zeiträumen
- Verschiedene Ergebnisanzeigen und interaktive Visualisierungen
- Zahlreiche Exportformate
- Verlinkung zu den Digitalisaten der Quellen

Das Korpus kann nun auch vergleichend oder gemeinsam mit weiteren digitalen Textkorpora des Zentrums Sprache der BBAW analysiert werden.

4. Befunde und Ausblick



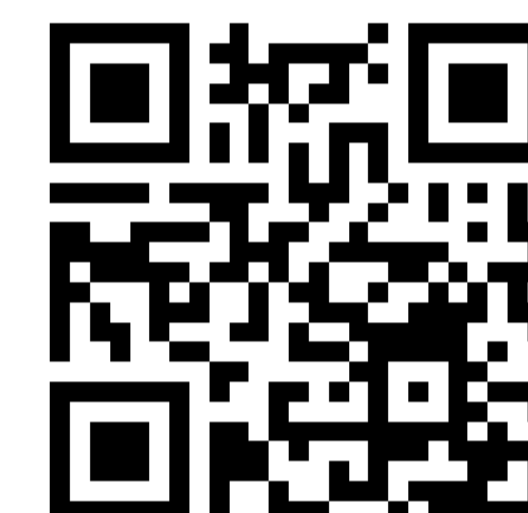
Die Möglichkeiten für digital gestützte Explorationen und Analysen historischer deutschsprachiger Schulbücher wurden erheblich erweitert.



Korpuszusammensetzung, Datenqualität und bestimmte Charakteristika von Schulbüchern haben sich als nicht optimal kompatibel mit den Logiken der Vorverarbeitungs- und Analysewerkzeuge erwiesen:

Ideal	GEI-Digital-2020 Korpus
Vollständig / Repräsentativ	Keine Gesamtbibliographie vorhanden
Gleichmäßige Verteilung über die Zeit	Sehr ungleiche Verteilung (cf. Schulpflicht, Demographie, Schulbuchproduktion)
Texttreue Volltexte	OCR-Fehler vorhanden
Monolingual	Fremdsprachliche Elemente
Ohne Dubletten	Texte mehrfach vorhanden durch mehrfache Auflagen, Regionalausgaben, Kanonbildung

Diese Besonderheiten sind bei der Interpretation von Frequenzanalysen und Kollokationsanalysen zu beachten. Grundsätzlich zeigten sich im Projekt nutzer:innenseitige Bedarfe für die ausführlich dokumentierte und auf unterschiedliche Zielgruppen abgestimmte Gestaltung von Forschungsinfrastrukturen, die ggf. auch projektspezifische Korpuszusammenstellungen und modulare, individuell durchführbare Datenkuratation erlauben.



Besuchen Sie unsere Projektseite!
<https://diacollo.gei.de/>

Kontakt: nielaender@leibniz-gei.de

Bibliographie

Hertling, Anke / Klaes, Sebastian (2018): "Historische Schulbücher als digitales Korpus für die Forschung: Auswahl und Aufbau einer digitalen Schulbuchbibliothek" in: Nieländer, Maret / De Luca, Ernesto William (eds.): Digital Humanities in der internationalen Schulbuchforschung. Göttingen: V&R unipress 21–44. DOI: [10.14220/9783737009539.21](https://doi.org/10.14220/9783737009539.21)

Hertling, Anke / Klaes, Sebastian (2018): "»GEI-Digital« als Grundlage für Digital-Humanities-Projekte: Erschließung und Datenaufbereitung" in: Nieländer, Maret / De Luca, Ernesto William (eds.): Digital Humanities in der internationalen Schulbuchforschung. Göttingen: V&R unipress 45–68. DOI: [10.14220/9783737009539.45](https://doi.org/10.14220/9783737009539.45)

Jurish, Bryan / Nieländer, Maret (2020): "Using DiaCollo for historical research" in: Simov, Kiril / Eskevich, Maria (eds.), Selected Papers from the CLARIN Annual Conference 2019, Linköping Electronic Conference Proceedings 172:5: 33–40. <https://doi.org/10.3384/ecp2020172005>

Nieländer, Maret / Jurish, Bryan (2021): D* für Anfänger:innen: Ein Tutorial. Einfache und komplexe Suchanfragen, Frequenzanalysen und diachrone Kollokationsanalysen in der D*-Korpusmanagement-Umgebung. [urn:nbn:de:0220-2021-0088](https://nbn-resolving.org/urn:nbn:de:0220-2021-0088)

Clip Art von <https://openmoji.org> CC BY-SA 4.0

Poster Layout: [ArsyaVisual](https://www.arsyavisual.com)

