

ABSCHLUSSBERICHT (Auszug, bearbeitet 9/2026)

Forschungsplattform Literarisches Feld DDR: Autor:innen, Werke, Netzwerke

Pilotprojekt: Die Student:innen des Instituts für Literatur „Johannes R. Becher“ Leipzig

gefördert aus Mitteln der DFG-Sachbeihilfe und aus Mitteln des Gottfried-Wilhelm-Leibniz-Preises (DFG) des Antragstellers Steffen Martus, 01.10.2019 – 31.12.2023

Beteiligte Institutionen: Humboldt-Universität zu Berlin – Institut für deutsche Literatur, Berlin-Brandenburgische Akademie der Wissenschaften und Berliner Hochschule für Technik (BHT) – Fachbereich Informatik und Medien

Gepris-Datenbank der DFG: <https://gepris.dfg.de/project/419244741>

Zusammenfassung / Summary

Das Projekt knüpft an Forschungen zum literarischen Feld der DDR an und zielt perspektivisch auf die prosopographische Gesamterschließung der Autor:innen in der SBZ/DDR (1945–1990). Es wurden 1.) empirische Grundlagen für quantitative Analysen geschaffen, 2.) Tools für einen niedrighwelligen Zugang zu den Daten entwickelt und 3.) digitale Verfahren für die Untersuchung literarischer Felder und Netzwerke erprobt. Das Projekt vereint so Grundlagenforschung zur bio-bibliographischen Erschließung von Autor:innen für die literaturwissenschaftliche Forschung mit dem Einsatz digitaler Forschungsansätze.

Das geförderte Pilotprojekt gilt den Direktstudent:innen des Leipziger Instituts für Literatur „Johannes R. Becher“ (IfL). Mit ihnen ist eine für die Literatur der DDR exemplarische Autor:innengruppe gegeben, die Einblicke zu Struktur und Beziehungen des literarischen Felds der DDR ermöglicht: Das 1955 gegründete IfL war als Hochschule für literarisches Schreiben eine Schlüsselinstitution im literarischen Feld der DDR. Es eröffnete seinen Absolvent:innen unterschiedliche Berufswege insbesondere in literarischen und kulturellen Bereichen, war Ausgangspunkt für die Etablierung von verschiedenartigen Netzwerken und trug zur Herausbildung diverser ästhetischer Positionen bei. Der Kreis der IfL-Studierenden kann als strukturell repräsentativ für das literarische Feld der DDR angesehen werden.

In Form einer digitalen Forschungsplattform wurden zu den 344 IfL-Direktstudent:innen Informationen zu Biografie, Werk, Rezeption und Quellen erstmals umfassend auf systematischer Quellengrundlage erschlossen und aufbereitet. Die Daten wurden durch 1.) Synthese veröffentlichter Quellen, 2.) Archivrecherchen und 3.) Autor:innen-Befragung erhoben. Insgesamt wurden rd. 14.300 biographische Ereignisse mit rd. 74.000 Elementaraussagen erfasst und mit rd. 191.000 Quellennachweisen aus insgesamt rd. 600 Quellenbeständen belegt (Stand 3/2024). Für Werk und Rezeption wurden Metadaten zu rd. 45.000 Publikationen und rd. 37.000 Rezensionen aus den *Bibliographischen Annalen. Literatur in der SBZ/DDR* importiert

und für die Nutzung in der Plattform aufbereitet. Im Rahmen der Aufbereitung der Daten wurden zu Institutionen, Berufen und (literarischen) Preisen eine DDR-spezifische Systematik und Normdaten erstellt, die für die Erforschung der DDR-Geschichte insgesamt nützlich sind. Technische Grundlage der Forschungsplattform ist ein im Rahmen des Projekts entwickeltes Softwaresystem, das Funktionen für die Dateneingabe, -abfrage, -analyse und -visualisierung bereitstellt. Auf Basis der gewonnenen Daten wurden exemplarisch Analysen zu Netzwerken, biographischen und bibliographischen Gruppenprofilen und Positionierungen im literarischen Feld DDR vorgenommen.

The project draws upon existing research on the literary field of the GDR and aims to provide a prosopographical overview of authors in the SBZ/GDR (1945–1990). We have 1) collected data from various sources as an empirical basis for quantitative analyses, 2) developed tools for low-threshold access to the data and 3) tested digital methods for the investigation of literary fields and networks. The project thus combines basic research on bio-bibliographical approaches to literary research with the use of digital research methods.

The pilot project funded by DFG is focused on those 344 students of the Leipzig Institute for Literature „Johannes R. Becher“ (IfL) who were enrolled as on-campus students („Direktstudierende“). They are an exemplary group of authors for the literature of the GDR that provides insights into the structure and relationships of its literary field: Founded in 1955, the IfL was a key institution in the literary field of the GDR as a university for literary writing. It prepared its graduates for various career paths, particularly in literary and cultural areas, was the starting point for networks between students and contributed to the development of diverse aesthetic positions. The group of IfL on-campus students can be seen as structurally representative for the literary field of the GDR.

As a digital research platform, which particularly allows for representing provenience information for all data, information on the biography, work and reception of the 344 IfL-direct-students was comprehensively and systematically gathered and prepared for quantitative analysis. The data was collected through 1) a synthesis of published sources, 2) archive research and 3) author surveys. In total, around 14.300 biographical events with around 74.000 elementary statements were recorded and documented with around 191.000 source references from a total of around 600 sources. For works and reception, metadata on around 45.000 publications and around 37.000 reviews were imported from the *Bibliographische Annalen. Literatur in der SBZ/DDR* and prepared for being used in the platform. As a part of the data preparation, a standardised GDR-specific classification system for institutions, professions and (literary) prices was created, which are useful for researching the history of the GDR as a whole. The technical basis of the research platform is a software system developed within the project, which provides functions for data input, retrieval, analysis and visualisation. Based on the data available in the platform, exemplary analyses on networks, on biographical and bibliographical group profiles and on positionings in the literary field of the GDR have been carried out.

1. Wissenschaftlicher Arbeits- und Ergebnisbericht

1.1 Ziele

Das Projekt knüpft an Forschungen zum literarischen Feld der DDR an (vgl. Kreutel u.a. 2024) und zielt auf die prosopographische Erfassung von Autor:innen, die in der SBZ/DDR 1945–1990 literarisch aktiv waren. Wir verknüpfen dabei das prosopographische Vorgehen einer auf Vollständigkeit zielenden Erfassung in der Breite mit einem auf biographische ‚Tiefe‘ ausgerichteten Ansatz und reagieren auf ein zweifaches Forschungsdefizit: Zum einen liegen zu den Autor:innen in der DDR Informationen zu Biografie, Werk und Rezeption bisher nur stark selektiv, zerstreut und unkoordiniert, widersprüchlich und in vielen Fällen nicht auf aktuellem Stand vor. Zum zweiten besteht für quantitative Forschungsansätze zur DDR-Literatur im Kontext der literaturwissenschaftlichen *Digital Humanities* bislang eine unzureichende Datenbasis. Es geht uns daher um den prototypischen Aufbau einer digitalen Datensammlung zu den Autor:innen in der DDR mit umfassenden Angaben zu Biografie (soziale Herkunft, Familie, Bildungsweg, berufliche Tätigkeiten, Mitgliedschaften, Lebens- und Wirkungsorte), Würdigungen, Rezeption und Werk. Zu diesen Daten sollen für unterschiedliche Erkenntnisinteressen und Nutzungsformen analytische Zugänge ermöglicht werden. Verbunden ist damit immer auch die Frage, wie das effiziente Erfassen der Daten gestaltet werden kann.

Als Pilotprojekt wurde eine beispielhafte Teilgruppe von Autor:innen in der DDR untersucht: die 344 Direktstudierenden des Instituts für Literatur „Johannes R. Becher“ (IfL), einer Schlüsselinstitution im literarischen Feld der DDR (zum IfL Lehn u.a. 2018). Der Kreis der IfL-Studierenden umfasst ein breites Spektrum an Feldpositionen und literarischen Karrierewegen. Unter ihnen befinden sich Autor:innen, die eine anerkannte Rolle im literarischen Feld der DDR wie auch der BRD spielten bzw. spielen (z.B. Kerstin Hensel, Thomas Rosenlöcher); Autor:innen, die den autonomen Pol des Feldes prägten (z.B. Adolf Endler); Autor:innen, die Opfer von Zensur und Verfolgung wurden (z.B. Gert Neumann, Heidemarie Härtl) und die DDR verließen (z.B. Kurt Bartsch, Sarah Kirsch, Katja Lange-Müller); Autor:innen mit engen Beziehungen zum politischen Feld (z.B. Günter Görlich); Autor:innen, die im Bereich der Unterhaltungsliteratur (z.B. Rudi Strahl) oder der Kinder- und Jugendliteratur (z.B. Gerhard Holtz-Baumert) erfolgreich waren, heute jedoch eher vergessen sind; sowie Autor:innen außereuropäischer Herkunft (z.B. Adel Karasholi, Hung Gurst). Die Werke der Absolvent:innen repräsentieren ein breites Spektrum an literarischen Gattungen und Genres: von Lyrik und Dramatik über Kinder- und Jugend-, Kriminal- und Science-Fiction-Literatur bis zu Essayistik, Reportage, Hörspiel und dem Schreiben für Film und Fernsehen. Am IfL studierten von 1955 bis 1993 insgesamt rd. 990 Personen in drei Studiengängen: in einem zuletzt dreijährigen Direktstudium für angehende Autor:innen, in einem berufsbegleitenden Fernstudium (ab 1969) und in einjährigen Sonderkursen für bereits etablierte Autor:innen (ab 1966). Das Direktstudium war der Hauptstudiengang, die anderen Optionen wurden als zusätzliche Studiengänge eingerichtet.

Ein aus unserer Sicht wegweisendes Projektergebnis liegt in der umfassenden Bestimmung des Korpus von literarischen Autor:innen in der DDR-Literatur: Vorliegende Bestandsaufnahmen sind enorm selektiv und nicht aktuell. In Vorbereitung auf das Folgeprojekt haben wir anhand von Lexika zur Literatur in der DDR, Mitgliederverzeichnissen von literarischen Organisationen und Institutionen in der DDR, Literaturzeitschriften, Anthologien zur DDR-Literatur und mit jungen DDR-Autor:innen, Publikationen der inoffiziellen Literaturszene sowie Bibliografien eine Gesamtübersicht der Autor:innen erarbeitet, die zwischen 1945 und 1990 in der SBZ/DDR – zumindest zeitweilig – lebten und schriftstellerisch tätig waren (einschließlich Autor:innen, die die DDR verließen oder nur im Westen bzw. erst nach 1990 veröffentlichten konnten). Nach unserer Quellenauswertung gehen wir insgesamt von ca. 3.400 Autor:innen aus.¹ Die Übersicht diene uns im Pilotprojekt nicht zuletzt dazu, in für Analysen genutzten zusätzlichen Datenbeständen (Bibliografien, Bibliothekskatalogen) DDR-Autor:innen zu identifizieren und als Vergleichsgruppe zu bestimmen. Wenn man zugesteht, dass es in einer solchen Bestandsaufnahme immer Grenzfälle geben wird, einzelne Personen hinzugefügt werden können und die Zugehörigkeit anderer womöglich kritisch zu diskutieren ist, liegt damit doch erstmals ein aktuelles Korpus vor, das den Anspruch auf strukturelle Repräsentation des gesamten Feldes erheben darf und den literaturhistorischen Horizont der künftigen Forschung gerade auch im Blick darauf definiert, was in jeweils aktuellen Untersuchungen nicht in den Blick genommen wird.

1.2 Erschließung der Autor:innen: Datenkonzeption und -modell

Datenkonzeption: Für die Konzeption der zu erfassenden Daten folgen wir den Standards zur Erschließung von Biografie, Werk, Rezeption und Quellen des „Deutschen Schriftsteller-Lexikons 1830–1880“ (Goedeke) und des „Deutschen Literatur-Lexikons“ (Kosch). Wir fokussieren Sachinformationen zu Biografie und Werk (vgl. Richter u.a. 2009: 138f.): Um ein möglichst breites Spektrum qualitativer und quantitativer Zugänge zu adressieren, erfassen wir für die betrachteten Autor:innen „Personenstammdaten“ wie Lebensdaten, Geburts- und Lebensorte und Geschlecht, detaillierte Informationen zur sozialen Herkunft und Familie, zum Bildungsweg, zu Berufstätigkeiten, Mitgliedschaften und Funktionen in literarischen, kulturellen und politischen Organisationen (z.B. Schriftstellerverband, Parteien) und Würdigungen; mit Blick auf spezifische Bedingungen für Schriftsteller:innen in der DDR werden ferner Aspekte politischer Verfolgung berücksichtigt (z.B. Observation durch das MfS). Für die Rezeption erfassen wir Rezensionen in ausgewählten Zeitschriften und Zeitungen; für das Werk erfassen wir die in DDR-Verlagen erschienenen Werke bibliographisch² sowie die in den Quellen genannte Zuordnung zu Gattungen und Genres.

¹ Zur Korpusbestimmung im Detail und Namensliste [hier](#). Wir öffnen deren Kuratierung als Public-Science-Projekt für Korrekturen und Ergänzungen von Seiten der Expert:innen und interessierten Öffentlichkeit.

² Datengrundlage für Werk und Rezensionen: *Literatur in der SBZ/DDR. Bibliographische Annalen 1945–1990* (Tanneberger u.a. 2021).

Datenmodell: In Anlehnung an Ontologien für die Modellierung biographischer Sachverhalte (u.a. an Tuominen u.a. 2018, ausführlich zum Datenmodell Kreutel u.a. 2023) wurden Biografien mittels „Ereignissen“ und „Entitäten“ formalisiert. Diese werden hinsichtlich eines Typs – z.B. Geburt, Ausbildung, Berufstätigkeit bzw. Person, Ort etc. – charakterisiert und im Detail mittels quellenbelegter „Elementaraussagen“ beschrieben. Auf diese Weise können alle in den verschiedenen Quellen vorgefundenen Angaben (z.B. für Berufstätigkeiten zu Inhalt, Zeit, Ort) jeweils einem Ereignis zugeordnet und sowohl übereinstimmende als auch widersprüchliche Angaben zu einem Ereignis (z.B. zum Geburtsdatum) mit ihren Quellen transparent repräsentiert werden.³ Diese Modellierung unterstützt, so das Ergebnis im Projekt, gleichermaßen quantitative wie qualitative, am Narrativ von Autor:innenbiografien orientierte, Zugänge.

1.3 Quellenauswertung und Datenerfassung

A) Zur Untersuchungsgruppe des Pilotprojekts: IfL-Direktstudierende

Für die 344 IfL-Direktstudierenden wurden systematisch Daten zu Biografie (soziale Herkunft, Bildungsweg, Berufstätigkeiten, Mitgliedschaften, Lebensorte), Werk (Bibliografie) und Würdigung (Literaturpreise u.a. andere Auszeichnungen) erhoben. Insgesamt wurden rd. 67.000 Elementaraussagen zu rd. 13.800 biografischen Ereignissen mit rd. 185.000 Quellennachweisen erfasst. Dabei wurden folgende Quellen ausgewertet:

Biographische Quellen

publizierte Quellen (print und digital/online):

- Lexika: Auswertung von 154 Lexika und Autor:innen-Handbüchern (inkl. verschiedener Auflagen), insgesamt rd. 2.400 Artikel;
- Autor:innenverzeichnisse von Anthologien und Literaturzeitschriften (rd. 500);
- Pressesammlungen des DDR-Schriftstellerverbands und der Akademie der Künste der DDR zu Autor:innen (AdK-Archiv): Autor:innenporträts, Interviews, Rezensionen;
- Webseiten und andere Online-Quellen von bzw. zu den Autor:innen;
- Forschungsliteratur zu den Autor:innen.

Archiv-Quellen – systematisch wurden folgende Archivbestände ausgewertet:

- Immatrikulations-Akten des IfL (Sächsisches Staatsarchiv Leipzig);
- Unterlagen des DDR-Schriftstellerverbands (SV) zur Förderung junger Autor:innen (Nachwuchsförderung), SV-Aufnahmeanträge für die Absolvent:innen des Literaturinstituts (AdK-Archiv);
- Reisekartei/-akten des SV zu den Reisen der Verbandsmitglieder für die Absolvent:innen des Literaturinstituts (AdK-Archiv);
- zusätzlich Einzelrecherchen in 69 Landesarchiven, kommunalen und institutionellen sowie in 18 ausländischen Archiven zur Schließung von Datenlücken und Validierung von Angaben.

³ Diese Modellierung ermöglicht auch eine Bereitstellung von Daten entsprechend einem aktuellen Vorschlag für eine generische API für prosopographische Daten, basierend auf dem Konzept von Faktoiden (vgl. Pasin u.a. 2015), vgl. Vogeler u.a. 2019.

3. Autor:innen-Befragung:

- Befragung von 102 Autor:innen per Fragebogen und Interviews.

Bibliographische Quellen

- Für die Autor:innen der Pilotgruppe und des Gesamtkorpus wurden die Daten von *Literatur in der SBZ/DDR. Bibliographische Annalen* (Tanneberger/Hillich 2021) zur autor:innen-bezogenen Nutzung für die Plattform aufbereitet; mit ihnen ist für die Autor:innen eine annähernd vollständige Bibliografie der in der SBZ/DDR 1945-1990 erschienenen Werke (inkl. Veröffentlichungen in Sammelpublikationen und Literaturzeitschriften) gegeben.

B) Zum Gesamtkorpus „Autor:innen in der DDR“

- Auswertung der Aufnahmeanträge zum DDR-Schriftstellerverband (AdK-Archiv): Erfassung biografischer Daten (soziale Herkunft, Bildungsweg, berufliche Tätigkeiten, Mitgliedschaften) zu rd. 1.200 Personen;⁴
- Auswertung Reiseakten des DDR-Schriftstellerverbands (AdK-Archiv): Daten zu Zeitraum, Zielen und Anlässen (Lesungen, Delegationen, Recherchen zu Buchprojekten etc.) von beruflichen Reisen der Verbandsmitglieder;⁵
- Auswertung der Datenbank der SLUB Dresden zu ihrer Sammlung inoffizieller Literatur- und Künstlerzeitschriften in der DDR mit Angaben zu Biografie, literarischer und künstlerischer Tätigkeit, Beteiligung an Ausstellungen und Auszeichnungen;⁶
- Auswertung der Autor:innen-Datenbank des Archivs Schreibende ArbeiterInnen: Angaben zu Biografie, literarischer Tätigkeit, Zirkel-Mitgliedschaften und Anthologie-Beteiligungen von Mitgliedern der Zirkel Schreibender Arbeiter;
- Ermittlung der relevanten Nachschlagewerke und Archivbestände zum Untersuchungskreis sowie von Schlüsselquellen zu Literaturpreisen in der DDR.

C) Data Cleansing und Klassifikation des erfassten Datenbestands

Datenbereinigung sowie Systematisierung der erfassten Berufs- und Tätigkeitsbezeichnungen, Institutionen und Würdigungen als Grundlage für Analysen; Verknüpfung der Personen mit der GND (so vorhanden) und der Orte mit Geodaten.

⁴ Erschließung der Aufnahmeakten des DDR-Schriftstellerverbands und Erfassung der biographischen Daten daraus: Zu Beginn der Archivrecherchen erwiesen sich die – bis dahin noch nicht erschlossenen – Aufnahmeakten des DDR-Schriftstellerverbands (SV) als eine Schlüsselquelle mit umfangreichen biographischen Angaben. Daher wurde die Entscheidung getroffen, diese Daten vollständig frühzeitig für die Plattform zu erfassen und ausgehend von diesem Datenbestand die weitere Datenerfassung aus Lexika und anderen Quellen durchzuführen. Die archivalische Erschließung der SV-Aufnahmeakten und Erfassung der biographischen Daten erfolgte in Kooperation mit dem Literaturarchiv der Akademie der Künste (AdK) im Rahmen eines zusätzlichen Werkvertrags (09/2019–05/2021, finanziert aus Mitteln des Leibniz-Preises des Projektleiters S. Martus).

⁵ Erschließung der Reiseakten des DDR-Schriftstellerverbands und Erfassung der Reisedaten: Ein weiterer noch nicht erschlossener Quellenbestand waren die Reiseakten des SV; diese wurden ebenfalls in Kooperation mit dem AdK-Literaturarchiv zum Teil erschlossen (ebenfalls im Rahmen eines zusätzlichen Werkvertrags, 02/2021–08/2021, finanziert aus Mitteln des Leibniz-Preises des Projektleiters S. Martus). Anhand dieser Daten wird u.a. einen genauerer und systematischerer Blick auf die westlichen und internationalen Beziehungen von Autor:innen der DDR möglich.

⁶ <https://www.deutschefotothek.de/cms/kuenstlerzeitschriften.xml>.

1.4 Forschungsdaten

Datenbeschreibung: Erhoben wurden Daten folgender Typen:

- strukturierte Repräsentationen von biographischen Ereignissen, benannten Entitäten, Publikationen und deren Verknüpfungen, durchschnittlich ca. 1 MB pro Autor:in in der verwendeten Datenbank, insgesamt ca. 340 MB;
- Textdaten mit Quelleninhalten (z. B. Lexikonartikel), insgesamt ca. 60 MB;
- Abfrageergebnisse für gezielte Datenauswertungen, inkl. Metadaten (derzeit ca. 1 GB).

Importiert wurden Daten aus den *Bibliographischen Annalen*, dem *Kallias*-Katalog des Deutschen Literaturarchivs Marbach, der GND und weiteren Normdatenbeständen wie *geonames.org*. Zur Datenhaltung wird eine relationale Datenbank verwendet; ein RDF-Datenexport wurde konzipiert und als Proof-of-Concept prototypisch implementiert.

Die Angaben aus den verschiedenen Quellen wurden quellenkritisch erfasst, koordiniert und synthetisiert.⁷ Die publizierten Angaben konnten durch die Archivquellen und Befragung konsolidiert und erweitert werden. Für die 344 IfL-Direktstudierenden und die weiteren Autor:innen wurden insgesamt rd. 14.300 biographische Ereignisse zu Herkunft und Familie, Bildungsweg, beruflichen Tätigkeiten, Mitgliedschaften, Lebens- und Wirkungsorten und Würdigungen mit rd. 74.000 Elementaraussagen erfasst und mit rd. 191.000 Quellennachweisen aus insgesamt rd. 600 Quellenbeständen belegt (Stand 3/2024). Für Werk und Rezeption wurden Metadaten zu rd. 45.000 Publikationen und Periodika und rd. 37.000 Rezensionen importiert und für die Nutzung in der Plattform aufbereitet. Die entwickelte Forschungssoftware hat einen Umfang von ca. 100.000 Quellcodezeilen.

Datenqualität: Die erfassten Daten wurden zunächst manuell verifiziert. Im nächsten Schritt wurden sie standardisiert, systematisiert und kategorisiert, so dass sie für quantitative Auswertungen genutzt werden können: Die aus den Quellen resultierende historische Vielfalt der Bezeichnungen von Berufstätigkeiten wurde dokumentiert und in einer Systematik (mit Berufsfeldern und Qualifikationsniveaus) geordnet.⁸ Zu den erfassten Institutionen und Auszeichnungen wurde eine Systematik erarbeitet, die Spezifika der DDR-Geschichte berücksichtigt, und die Entitäten jeweils kategorisiert (Institutionstypen, Auszeichnungstypen etc.).

Speicherung und technische Sicherung während des Projektverlaufs: Die Datensicherung wird für die gesamten Bestände durch tägliche Sicherungskopien durchgeführt, die getrennt von der Infrastruktur sind, auf der die Plattform betrieben wird, und für die durch den Hosting-Anbieter Backups seinerseits durchgeführt werden.

⁷ Die Angaben wurden in der Datenbank je Person unter Angabe der jeweiligen Quellen erfasst und gemäß Datenmodell biographischen Ereignissen zugeordnet (inkl. widersprüchlicher Angaben). Die Daten für Personen und Orte wurden mit Normdaten (u.a. GND, geonames.org) verknüpft.

⁸ Wir stützen uns auf die Berufsklassifikation der Bundesagentur für Arbeit, die ihrerseits auf der *International Standard Classification of Occupations* als international in den Sozialwissenschaften etabliertes Klassifikationsschema für Berufe basiert. Zur Vereinheitlichung der Berufsbezeichnungen und der damit verbundenen Problematik, historische Kontexte zu überschreiben, vgl. Kreutel u.a. 2023, 2024.

Sicherheit sensibler Daten: Der Zugriff auf die Plattform ist auf die Mitglieder der Projektgruppe und Kooperationspartner beschränkt und passwortgeschützt. Sensible Inhalte können auf verschiedenen Ebenen – Elementaraussagen, Ereignisse, einzelne Quelldokumente, Bestände von Quelldokumenten – markiert und ggf. ausgeblendet werden.

Rechtliche Verpflichtungen und Rahmenbedingungen: Im Rahmen der biographischen Daten zu den Autor:innen wurden auch personenbezogene Daten erhoben, die dem Persönlichkeits- und Datenschutz unterliegen. Die Verwendung dieser Daten unterliegt einer individuellen Prüfung. Bei Daten aus Archiven werden die geltenden Schutzfristen eingehalten. Für die Nutzung und Veröffentlichung wurde mit den betreffenden Archiven eine Verkürzung der Schutzfrist vereinbart (Veröffentlichung zehn Jahre nach Tod einer betreffenden Person). Bei Daten zu vor der 10-Jahres-Frist verstorbenen oder noch lebenden Personen wurde versucht, von diesen bzw. deren Erben die Einwilligung zur Veröffentlichung einzuholen. Daten ohne Veröffentlichungsgenehmigung werden bis zum Ablauf der Schutzfrist in der Plattform als vertraulich gekennzeichnet und für den externen Zugriff gesperrt, stehen jedoch für Analysen in anonymisierter Form zur Verfügung (die Anonymisierungsregeln folgen den Konventionen des Literaturarchivs der Akademie der Künste). Rechteinhaber:innen sind generell durch die Quellenzuweisung ausgewiesen.

Datenaustausch und dauerhafte Zugänglichkeit der Daten: Für die Datennutzung existieren folgende Schnittstellen:

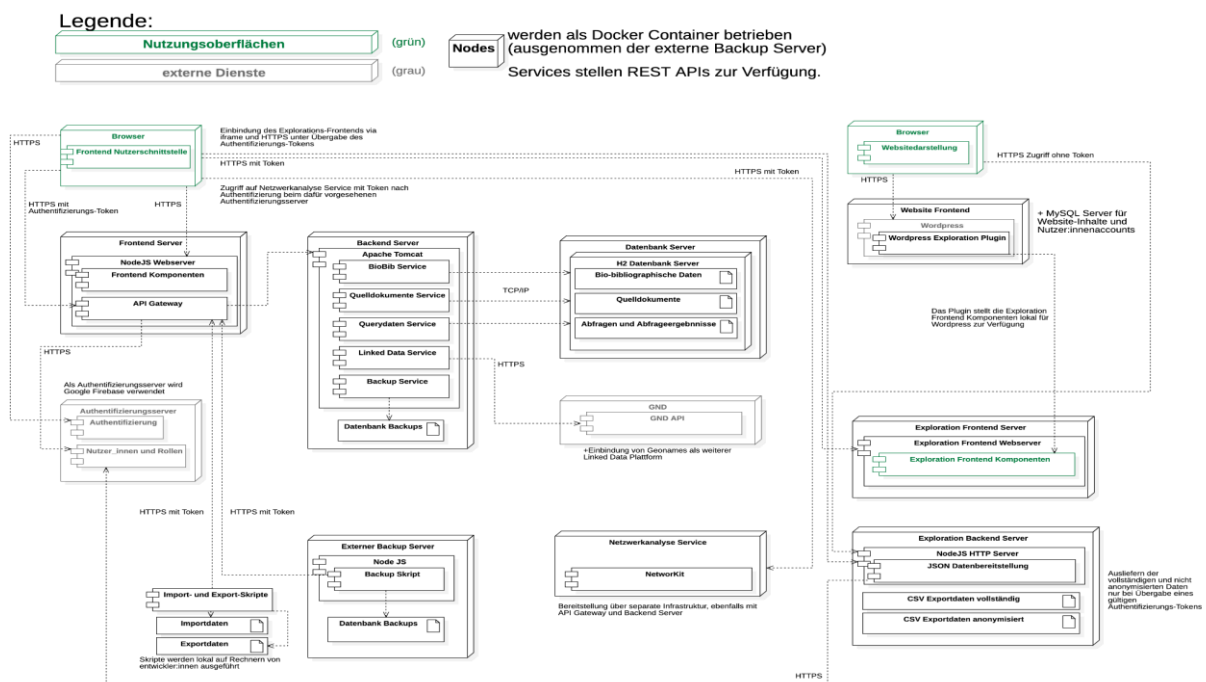
- Eine in die Projektwebsite eingebundene und am *Information Seeking* Mantra (Shneiderman 1996) orientierte niedrigschwellige Nutzungsoberfläche zur Exploration der anonymisiert bereitgestellten Daten;
- die zur Datenerfassung verwendete webbasierte Nutzerschnittstelle der Plattform und die darin integrierten Funktionen für Datenabfrage, Analyse und Visualisierung (inkl. Exportfunktion von Abfrageergebnissen im CSV-Format);
- eine REST API (Fielding 2000) für den Zugriff auf die erfassten Daten im JSON-Format für registrierte Nutzer:innen.

Darüber hinaus ist ein Export der Daten im CSV-Format möglich. Der Export im RDF-Format wurde konzipiert und als Proof-of-Concept prototypisch implementiert (Bulaj 2022). RDF als Exportformat soll bereitgestellt werden, sobald die Daten in der NFDI-Infrastruktur bereitgestellt werden können. Aufgrund des Gewichts biographischer Daten als dem originär im Projekt erarbeiteten Datenbestand ist – in Absprache mit den NFDI-Verantwortlichen – das NFDI4Memory-Konsortium die geeignete Infrastruktur. Die Veröffentlichung der Daten hängt somit von der Weiterentwicklung dieses Konsortiums ab. Unabhängig von der NFDI können die Daten für wissenschaftliche Anfragen unter Berücksichtigung der oben genannten Rechtefrage als CSV-Export im Zenodo-Repository des CERN bereitgestellt werden (<https://zenodo.org>).

Verantwortlichkeiten und Ressourcen: Die Verantwortung für Datenqualität und die Markierung der Daten hinsichtlich Freigabebeschränkungen liegt bei den Fachwissenschaftler:innen; technisch wird der Datenzugriff gemäß dem Berechtigungskonzept durch die BHT gewährleistet. Hosting und Betrieb des Softwaresystems erfolgen in einer Container-Infrastruktur (Merkel 2014) und werden durch J. Kreutel verantwortet.

1.5 Die Plattform: Softwareinfrastruktur, Analyse- und Visualisierungsfunktionen

Für die Forschungsplattform wurde eine erweiterbare, am *Microservices*-Ansatz orientierte (Newman 2015) und mittels Docker Containern (Merkel 2014) betriebene webbasierte Infrastruktur mit relationalen Datenbanken, REST APIs (Fielding 2000) in Java für den Datenzugriff und einer modular organisierten, browsergestützten Nutzungsoberfläche als *Progressive Web Application* (Biørn-Hansen u.a. 2017) aufgebaut. Ein Überblick über die Softwareinfrastruktur der Plattform und deren Softwarekomponenten ist in der folgenden Abbildung zu sehen. Neben dem Kernsystem aus Frontend, Backend und Datenbank umfasst die Plattform die Projektwebsite, ein Backup-System sowie Softwarekomponenten für Datenimport und -export und für Datenkonsolidierung und -anreicherung auf Grundlage von Excel-Tabellen und anderen Formaten, inkl. des oben genannten JSON-Exportformats.



Zum gegenwärtigen Zeitpunkt (Frühjahr 2024) ist die Verwendung der Software auf den konkreten Rahmen des Projekts und des dafür entwickelten Datenmodells beschränkt, das auf die fachlichen Projektanforderungen abgestimmt ist. Auf Grundlage des im Projekt praktizierten modellbasierten Entwicklungsansatzes dient das in Java erstellte fachliche Datenmodell nicht nur als Vorlage für die automatische Generierung des Datenbankschemas, sondern steuert in Form von Metadaten auch den Aufbau der Nutzungsoberfläche. Hinsichtlich dieser Aspekte

ist die Software generisch und grundsätzlich auch von anderen Projekten mit alternativem Gegenstand und Fokus auf quellenbasierter Datenerfassung verwendbar (unter der Voraussetzung, dass die beiden Quellcoderepositorien für Nutzungsoberfläche und Backend in jeweils ein generisches und ein projektspezifisches Teilprojekt getrennt werden).⁹

Für die Weiterentwicklung der Plattform und für eine Ausweitung des Kreises der zu erfassenden Personen über die IfL-Studierenden hinaus wurde für die Unterstützung der Datenerfassung durch *Natural Language Processing* (NLP) eine prototypische Integration mit NLP-Werkzeugen implementiert und verifiziert.

Analyse- und Visualisierungsmöglichkeiten: Die in die Plattform integrierten Funktionen für Datenzugriff und Visualisierung ermöglichen quantitative Analysen und qualitative Betrachtungen. Sie bieten für unterschiedliche Betrachtungsperspektiven und Detailebenen Zugang zu den erhobenen Daten, beginnend mit der Detailebene der individuellen Autor:innenbiografien über Zugänge anhand ausgewählter Merkmale, Ereignisse, Institutionen und spezifischer Autor:innengruppen bis hin zu Netzwerken, die aus den erfassten Daten generiert und dann analysiert und visualisiert werden können. Für analytische Zugänge, deren Ergebnisse mittels einer Konfigurationsoberfläche nach Interessen der Nutzer:innen in verschiedener Weise visualisiert werden können, steht ein visueller Editor für strukturierte Abfragen zur Verfügung (zu den Möglichkeiten für Datenzugriff und Visualisierung vgl. den [Überblick](#) auf der Projektwebseite; Anwendungsbeispiele Kreutel u.a. 2023, 2024). Darüber hinaus wurde im Rahmen der Projektförderung der Prototyp der o.g. explorativen Nutzungsoberfläche entwickelt, mittels derer der Datenbestand in seiner Gesamtheit und ohne vorherige Suchanfragen, sondern in direkter Interaktion mit der Datenvisualisierung erschlossen werden kann; dieser intuitive Zugang wird perspektivisch auch für nicht authentifizierte Nutzer:innen über die Projektwebseite zugänglich sein.

1.6 Heuristische Feld- und Netzwerkanalysen

Auf Basis der erhobenen Daten haben wir exemplarisch Analysen zu Netzwerken und Positionierungen im literarischen Feld DDR sowie Profilen von Autor:innen-Gruppen vorgenommen und die Funktionalität und Nutzerfreundlichkeit der Analysetools getestet (Beispiele auf der Projektwebseite). In Fallstudien zu einzelnen Autor:innen wurde dabei auch erprobt, wie sich die Selektivität der Quellen offenlegen lässt, um die in diesen tradierten biographischen Konstruktionen aufzubrechen, und wie sich mit Visualisierungen alternative Betrachtungsperspektiven generieren lassen (Kreutel u.a. 2023, 2024). In vertiefenden Pilotstudien haben wir die Analysewerkzeuge auf das Handlungssystem Kinder- und Jugendliteratur (KJL) in der DDR angewendet, um etablierte Ansichten zu charakteristischen Struktureigenschaften des literarischen Felds der DDR insgesamt einer Überprüfung zu unterziehen. Damit fragten wir zugleich

⁹ Der Quellcode aller Komponenten der Plattform wird in privaten GitHub-Repositorien verwaltet (<https://github.com>). Eine individuelle Freigabe ist auf Anfrage (Kontakt: plattform@ddr-literatur.de) möglich. Die Bereitstellung des Quellcodes als Open Source soll im Zuge eines Anschlussprojekts erfolgen.

nach dem Status von KJL sowohl innerhalb des literarischen Felds DDR als auch innerhalb individueller Autor:innen-Karrieren. Für Autor:innen des Kinderbuchverlags Berlin in den 1970er Jahren wurde das biographische und bibliographische Profil dieser Autor:innengruppe nachgezeichnet und gefragt, inwieweit mit dieser Gruppe eine neue KJL-Autor:innengeneration auftrat, wie sie sich als KJL-Autor:innen etablierte und welche Stellung ihr in der KJL zukam.¹⁰ Die Untersuchung zu den 1970er Jahren dient uns als Grundlage für eine weitergehende Analyse der Gesamtheit der KJL-Autor:innen in der DDR und ihrer Verortung im literarischen Feld der DDR, einen Schwerpunkt bildet dabei die Differenzierung der KJL-Autor:innen anhand bibliographischer und biographischer Profile.

1.7 Durchgeführte wissenschaftliche Veranstaltungen, Maßnahmen zur Wissenschaftskommunikation

Projektworkshop „Forschungsplattform Literarisches Feld DDR“, 26.01.2024, mit Literaturwissenschaftler:innen, Historiker:innen, Informatiker:innen und Vertreter:innen des Literaturarchivs der Akademie der Künste Berlin.

Vortrag auf der Tagung „Kinder- und Jugendliteratur in der DDR – die 1970er Jahre“ zum Thema „Die DDR-Kinder- und Jugendliteratur der 1970er Jahre aus bio-bibliographischer Sicht“, 07.09.2023.

Öffentliche Projektpräsentation beim „Salon Sophie Charlotte“ an der Berlin-Brandenburgischen Akademie der Wissenschaften, 13.05.2023.

Vortrag auf der „Conference for Research Software Engineering in Germany“ zum Thema „User centering and model-based development – a perfect match?“, 21.02.2023.

Vortrag im Rahmen der Reihe „Digital*Humanities im Gespräch“ des Dahlem Humanities Center: „Erfassung und Analyse bio-bibliographischer Daten von Autor:innen am Beispiel der Forschungsplattform Literarisches Feld DDR“, 20.10.2022.

Vortrag auf der Jahrestagung des Forschungsschwerpunkts digitale_Kultur „Digitale Hermeneutik: Maschinen, Verfahren, Sinn“ zum Thema „(Netzwerk-)Visualisierungen als biographische Sinnstiftung“, 29.06.2022.

Öffentliche Projektpräsentation beim „Salon Sophie Charlotte“ an der Berlin-Brandenburgischen Akademie der Wissenschaften, 21.05.2022.

Präsentation der Forschungsplattform an der Berlin-Brandenburgischen Akademie der Wissenschaften, 10.12.2021.

Online-Präsentationen auf der „EADH 2021“-Konferenz der European Association for Digital Humanities zu den Themen „User-centered design of a query editor for bio-bibliographical data retrieval by domain experts“ und „Evaluation of a text-based data entry interface for domain experts“, Krasnojarsk/ Russland, 21.-25.9.2021.

Online-Poster-Präsentation auf der „ACH2021“-Konferenz der Association for Computers and the Humanities zum Thema „Evaluation of a query editor for bio-bibliographic data retrieval by domain experts“, 21.-23.7.2021.

Online-Video-Präsentationen beim „Digital Humanities Summer Institute“ der University of Victoria, British Columbia/ Kanada, zu den Themen „Accessing bio-bibliographic data with a user friendly query editor“ und „Employment of a text-based user interface for manual data collection“, 12.05.2021.

¹⁰ [Präsentation](#) auf der Tagung „Kinder- und Jugendliteratur in der DDR – die 1970er Jahre“, 07.-09.09.2023, Potsdam.

Online-Vortrag für die Projektgruppe des SoNAR (IDH) Projekts (Interfaces to Data for Historical Social Network Analysis and Research), 19.04.2021.

Online-Vortrag im Rahmen des „Networkit Day 2020“ zum Thema „Integrating NetworkKit into a web-based Environment for Network Analysis“, 15.10.2020.

1.8 Veröffentlichte Projektergebnisse

Für die öffentliche Vermittlung der Projektergebnisse wurde eine Projektwebsite aufgesetzt (www.ddr-literatur.de). Dort werden die im Pilotprojekt untersuchten Autor:innen sowie das Gesamtkorpus von Autor:innen in der DDR gelistet (s. 1.1; die Kuratierung der Gesamtliste ist als Public-Science-Projekt für die Öffentlichkeit geöffnet). Weiterhin finden sich dort Analysenproben und Präsentationen, Arbeitsberichte, Fundstücke aus den Archivrecherchen sowie Hinweise zur aktuellen Forschung zur Literatur in der DDR.

Dahlke, Birgit (2020): Sozialistische Kaderschmiede oder Exterritorium? Ein Blick auf das Leipziger Literaturinstitut „J. R. Becher“ und einige seiner Absolventinnen. In: *Generationalität – Gesellschaft – Geschichte*, hrsg. von Norman Ächtler u.a. Berlin, S. 301-322.

Dahlke, Birgit/ Kreutel, Jörn/ Martus, Steffen/ Möbius, Thomas (i.E.): Die Kinder- und Jugendliteratur der DDR der 1970er Jahre aus bio-bibliographischer Sicht. In: *Die Kinder- und Jugendliteratur in der DDR*, hrsg. von Carsten Gansel, Monika Hernik, José Fernández Pérez. Berlin u.a.: de Gruyter.

Frenzel, Tabea/ Philipp, Luisa/ Kreutel, Jörn (i.E.): Evaluation of a text-based data entry interface for domain experts. In: *Book of Abstracts zur EADH2021-Konferenz der European Association for Digital Humanities*.

Kreutel, Jörn/ Möbius, Thomas/ Dahlke, Birgit/ Martus, Steffen (2023): Forschungsplattform Literarisches Feld DDR: Ein Werkstattbericht zur prosopographischen Erfassung von Schriftsteller:innen in der DDR. In: *Historische Biographik und kritische Prosopographie als Instrumente in den Geschichtswissenschaften*, hrsg. von Helmuth Albrecht, Michael Farrenkopf, Helmut Maier, Torsten Meyer. Berlin u.a.: de Gruyter Oldenbourg, S. 141-166, <https://doi.org/10.1515/9783111137063-008>.

Kreutel, Jörn/ Möbius, Thomas/ Dahlke, Birgit/ Martus, Steffen (2024): Visualisierung und Sinnstiftung. Repräsentation biographischer Daten in der Forschungsplattform Literarisches Feld DDR. In: *Digitale Hermeneutik: Maschinen, Verfahren, Sinn*, hrsg. von Thomas Bedorf, Peter Risthaus. Hagen: Hagen University Press.

Philipp, Luisa/ Kreutel, Jörn (i.E.): User-centered design of a query editor for bio-bibliographical data retrieval by domain experts. In: *Book of Abstracts zur Konferenz der European Association for Digital Humanities 2021*.

1.9 Literaturverzeichnis zum Bericht

Biørn-Hansen, Andreas u.a. (2017): Progressive web apps: The possible web-native unifier for mobile development. In: International Conference on Web Information Systems and Technologies. Vol. 2, S. 344-351.

Bulaj, Sebastian (2022): Abbildung objektorientierter Daten auf RDF – Konzeption und prototypische Entwicklung eines Abbildungsverfahrens für ein komplexes Datenmodell. Masterarbeit, Berliner Hochschule für Technik.

Fielding, Roy Thomas (2000): Architectural styles and the design of network-based software architectures. PhD Dissertation, University of California.

Gruber, Christine u.a. (2020): The Austrian Prosopographical Information System (APIS) – Vom gedruckten Textkorpus zur Webapplikation für die Forschung. Wien u.a.

Vogeler, Georg u.a. (2019): Data exchange in practice: Towards a prosopographical API. In: Proceedings of the Third Conference on Biographical Data in a Digital World (BD2019).

Lehn, Isabelle/ Macht, Sascha/ Stopka, Katja (2018): Schreiben lernen im Sozialismus. Das Institut für Literatur „Johannes R. Becher“. Göttingen.

Merkel, Dirk (2014): Docker: lightweight Linux containers for consistent development and deployment. In: Linux Journal 239, Aufsatz Nr. 2.

Newman, Sam (2015): Building Microservices: Designing Fine-Grained Systems. Sebastopol/ California.

Pasin, Michele u.a. (2015): Factoid-based prosopography and computer ontologies: towards an integrated approach. In: Digital Scholarship in the Humanities 30:1, S. 86-97.

Richter, Myriam u.a. (2009): Biographische Kleinformen. In: Handbuch Biographie, hrsg. von Christian Klein. Stuttgart u.a., S. 137-142.

Schlögl, Matthias u.a. (2020): Die APIS-(Web-)Applikation, das Datenmodell und System. In: The Austrian Prosopographical Information System (APIS) – Vom gedruckten Textkorpus zur Webapplikation für die Forschung, hrsg. von Christine Gruber u.a. Wien u.a., S. 31-48.

Shneiderman, Ben (1996): The eyes have it: A task by data type taxonomy for information visualizations. In: Proceedings of the 1996-IEEE-Symposium on Visual Languages, S. 336-343.

Tanneberger, Horst u.a. (2021): Literatur in der SBZ/DDR. Bibliographische Annalen 1945–1990. Berlin u.a.

Tuominen, Jouni u.a. (2018): Bio CRM: A data model for representing biographical data for prosopographical research, in: Proceedings of the Second Conference on Biographical Data in a Digital World 2017 (BD2017).

2 Weitere Angaben

2.1 Selbstevaluation und Effizienzverbesserung

Die Projektorganisation war durch engen und kontinuierlichen Austausch der beteiligten Fachwissenschaftler:innen und Softwareentwickler:innen, inkl. der studentischen Mitarbeiter:innen, gekennzeichnet. In der ersten Phase des Projekts von Oktober 2019 bis Sommer 2020 wurden im Rhythmus von ein bis zwei Wochen Jour fixes durchgeführt und dann über den gesamten Projektzeitraum hinweg in der Regel zweiwöchentlich. Die initialen Prototypen des Datenmodells und der Eingabeoberfläche, die von J. Kreutel zum Projektbeginn auf Basis des Datenschemas der Vorhabensbeschreibung entworfen worden waren, wurden so hinsichtlich der spezifischen fachwissenschaftlichen Anforderungen und anhand von Beispielmateriale aus den zu erfassenden Quellen kontinuierlich weiterentwickelt. Die Jour fixes dienten darüber hinaus dazu, die Fachnutzer:innen mit den von ihnen zu bedienenden Oberflächen vertraut zu machen und frühzeitig Optimierungsvorschläge aufzunehmen. Auch im weiteren Projektverlauf wurden unter Berücksichtigung der sich ab Sommer 2020 herausbildenden Arbeitsabläufe mit der Plattform Anpassungen und Erweiterungen der zur Erfassung bereitgestellten Nutzungsoberflächen vorgenommen. Die gemeinsame Arbeit im Projekt konnte auf dieser Basis auch unter Pandemiebedingungen ab dem Frühjahr 2020 weitgehend bruchlos fortgeführt werden. Zur Unterstützung der Datenerfassung über die im Pilotprojekt betrachteten IfL-Direktstudierenden hinaus erscheint der Einsatz von Verfahren des *Natural Language Processing* (NLP) nicht nur naheliegend und notwendig, sondern auch realistisch umsetzbar. Einfache Verfahren auf Basis von Wortarterkennung, Lemmatisierung und *Named Entity Recognition* für Quellentexte wurden während des Pilotprojekts prototypisch in die Softwareinfrastruktur und Nutzungsoberfläche der Projektplattform integriert, jedoch noch nicht mit Endnutzer:innen getestet. Zum gegenwärtigen Zeitpunkt (Frühjahr 2024) werden im Rahmen einer Kooperation mit M. Stede von der Universität Potsdam Möglichkeiten des Einsatzes von großen Sprachmodellen und generativer KI für die Extraktion biographischer Daten aus einer Menge von Quellentexten zu einem/r gegebenen Autor:in erprobt.

2.2 Forschung und Lehre

Der Forschungsprozess wurde von Beginn des Projekts an mit der Lehre verbunden: In vier Projektseminaren im Wintersemester 2019/20, Sommersemester 2020, Sommersemester 2021 und Wintersemester 2021/22 wurden Studierende an der HU Berlin in das Forschungskonzept und in Methoden der Auswahl und Auswertung projektrelevanter Archivquellen und der Quellenkritik eingeführt. Quellen für 40 IfL-Direktstudierende wurden durch die Teilnehmer:innen in ein Datenschema überführt, das sich am Datenmodell des Projekts orientierte. Aus dem Kreis der Seminarteilnehmer:innen wurden studentische Hilfskräfte rekrutiert und längerfristig im Projekt beschäftigt. Ausgewählte studentische Ergebnisse aus den Seminaren wurden auf der Website des Projekts veröffentlicht.

Im Projektzeitraum sind an der BHT die folgenden durch J. Kreutel betreuten Abschlussarbeiten mit Bezug zur Projektplattform entstanden:

- „Vergleich existierender Frameworks zur Visualisierung von Graphen auf Basis von Webtechnologien“ (BSc, Nives Boni, 2019)
- „Benutzungsoberflächen für Abfragen über komplexen Datenbeständen“ (MSc, Luisa Philipp, 2021)
- „Erstellung, Visualisierung und Exploration von Datenanalysen“ (MSc, Jacqueline Hofmann, 2021)
- „Kontrollierte Sprache als alternative Eingabemodalität für Formulardaten – Konzeption und prototypische Implementierung eines Texteditors auf Basis von Webtechnologien“ (MSc, Tabea Frenzel, 2021)
- „Visualisierung von Graphen für die Analyse sozialer Netzwerke: Recherche existierender Lösungen und prototypische Eigenentwicklung auf Basis von Webtechnologien“ (MSc, Cinzia Fabiana Errichiello, 2021)
- „Visualisierung und Exploration von Graphen: Konzeption einer API zur Graphexploration und prototypische Implementierung“ (MSc, Marie Schumacher, 2021)
- „Data Visualization Based on Web Technologies – Analysis of Existing Visualization Libraries and Prototypical Development of a Web Application“ (MSc, Farah Cahya Artika, 2022)
- „Visualisierung und Exploration sozialer Netzwerke mit 3D Webtechnologien – Existierende Lösungen und Prototypische Eigenentwicklung“ (MSc, Gero Bader, 2022)
- „Abbildung objektorientierter Daten auf RDF – Konzeption und prototypische Entwicklung eines Abbildungsverfahrens für ein komplexes Datenmodell“ (MSc, Sebastian Bulaj, 2022)
- „Visuelle vs. codebasierte Werkzeuge für automatisierte Tests von Webanwendungen – Vergleichende Analyse existierender Lösungen und Entwicklung einer Testsuite für eine komplexe Benutzeroberfläche“ (MSc, Sali Hassan, 2022)

2.3 Qualifikation von Forschenden

Seit Oktober 2023 ist an der BHT eine Promotionsstelle mit Bezug zur Projektplattform angesiedelt. Betreut wird die Promovendin, Axelle Lecroq, durch J. Kreutel in Kooperation mit M. Burghardt von der Universität Leipzig (Computational Humanities). Im Rahmen dieses Promotionsprojekts erfolgt die Weiterentwicklung und Erprobung der im Projekt prototypisch umgesetzten explorationsbasierten Nutzer:innenoberfläche, welche für interne und externe Fachwissenschaftler:innen sowie für die interessierte Öffentlichkeit einen niedrigrschwelligen Zugang zu den erfassten Daten und deren Auswertung ermöglichen soll.