

# Normdaten geben und nehmen, oder: Frag nicht, was die GND für Dich tun kann, sondern was Du für die GND tun kannst

## Jurst-Görlach, Denise

jurst-goerlach@em.uni-frankfurt.de  
Goethe-Universität Frankfurt, Deutschland  
ORCID: 0000-0002-7130-1121

## Kollatz, Thomas

thomas.kollatz@adwmainz.de  
Akademie der Wissenschaften und der Literatur | Mainz, Deutschland  
ORCID: 0000-0003-1904-1841

## Jarosch, Julian

Julian.Jarosch@adwmainz.de  
Akademie der Wissenschaften und der Literatur | Mainz, Deutschland  
ORCID: 0000-0001-8483-8123

## Ausgangslage

Waren Normdaten ursprünglich ein von und für bibliothekarische Erschließung entwickeltes Werkzeug, sind sie heute aus einem breiten Feld von digitalen Forschungs- und Editionsprojekten nicht mehr wegzudenken. Die eindeutige Referenzierbarkeit von Personen, Körperschaften und Orten spielt sowohl für die Erfassung und Auswertung innerhalb des einzelnen Projekts eine wichtige Rolle als auch für die Verknüpfung mit anderen Projekten und damit ihrer Nutzbarkeit (Stadler, 2012). Auch forschungsfördernde Institutionen wie die DFG legen zunehmend Wert auf einen “fachspezifisch adäquate[n] Umgang mit Forschungsdaten”<sup>1</sup>. Dies umfasst im geisteswissenschaftlichen Bereich nicht zuletzt die Anbindung der im Projekt erhobenen Daten an kontrollierte Vokabularien, also Normdateien.

Während Projekte zunehmend versuchen, ihre Daten mit den Diensten von GND, VIAF, GeoNames oder TGN abzugleichen, stoßen sie immer wieder an die Grenzen dieser Angebote, die domänenspezifische ebenso wie historische Gründe haben (Busch und Müller, 2023). So waren etwa die Personeneinträge der GND ursprünglich an mindestens eine selbständige Publikation dieser Person im deutschsprachigen Raum gebunden, was die Aufnahme marginalisier-

ter Personen<sup>2</sup> und bestimmter Berufsgruppen in der Vergangenheit signifikant unwahrscheinlicher machte. Auch wurden und werden Datensätze von Personen des 19. und 20. Jahrhunderts, die nicht in der ‘ersten Reihe’ der öffentlichen Aufmerksamkeit stehen, häufig nicht aktualisiert, so dass wesentliche Angaben zu ihnen (etwa das Sterbedatum) fehlen. Dies muss für die eindeutige Referenzierbarkeit einer Person nicht zwangsläufig ein Problem darstellen, allerdings werden Normdateien in Projekten häufig auch zur Anreicherung eigener Daten – etwa in Registern – genutzt, in denen das Fehlen solch grundlegender Informationen mindestens zu Irritationen, wenn nicht zu groben Irrtümern und Fehlinterpretationen führt.

Schon früh wurde daher von Forschenden der Bedarf formuliert, auch eigene Ergebnisse in die großen Normdatendienste einfließen zu lassen. Spätestens mit der Propagierung des Semantic Web wurde daraus die Forderung, die “Holschuld muss in eine Bringschuld umgewandelt werden, der Fetch- in einen Pushdienst” (Kamzelak, 2016, S. 434). Die Umsetzung dieses – durchaus als berechtigt erkannten – Wunsches scheiterte im Fall GND bislang jedoch weitestgehend an den hierarchischen Strukturen und der personellen Ausstattung der beteiligten Institutionen.

Während aktuell ein neuer Anlauf in Sachen GND-Agentur unter dem Dach des NFDI-Konsortiums Text+ läuft (Kett et al., 2022), hat die wissenschaftliche Community einstweilen andere Wege und Möglichkeiten gesucht, ihre Daten strukturiert, nachnutzbar und interoperabel zu gestalten, wobei die freie Wissensdatenbank Wikidata eine zunehmend zentrale Rolle spielt.

## Der BKD-Workflow

Das Poster soll ausgehend vom Editionsprojekt *Buber-Korrespondenzen Digital* (BKD)<sup>3</sup> und am Beispiel von Personennormdaten zeigen, wie Workflows im Umgang mit Normdaten entwickelt werden können, die sowohl die konkreten und oft akuten Bedarfe des Projektes im Blick haben als auch die Möglichkeit einer (nachträglichen) Aufbereitung und nachhaltigen Integration der qualitativen Forschungsdaten in die GND.

Das Quellenkorpus von BKD umfasst mehr als 43.000 Briefe von und an den bedeutenden Religionsphilosophen Martin Buber (1878–1965), die in rund 70 Jahren und mindestens 16 Sprachen gewechselt wurden. Im Zuge der Projektarbeit wurden bislang etwa 7.000 Korrespondenzpartner:innen und über 600 weitere individualisierte Personen (Erwähnungen) identifiziert. Diese werden zunächst mit einem projektinternen Identifier versehen, in einem nächsten Schritt – sofern möglich – mit einem GND- und Wikidata-Eintrag verknüpft. Hierfür kommen neben der manuellen Eingabe zwei Werkzeuge zum Einsatz: Zum einen nutzen wir für die Datenharmonisierung den Reconciliation-Service von OpenRefine<sup>4</sup>, zum anderen führen wir Datenanreicherung – konkret die wechselseitige Ergänzung von GND- und Wikidata-Identifikatoren in unserer Personenliste – mit dem Datenanalyse-Tool KNIME<sup>5</sup> durch. Für

den im Projekt häufigen Fall, dass eine Person nicht oder mit nur rudimentären Angaben in der GND vorhanden ist, sich über Recherchen jedoch ausreichend gesicherte Informationen ermitteln lassen, werden diese in bestehenden oder neu angelegten Wikidata-Einträgen gesammelt und belegt. Über die Q-ID der Wikidata lassen sich die – vom eigenen Projekt oder anderen Akteur:innen – erhobenen Informationen schließlich vielfältig weiternutzen: einmal für das Enrichment eigener Registerdaten – im Poster exemplarisch anhand der für Netzwerkanalysen relevanten Kategorie *Gender* demonstriert –, in Zukunft aber auch für die Extraktion und Aufbereitung der Daten im Austausch- und Speicherformat entityXML<sup>6</sup>, das eigens dafür konzipiert wurde, der Text+ GND-Agentur die effiziente Einarbeitung von wissenschaftlich gesicherten Daten in die GND zu ermöglichen.

Ob und inwiefern dieser Workflow den unterschiedlichsten Bedarfen von Edition und Forschung (projektintern und darüber hinaus) gerecht wird und wo es noch Ver- und Nachbesserungsbedarf gibt, erprobt das Projekt nicht nur in der täglichen Editionsarbeit, sondern diskutiert es auch mit der Fachcommunity, zuletzt im Rahmen zweier Workshops zu “Potentialen von Normdaten in den Jüdischen Studien” (Frankfurt, 30.09.–01.10.2024)<sup>7</sup> sowie zum Thema “Briefe von Frauen” und wie diese erforscht werden können (Darmstadt, 12.–13.11.2024)<sup>8</sup>. Die Ergebnisse dieser Workshops bestätigen die gewählten Vorgehensweisen als zukunftsfruchtig und für zahlreiche Projekte anschlussfähig.

## Ausblick

Aus den Erfahrungen unserer Projektarbeit ergeben sich die – potenziell projektübergreifend übertragbaren – Prinzipien, strukturierte Forschungsdaten erstens sofort in Wikidata ohne bürokratische Einstiegshürde zu veröffentlichen und diese zweitens mit Blick auf die langfristige Verfügbarkeit in das Format entityXML zu transformieren. Dies zielt darauf ab, an Normdateien anschlussfähige Forschungsdaten parallel auf mehreren Wegen offen zur Verfügung zu stellen. Wir hoffen, so nicht nur GND-Daten in unser Projekt einfließen zu lassen, sondern auch einen Beitrag zu leisten, qualitativ hochwertige fachwissenschaftliche Recherchen in die GND zurückfließen zu lassen.

## Fußnoten

1. <https://www.dfg.de/de/grundlagen-themen/grundlagen-und-prinzipien-der-foerderung/forschungsdaten> (29.11.2024).
2. Neben einem signifikanten Gender-Data-Gap (Juen, 2021) stößt unser Projekt zudem auf die Marginalisierung von Jüd:innen, Emigrant:innen sowie Personen osteuropäischer Herkunft.

3. Informationen zum Projekt, Beteiligten, Laufzeit etc.: <https://www.adwmainz.de/projekte/buber-korrespondenzen-digital> (29.11.2024); Repositorium: <https://gitlab.rlp.net/adwmainz/digicademy/bkd/correspondences> (29.11.2024).
4. <https://openrefine.org/> (29.11.2024).
5. <https://www.knime.com/> (29.11.2024); siehe auch: Berthold, Michael R., u. a. „KNIME – the Konstanz Information Miner: Version 2.0 and Beyond“. *SIGKDD Explor. Newsl.*, Bd. 11, Nr. 1, November 2009, S. 26–31, <https://doi.org/10.1145/1656274.1656280> (29.11.2024).
6. <https://entities.pages.gwdg.de/entityxml> (29.11.2024).
7. Organisation gemeinsam mit dem *FID Jüdische Studien*, Beitrag zum Normdaten-Workflow in BKD.
8. Beiträge von Heike Breitenbach & Denise Jurst-Görlach: *„Unsere Gedanken kreisen immer um die Lage bei Euch“. Frauen und ihre Briefe im Umfeld von Buber-Korrespondenzen Digital*; Thomas Kollatz: *Kodierungsvorschlag zur projektübergreifenden Auszeichnung von Frauenbriefen*. Weitere Informationen zur Tagung: <https://www.digitalhumanities.tu-darmstadt.de/workshop-frauenbriefe> (29.11.2024).

## Bibliographie

- Berthold, Michael R., u. a.** „KNIME – the Konstanz Information Miner: Version 2.0 and Beyond“. *SIGKDD Explor. Newsl.*, Bd. 11, Nr. 1, November 2009, S. 26–31, <https://doi.org/10.1145/1656274.1656280>.
- Busch, Nathanael, und Diana Müller.** „Normdaten in den Geisteswissenschaften. Frequently Asked Questions“. *Zeitschrift für Literaturwissenschaft und Linguistik*, Bd. 53, Nr. 3, September 2023, S. 781–796. <https://doi.org/10.1007/s41244-023-00295-1>.
- Juen, Sara.** „Feminismus, Algorithmen, Gender-Data-Gap und was das alles mit Bibliotheks- und Informationswissenschaft zu tun hat“. *Library Ideas*, 39, 2021. <https://doi.org/10.18452/23448>.
- Kamzelak, Roland S.** „Digitale Editionen im semantic web. Chancen und Grenzen von Normdaten, FRBR und RDF“, in: Richts, Christina und Stadler, Peter (Hrsg.), *„Ei, dem alten Herrn zoll’ ich Achtung gern“*. *Festschrift für Joachim Veit zum 60. Geburtstag*. München 2016, S. 423–435.
- Kett, Jürgen, u. a.** „Text+ und die GND – Community-Hub und Wissensgraph“. *Zeitschrift für Bibliothekswesen und Bibliographie*, Bd. 69, Nr. 1–2, April 2022, S. 37–47. <https://doi.org/10.3196/1864295020691262>.
- Stadler, Peter.** „Normdateien in der Edition“. *Editio*, Bd. 26, Nr. 1, Dezember 2012, S. 174–183. <https://doi.org/10.1515/editio-2012-0013>.
- Trzeciak, Debby.** *Normdaten und Liebesbriefe. Möglichkeiten und Grenzen der Auszeichnung am Beispiel einer Korrespondenz aus dem Liebesbriefarchiv*. Hrsg. von Sabine Bartsch u. a., Bd. 5, Universitäts- und Landesbibliothek Darmstadt, 2023. <https://doi.org/10.26083/tuprints-00024477>.