

Zuweisung von Katalogdatensätzen an Personennormdatensätze mittels Wahrscheinlichkeiten

Frank Förster

Im Juni 2020 werden die Tn-Sätze in der Gemeinsamen Normdatei (GND) gelöscht. Die Tp-Sätze für eindeutig identifizierbare Personen bleiben im Bereich des Personenbestandes übrig. Dieser Beitrag soll eine Anreicherung und Bereinigung der Personennamensdatensätze mittels Wahrscheinlichkeiten auf der Datenbasis von GND und k10plus anregen. Zu jedem Tp-Satz kann ein Profil aus verknüpften Informationen erstellt werden: z. B. über Stichwörter, fachliche Schwerpunkte, Ko-Autoren, Zeiten und Orte usw. Im gleichen Maß können abgrenzbare Profile für Tn-Sätze per Algorithmus erkannt werden. Zusätzlich könnten bestehende Verknüpfungen von Personen- in Titeldatensätzen Fehlzusweisungen aufspüren. Die Folgen eines solchen Verfahrens wären eine retrospektive Anreicherung des Altbestandes und eine präzisere Ausgestaltung des Katalogs.

Einleitung

Die Gemeinsame Normdatei (GND) und der k10plus beinhalten mehrere Millionen Datensätze. Das ist bekannt. Es gibt in GND und k10plus mehrere Millionen Personendatensätze. Das ist ebenfalls bekannt. Auch hinreichend bekannt sollte sein, dass es Personennormdatensätze (Tp-Sätze) zur Verzeichnung individualisierter Personen und Namensdatensätze (Tn) zur Verzeichnung von Personennamen gibt.¹ Möglicherweise übersehen wurde hingegen die Mitteilung der Deutschen Nationalbibliothek (DNB) vom 23. August 2019, dass die Tn-Sätze im Juni 2020 von der GND ausgeschlossen werden sollen.² In der Katalogisierungspraxis werden Tn-Sätze als „Sammeltopfe“ für Personennamen genutzt und haben

als zusätzliche Merkmale vor allem Verweisungsformen.³ Dabei enthalten einige Tn-Sätze durchaus eine Art semi-individualisierte Kennzeichnung eines bestimmten Verfassers, was im Sinne der Erfassungsrichtlinien für den Status als brauchbarer Tp-Satz angesehen werden könnte, wenn auch mit wenigen Informationen. Die Erfassungshilfen legen die Hürden hierfür niedrig genug, die Hemmschwelle der Bibliotheksbeschäftigten scheint in Teilen dennoch zu hoch. Der Wunsch der GND-Kooperative ist es, dass bei wenigen individualisierenden Merkmalen der Schritt zum Anlegen eines Tp-Satzes gewagt wird. Dies sollte mit einem verstärkten Einsatz von Algorithmen unterstützt, analysiert und nachverfolgt werden, um Kandidaten für Personennormdatensätze zu ermitteln. Doch mit welchen Verfahren könnte das konkret gelingen?

Der vorliegende Beitrag soll hierfür ein paar Ideen geben. Er beschreibt Vorschläge für a) eine maschinelle Ersetzung der Tn-Sätze⁴ durch Verlinkungen zu einem vorhandenen Tp-Satz, b) die Ausdifferenzierung einzelner Tn-Sätze in mehrere Tn-Sätze, c) die automatische Erzeugung von neuen Tp-Sätzen über den Umweg angereicherter Tn-Sätze und d) die Anreicherung bestehender Tp-Sätze. Da sich Beispiele hierfür leicht finden lassen, wird die Erfolgsquote nicht unerheblich sein. In vielen Datensätzen in der GND steht bereits zu lesen: „Zeile XXX maschinell ergänzt“, manchmal sogar „maschinell aus ähnlichen Titelaufnahmen ergänzt“. Wird der vorliegende Datensatz per Autopsie mit einem vorliegenden Werk abgeglichen, sollten maschinell erzeugte Informationen überprüft werden.

¹ Vgl. EH-P-01, EH-P-16 und EH-P-17.

² „Ein Namensdatensatz (Tn) kann bisher für verschiedene Personen mit demselben Namen gelten. Diese Datensätze wurden bisher immer dann angelegt, wenn die individualisierenden Merkmale für das Anlegen eines Personennormdatensatzes (Tp) nicht ausreichend vorhanden bzw. bekannt waren. Die Namensdatensätze entsprechen aber nicht den heutigen Anforderungen an Normdaten, die eindeutig und zuverlässig Entitäten identifizieren und beschreiben.“ (DNB-Datendienste 2019, S. 1) Die Datensätze werden *logisch gelöscht* (Telefonat des Verfassers mit Jürgen Kett, DNB). Bereits ab 1.9.2019 werden keine Tn-Sätze mehr in der GND angelegt (vgl. DNB-Datendienste 2019, S. 1). Nach Recherchen beträgt die Zahl der Tn-Sätze in der GND ca. 7,2 Millionen gegenüber 5,5 Millionen Tp-Sätzen. Im k10plus finden sich „nur“ ca. 746.000 Tn-Sätze mit Normdatentyp Tnx, d. h. ohne Katalogisierungslevel, gegenüber 5,3 Millionen Tp-Sätzen. Das bedeutet, im k10plus hat die großflächige Löschung bereits stattgefunden. (Stand: 23.4.2020)

³ Vgl. EH-P-16 bzw. Katalogisierungshandbuch Normdaten 2018, S. 17f.

⁴ Im Folgenden wird der Terminus „Tn-Satz“ sowohl für Tn-Sätze im bisherigen Verständnis als auch für jene Textstrings in den 30xx-Feldern der Titeldatensätze, die dort den Namen einer Person repräsentieren, verwendet.

Profiling

Welche Elemente in Personen- und Titeldatensätzen sind mit einem Verfasser⁵ potentiell verknüpft? Im Bereich der Katalogisierung sind dies Name, Verweisformen, Lebensdaten, Beruf, Titel, Titelwörter, Ko-Autoren, Wirkungszeiten, Körperschaften, Zeitschriften, Länder, Sprachen und sicher noch ein paar mehr. Im Bereich der Sacherschließung sind dies Fachgebiete, Schlagwörter, Schlagwort-Ketten oder Schlagwort-Kombinationen, Klassifikationen und andere Formen der inhaltlichen Beschreibung.

Anhand der eingesammelten Informationen kann für den einzelnen Verfasser ein Profil erstellt werden. In Summe erscheinen die Veröffentlichungen in einem abgrenzbaren Zeitraum, in einem bestimmten fachlichen Spektrum, in einer gewissen Spannweite an Veröffentlichungsformen und Publikationsformaten, mit einer eingrenzenden Menge an Ko-Autoren und in einem vorgegebenen Sprachraum. Mit Hilfe der durch Katalogisierung und Sacherschließung erhobenen Merkmale und Informationen kann ein Profil erstellt werden. Im Fachjargon heißt dies Profiling, also „die nutzbare Erstellung des Gesamtbildes einer Persönlichkeit für bestimmte Zwecke“⁶, im vorliegenden Fall für die Ermittlung von Autorenschaften oder anderer Beteiligungen an Veröffentlichungen.⁷

Klar ist, dass die Sacherschließung eine erhebliche Bedeutung hat und keinesfalls unterschätzt werden darf! Durch die wissenschaftliche Durchdringung eines Werks mit Ziel auf ein standardisiertes Exzerpt in Form einer Klassifikation oder verbalen Sacherschließung werden Beziehungen zwischen Titel- und Normdaten belastbarer. Deshalb wird eindringlich für eine im gemeinsamen Konsens ausgearbeitete Form der Sacherschließung für die jeweilige Einrichtung und im Weiteren für eine kooperative Zusammenarbeit und die Erarbeitung von Konkordanzen plädiert.⁸ Über formale Merkmale (Titel, Verlag, Zeitschrift usw.) werden Themen ebenfalls sichtbar, falls es an Informationen aus dem Bereich der Sacherschließung mangelt.⁹

Die Profilbildung eines Personendatensatzes muss in einem iterativen Verfahren geschehen und findet seinen Ausgangspunkt in einem nach Methoden von

Vorkommenswahrscheinlichkeiten gewählten Namen. Alle Tp- und Tn-Sätze mit dem gewählten Namen (und seinen Varianten) als Hauptansetzung werden vom Algorithmus erfasst und durchlaufen das Profiling. Folgende Prozesse können hierbei parallel für jeden ausgewählten Tp- bzw. Tn-Satz ablaufen:

- Bildung von 1 bis n Profilen aus einem Tn-Satz anhand der gesammelten Informationen plus automatisches Erzeugen von 1 bis n+1 Tn-Sätzen als Hilfskonstrukt zur Profilbildung durch Anreicherung um Merkmale gemäß EH-P-16,
- Bildung eines Profils zu jedem Tp-Satz anhand der zusammengetragenen Informationen plus Kennzeichnung von Titeldatensatz-Ausreißern auf der Basis von Wahrscheinlichkeiten,
- Abgleich der Profile der vorhandenen bzw. erzeugten Tn-Sätze mit den Profilen der vorhandenen Tp-Sätze mit dem Ziel der a) Zuweisung bzw. Ersetzung der Verlinkung der Tn-Sätze durch die Verlinkung der Tp-Sätze in Titeldatensätzen oder b) Bildung eines neuen Tp-Satzes mit entsprechender Verlinkung in den betreffenden Titeldatensätzen,
- Ausschärfung des Profils eines vorhandenen Tp-Satzes durch a) Anreicherung von standardisierten Informationen im Datensatz, b) Hinzufügung der Verlinkung zu vorhandenen Titeldatensätzen oder c) Ersetzung einer Verlinkung des Tp-Satzes im Titeldatensatz durch einen Tn-Satz, wenn die Wahrscheinlichkeit für eine Verbindung von Verfasser und Werk zu gering erscheint.

Bildlich gesprochen ist ein Profil das Gesamtbild aller Informationen aus allen verlinkten Daten und der Personendatensatz das standardisierte Kondensat davon. Jede Veränderung im Verlaufe des Prozesses schärft oder „verunschärft“ das jeweilige Profil. Ein von einem klaren Profil gestärkter Personendatensatz steigert dessen Identifizierbarkeit gegenüber parallel erarbeiteten Systemen, wie zum Beispiel Wikipedia oder ORCID¹⁰. Die aus der GND in andere Systeme erfolgende Verlinkung steigert die Genauigkeit und offenbare Zuordnung zu einer einzelnen Person noch weiter. Schlussendlich werden die Nut-

5 Der Vereinfachung halber wird die männliche Form gewählt, es ist jedoch immer die weibliche Form mit gemeint. Unter Verfasser werden hier weitere Personenfunktionen in Bezug auf Publikationen subsummiert: Herausgeber, Bearbeiter usw.

6 Vgl. Seite „Profiling“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 29. Juli 2019. URL: <https://de.wikipedia.org/w/index.php?title=Profiling&oldid=190863590> (Abgerufen: 20. April 2020)

7 Wird der Ansatz weitergedacht, lässt sich ein Profil für nahezu jedes Element und Elementkombinationen schaffen: das Profil einer Zeitschrift, das Profil eines Jahres oder eines Jahrzehnts, das Profil einer Körperschaft oder das Profil einer Konferenz in Hannover.

8 Hier sei auf die Initiative coli-conc („Verwaltung und Bereitstellung von Konkordanzen zwischen bibliothekarischen Wissensorganisationssystemen“) des GBV verwiesen, siehe <https://coli-conc.gbv.de/>

9 Maschinelle sacherschließende Maßnahmen werden vielerorts entwickelt und könnten hier zweckdienlich zusätzlich eingesetzt werden, um Personendatensätze um fachliche Aspekte anzureichern.

10 Das Projekt ORCID DE, siehe <https://www.orcid-de.org/>, ist eines der derzeitigen GND-Wiki-Projekte, das sich der Vernetzung der ORCID mit anderen Informationsinfrastrukturen verschrieben hat.

zer der bibliographischen Nachweissysteme dankbar sein, wenn die den individualisierten Personendaten-sätzen zugeordnete Literatur vollständiger ist und biographische und bibliographische Informationen nicht über mühsame Suchen zusammengetragen werden müssen.

Auswahl

Die Bibliothek des Geozentrums Hannover (GZH)¹¹ ist eine der größten Spezialbibliotheken für die Geowissenschaften weltweit. Es gibt einen klar umgrenzten Bestandsaufbau. Als Bibliothekssystem wird seit 1990 aDIS bzw. seit 2006 aDIS/BMS verwendet. Der Bestand umfasst über 600.000 selbständige Werke (vor allem Monographien, Sammelwerke, Zeitschriftenhefte, Zeitschriftenbände, Fachberichte und Karten) und über 700.000 unselbständige Werke (vor allem Aufsätze). Seit kurzem ist die Bibliothek Teilnehmerin am Südwestdeutschen Bibliotheksverbund (SWB). Mit dem Fremddatenaustausch wird in nächster Zeit begonnen.

Ungefähr 755.000 Personendatensätze gibt es im System. Es gibt die Möglichkeit, diese als individualisiert zu kennzeichnen. Ebenso können Verweisungsformen angegeben werden. Auch andere, aus der GND bekannte Möglichkeiten der individualisierenden Kennzeichnung sind vorhanden oder können nachgeführt werden. Allerdings wurde dies über die Jahre vor allem für Personen genutzt, die entweder wegen Namensgleichheit unterschieden werden mussten oder die als Wissenschaftler im Bereich der Geo- und Naturwissenschaften einen erheblichen Bekanntheitsgrad haben. Weitere Informationen wurden eher in einem Feld für interne Bemerkungen untergebracht (in knapp 1.600 Datensätzen), sodass das vorgefertigte Gerüst für detaillierte Informationen weitestgehend leer geblieben ist.

Eine Statistik zu individualisierenden Merkmalen im GZH-Katalog (siehe Tabelle 1) zeigt, dass die Personendatenbank zu über 99 % aus Tn-Sätzen besteht. An dieser Stelle wird das Anliegen konkret, die Personendaten erstens zu individualisieren und zweitens unter Verknüpfung mit der GND die Tp-Sätze wechselseitig anzureichern. Das ermöglicht es der GZH-Bibliothek, die fachliche Nische der Geowissenschaften innerhalb der GND zu pflegen und auszubauen.

Ausgangspunkt des Algorithmus ist ein geeigneter Personennamen, dessen Auswahl nach ansteigender

Merkmal	Anzahl ¹²
Insgesamt	Ca. 755.000 Datensätze
Geburtsjahr	In ca. 2.400 Datensätzen
Sterbejahr	In ca. 1.700 Datensätzen
Ort (Geburt, Wirkung)	In ca. 100 Datensätzen
Land	90 Einträge in 88 Datensätzen
Sprache	52 Einträge in 38 Datensätzen
Beruf	10 Einträge in 6 Datensätzen
GND-Nummer	In 6 Datensätzen
Verweisungsform	Ca. 64.000 Einträge in 40.000 Datensätzen
Interne Bemerkung	In 1.585 Datensätzen

Tabelle 1

Wahrscheinlichkeit von selten nach viel hinsichtlich folgender Parameter erfolgt:

1. Anzahl der Vornamen,
2. Doppelnamen,
3. Länge des Namens,
4. Adelstitel und Präfixe,
5. Vorkommenswahrscheinlichkeit von Vornamen und Nachnamen,
6. Vorkommenswahrscheinlichkeit einzelner Buchstaben.

Hinsichtlich Parameter 1 zeigt der GZH-Katalog folgendes Muster: Einen abgekürzten Vornamen gibt es in ca. 396.000 Datensätzen, zwei in ca. 142.000, drei in ca. 14.000, vier in ca. 500, fünf in 34 und sechs abgekürzte Vornamen in drei Datensätzen. Entsprechend sollte hier mit den Datensätzen mit mehreren Vornamen begonnen werden.

Hinsichtlich Parameter 2 sollte mit Parameter 3 kombiniert werden, sodass längere Doppelnamen zuerst dran wären. Im Personenbestand gibt es etwa 20.000 Doppelnamen als Nachnamen.

Hinsichtlich Parameter 3 kann zwischen Länge des gesamten Namens inkl. Nachname und Vorname(n) und der Länge von Vor- oder Nachname unterschieden werden. Mit den längsten Namen sollte begonnen werden. Im GZH-Katalog folgt die Verteilung der Länge der Nachnamen der Gaußverteilung, wobei fünf- bis achtbuchstabige Namen den Großteil von etwa 60% ausmachen.

¹¹ Im GZH sind die Bundesanstalt für Geowissenschaften und Rohstoffe, das Landesamt für Bergbau, Energie und Geologie und das Leibniz-Institut für Angewandte Geophysik untergebracht. Die Bibliothek stellt die Literaturversorgung im Rahmen einer Vereinbarung für alle drei Einrichtungen sicher. Darüber hinaus ist sie für die Öffentlichkeit zugänglich.

¹² Weil es sowohl multiple als auch singuläre Felder gibt, ist für singuläre Felder die Anzahl der Datensätze, für multiple Felder die Anzahl an verschiedenen Einträgen plus die Anzahl in wie vielen Datensätzen genannt. (Stand: April 2020)

ZAMBELLI PAZIO

Room for books and readers.



GEÖFFNET = SCHALLGESCHÜTZTER LESEPLATZ
KOMPAKTMODUS = FLEXIBLE RAUMNUTZUNG

Zambelli Pazio ist die innovative Antwort auf die gestiegenen Anforderungen in der modernen Bibliothek. Mit Zambelli Pazio schaffen Sie Platz für ungestörtes Verweilen ohne Kapazitätsverlust. Sie erhalten neue Möglichkeiten bei der Raumplanung und können konzeptionelle Aufgaben völlig neu interpretieren. Das ist innovativ und einzigartig. Das ist Zambelli Pazio.

www.zambelli.com

zambelli

EINFACH INTELLIGENT AUFBEWAHRT

MADE IN GERMANY

Hinsichtlich Parameter 4 wird eine kleinere Gruppe im Bestand identifiziert. Im GZH-Katalog ist „de“ etwa 3.000 Mal als Namensbestandteil belegt, „van“ etwa 2.000 Mal, „von“ etwa 1.200 Mal und „van der“ etwa 600 Mal, um nur einige Beispiele zu nennen. Bei der Suche nach den Personendatensätzen sollte mit dem „sinntragenden“ Wort des Namens, also ohne Präfix, aufgrund seiner höheren Güte gesucht werden.

Bei Parameter 5 und 6 wird es vorkommen, dass mehrere Tp-Sätze existieren, weshalb die Näherung über die Vorkommenswahrscheinlichkeit der jeweiligen Vor- und Nachnamen bzw. einzelner Buchstaben gehen sollte. Das ist zu einem erheblichen Teil sprachabhängig, d. h. im deutschen Sprachraum sieht die Hitliste der Vor- bzw. Nachnamen anders aus als z. B. im englischen. Um eine übergreifende Hitliste zu erstellen, sollte der Katalog als Grundlage dienen. Über eine Abfrage lässt sich die Hitliste der Nachnamen ermitteln.¹³ Für die Vornamen geht das auch¹⁴, ist nur etwas komplizierter aufgrund der Syntax des Datenbankfeldes, der möglichen Namenshäufung bzw. der eventuellen Abkürzungspraxis. Die Wahrscheinlichkeit eines Namens erlaubt bereits eine hinreichende Eingrenzbarkeit in Kombination mit anderen Parametern: mehrere Vornamen und Doppel-Vornamen vor einzelnen Namen, lange vor kurzen Namen, Adelstitel und Präfixe in Namen vor solchen ohne, seltene Namen und Buchstaben¹⁵ vor häufigen. Bildsprachlich muss das Mesh (die Maschenweite bzw. Siebgröße) angepasst werden, damit die klaren Treffer ausgesiebt werden und die falschen hindurch fallen. Das ist ein wesentlicher Punkt, weswegen mit Unwahrscheinlichkeiten begonnen werden sollte, bevor sich der Algorithmus zu den Wahrscheinlichkeiten durchrechnet. Schlussendlich dienen Parameter 5 und 6 auch für die Suche in der „großen Masse“, also Namen aus den ermittelten Hitlisten. Hinzu kommt die Möglichkeit von Tipp-

fehlern, die einen anderen Namen vorgaukeln (z. B. Neumann statt Naumann).

Nach der Auswahl zu analysierender Personendatensätze erfolgen die Profilbildung und die Zuordnung der Identifier in den Titeldatensätzen wie beschrieben. Die praktische Umsetzung wird im folgenden Abschnitt anhand GND und k10plus erklärt.

Analyse

Für das Profil eines Personendatensatzes sind folgende Kategorien denkbar:

- Person (Hauptansetzungen, Verweisungsformen, Lebensdaten, Geschlechter),
- Netzwerk (Ko-Autoren, Zeitschriften, Körperschaften),
- Sprachraum (Sprachen, Länder) und
- Thema (Sacherschließungen, Titelstichwörter).

Die hier gewonnenen Merkmale werden als Informationen ersten Grades eingestuft, wenn sie aus dem Datensatz selbst und den direkt damit verknüpften Datensätzen stammen.

Über die Merkmale des Netzwerkes wird die statistische Erhebung um weitere Merkmale aus den Bereichen Sprachraum und fachliches Spektrum angereichert: Zeitschriften erscheinen in bestimmten Sprachräumen und haben fachliche Schwerpunkte; Körperschaften sind raumzeitlich verankert und thematisch umgrenzt; Ko-Autoren haben ein eigenes Profil. Diese Merkmale werden als Informationen zweiten Grades behandelt, da sie aus indirekt verknüpften Datensätzen stammen, d. h., aus Datensätzen, die mit den direkt verknüpften Datensätzen darüber hinaus verknüpft sind, und werden mit reduzierter Merkmalsbedeutung hinterlegt.

Die Kategorien können in einer Mischung aus Schlagwortwolke und Vergleichsdiagramm befüllt und in Form eines „Relevance Rankings“ visualisiert wer-

13 Im GZH-Katalog ist die Top 50 der etwa 240.000 verschiedenen belegten Nachnamen folgende: Wang, Li, Zhang, Chen, Liu, Smith, Lee, Kim, Yang, Jones, Johnson, Wu, Brown, Miller, Williams, Anderson, Singh, Huang, Schmidt, Wilson, Martin, Xu, Zhao, Thomas, Müller, Lin, Taylor, Silva, Zhou, Meyer, Schneider, Moore, Sun, Davis, White, Rao, Yu, Clark, Thompson, Park, Zhu, Weber, Fischer, Hansen, Evans, Young, Chang, Lu, Wagner, Hu. Deutlich wird, dass hier Namen aus dem chinesischen und englischen Sprachraum dominieren. Sucht man nach Nachnamen aus dem deutschen Sprachraum, findet man Schmidt (Rang 19), Müller (25), Meyer (30), Schneider (31), Weber (42), Fischer (43), Wagner (49). Diese Nachnamen bilden in etwas anderer Reihenfolge ebenfalls die Top 7 der häufigsten Nachnamen in Deutschland, vgl. Seite „Liste der häufigsten Familiennamen in Deutschland“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 27. März 2020. URL: https://de.wikipedia.org/w/index.php?title=Liste_der_häufigsten_Familiennamen_in_Deutschland&oldid=198159959 (Abgerufen: 23. April 2020). Das beweist, dass selbst in kleinerem Maßstab die Wahrscheinlichkeiten größerer Stichproben schon sichtbar zutreffen.

14 Die ermittelte Top 50 der etwa 118.000 verschiedenen belegten Vornamen im Personenbestand ist nur eine Annäherung, da alle abgekürzten Vornamen aus dieser Aufstellung herausgenommen wurden und die Hausregeln vor etlichen Jahren vorsahen, keine ausgeschriebenen Vornamen aufzunehmen, weswegen eine ganze Reihe von abgekürzten Vornamen im Bestand ist, die zahlenmäßig die ausgeschriebenen Namen dominieren. Die 21 häufigsten „Vornamen“ sind eigentlich auf einen Buchstaben abgekürzte Vornamen: M., A. und J. mit jeweils über 10.000 Treffern. Die Top 100 der ausgeschriebenen Vornamen ist demnach: Peter (ca. 2.700 Treffer), Michael (ca. 2.600), Thomas (ca. 2.200), David, Martin, John, Christian, Wolfgang, Robert, Andreas, Klaus, Paul, Daniel, Stefan, Richard, Jürgen, Frank, Jan, Hans, Gerhard, Ulrich, Alexander, Werner, Dieter, Andrea, Matthias, Manfred, Walter, Mark, Bernd, Helmut, Pierre, Joachim, James, Christoph, Maria, Anna, Markus, Rainer, Barbara, Michel, Andrew, Jörg, Patrick, Marco, Karl, Johannes, Antonio, Günter, Rolf. Es fällt auf, dass in den Top 100 männliche Vornamen im Verhältnis 9:1 gegenüber weiblichen Vornamen dominieren.

15 Die Reihenfolge einzelner Buchstaben (ohne Sonder-Buchstaben) in der Hauptansetzung der Personendatei im GZH-Katalog ist von viel nach wenig: a, e, n, r, i, o, s, l, h, t, m, u, d, c, g, k, b, j, p, y, v, w, f, z, x, q. Die Sonderzeichen-Buchstaben (Umlaute, Ligaturen usw.) sind eher selten und daher in einer anfänglichen Auswahlrunde zu beachten.

den. Der Personendatensatz ist das standardisierte Kondensat dieser Informationen. Häufigkeiten sind ebenso wichtig wie Seltenheiten, um Schärfe und Unschärfe zu erkennen. Sobald eine Verlinkung zwischen Personendatensatz und Titeldatensatz getilgt wird, verändert sich die Zusammenstellung des Profils und eventuell der Personendatensatz. Wird ein statistischer „Ausreißer“ entfernt, wird sich deut-

lich zeigen, dass das Profil geschärft wird. Deshalb sei der Verweis auf die Wichtigkeit eines iterativen Verfahrens hier noch einmal betont.¹⁶ Tabelle 2 zeigt eine Auswahl von Feldern, die als Informationsquellen für das Profiling dienen können. Eine komplette Aufstellung über alle Felder, die in GND und k10plus zur Verfügung stehen, würde in diesem Rahmen zu weit führen.

Information	Kategorien	GND (Feld, Unterfeld)	k10plus (Feld, Unterfeld)
Individualisierungsmerkmale der Gruppe 1			
Verfasser	Person	100 [Personenname, bevorzugt] 400 [Personenname, abweichend] 700 [Personenname, anderer Datenbestand]	3000 3010 3050
Beruf	Thema	550 + \$4berc [charakteristischer Beruf] 550 + \$4beru [sonstiger Beruf]	
Individualisierungsmerkmale der Gruppe 2			
Erscheinungsjahre	Person	548 + \$4datw [Wirkungsjahre ¹⁷] 548 + \$4datz [exakte Wirkungsjahre]	4030 \$d 4030 \$h
Ort	Person	551 + \$4ortg [Geburtsort] 551 + \$4orts [Sterbeort] 551 + \$4ortw [Wirkungsort] 551 + \$4ortx [Exilort]	4030 \$p [Erscheinungsort]
Adelstitel	Person	550 + \$4adel	
Sprachen	Sprachraum	337	1500
Verlag	Netzwerk		4030 \$n
Affiliation	Netzwerk	510 + \$4affi 511 + \$4affi	
Fachgebiete und Schlagwörter	Thema	065 [GND-Systematik] 550 + \$4them	5XXX
Studienfach	Thema	550 + \$4stud	
keine Bedeutung als Individualisierungsmerkmal			
Titelstichwörter	Thema		4000 4002 4010 4011
Ko-Autoren	Netzwerk	500 + \$4bezb [berufliche Beziehung]	3010 Person
Länder	Sprachraum	043 [Ländercode]	1700
Zeitschriften	Netzwerk		4241
Serien	Netzwerk		417X 418X
Körperschaften	Netzwerk		3100 3110 3150
Konferenzen	Netzwerk		3160
Bemerkungen	alle	667 [redaktionelle Bemerkungen] 678 [weitere biographische Angaben] 909 [Kommentar]	

Tabelle 2

¹⁶ Im geowissenschaftlichen Verständnis würde man wohl von „Kreislaufwirtschaft“ sprechen.

¹⁷ Hier wäre zulässig, die Wirkungszeit mit höherer Wahrscheinlichkeit pauschal 20 Jahre nach Geburtsjahr beginnen zu lassen und pauschal zwei Jahre nach dem Todesjahr enden zu lassen. Die Zeit, die darüber hinaus geht, würde mit absteigender Wahrscheinlichkeit in die Statistik mit einfließen.

Schwierigkeiten werden sich dort zeigen, wo das zu erstellende Profil aufgrund mangelnder Informationen in den Personen- und Titeldatensätzen offenbleibt. Eine Abfrage über die Verteilung und Befüllung einzelner Felder in GND und k10plus könnte Abhilfe schaffen. Welche Datenfelder „sachdienliche“ Hinweise wofür liefern, muss im Algorithmus exakt ausformuliert werden. Strukturdiagramme können hierbei hilfreich sein. Mit einer ausgeführten Sacherschließung besteht auf jeden Fall ein enormer Vorteil. Zu klären wäre die Frage, welchen Einfluss Satzstatus und Katalogisierungslevel auf den Algorithmus haben sollen.

Ungenauigkeiten und unscharfe Suche

Genauigkeit ist Zielvorgabe in GND und k10plus, aber nicht immer Realität. Anhand weniger Beispiele soll gezeigt werden, warum eine unscharfe Suche (probabilistisches bzw. Fuzzy-Retrieval) bei der Verlinkung von Tp-Sätzen zu zugehörigen Titeldatensätzen sinnvoll erscheint. Nebenbei bemerkt, Kommentar und Mailboxfelder ermöglichen stets den Hinweis auf zu überprüfende Sachverhalte.

Veraltete Quelle: Ein Beispiel einer nicht mehr aktuellen Quelle ist der Normdatensatz des Verfassers (GND-PPN 138068224). Die Quelle – eine Internetseite der Universität Kiel – ist nur noch im Internetarchiv auffindbar. Sie ist mehr als fünf Jahre alt, dennoch stehen aktuellere Informationen im Tp-Satz, was nicht ganz schlüssig wirkt.

Tippfehler: Tippfehler können dazu führen, dass ein nicht ausreichend programmierter Algorithmus Zusammenhänge nicht erkennt. Beispiel: Der Tn-Satz von Marie-Luise Favreau (GND-PPN 108368505) ist mit einer Hochschulschrift verlinkt, die im normalen Wissenschaftlerleben eine der ersten Publikationen ist und mit einer etwas höheren Wahrscheinlichkeit auch vor einer Namensänderung durch Eheschließung veröffentlicht wird. Der Tp-Satz von Marie-Luise Favreau-Lilie (GND-PPN 110057465) führt als früheren Namen „Favreau, Marie-Luise“ (sic!) auf. Es handelt sich um dieselbe Person.¹⁸

Transliteration¹⁹: Die Transliteration nichtlateinischer Schriftzeichen in Buchstaben der lateinischen Schrift ist je nach Sprachraum unterschiedlich und kann zu Namensvarianten mit unbeabsichtigten Fehlern führen. Das wird im GZH-Katalog deutlich, wenn Datensätze aus fremdsprachigen Quellen eingespielt wurden. So gibt es Titeldatensätze von einem ame-

rikanischen Partner, der das slawische, kyrillische Alphabet in lateinischen Buchstaben auf Grundlage englischer Vorgaben umschrieben hat, und es gab einen französischen Partner, der nach französischen Vorgaben kyrillische in lateinische Buchstaben transliteriert hat. Hilfreich ist, wenn die entsprechenden transliterierten Formen als Verweisungsformen in den 4XX-Feldern abgelegt werden.

Fachliche Nähe: Dieses Verfahren produziert Fehler, wenn Personendatensätze vom Namen her gleich und fachlich-thematisch zu eng sind und der Algorithmus nicht ausgeschärft genug ist. Daher ist es wesentlich, dass der Algorithmus einen bezüglich der Güte interpretierbaren Wert liefert. Selbstverständlich bleibt dabei das Risiko, dass dieser Wert in die Irre führt. Die Konzentration muss auf den Tp-Sätzen liegen und sollte sich bei Bedarf auch externer Web-Quellen bedienen. Mit Sicherheit sind zudem Recherchen in gedruckten Nachschlagewerken notwendig, oft wird erst eine Autopsie Klarheit erbringen. Digitale Quellen reichen nicht tief genug in die Vergangenheit. So gibt es im GZH-Katalog einen Helmut und einen Hans Materna, beide Denkmalpfleger in Halle/Saale in derselben Wirkungszeit. Weder GND noch k10plus verzeichnen Veröffentlichungen dieses Verfassers bzw. den Verfasser selbst. Die Frage stellt sich, ob es die gleiche Person ist, die mit verschiedenen Vornamen publiziert hat, oder ob es zwei verschiedene Personen sind. Hilfreiche Quelle ist hierbei auch nicht Wikipedia, sondern erst die Autopsie der Werke, Nachrufe, Biographien oder Bibliographien erbringen Klarheit.

Namensdubletten: Bei Personennamen aus Kombination häufig vorkommender Vor- und Nachnamen wird es auf die Genauigkeit des Algorithmus mit Ziel auf eine klare Abgrenzung verschiedener Personen mit dem gleichen Namen ankommen. Im GZH-Katalog gibt es lediglich einen Personendatensatz für den Namen Christian Müller, keiner dieses Namens liegt individualisiert vor. Eine erste Überprüfung der mit dem Personendatensatz verlinkten Titel ergibt fünf Hochschulschriften: Eine Dissertation 1984 im Fachbereich Hydrogeologie, eine Dissertation 2000 in Geodynamik, eine Dissertation 2001 in Meeresgeologie (arbeitet aktuell im GZH), eine Dissertation 2009 in Sedimentologie und eine Masterarbeit 2014 in Paläontologie. Allein hieran wird deutlich, dass mindestens fünf verschiedene Christian Müllers im Bestand in einem Datensatz zusammengefasst sind, im ersten Schritt also sechs²⁰ „Hilfs-Tn-Sätze“ als

¹⁸ Der Verfasser kennt sie persönlich.

¹⁹ Siehe auch EH-P-14.

²⁰ Der sechste Tn-Satz ist „Sammeltopf“ für alle weiteren Christian Müllers.

Programmieren mit
dem Filzstift:
Mini-Roboter Ozobot



Klassensatz
Ozobot BIT



Klassensatz BeeBot Plus



Ihr Makerspace aus der Kiste

Unsere MakerBoxen bieten einen spielerischen Einstieg in die Welt des Programmierens. Jetzt kommen sie als starke Helfer aus der Bibliothek in die Schulen: Wir modifizieren unsere beliebten Roboter-MakerBoxen, sodass sie als Arbeitsmittel in Gruppen eingesetzt werden können.

Mehr erfahren und bestellen unter bit.ly/makerboxen



Wir machen Bibliothek einfach – und beraten Sie gerne.
Service und Vertrieb Bibliotheksausstattung • Telefon 07121 144-420
Bibliotheksausstattung@ekz.de • www.ekz.de

ekz
bibliotheks
service

Ausgangspunkt für den Algorithmus erzeugt werden müssten.

Beispiel: Frank Förster

Als genaueres Beispiel bietet der Personennamen Frank Förster eine hinreichend überschaubare und auch differenzierbare Gruppe.²¹ Die 16 Personendatensätze teilen sich auf 14 Tp-Sätze und 2 Tn-Sätze auf. Von den Tp-Sätzen haben zwei das Katalogisierungslevel 1, elf Datensätze das Level 3 und einer das Level 6 und bei dreizehn Datensätzen sind Berufe angegeben (siehe Abbildung 1).

den. Auch Falschzuordnungen in Tp-Sätzen sind vorhanden.

Formalkatalogisierung: Der Abgleich des Hauptsachtitels erbringt bereits Übereinstimmungen. Die Textstrings der Titel werden als Ganzes (oder als Hashwert) genommen und miteinander abgeglichen. **Beispiel:** Beim Buch „Tafeltrauben für den Hausgarten“ (3. Auflage) ist Frank Förster in einem Fall als Tp-Satz (GND-PPN 1191128032) verknüpft, aber in drei anderen Fällen unspezifisch über einen Tn-Satz (GND-PPN 106294261) verlinkt. Es stellt sich die

1	Tpv3	f	Förster, Frank	piz	Autör	
2	Tnx	f	Förster, Frank M.		Förster, Frank	
3	Tnx	f	Förster, Frank		Foerster, Frank	
4	Tpv6	f	Förster, Frank	piz	Förster, Frank Michael	
5	Tpv3	f	Förster, Frank	piz	Mathematiker	
6	Tpv3	f	Förster, Frank	piz	19XX- Archäologe	
7	Tpv3	f	Förster, Frank	piz	1980- Biochemiker	
8	Tpv3	f	Förster, Frank	piz	1958- Ingenieur	
9	Tpv3	f	Förster, Frank	piz	1969- Maschinenbauingenieur	
10	Tpv1	s,f	Förster, Frank	piz	1962- Groß- und Außenhandelskaufmann	
11	Tpv3	f	Foerster, Frank	piz	1962 Theologe	
12	Tpv3	f	Förster, Frank	piz	1978- Germanist	
13	Tpv3	f	Foerster, Frank	piz		
14	Tpv3	f	Förster, Frank	piz	1958- Arzt	Förster, F., 1958-; Förster, Frank Gerhard, 1958-
15	Tpv3	f	Förster, Frank	piz	Lebensmittelchemiker	Förster, Frank Oliver
16	Tpv1	f,s	Förster, Frank	piz	1937-2011 Historiker	

Abbildung 1

Das fachliche Spektrum zeigt, dass unspezifisch verknüpfte Autoren mit dem Namen Frank Förster vergleichsweise einfach über thematische Schwerpunkte eingegrenzt und zugeordnet werden könnten, also ein über Wahrscheinlichkeiten eingerichteter Algorithmus mit relativer Gewissheit die Titeldatensätze dem „richtigen“, individualisierten Frank Förster zuweisen könnte.

Beim Analysieren der mit dem Tn-Satz von Frank Förster (GND-PPN 106294261) verlinkten 21 Titel (siehe Abbildung 2) können mit Maßnahmen aus Formalkatalogisierung und Sacherschließung wahrscheinliche Verbindungen zu Tp-Sätzen erfasst wer-

den. Frage, ob nicht derselbe Frank Förster im Tp-Satz mit GND-PPN 13935073X gemeint ist, denn dort finden sich als Hinweis der übereinstimmende akademische Grad und der thematische Bezug zum Weinbau. Hier wird deutlich, dass bei einem Durchlauf mehrere Katalogdatensätze einem Tp-Satz zugeordnet werden können, sofern die Wahrscheinlichkeit es zulässt, zum Beispiel bei gleichlautenden Titeln eines Werks aufgrund von Folgeauflagen.

Sacherschließung: Die fachliche Eingrenzung über Wörter aus den Bereichen Titel, Abstract, Klassifikation und/oder Verschlagwortung erbringt hohe

1	Tnx	f	Förster, Frank			Förste.
2	*Aau		Tafeltrauben für den Hausgarten	Ulrich, Gerd	3., aktualisierte Auflage	Stuttgart (Hohenhei... [2016]
3	*Asn	31(2011,2) S. 118-120	Imgard Siebert (Hrsg.), Bibliothek und Forschung : die Bedeutung von Sammlungen für die Wissenschaft, Frankfurt am Main, ...	Förster, Frank	In: ABI-Technik	2011
4	*Aau		Die Erschließung belletristischer Literatur in Sachkatalogen wissenschaftlicher Bibliotheken im deutschsprachigen Raum	Förster, Frank		Leipzig 2009
5	*Asn	(2006) S. 45-61	Persönlichkeitsmerkmale von Studienanfängerinnen des Lehramts an Grundschulen - ein Vergleich verschiedener Wege des...	Förster, Frank	In: Empirische Lehrerbildungsforschung	2006
6	*Asn	(2004,12) S. 33-45	Untersuchungen zur Entwicklung der Stickstoffbilanzsalden und der Stickstoff-Effizienz in Sachsen	Albert, Erhard	In: Infodienst ... für Beratung und Schule der Sächsi...	2004
7	*Asn	95(2003,3) S. 329-338	Auswahlgespräche zur Vergabe von Studienplätzen im Lehrerstudium : Erfahrungen im Fach Grundschulpädagogik in Bamberg	Faust, Gabriele	In: Die deutsche Schule	Hannover 2003
8	*Aan		Die nationalsozialistische Wendenzählung von 1939 und der Einfluss ihres Ergebnisses während des Zweiten Weltkriegs	Förster, Frank		2003
9	*Aan		Lebensraumkundlich zur Wendenfrage : Wege und Stationen eines Arbeitsgemeinschaftsleiters des Nationalsozialistischen L...	Förster, Frank		2002
10	*Asn	(2001) S. 291-302	Der Gesandte Bunsen - zum Briefmachlass eines Vormärz-Politikers : Forschungsbericht über eine Biographie	Förster, Frank	In: Literaturkonzepte im Vormärz	2001
11	*Asn	(1997,68) S. 243-250	Die Wendenzeit des Bundes Deutscher Osten	Förster, Frank	In: Neues Archiv für sächsische Geschichte	1997
12	*Aau		Carl Pfeiffer : Förderer des Sächsischen Weinbaus	Schließer, Liselotte		[Leipzig] : Reintzsch [1996]
13	*Asn	71(1995) S. 5-12	Fünf Brote und zwei Fische : eine Spielidee zu Lukas	Förster, Frank	In: Braunschweiger Beiträge für Theorie und Praxis ...	1995
14	*Ask	(1993) S. 175 - 184	Das Wendenbild der NS-Wissenschaft	Förster, Frank	In: Neues Archiv für sächsische Geschichte	1993
15	*Aan		Das Ende des Braunkohlenwerkes "Olba" bei Kleinsaubernitz	Förster, Frank	Kopie aus Bautzener Kulturschau	1987
16	*Aau		Untersuchungen zum Mechanismus der Haftung zwischen Polymeren und aussenstromlos abgeschiedenen Metallschichten	Förster, Frank		1985
17	*Asr	17(1983) S. 27-31	Das demokratische, soziale und ethnokulturelle Bild des gemischtnationalen Gebietes der Lausitz	Förster, Frank	In: Geschichte und Gegenwart des Bezirkes Cottbus	1983
18	*AaB		Senftenberger Revier 1890 - 1914 : zur Geschichte der Niederlausitzer Braunkohlenindustrie vom Fall des Sozialistengesetz...	Förster, Frank		Bautzen : Domowina-... 1968
19	*Aan		Die Entwicklung der Sorbischen Zentralbibliothek in Bautzen : Aus 3000 wurden 4400 Bände	Förster, Frank		1966
20	*Asn	23(19XX)	Prof. Dr. Dr. h. c. Paul Nowotny 65 Jahre	Förster, Frank	In: Sächsische Heimatblätter. - 23 (1977) H. 3, S. 106	19XX
21	*AsB	68(19XX)	Die Wendenzeit des Bundes Deutscher Osten	Förster, Frank	In: Neues Archiv für Sächsische Geschichte. - 68 (1...	19XX
22	*Aan		Weggang eines Wendenbekämpfers : Dr. Walter Frenzels scheinbar überraschender Wechsel von Bautzen nach Frankfurt (...)	Förster, Frank		[o.J.]

Abbildung 2

21 GND: 110 Treffer, DNB-Katalog: 124 Treffer, k10plus: 231 Treffer (Stand: 20.4.2020).

Wahrscheinlichkeiten bei der Zuordnung weiterer Personen- zu Titeldatensätzen. Die Sacherschließung von Titeldaten spielt hier eine entscheidende Rolle, aber weniger bekannt ist, dass auch Personendatensätze eine Sacherschließung erlauben (GND-Feld 550 \$4them). Deshalb wird der Abgleich entweder direkt möglich oder über Konkordanzen ermittelt. Beispiele: 1) Die Hochschulschrift „Untersuchungen zum Mechanismus der Haftung zwischen Polymeren und außenstromlos abgeschiedenen Metallschichten“ (Dissertation, Berlin 1985) von Frank Förster (Tn-Satz GND-PPN 106294261) ist im k10plus der Sachgruppe „42 Technische Chemie, Lebensmitteltechnologie, Textiltechnik und andere Technologien“ zugeteilt. Unter den Tp-Sätzen findet sich ein Lebensmittelchemiker (GND-PPN 132602047) mit einer Dissertation (Heidelberg 2006). Wirkungszeit und bereits vorhandene Dissertation schließen mit hoher Wahrscheinlichkeit aus, dass es sich um denselben Frank Förster handelt. 2) Zwei Aufsätze im k10plus aus dem Bereich Lehrerbildung („Auswahlgespräche zur Vergabe von Studienplätzen im Lehrerstudium: Erfahrungen im Fach Grundschulpädagogik in Bamberg“, 2003, und „Persönlichkeitsmerkmale von Stu-

dienanfängerinnen des Lehramts an Grundschulen“, 2006) sind dem Tp-Satz GND-PPN 137053525 zuzuweisen, wobei sich der Bezug nicht über eine Berufsbezeichnung, sondern über die fachliche Zuordnung der angegebenen Dissertation und der verknüpften Titel aus dem Bereich der Grundschulpädagogik bzw. aus dem Wirkungszeitraum mit relativ hoher Wahrscheinlichkeit ergibt. Nach diesem Muster sind die Zuordnungen zum Professor für sorbische Volkskunde und dem thematischen Schwerpunkt der Wenden und Sorben (10 Treffer, GND-PPN 121835693), dem Germanisten und Bibliothekar (2 Treffer, GND-PPN 138068224) und dem Theologen mit mehreren biographischen Schriften zum preußischen Diplomaten Christian Carl Josias Bunsen, der in der Wikipedia einen Eintrag hat (1 Treffer, GND-PPN 143321625), für einen gut programmierten Algorithmus problemlos zu finden. Am Schluss des iterativen Verfahrens sollte ein Protokoll der Analysen und Veränderungen erfolgen. Hier beginnt die händische Arbeit der Bibliotheksbeschäftigten.

Falschzuordnung: Über das Profiling der Personendatensätze werden Falschzuordnungen bereits aus

Dietmar Dreier

International Library Suppliers

Seit 1981 für europäische Bibliotheken erfolgreich tätig.



dreier
ASPECTUS

Die Rechercheplattform für Bibliotheken
Sämtliche erwerbungsrelevante Fakten auf einen Blick

- Abbildung aller verfügbaren Kaufoptionen:
 - E-Book: Verlagscampuspreise (Pick&Choose)
 - E-Book: ProQuest Ebook Central™
 - Print: Gebundene und kartonierete Ausgabe (inkl. Sondernachlässe)
- Mindestbedingungen für einen Pick & Choose-Kauf
- Hinweis auf Paketzugehörigkeit
- Angabe der URL zum E-Book
- Bereitstellung von MARC-Daten für Einzeltitel oder Titellisten
- Kennzeichnung der Knowledge Unlatched Open Access-Titel

Sind Sie interessiert an EBA/EBS Modellen?
An Umwandlungen von Print zu E-Book?
An Sondernachlässen und frei verfügbaren Inhalten?
Kontakt: team-e@dietmardreier.de

Informationen ersten Grades offenbar. Beispiel: Dem 1962 geborenen Groß- und Außenhandelskaufmann (GND-PPN 11882662X), der in den 1980er Jahren in Malaysia wegen Drogenbesitzes festgenommen worden war und darüber einen Erfahrungsbericht veröffentlicht hat („Angst ist schlimmer als der Tod“, 1987) und nach Internetrecherche später als Web-Designer in Hessen gearbeitet hat, sind im k10plus drei Titeldatensätze vermutlich fälschlicherweise zugeordnet worden. Demnach dürfte das Buch „Entwicklung der P/K-Bilanzsalden und der Nährstoff-Versorgung in Sachsen“ (2001) nicht von ihm verfasst worden sein. Zudem gibt es im k10plus einen mit dem Tn-Satz verlinkten Aufsatz „Untersuchungen zur Entwicklung der Stickstoffbilanzsalden und der Stickstoff-Effizienz in Sachsen“ (2004), der dem anderen fachlich so nahe scheint, dass eine hohe Wahrscheinlichkeit besteht, dass es vom selben Verfasser stammt. Die beiden Titeldatensätze könnten eine Verlinkung zu einem neu anzulegenden Tn-Satz bekommen. Durch Autopsie werden sich möglicherweise genügend individualisierende Merkmale für einen neuen oder einen vorhandenen Tp-Satz ergeben. Für die Benennung des charakteristischen Berufes als Merkmal der Gruppe 1 (und somit hinreichend für einen Tp-Satz) gelingt anhand der Titel schon eine recht genaue Annäherung im Bereich der Agrochemie. Eine statistische Auswertung aller Einträge für die charakteristischen Berufe (Feld 550 + \$4berc) oder

weiterer GND-Bezeichnungen für den Beruf bzw. Tätigkeitsbereich (Feld 550 + \$4beru) und ein entsprechendes Profiling jedes Eintrags nach dem hier vorgeschlagenen Muster könnte relevante Vorschlagslisten für dieses Feld anhand der eingesammelten Informationen erzeugen. Das wäre ein Analysemuster aus dem Bereich der künstlichen Intelligenz.

Es gilt als sicher, dass es in künftigen Generationen Mittel und Techniken geben wird, die mit derartigen Analysemustern Zusammenhänge im Katalog aufdecken und ergänzen können. Insofern ist es zulässig, nicht absolute Genauigkeit anzustreben, sondern eine der Richtigkeit angenäherte Genauigkeit und eine „gelassene Haltung gegenüber Fehlern“ zuzulassen, wie sie der neue Generaldirektor der DNB, Frank Scholze, einfordert.²² Die Bibliotheksbenutzer können mit Falschzuordnungen und Unschärfen leben und werden diese bestenfalls noch an das Bibliothekspersonal weitergeben. Bei den anfangs genannten umfangreichen Mengen mehrerer Millionen Datensätze gilt als sicher, dass jeder hinzugekommene Datensatz, jedes neue Detail einer im Ansatz möglichst präzisen Beschreibung dazu beiträgt, dass künftige Auswerte- und Analysemechanismen besser arbeiten können und Wahrscheinlichkeiten klarer hervortreten lassen. Eine gelungene, bestehende Sacherschließung trägt einen wesentlichen Teil hierzu bei. ■

Literaturverzeichnis

- DNB-Datendienste 2019.** Datendienste: Abschaffung von Namensdatensätzen der GND mit Auswirkung auf die Titeldaten der DNB. Datum der Veröffentlichung: 23.8.2019.
- EH-P-01.** GND-Erfassungshilfe: Bildung von Sucheinstiegen, die Personen repräsentieren und Erfassung von identifizierenden Merkmalen. Stand: 24.2.2020.
- EH-P-14.** GND-Erfassungshilfe: Transliteration von Personennamen. Stand: 5.11.2014.
- EH-P-16.** GND-Erfassungshilfe: Individualisierung von Personennamen. Stand: 15.8.2019.
- EH-P-17.** GND-Erfassungshilfe: GND-Codes für Beziehungen in Tn- und Tp-Sätzen. Stand: 16.11.2017.
- Katalogisierungshandbuch Normdaten 2018.** Annabel Feuerstein, Armin Kühn: Normdaten der GND für Formal- und Sacherschließung (Katalogisierungshandbuch). Stand: 3.8.2018.



Frank Förster

Wissenschaftlicher Bibliothekar
Geozentrum Hannover
Stilleweg 2, D-30655 Hannover
frank.foerster@bgr.de
ORCID: 0000-0002-4373-7235

²² Sonntagsfrage: Warum wollen Sie bei der Deutschen Nationalbibliothek eine andere Fehlerkultur leben, Herr Scholze? In: Börsenblatt, 22.12.2019.