

# Die Recherche nach Pseudonymen in deutschen Katalogen nach dem RDA-Umstieg

## Bestandsaufnahme und Verbesserungspotenzial

Clara Schilling, Regierungspräsidium Tübingen, Fachstelle für das öffentliche Bibliothekswesen  
Heidrun Wiesenmüller, Hochschule der Medien, Stuttgart

### Zusammenfassung:

Nach RDA werden unter Pseudonym geschriebene Werke anders behandelt als nach RAK: Liegen mehrere bibliografische Identitäten vor, müssen nach dem Regelwerkumstieg nicht nur die Normdatensätze gesplittet, sondern auch Teile der Titeldatensätze umgehängt werden. Dies vermeidet Rechercheprobleme, bedeutet jedoch einen erheblichen Aufwand. In einer Befragung wurde ermittelt, wie sich die aktuelle Situation bei den Pseudonymen darstellt und als wie schwerwiegend die Rechercheprobleme eingeschätzt werden. Dabei wurde deutlich, dass eine umfassende intellektuelle Aufarbeitung der Pseudonymfälle in den Altdaten nicht zu erwarten ist. Auch eine anfangs diskutierte recherchetechnische Lösung kommt nur selten zum Einsatz. Eine interessante Alternative wäre deshalb eine maschinelle Zuordnung von Titeldatensätzen zu Normdatensätzen. Ein Test ergab, dass sich schon mit einem relativ einfachen Python-Skript ein großer Teil der Fälle lösen lässt. Mit einigen Erweiterungen und optimalerweise dem Einbezug von Werkclustern könnte eine sehr zuverlässige maschinelle Lösung implementiert werden.

### Summary:

Works written under pseudonym are treated differently according to RDA than according to RAK: If there are several bibliographic identities, not only must the authority records be split after the changeover to the new rules, but also some of the title records must be reassigned. This avoids problems in retrieval, but means considerable effort. A survey was conducted to determine the current situation with pseudonyms and how bad the consequences for the retrieval are considered to be. It became clear that a comprehensive intellectual processing of the pseudonym cases in the legacy data is unrealistic. In addition, a technical search solution which had originally been discussed has only rarely been implemented. Therefore, a machine assignment of title records to authority records would be an attractive alternative. A test showed that even with a relatively simple Python script, a large number of cases can be solved. With some extensions and optimally the inclusion of work clusters, a very reliable machine solution could be implemented.

**Zitierfähiger Link (DOI):** <https://doi.org/10.5282/o-bib/5985>

**Autorenidentifikation:** Schilling, Clara: ORCID: <https://orcid.org/0009-0001-4230-3451>,  
Wiesenmüller, Heidrun: ORCID: <https://orcid.org/0000-0002-9817-5292>, GND: 122087801

**Schlagwörter:** Resource Description and Access (RDA), RDA-Umstieg, Katalogisierung, Pseudonym, Katalogrecherche

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0](https://creativecommons.org/licenses/by/4.0/).

# 1. Der Regelwerksumstieg und die Problematik der Pseudonyme

Seit Ende 2015 wird das aus der angloamerikanischen Tradition stammende Katalogisierungsregelwerk „Resource Description and Access“ (RDA) auch an deutschen Bibliotheken angewendet. Neben vielen anderen, auffälligeren Regelwerksänderungen war mit dem Umstieg auch ein veränderter Umgang mit unter Pseudonym schreibenden Personen verbunden.

## 1.1 Pseudonyme in RAK und RDA

Gemäß den „Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken“ (RAK-WB) wurde, sofern das Pseudonym gelüftet war, immer nur ein einziger Normdatensatz für eine Person angelegt. In diesem wurden alle Namen – ggf. auch ein oder mehrere Pseudonyme – erfasst. Es war dann nur die Frage, ob der wirkliche Name oder ein Pseudonym als bevorzugter Name zu verwenden war – prinzipiell sollte es der bekanntere Name sein.<sup>1</sup>

Nach RDA hingegen werden für eine Person u.U. mehrere Normdatensätze angelegt – und zwar immer dann, wenn diese bewusst unter verschiedenen „bibliografischen Identitäten“ auftritt, indem sie teilweise unter ihrem wirklichen Namen, teilweise unter einem (oder mehreren) Pseudonymen publiziert.<sup>2</sup> Je nachdem, unter welchem Namen ein Werk veröffentlicht wird, wird eine Beziehung zum jeweils passenden Normdatensatz hergestellt; auch werden die entsprechenden Normdatensätze in der Gemeinsamen Normdatei (GND) miteinander verknüpft.<sup>3</sup>

Hinter dieser Praxis steht die Idee, dass sich mit unterschiedlichen Namen häufig auch unterschiedliche Arten von Werken verbinden. Benutzende sollen deshalb im Katalog gezielt nach Titeln suchen können, die unter einem bestimmten Namen veröffentlicht wurden. Um das klassische Beispiel zu bemühen: Wer nach den „Alice“-Büchern von Lewis Carroll (Pseudonym) sucht, ist vermutlich nicht an der von Charles Lutwidge Dodgson (wirklicher Name) verfassten mathematischen Fachliteratur interessiert.

Hat hingegen eine Person nur unter einem einzigen Namen publiziert, so gibt es auch unter RDA nur einen einzigen Normdatensatz. Beispielsweise hat der Autor von „1984“ alle seine Werke unter dem Pseudonym George Orwell veröffentlicht, weshalb sein echter Name Eric Arthur Blair nur als abweichender Name erfasst wird.

- 1 Vgl. Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken (RAK-WB), Stand: April 2006, Leipzig 2007. Online: <<https://nbn-resolving.org/urn:nbn:de:101-2007072711>>, § 308, 1–4. Der Paragraph ist etwas verquast formuliert: Eine Person wird „im allgemeinen unter ihrem wirklichen Namen angesetzt“, aber nicht, wenn sie unter dem Pseudonym „sehr viel bekannter“ ist. Dies bezieht sich sowohl auf Personen, die vor dem 20. Jahrhundert gelebt haben, als auch auf Personen des 20. Jahrhunderts, „die unter einem Namen bekannt geworden und darunter in gängigen Nachschlagewerken oder einer Ausgabe eines eigenen Werkes mit Nennung eines oder mehrerer benutzter Pseudonyme bzw. des wirklichen Namens zu ermitteln“ sind.
- 2 Oder auch unter mehreren Pseudonymen, ohne den echten Namen zu benutzen.
- 3 Vgl. Wiesenmüller, Heidrun; Horny, Silke: Basiswissen RDA. Eine Einführung für deutschsprachige Anwender, Berlin; Boston 2017<sup>2</sup>, S. 99f. und 167. Für eine detaillierte vergleichende Darstellung der Pseudonym-Regeln von RAK, AACR2 und RDA vgl. Schilling, Clara: Die Konsequenzen des RDA-Umstiegs für die Recherche nach Pseudonymen – aktuelle Situation, Probleme, Lösungsansätze, Bachelorarbeit, Hochschule der Medien, Stuttgart 2022. Online: <<https://nbn-resolving.org/urn:nbn:de:bsz:900-opus4-67622>>, S. 10–16.

## 1.2 Folgen des Regelwerksumstiegs

Der Regelwerkswechsel ist an dieser Stelle nicht so einfach umzusetzen. Zunächst müssen die betroffenen Normdatensätze in der GND gesplittet werden. Aus dem ehemals einzigen Normdatensatz werden zwei – in manchen Fällen auch mehr – Normdatensätze gemacht. Der ursprüngliche Normdatensatz bleibt erhalten, doch werden bestimmte Informationen dort gelöscht und in einen neu angelegten Normdatensatz verlagert.<sup>4</sup> Diese Arbeiten sind enorm aufwendig und noch keineswegs abgeschlossen, können aber zumindest kooperativ und zentral in der GND erledigt werden.

Um eine korrekte Recherche zu ermöglichen, genügt es jedoch nicht, nur die Normdatensätze zu überarbeiten. Vielmehr müsste man auch dafür Sorge tragen, dass alle Titeldatensätze mit dem passenden Normdatensatz verknüpft sind. Logischerweise sind jedoch Titeldatensätze aus RAK-Zeiten mit dem ursprünglichen Normdatensatz verknüpft – der nunmehr aber nur noch für eine von zwei oder mehr Identitäten der Person steht. Deshalb müsste man in Pseudonym-Fällen alle zugehörigen Titel durchsehen und sie ggf. an den neuen Normdatensatz „umhängen“. Dies kann leider nicht zentral erfolgen, sondern muss für jeden (Verbund-)Katalog separat durchgeführt werden.

Werden die Zuordnungen nicht bereinigt, führt dies zu Rechercheproblemen, wie man sich am folgenden Beispiel klarmachen kann: Für eine Person mit dem echten Namen A und dem Pseudonym B gab es ursprünglich nur einen einzigen Normdatensatz N1. Mittlerweile steht jedoch N1 nur noch für den echten Namen A, während für das Pseudonym B ein zweiter Normdatensatz N2 angelegt wurde. Sucht man nun nach B, erhält man alle mit N2 verknüpften Titeldatensätze. Bei den seit dem RDA-Umstieg angelegten Titeldatensätzen für Werke unter Pseudonym B ist die Zuordnung korrekt; diese werden richtig ausgegeben. Aber alle noch unter RAK katalogisierten Werke unter dem Pseudonym B erscheinen nicht als Treffer, da diese Titeldatensätze noch mit N1 verknüpft sind – die Trefferliste ist hier also unvollständig. Umgekehrt gibt es bei einer Recherche nach A einen gewissen Ballast, da hier auch alle unter B verfassten Werke erscheinen, die noch zu RAK-Zeiten katalogisiert wurden.

Deshalb wurde festgelegt, dass bei einer Aufspaltung in mehrere Normdatensätze in jedem davon der Benutzungshinweis „Weitere Titel ggf. auch unter dem Pseudonym bzw. dem wirklichen Namen“ anzubringen ist. Auch wird in diesen Fällen zur Information eine Mailbox geschrieben, damit jeder Verbund im eigenen Katalog die betroffenen Titel an den neuen Normdatensatz umhängen kann.<sup>5</sup>

Mit Blick auf die Recherche wurde außerdem eine Anpassung der Indexierung bei miteinander verknüpften Normdatensätzen wie N1 und N2 diskutiert: In einem solchen Fall könnten – so die Überlegung – stets alle Titeldatensätze ausgegeben werden (also sowohl die mit N1 als auch die mit N2 verknüpften), egal wonach gesucht wurde. Man könnte dann zwar nicht gezielt nach A oder B recherchieren, würde aber zumindest keine relevanten Treffer verlieren.

---

4 Für Details vgl. Schilling, Clara: Die Konsequenzen des RDA-Umstiegs für die Recherche nach Pseudonymen, 2022, S. 17f.

5 Vgl. GND-Erfassungshilfe „Pseudonyme“ (EH-P06), <<https://wiki.dnb.de/download/attachments/90411361/EH-P-06.pdf>>, Stand: 30.07.2023.

Knapp sieben Jahre nach dem RDA-Umstieg untersuchte Clara Schilling in ihrer Bachelorarbeit<sup>6</sup>, wie sich die aktuelle Situation bei den Pseudonymen darstellt und als wie schwerwiegend die Rechercheprobleme eingeschätzt werden. Ebenso wurde geprüft, ob eine maschinelle Zuordnung von Titeldatensätzen zu Normdatensätzen möglich wäre. Dies war bereits 2015 von Heidrun Wiesenmüller vorgeschlagen worden.<sup>7</sup> Im Folgenden werden wichtige Ergebnisse und weitere Überlegungen zu dem Themenkomplex vorgestellt.

## 2. Aktuelle Praxis und Wahrnehmung der Problematik

Um die aktuelle Praxis in Deutschland zu ermitteln und herauszufinden, wie das Problem von Praktiker\*innen wahrgenommen wird, wurden Vertreter\*innen der Verbundzentralen, der Deutschen Nationalbibliothek (DNB) sowie von je drei Universitätsbibliotheken und Öffentlichen Bibliotheken befragt.<sup>8</sup>

Die Befragung ergab, dass die Umarbeitung von älteren Normdatensätzen nirgends systematisch geschieht, sondern stets nur bei Bedarf. Auch kann man nicht davon ausgehen, dass die Mailboxen über gesplittete Normdatensätze überall vollständig abgearbeitet werden. Bei der DNB und im Bibliotheksverbund Bayern (BVB) bemüht man sich darum. Im K10plus beispielsweise kann hingegen nur ein Teil bearbeitet werden – am ehesten sind dies Fälle mit überschaubaren Titelmengen. Auch ist bei der Bearbeitung mit einem größeren Backlog zu rechnen; in der DNB wurde dieses zum Zeitpunkt der Befragung auf etwa 1.000 Fälle geschätzt.

Die Rechercheproblematik war fast überall bekannt; Ausnahme war eine der befragten Stadtbibliotheken. Für die Öffentlichen Bibliotheken spielt allerdings das Thema Normdaten generell eine untergeordnete Rolle: Teilweise wird überhaupt nicht mit Normdaten gearbeitet oder es wird nicht die GND, sondern eine eigene Normdatei verwendet, die im Bereich der Pseudonyme nicht RDA-gerecht ist.

Zum Zeitpunkt der Befragung hatten nur der K10plus und hebis eine erweiterte Indexierung implementiert (s.o. Kap. 1.2). In beiden Verbänden werden also beispielsweise bei der Suche nach „King, Stephen“ oder seinem Pseudonym „Bachman, Richard“ alle Titel angezeigt, die mit einem der beiden Normdatensätze verknüpft sind.<sup>9</sup> Allerdings ist die beschriebene technische Lösung natürlich nur auf der Ebene der Verbundkataloge realisiert und nicht automatisch auch in den lokalen Katalogen

- 6 Schilling, Clara: Die Konsequenzen des RDA-Umstiegs für die Recherche nach Pseudonymen. 2022. Die Bachelorarbeit wurde an der Hochschule der Medien in Stuttgart erstellt und von Heidrun Wiesenmüller (HdM) und Annabel Feuerstein (BSZ) betreut.
- 7 Vgl. Wiesenmüller, Heidrun: Der RDA-Umstieg in Deutschland – Herausforderungen für das Metadatenmanagement, in: o-bib 2 (2), 2015, S. 43–60. Online: <<https://doi.org/10.5282/o-bib/2015H2S43-60>>, hier S. 54–58.
- 8 Bei den Einzelbibliotheken handelte es sich um die Universitätsbibliotheken in Stuttgart, Tübingen und Heidelberg sowie die Stadtbibliotheken Reutlingen und Tübingen und die Bücherhallen Hamburg. Für die Methodik und Details der Befragung vgl. Schilling, Clara: Die Konsequenzen des RDA-Umstiegs für die Recherche nach Pseudonymen, 2022, S. 20–32.
- 9 Bei der Anzeige gibt es einen interessanten Unterschied in den beiden Verbundkatalogen: Unabhängig davon, nach welchem Namen man sucht, erhält man in hebis dieselbe gemischte Ergebnisliste (nach Erscheinungsdatum sortiert). Im K10plus erscheinen hingegen zunächst alle Titel unter dem Namen, nach dem man tatsächlich gesucht hat; anschließend folgen die Titel, die mit dem anderen Normdatensatz verlinkt sind.

der Verbundteilnehmer. Mittlerweile hat auch die DNB eine erweiterte Indexierung in ihrem neuen Katalog umgesetzt.<sup>10</sup> Bei den Bücherhallen Hamburg wäre eine entsprechende Indexierung gewünscht gewesen, ließ sich aber technisch nicht realisieren. An der UB Tübingen entschied man sich hingegen bewusst dagegen, da man den dabei entstehenden Ballast für problematischer hielt als das mögliche Fehlen relevanter Titel.

Öfter wurde auf den am Normdatensatz angebrachten Hinweis (s.o. Kap. 1.2) verwiesen. Dessen praktischer Nutzen ist allerdings fraglich, da die Normdatensätze in den meisten Katalogen gar nicht angezeigt werden können. Und selbst dort, wo dies möglich ist, müssten Nutzende zunächst aktiv den Normdatensatz aufrufen, um von den weiteren bibliografischen Identitäten zu erfahren – was auch Kenntnisse über Normdaten voraussetzt. Dass ihre Trefferliste möglicherweise unvollständig ist, dürfte Nutzenden deshalb zumeist entgehen. Da keine einschlägigen Nachfragen oder Beschwerden bekannt sind, betrachten die befragten Praktiker\*innen das Problem als wenig bis gar nicht relevant aus Nutzendensicht. Allerdings können sich Nutzende ja auch nicht über unvollständige Trefferlisten beschweren, wenn ihnen gar nicht klar ist, dass Treffer fehlen.

Darüber hinaus ist der Umfang der Pseudonymfälle in den Bibliotheken recht unterschiedlich: So wies die UB Stuttgart darauf hin, dass bei ihrem Bestand kaum Pseudonyme vorkämen. Und bei Öffentlichen Bibliotheken wie den Bücherhallen Hamburg führt das regelmäßige Aussondern älterer Titel dazu, dass sich das Problem mit der Zeit sozusagen von selbst löst.

Insgesamt wurde deutlich, dass mit einer umfassenden intellektuellen Aufarbeitung der Pseudonymfälle in den Altdaten auch in Zukunft nicht zu rechnen ist. Dies liegt zum einen an fehlenden Kapazitäten, zum anderen an der als eher gering eingeschätzten Relevanz der Problematik. Auch eine technische Lösung in Form einer erweiterten Indexierung ist eher die Ausnahme als die Regel.

### **3. Maschinelle Zuordnung von Normdatensätzen**

Umso interessanter ist die Frage, ob eine korrekte Zuordnung der noch unter RAK erfassten Titeldatensätze zu nachträglich gesplitteten Normdatensätzen nicht auch maschinell erreicht werden kann. Die Grundidee dabei ist, die in den Titeldatensätzen erfassten Verantwortlichkeitsangaben zu parsen und entsprechend der darin aufgefundenen Namensform den passenden Normdatensatz zu ermitteln.

Um Möglichkeiten und Grenzen dieser Idee besser abschätzen zu können, erfolgte im Rahmen der Bachelorarbeit eine Analyse anhand eines vom Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) zur Verfügung gestellten Datenabzugs aus dem K10plus.<sup>11</sup> Entsprechend basiert die folgende Darstellung auf dem Pica-Format, wäre aber ohne Weiteres auch in einem anderen Datenformat umzusetzen.

---

<sup>10</sup> Verfügbar in einer Beta-Version unter <https://katalog.dnb.de/DE/home.html>, Stand: 30.07.2023.

<sup>11</sup> Der für die Analyse verwendete Datenabzug wurde im September 2022 erstellt und unter Einbezug der GND und des K10plus auf dem Stand von Oktober und November 2022 analysiert. Einige Veränderungen, die sich seither im Datenbestand ergeben haben, wurden für die vorliegende Aufsatzfassung berücksichtigt.

### 3.1 Methodik der Analyse

Der für die Analyse verwendete Datenabzug enthielt Titeldatensätze, die zwischen den Jahren 2000 und 2014 – also in jedem Fall noch unter RAK – angelegt wurden und mit einem Pseudonymnormdatensatz<sup>12</sup> in einem der RAK-Verfasserfelder (in Pica 3000, 3001 oder 3002) verlinkt waren. Darüber hinaus sollten die selektierten Datensätze an einem Normdatensatz hängen, der mit mindestens fünf Titeldatensätzen im K10plus verknüpft ist, um die Analysedaten nicht unnötig mit „Eintagsfliegen“ aufzuschwemmen. Dies ergab zunächst 169.861 Datensätze.

Als Grundlage für die Analyse wurde ein Python-Skript geschrieben, mit dem geistige Schöpfer\*innen und Verantwortlichkeitsangaben – in einer späteren Version des Skripts auch Haupttitel und Titelzusatz – aus den Titeln extrahiert und miteinander abgeglichen werden konnten.<sup>13</sup> Folgende Überlegungen lagen zugrunde: In den meisten Fällen wird der\*die geistige Schöpfer\*in eines Werkes in der Verantwortlichkeitsangabe genannt und entsprechend im Unterfeld \$h des Pica-Felds 4000 erfasst. Außerdem wird der entsprechende Normdatensatz im dafür vorgesehenen Feld verknüpft (Feld 3000 für den\*die erstgenannte\*n geistige\*n Schöpfer\*in). Generell vergleicht das Skript nur Nachnamen, da die Einbeziehung von Vornamen weitere Variablen einführen würde.<sup>14</sup>

Unter den Bedingungen von RDA entspricht der im Verknüpfungsfeld angezeigte Name (bzw. eine abweichende Form davon) grundsätzlich dem Namen in der Verantwortlichkeitsangabe, da jeweils die bibliografische Identität verknüpft wird, die das vorliegende Werk verfasst hat. Wenn aber der verlinkte Normdatensatz nach RAK erstellt wurde, so gibt es zwei Möglichkeiten: Entweder der angezeigte Name entspricht dem Namen in der Verantwortlichkeitsangabe (bzw. einer abweichenden Form davon) – d.h. das Werk wurde unter dem Namen veröffentlicht, der auch im Normdatensatz als bevorzugter Name erfasst ist. Oder der angezeigte Name entspricht nicht der Verantwortlichkeitsangabe – d.h. das Werk wurde unter einem Namen veröffentlicht, der im Normdatensatz nicht als bevorzugter Name erfasst ist. Dies ist z.B. der Fall, wenn das Werk unter einem Pseudonym veröffentlicht wurde, aber der echte Name als bevorzugter Name gewählt wurde.

Entspricht der Name (bzw. eine leicht abweichende Form davon) in der Verantwortlichkeitsangabe dem Namen in Feld 3000, so würde eine Zuordnung des passenden Normdatensatzes über ein Parsing in jedem Fall funktionieren. Denn wenn bei einem Split der Normdatensätze der nach RAK gewählte bevorzugte Name im „alten“ Datensatz, an dem alle Titeldatensätze hängen, erhalten bleibt, müssen die Titeldatensätze, die diesen Namen in der Verantwortlichkeitsangabe enthalten, auch nicht an einen neuen Normdatensatz umgehängt werden. Sie können von einem Programm als bereits korrekt zugeordnet sozusagen „abgehakt“ und bei der Bearbeitung übersprungen werden.

12 Für das Skript wurden Normdatensätze aus der GND berücksichtigt, die den Code „pip“ im Feld 008 besitzen. Es war dabei unerheblich, ob der Normdatensatz gesplittet worden war oder nicht. Im Falle eines Splits bekommt allerdings der Datensatz für den wirklichen Namen den Code „piz“; diese Datensätze waren deshalb im Abzug nicht enthalten.

13 Für Details zur Programmierung und das vollständige Skript vgl. Schilling, Clara: Die Konsequenzen des RDA-Umstiegs für die Recherche nach Pseudonymen, 2022, S. 35–42 und 52–54.

14 Beispielsweise sind Personen häufig nur mit abgekürzten oder ganz ohne Vornamen angegeben, was das Skript nicht als gleichen Namen erkennen würde.

Folglich stellen alle Titeldatensätze, bei denen der Name in der Verantwortlichkeitsangabe und der Name im Verfasser\*innenfeld übereinstimmen, keine Probleme für eine maschinelle Lösung dar. Entsprechend mussten sie in der Analyse nicht näher betrachtet werden. Deshalb sollte das Python-Skript solche Fälle aus dem Datenabzug herausfiltern. Mit der ersten Version des Skripts ergaben sich dabei noch verschiedene Probleme, da nur auf exakte Übereinstimmung geprüft wurde – folglich wurden Fälle mit Namensvarianten, abweichenden Transkriptionen u.Ä. zunächst nicht korrekt herausgefiltert. Das Skript wurde deshalb um eine Prüfung auf ähnliche Zeichenstrings erweitert, damit auch solche Fälle erkannt wurden.<sup>15</sup>

Datensätze ohne eine Beziehung zu einem bzw. einer geistigen Schöpfer\*in wurden ebenfalls herausgefiltert, da in diesen Fällen logischerweise kein Abgleich möglich ist. Anfangs wurden auch alle Datensätze ohne Verantwortlichkeitsangaben entfernt; dies wurde später revidiert, da geistige Schöpfer\*innen teilweise auch im Titel genannt werden (s.u. Kap. 3.3).

Um die Chancen für eine maschinelle Bearbeitung zu prüfen und mögliche Fehlerquellen zu identifizieren, sollten also zunächst diejenigen Datensätze analysiert werden, die sowohl eine Verantwortlichkeitsangabe als auch eine Beziehung zu einem geistigen Schöpfer bzw. einer geistigen Schöpferin enthalten, welche sich beim Namen nicht entsprechen. Über das Parsing der Verantwortlichkeitsangabe sollte es in diesen Fällen nach einem Split des ursprünglichen Normdatensatzes möglich sein, den korrekten Satz – also den, bei dem der Name aus der Verantwortlichkeitsangabe (bzw. eine abweichende Form davon) als bevorzugter Name erfasst ist – zu ermitteln. Dabei geht es natürlich nur darum, unter mehreren bibliografischen Identitäten derselben Person auszuwählen (d.h. unter zwei oder mehr miteinander verknüpften Personendatensätzen). Andere, zufällig namensgleiche Personen dürfen in einen solchen Abgleich nicht mit einbezogen werden.

Gemäß den beschriebenen Kriterien wurden die ursprünglich 169.861 Datensätze auf 9.040 reduziert. Diese verblüffend starke Reduktion zeigt, dass die von einer Person verwendeten Namen typischerweise nicht gleichmäßig verteilt sind. Eher ist davon auszugehen, dass zumeist ein Name deutlich überwiegt – welcher nach den Prinzipien von RAK auch der bevorzugte Name im ursprünglichen Normdatensatz ist. Zu den unproblematischen Datensätzen gehören außerdem die Fälle, bei denen zwar ein Pseudonym existiert, die betreffende Person aber nur einen einzigen Namen tatsächlich verwendet (s.o. Kap. 1.1).

50 der verbliebenen Datensätze wurden detailliert daraufhin untersucht, ob eine maschinelle Zuordnung grundsätzlich funktionieren würde, ob ggf. Anpassungen oder Erweiterungen beim Skript nötig wären oder nicht lösbare Fehler auftreten würden. Beim Durchsehen weiterer Datensätze wurde klar, dass die Überprüfung von 50 Beispielen ausreichend war, da sich anschließend die Fälle nur noch wiederholten und keine neuen Aspekte mehr zutage traten. Im Folgenden werden die wichtigsten Fallgruppen vorgestellt und mit Beispielen illustriert. Angegeben werden dabei jeweils

---

15 Dafür wurde die Pythonklasse `Sequencematcher` verwendet, die die Ähnlichkeit (Ratio) zweier Strings vergleicht. Bei einer Ratio von über 0.5 wurden die Strings als gleicher Name gewertet. Für Details vgl. Schilling, Clara: Die Konsequenzen des RDA-Umstiegs für die Recherche nach Pseudonymen, 2022, S. 53.

die Nummer des Titeldatensatzes im K10plus (PPN) sowie die relevanten Felder bzw. Teile davon – insbesondere 3000 (erste\*r geistige\*r Schöpfer\*in mit PPN-Nummer im K10plus) und 4000 (Titel und Verantwortlichkeitsangabe).<sup>16</sup>

### 3.2 Sofort funktionierende Fälle

Bei sechs der im Detail analysierten Fälle würde das Skript bereits in der bestehenden Form zum korrekten Ergebnis kommen, wie im folgenden Beispiel:

(A) *Keine Probleme*

Beispiel: PPN 1095169378  
3000 !136307949!Beets, Nicolaas  
4000 Camera obscura\$hvan Hildebrand

Der Autor hat sowohl unter seinem echten Namen Nicolaas Beets als auch unter seinem Pseudonym Hildebrand publiziert; der Normdatensatz müsste deshalb gesplittet werden. Wäre dies schon geschehen, dann würde die maschinelle Zuordnung folgendermaßen funktionieren: Beim Parsing der Verantwortlichkeitsangabe würde zuerst geprüft, ob eine Namensform aus dem Feld 100 (bevorzugter Name, hier „Beets, Nicolaas“) oder einem 400er-Feld (abweichende Namen, hier u.a. „Beets, Nikolaus“, „Beets, Nicolas“) des Normdatensatzes enthalten ist. Dies ist nicht der Fall, weshalb das Programm zu den 500er-Feldern (in Beziehung stehende Personen) weitergeht, wo die Normdatensätze der anderen Identitäten verknüpft sind. Im Normdatensatz für das Pseudonym ist die gesuchte Form „Hildebrand“ enthalten, sodass dieser Datensatz zugeordnet werden kann.

### 3.3 Nach Anpassung funktionierende Fälle

Bereits von Anfang an war klar, dass das Skript bestimmte Fälle nicht finden konnte. Denn zur Vereinfachung wurde bei der Programmierung nur Feld 3000 berücksichtigt, in dem der oder die erste Verfasser\*in erfasst wird. Folglich fielen alle Titeldatensätze unter den Tisch, in denen die Person mit dem Pseudonym in einem der anderen Personfelder verknüpft ist (als weitere\*r geistige\*r Schöpfer\*in, als Herausgeber\*in etc.). Bei den im Detail analysierten Fällen trat dieser Fall dreimal auf. Im Falle einer Echanwendung des Mechanismus wäre das Skript entsprechend zu erweitern.

Auch die Berücksichtigung abweichender Namensformen wäre noch zu verbessern. Denn es wurde mit einer einfachen Ähnlichkeitssuche gearbeitet, welche abweichende Namen in den meisten, aber nicht in allen Fällen richtig erkannte. So wurde in einem Fall die in der Verantwortlichkeitsangabe enthaltene Namensform „G. Sinowjew“ nicht als Variante des bevorzugten Namens „Zinov’ev, Grigorij Evseevič“ erkannt. Im zugehörigen Normdatensatz ist jedoch „Sinowjew, G.“ als abweichender Name erfasst, sodass der Fehler bei einer etwas aufwendigeren Programmierung nicht aufgetreten wäre.

---

16 Für die vollständige Darstellung der Analyseergebnisse vgl. Schilling, Clara: Die Konsequenzen des RDA-Umstiegs für die Recherche nach Pseudonymen, 2022, S. 38–46.

Dennoch wird eine gewisse Unschärfefunktion auch in einer Echtenwendung nötig sein, um kleine Unterschiede gegenüber den im Normdatensatz erfassten Namensformen auszugleichen. Solche ergeben sich z.B. häufig durch ein Genitiv-s am Ende des Verfassernamens (wenn dieser im Titel genannt ist, s.u.), manchmal auch durch banale Tippfehler. Ein weiteres Beispiel war eine Verantwortlichkeitsangabe, in der ein teilweise ermittelter Name durch eckige Klammern gekennzeichnet worden war: [Verf.: Franz Hermann] H[egewisch].<sup>17</sup>

Darüber hinaus wurden zwei weitere Fallgruppen identifiziert, für die eine Erweiterung des Skripts erforderlich wäre. Ein gar nicht so seltener Fall ist, dass der Name des Autors bzw. der Autorin nicht in der Verantwortlichkeitsangabe steht, sondern im Haupttitel oder Titelzusatz wie im folgenden Beispiel:

*(B) Geistiger Schöpfer im Titelbereich*

Beispiel: PPN 014957582

3000 !014957582!Kishon, Ephraim

4000 Kishons beste Tiergeschichten\$Hmit Zeichn. von Rudolf Angerer.

[Ins Dt. übertr. von Friedrich Torberg u. Gerhard Bronner]

In der Verantwortlichkeitsangabe sind in diesem Fall nur andere Personen genannt, während der Autor Ephraim Kishon (Pseudonym von Ferenc Hoffmann) als Teil des Haupttitels erfasst wurde. In der derzeitigen Fassung würde das Skript anhand der Verantwortlichkeitsangabe also keine Namensform von Kishon (weder den echten Namen noch das Pseudonym) zuordnen können. Im konkreten Beispiel ist dies zwar unproblematisch, da der Normdatensatz ohnehin nicht gesplittet werden muss (der Autor hat nie unter seinem echten Namen publiziert).<sup>18</sup> Dieselbe Sachlage taucht aber natürlich auch in anderen Fällen auf, wo ein Normdatensatz bereits gesplittet wurde oder noch gesplittet werden müsste.

Zur Lösung des Problems müsste das Skript nicht nur die Verantwortlichkeitsangaben, sondern auch das Titelfeld absuchen. Entsprechend wurden die restlichen Analysen mit dieser Anpassung durchgeführt. Der beschriebene Fall trat insgesamt 27-mal auf (25-mal im Haupttitel und zweimal im Titelzusatz) – also in mehr als der Hälfte der analysierten Datensätze.

Eine weitere Fallgruppe betrifft mehrteilige Monografien. Hier wird typischerweise die Verantwortlichkeitsangabe, die den geistigen Schöpfer bzw. die geistige Schöpferin nennt, nur im Datensatz für die übergeordnete Aufnahme erfasst:

---

17 Dieser Fall könnte alternativ durch eine entsprechende Indexierung gelöst werden, bei der Klammern u.Ä. entfernt werden.

18 Bei den im Folgenden verwendeten Beispielen wird nicht in jedem Fall darauf hingewiesen, ob es sich um einen tatsächlich zu splittenden Normdatensatz handelt oder nicht, da dies bei der Frage, ob das Parsing funktionieren würde oder nicht, letztlich irrelevant ist.

(C) Geistiger Schöpfer bei mehrteiliger Monografie

Beispiel: PPN 069147841

3000 !069147841!Borgersrud, Lars

4000 I tilfelle opprøret i det norske militærapparatet fra 1814 til 1905

4150 Våpen mot folket \$hOttar Strømme

Lars Borgersrud benutzte auch das Pseudonym Ottar Strømme; sein Normdatensatz müsste deshalb gesplittet werden. Die Verantwortlichkeitsangabe „Ottar Strømme“ findet sich im gezeigten Titeldatensatz für Band 1 allerdings nicht in Feld 4000, sondern nur in 4150 beim Gesamttitel der mehrteiligen Monografie. Damit das Skript diesen Fall richtig zuordnen könnte, müsste auch Feld 4150 (hier die Verantwortlichkeitsangabe in \$h, ggf. auch der Titel selbst) mit berücksichtigt werden.<sup>19</sup> Dieser Fall kam siebenmal vor.

### 3.4 Problematische und komplexe Fälle

In zwei der analysierten Datensätze kamen Erfassungsfehler vor. So wurde im folgenden Beispiel eine falsche Person verknüpft:

(D) Erfassungsfehler

Beispiel: PPN 065485459

3000 !793468132!Niebuhr, Carl

4000 Über Condensationen des Anhydrides der

1-Phenyl-3-methyl-pyrazolon-Bz.-o.-Carbonsäure \$hCarl Krug

Carl Niebuhr ist ein Pseudonym von Karl Krug (abweichender Name: Carl Krug), doch war Niebuhr Orientalist. Hier handelt es sich um die Dissertation eines anderen Carl Krug mit ähnlichen Lebensdaten (PPN 1005755604); bei zwei weiteren Titeldatensätzen im K10plus ist er korrekt verknüpft. Das Parsing würde hier zwar funktionieren, allerdings natürlich weiter die falsche Person zuordnen.

Ausgehend von einem analysierten Titel des nationalsozialistischen Schriftstellers Gottfried Rothacker, der teilweise unter seinem Pseudonym Bruno Nowak schrieb, fiel eine Unregelmäßigkeit bei mehreren Ausgaben desselben Werks auf. Die folgenden Beispiele zeigen die Situation:

---

<sup>19</sup> Die Alternative, von vornherein nur den übergeordneten Datensatz zu betrachten, wäre vermutlich nicht ausreichend, da auf Ebene der untergeordneten Datensätze weitere Personen (z.B. Herausgebende) mit einem Pseudonym vorkommen könnten. Evtl. könnten jedoch übergeordnete und untergeordnete Ebene zusammen betrachtet werden.

(E) Name in Verantwortlichkeitsangaben uneinheitlich

Beispiel 1: PPN 1502288435  
1100 1939\$n[1939]  
3000 !078523575!Nowak, Bruno  
4000 Die @Kinder von Kirwang\$hBruno Nowak  
4020 27. – 32. Tsd.  
4030 Berlin\$nJunge Generation Verl.

Beispiel 2: PPN 1512284076  
1100 1941\$n[1941]  
3000 !078523575!Nowak, Bruno  
4000 Die @Kinder von Kirwang\$hGottfried Rothacker  
4020 38. –47. Tsd.  
4030 Berlin\$nJunge-Generation-Verl.

Beim ersten Beispiel würde das Skript den Datensatz für das Pseudonym zuordnen und beim zweiten den für den echten Namen. Wahrscheinlich handelt es sich aber nur um eine unterschiedliche Erfassungs- bzw. Regelwerkspraxis, die sich in den vermutlich retrokonvertierten Aufnahmen widerspiegelt. Die Gestaltung des Titelblatts hat sich bei den unterschiedlichen Drucken wohl nicht tatsächlich geändert; durchgehend wird Gottfried Rothacker genannt.

Uneinheitliche Verantwortlichkeitsangaben kamen in den analysierten Datensätzen insgesamt dreimal vor. Besonders bemerkenswert ist dabei der Fall der dänischen Autorin Karen Blixen, die ihre Werke je nach Sprachraum unter verschiedenen Pseudonymen publizierte. Beispielsweise erschien „Out of Africa“ zunächst auf Englisch unter dem Namen Isak Dinesen, die dänische Version dann unter Karen Blixen und die deutsche Fassung unter Tania Blixen. Die Normdatensätze sind in der GND bereits separiert, doch ist unabhängig von der Sprache das Gros der Titel im K10plus derzeit mit dem Datensatz für das „deutsche“ Pseudonym verknüpft. Die folgenden Beispiele zeigen Ausgaben desselben Werks:

Beispiel 3: PPN 1615537236  
3000 !123941180!Blixen, Tania  
4000 Winter's tales\$hIsak Dinesen (Karen Blixen)

Beispiel 4: PPN 33655334X  
3000 !123941180!Blixen, Tania  
4000 Wintergeschichten\$hTania Blixen. Dt. von Jürgen Schweier

Um eine RDA-gerechte Lösung zu erreichen, sollten diese Titel konsequent entsprechend der Verantwortlichkeitsangaben zugeordnet werden, was mit einem Parsing grundsätzlich möglich ist.

Beispiel 3 zeigt noch eine zusätzliche Schwierigkeit, nämlich das Vorkommen eines Pseudonyms und des echten Namens in derselben Verantwortlichkeitsangabe. Anstatt einfach den erstgenannten Namen zu benutzen, wäre es in einem solchen Fall vermutlich besser, den Datensatz für eine intellektuelle Überprüfung zurückzustellen.

Ein ähnlicher Typus ist das bekannte „Stephen King writing as Richard Bachman“ bei Werken, die anfangs unter dem Pseudonym erschienen waren. Nach dessen Lüftung wollte der Verlag vermutlich zur besseren Vermarktung auch den bekannteren Namen auf der Titelseite sehen. Die Bachman-Bücher sind zugleich Beispiele für Werke, die zu unterschiedlichen Zeiten unter unterschiedlichen Namen präsentiert werden.

## 4. Nutzung von Werkclustern

Mit dem für die Bachelorarbeit erstellten Skript wurde jeder Titeldatensatz einzeln betrachtet. Es liegt jedoch nahe, ein entsprechendes Verfahren nicht nur auf der Manifestationsebene aufzusetzen, sondern stattdessen mit Werkclustern bzw. Werkbündeln zu arbeiten. Dies konnte im Rahmen der Bachelorarbeit allerdings nicht tiefergehend untersucht werden, sodass hier nur einige allgemeine Überlegungen angestellt werden können.

In einem solchen Werkcluster oder Werkbündel werden Titeldatensätze zusammengebracht, die Manifestationen desselben Werks sind (unterschiedliche Ausgaben, Auflagen und Übersetzungen). Ein entsprechendes Verfahren, das auf einem Abgleich von wenigen Feldern beruhte, wurde bereits 2010 erfolgreich von Magnus Pfeffer eingesetzt.<sup>20</sup> Auch die von der DNB betriebene Culturegraph-Plattform, welche Metadaten für Datenanalysen und -vernetzungen bereitstellt, besitzt eine entsprechende Funktionalität, die im Vergleich zur Pfeffer'schen Methode deutlich weiter entwickelt wurde.<sup>21</sup> Bei der Erstellung von Werkclustern im Bereich der Pseudonyme wäre natürlich darauf zu achten, dass alle bibliografischen Identitäten einer Person berücksichtigt werden.

Die Nutzung von Werkclustern kam bisher insbesondere im Bereich der Inhaltserschließung zum Einsatz. Doch schon 2013 schlugen Heidrun Wiesenmüller und Magnus Pfeffer eine Anwendung auch bei Verknüpfungen mit Personennormdatensätzen vor.<sup>22</sup> Denn genau wie die Inhaltserschließung, so sind auch geistige Schöpfer\*innen auf der Ebene des Werks anzusiedeln. Tatsächlich wurde

20 Vgl. Pfeffer, Magnus: Using clustering across union catalogues to enrich entries with indexing information, in: Spiliopoulou, Myra; Schmidt-Thieme, Lars; Janning, Ruth (Hg.): Data analysis, machine learning and knowledge discovery. Proceedings of the 36th Annual Conference of the Gesellschaft für Klassifikation e.V. in Hildesheim, Germany. Berlin; Heidelberg 2013, S. 437–445.

21 Vgl. Vorndran, Angela: Hervorholen, was in unseren Daten steckt! Mehrwerte durch Analysen großer Bibliotheksdatenbestände, in: o-bib 5 (4), 2018, S. 166–180. Online: <<https://doi.org/10.5282/o-bib/2018H4S166-180>>, sowie Vorndran, Angela; Grund, Stefan: Metadata sharing. How to transfer metadata information among work cluster members, in: Cataloging & classification quarterly 59 (8), 2021, S. 757–774. Online: <<https://doi.org/10.1080/01639374.2021.1989101>>.

22 Vgl. Wiesenmüller, Heidrun; Pfeffer, Magnus: Abgleichen, anreichern, verknüpfen. Das Clustering-Verfahren – eine neue Möglichkeit für die Analyse und Verbesserung von Katalogdaten, in: BuB, 65 (9), 2013, S. 625–629. Online: <[https://b-u-b.de/fileadmin/archiv/imports/pdf\\_files/2013/bub\\_2013\\_09\\_625\\_629.pdf](https://b-u-b.de/fileadmin/archiv/imports/pdf_files/2013/bub_2013_09_625_629.pdf)>, Stand: 30.07.2023, hier S. 626–628.

auch das Clustering von Werken auf der Culturegraph-Plattform schon dazu genutzt, um fehlende Verknüpfungen zu Personennormdatensätzen nachzuführen.<sup>23</sup> Evtl. könnten die dafür entwickelten Abläufe für unter Pseudonym geschriebene Werke angepasst werden.

Bei Pseudonymfällen kann man im Normalfall davon ausgehen, dass alle Manifestationen eines Werks mit derselben bibliografischen Identität zu verknüpfen sind. Die Vorteile liegen auf der Hand: Fasst man die zusammengehörigen Titeldatensätze zu einem Cluster zusammen, so genügt es, wenn einer der Titeldatensätze die nötigen Informationen enthält, um für alle eine korrekte Zuordnung zum entsprechenden Normdatensatz zu ermöglichen. So wären beispielsweise fehlende Verantwortlichkeitsangaben bei manchen Ausgaben oder einzelne „Ausreißer“, die z.B. durch einen Tippfehler oder eine Fehlerfassung entstanden sind, unproblematisch.

Eine Besonderheit sind diejenigen Fälle, bei denen sich die Verantwortlichkeitsangaben innerhalb des Clusters von einem bestimmten Zeitpunkt an systematisch ändern (was durch einen entsprechenden Algorithmus feststellbar sein müsste). Beim Typ „Stephen King writing as Richard Bachman“ erscheint es sinnvoll, mit der ursprünglichen Zuordnung – hier also dem Pseudonym Richard Bachman – zu verknüpfen. Denkbar sind außerdem Fälle, in denen ein Werk z.B. aus politischen Gründen zunächst unter Pseudonym erschien und ab einem bestimmten Zeitpunkt unter dem wirklichen Namen veröffentlicht wurde. In einem solchen Fall wäre es eher sinnvoll, einheitlich mit dem wirklichen Namen zu verknüpfen. Die Entscheidung sollte in dieser sicher nicht allzu häufigen Situation am besten intellektuell getroffen werden.

Bei den sprachabhängigen Pseudonymen von Karen Blixen (s.o. Kap. 3.4) funktioniert die Werkcluster-Methode nicht; hier müsste man innerhalb eines Werks zusätzlich nach der Sprache clustern. Diese Praxis dürfte aber mindestens sehr selten, wenn nicht gar singular sein. Die beschriebene Methode stößt außerdem an ihre Grenzen, wenn es um Pseudonyme bei Personen geht, die nicht auf der Werk-Ebene anzusiedeln sind (z.B. bei Herausgebenden oder Illustrator\*innen).

## 5. Mögliche Umsetzung

Eine maschinelle Methode zur korrekten Zuordnung von Titeldatensätzen könnte zunächst über den gesamten Datenbestand in einem Verbundkatalog laufen, um die bestehenden Verknüpfungen zu prüfen und wo nötig zu korrigieren. Unter Einbezug von Werkclustern wären dafür folgende Schritte nötig:

1. Selektion aller Titeldatensätze, die an einem gesplitteten Pseudonymdatensatz hängen (d.h. einem Normdatensatz, der mit mindestens einem weiteren Normdatensatz für eine andere bibliografische Identität der Person verknüpft ist)
2. Bildung von Werkclustern

<sup>23</sup> Vgl. Vorndran, Angela; Grund, Stefan: Metadata sharing, S. 769–771, sowie Grund, Stefan; Vorndran, Angela: Personen in GND und Titeldaten. Vorschläge, Anreicherungen und Verknüpfungen, Folienpräsentation vom KIM-Workshop, 09.05.2023, <[https://wiki.dnb.de/download/attachments/266466762/2023\\_05\\_09\\_Vorndran\\_Angela\\_Grund\\_Stefan\\_PersonenInGNDundTiteldaten.pdf](https://wiki.dnb.de/download/attachments/266466762/2023_05_09_Vorndran_Angela_Grund_Stefan_PersonenInGNDundTiteldaten.pdf)>, Stand: 18.09.2023.

3. Zuordnung der Cluster zum korrekten Normdatensatz auf der Basis von Analyse und Abgleich der Verantwortlichkeitsangaben und Titelfelder
4. Unter bestimmten Bedingungen keine maschinelle Zuordnung, sondern Vorlage für eine intellektuelle Entscheidung
5. Ggf. Umhängen der Titeldatensätze aus dem Cluster, die noch nicht mit dem korrekten Normdatensatz verknüpft sind

Schritt 2 sollte möglichst auf der Culturegraph-Plattform geschehen, um die Ergebnisse leicht nachnutzbar zu machen. Schritt 5 sollte möglichst maschinell geschehen.

In bestimmten Abständen könnte dieselbe Routine erneut durchgeführt werden – dann jeweils für die Titeldatensätze, die mit seit dem letzten Durchgang neu gesplitteten Pseudonymdatensätzen verknüpft sind. Alternativ wäre es auch denkbar, ein Verfahren zu entwickeln, das jeweils nach dem Aufsplitten eines Pseudonymdatensatzes angestoßen wird und sozusagen „on the fly“ für die korrekten Verknüpfungen sorgt.

Grundsätzlich wäre eine Umsetzung auch ohne die Nutzung von Werkclustern möglich; hierzu müsste das in der Bachelorarbeit erarbeitete Skript entsprechend erweitert werden.<sup>24</sup>

Zu überlegen wäre außerdem, ob nicht auch das Aufsplitten von Pseudonymnormdatensätzen maschinell oder zumindest maschinell unterstützt erfolgen könnte. Denn bisher wurde nur ein Teil der Pseudonymdatensätze bearbeitet. Mit Stand Juli 2023 finden sich in der GND 29.840 Normdatensätze mit dem Entitätencode „pip“. Dieser wird sowohl in ungesplitteten Pseudonymdatensätzen verwendet als auch bei Datensätzen für Pseudonymidentitäten in Split-Fällen. Nur in 6.452 davon ist der Benutzungshinweis zu weiteren bibliografischen Identitäten erfasst; hier handelt es sich also um Split-Fälle. Von den verbleibenden 23.388 Fällen sind noch diejenigen abziehen, die nicht aufgespalten werden müssen, da die jeweilige Person nur einen einzigen Namen verwendet hat. Aber auch dann verbleibt noch eine beträchtliche Menge unbearbeiteter Pseudonymfälle.

Zwar wird es nur in einem Teil davon zur Veröffentlichung neuer Ausgaben kommen; dennoch könnte eine maschinell gestützte systematische Bearbeitung effizienter sein als das intellektuelle Splitten im Bedarfsfall. Bei der Identifikation der zu splittenden Normdatensätze würde wiederum ein Parsing helfen: Die Verwendung unterschiedlicher Namen in den Verantwortlichkeitsangaben weist darauf hin, dass mehrere bibliografische Identitäten vorliegen.

## 6. Fazit

Nicht korrigierte Titeldatenverknüpfungen nach dem Split von Pseudonymdatensätzen stellen ein zwar quantitativ begrenztes, aber dennoch reales Problem für die Recherche dar. Mit fortschreitender

---

<sup>24</sup> Für Vorschläge zu konkreten Workflows vgl. auch Schilling, Clara: Die Konsequenzen des RDA-Umstiegs für die Recherche nach Pseudonymen, 2022, S. 46–48.

Aufarbeitung der Normdatensätze in der GND wird sich außerdem die Zahl der Titel, die bei einer entsprechenden Suche nicht gefunden werden, tendenziell erhöhen.

Eine rechnerseitige Lösung, wie sie ursprünglich diskutiert worden war, ist nicht wirklich erfolgreich gewesen – sicher auch deshalb, weil die Indexierung nicht nur in den Verbänden, sondern auch in allen lokalen Katalogsystemen entsprechend angepasst werden müsste. Auch hebt dieser Weg die von RDA intendierte Möglichkeit aus, gezielt nach Titeln einzelner bibliografischen Identitäten zu recherchieren.

Es bleibt daher nur die Korrektur der Metadaten selbst, die aber bei einer händischen Bearbeitung schon aus Kapazitätsgründen nicht leistbar ist. Auch würden bei einer rein intellektuellen Bearbeitung Aufwand und Nutzen in keinem guten Verhältnis zueinander stehen.

Sinnvoll ist deshalb eine maschinelle Lösung. Dass eine solche möglich ist und zu guten Ergebnissen führen würde, wurde in der Bachelorarbeit von Clara Schilling demonstriert. Schon mit einem relativ einfachen Python-Skript lässt sich ein großer Teil der Fälle lösen. Mit einigen Erweiterungen und optimalerweise dem Einbezug von Werkclustern könnte eine maschinelle Lösung implementiert werden, die die korrekten Normdatensätze sehr zuverlässig zuordnen könnte. Mit der Nutzung der Culturegraph-Plattform wäre es außerdem möglich, entsprechende Analyseergebnisse sehr leicht zwischen Verbänden und Institutionen auszutauschen. Wünschenswert wäre in diesem Zusammenhang außerdem die Entwicklung einer maschinellen oder halb-maschinellen Methode für das Aufsplitten der Normdatensätze.

## Literaturverzeichnis:

- GND-Erfassungshilfe „Pseudonyme“ (EH-P06), <<https://wiki.dnb.de/download/attachments/90411361/EH-P-06.pdf>>, Stand: 30.07.2023.
- Grund, Stefan; Vorndran, Angela: Personen in GND und Titeldaten. Vorschläge, Anreicherungen und Verknüpfungen, Folienpräsentation vom KIM-Workshop, 09.05.2023, <[https://wiki.dnb.de/download/attachments/266466762/2023\\_05\\_09\\_Vorndran\\_Angela\\_Grund\\_Stefan\\_PersonenInGNDundTiteldaten.pdf](https://wiki.dnb.de/download/attachments/266466762/2023_05_09_Vorndran_Angela_Grund_Stefan_PersonenInGNDundTiteldaten.pdf)>, Stand: 18.09.2023.
- Pfeffer, Magnus: Using clustering across union catalogues to enrich entries with indexing information, in: Spiliopoulou, Myra; Schmidt-Thieme, Lars; Janning, Ruth (Hg.): Data analysis, machine learning and knowledge discovery. Proceedings of the 36th Annual Conference of the Gesellschaft für Klassifikation e.V. in Hildesheim, Germany. Berlin; Heidelberg 2013, S. 437–445.
- Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken (RAK-WB), Stand: April 2006, Leipzig 20072. Online: <<https://nbn-resolving.org/urn:nbn:de:101-2007072711>>.
- Schilling, Clara: Die Konsequenzen des RDA-Umstiegs für die Recherche nach Pseudonymen – aktuelle Situation, Probleme, Lösungsansätze, Bachelorarbeit, Hochschule der Medien, Stuttgart 2022. Online: <<https://nbn-resolving.org/urn:nbn:de:bsz:900-opus4-67622>>.

- Vorndran, Angela: Hervorholen, was in unseren Daten steckt! Mehrwerte durch Analysen großer Bibliotheksdatenbestände, in: o-bib 5 (4), 2018, S. 166–180. Online: <<https://doi.org/10.5282/o-bib/2018H4S166-180>>.
- Vorndran, Angela; Grund, Stefan: Metadata sharing. How to transfer metadata information among work cluster members, in: Cataloging & classification quarterly 59 (8), 2021, S. 757–774. Online: <<https://doi.org/10.1080/01639374.2021.1989101>>.
- Wiesenmüller, Heidrun: Der RDA-Umstieg in Deutschland – Herausforderungen für das Metadatenmanagement, in: o-bib 2 (2), 2015, S. 43–60. Online: <<https://doi.org/10.5282/o-bib/2015H2S43-60>>.
- Wiesenmüller, Heidrun; Horny, Silke: Basiswissen RDA. Eine Einführung für deutschsprachige Anwender, Berlin; Boston 2017<sup>2</sup>.
- Wiesenmüller, Heidrun; Pfeffer, Magnus: Abgleichen, anreichern, verknüpfen. Das Clustering-Verfahren – eine neue Möglichkeit für die Analyse und Verbesserung von Katalogdaten, in: BuB, 65 (9), 2013, S. 625–629. Online: <[https://b-u-b.de/fileadmin/archiv/imports/pdf\\_files/2013/bub\\_2013\\_09\\_625\\_629.pdf](https://b-u-b.de/fileadmin/archiv/imports/pdf_files/2013/bub_2013_09_625_629.pdf)>, Stand: 30.07.2023.