
Der RDA-Umstieg in Deutschland Herausforderungen für das Metadatenmanagement

Heidrun Wiesenmüller, Hochschule der Medien Stuttgart

Zusammenfassung:

Der Umstieg der deutschsprachigen Länder auf den neuen Katalogisierungsstandard RDA wirft die Frage auf, wie mit den gemäß RAK erschlossenen Altdaten umgegangen werden soll. Der vorliegende Beitrag untersucht aus datentechnischer Sicht, ob bzw. inwieweit RAK-Katalogisate (Titel- und Normdatensätze) mit maschinellen Mitteln auf RDA angehoben werden könnten. Beispielhaft betrachtet werden dabei u.a. die zu übertragenden Elemente in der bibliografischen Beschreibung, der Inhaltstyp sowie die Aufspaltung in getrennte bibliografische Identitäten im Fall von Pseudonymen. Es wird gezeigt, dass RDA-Upgrades möglich, aber hochkomplex sind.

Summary:

The switch of the German-speaking countries to the new cataloguing standard RDA raises the question how the legacy data should be handled. Based on an analysis of the available data, the paper discusses whether and to what extent records (bibliographic and authority), which were created according to the former cataloguing code RAK, could be upgraded to RDA. The examples discussed include the transcribed information in the bibliographic description, the content type, and the treatment of separate bibliographic identities in the case of pseudonyms. The study shows that upgrades to RDA are possible, but highly complex.

Zitierfähiger Link (DOI): <http://dx.doi.org/10.5282/o-bib/2015H2S43-60>

Autorenidentifikation: Wiesenmüller, Heidrun: GND 122087801

1. Vorüberlegungen

Im Laufe des vierten Quartals 2015 werden die Bibliotheksverbände in Deutschland und Österreich, die Zeitschriftendatenbank (ZDB), Teile der deutschsprachigen Schweiz sowie große Bibliotheken wie die Deutsche Nationalbibliothek auf den neuen internationalen Katalogisierungsstandard „Resource Description and Access“ (RDA) umsteigen.¹ Das bisherige deutsche Regelwerk, die „Regeln für die alphabetische Katalogisierung“ (RAK), wird dadurch abgelöst.² Dieser Umstieg bedeutet auch, dass Millionen von bibliografischen Datensätzen in unseren Katalogen mit einem Schlag zu „Altdaten“

1 Beim vorliegenden Beitrag handelt es sich um die erweiterte und aktualisierte Fassung eines Vortrags, der am 5. Dezember 2014 auf dem Symposium „Forschung für die Praxis – Perspektiven für Bibliotheks- und Informationsmanagement“ an der Hochschule der Medien in Stuttgart gehalten wurde. Vortragsfolien unter https://www.hdm-stuttgart.de/bi/symposium/skripte/Wiesenmueller_RDA-Umstieg_Forum1_14-12-05.pdf (10.05.2015). Vgl. auch den Veranstaltungsbericht: Vonhof, Cornelia; Stang, Richard; Wiesenmüller, Heidrun: Forschung für die Praxis – Perspektiven für Bibliotheks- und Informationsmanagement. In: o-bib 2 (2015), H. 1, S. 68-74. <http://dx.doi.org/10.5282/o-bib/2015H1S68-74>.

2 Für die deutschsprachige Schweiz ist die Situation etwas anders gelagert; dort wurde bisher AACR2 angewendet.

werden. Allein im Katalog des Südwestdeutschen Bibliotheksverbunds (SWB), der im Folgenden beispielhaft betrachtet wird, handelt es sich um etwa 21 Millionen Titeldatensätze.³

Schon seit langem ist klar, dass es keinen Katalogbruch geben kann; vielmehr werden alte und neue Daten im selben Datenpool koexistieren müssen. Allerdings ist es nicht immer ohne weiteres möglich, einfach einen Schnitt zu machen – also bis zu einem bestimmten Zeitpunkt nach RAK zu katalogisieren und danach nach RDA. So wird ein großer Teil der existierenden Normdaten regelmäßig wieder verwendet und mit neuen Datensätzen in Beziehung gesetzt. In diesen Fällen kommt man um eine Umarbeitung kaum herum. Ein weiterer Problembereich sind Ressourcen, die über einen längeren Zeitraum laufen; neben fortlaufenden Ressourcen wie Zeitschriften betrifft dies auch manche mehrteilige Monografien. Hier sind die Katalogisate nicht statisch, sondern müssen regelmäßig wieder angefasst werden, wenn z.B. eine Zeitschrift den Verlag wechselt oder wenn ein neuer Band dazu kommt. Viele solcher Titelaufnahmen wurden unter RAK begonnen, müssen aber in der RDA-Ära weitergeführt werden. Und selbst bei einbändigen Monografien sind Kontinuitäten zu beachten, etwa wenn alte und neue Ausgaben desselben Werks aufgrund des Regelwerksumstiegs unterschiedlich behandelt werden (vgl. Kap. 2.3).

Angesichts der begrenzten Ressourcen in den Bibliotheken sind intellektuell durchgeführte Anpassungen in größerem Umfang wenig realistisch. Umso wichtiger ist die Frage, ob bzw. welche maschinellen Änderungen an den Altdaten möglich sein werden. Genauere Untersuchungen dazu wurden bisher jedoch zurückgestellt, da die Vorbereitung des Umstiegs und die technische Implementierung von RDA alle verfügbaren Kräfte binden. Ein grundsätzliches Hindernis für maschinelle Änderungen lässt sich freilich auch ohne detaillierte Forschungen identifizieren: Wo nach RAK vorzugsweise rein formale Kriterien angelegt wurden, fordert RDA von den Katalogisierenden oftmals inhaltliche Entscheidungen.⁴ Diese können nur intellektuell getroffen werden und sind daher einer maschinellen Umsetzung weitgehend entzogen.

Bei der Bewertung der durch den Umstieg entstehenden Situation sollte man sich klar machen, dass unsere Kataloge natürlich auch bisher nicht völlig einheitlich waren. Beispielsweise kann man in älteren Katalogisaten manchmal noch frühere Regelwerksstände erkennen,⁵ und Fremddaten angloamerikanischer Herkunft werden oft nur unvollkommen angepasst: Umfangs- und Illustrationsangabe, Fußnoten u.ä. erscheinen dann auf Englisch anstatt auf Deutsch. Man darf mit gutem Recht in Frage stellen, ob Benutzerinnen und Benutzer solche Inkonsistenzen überhaupt bewusst wahrnehmen – und wenn ja, ob sie sich tatsächlich daran stören.

3 Vgl. die Bestandsdaten in den Statistiken des SWB:

<https://wiki.bsz-bw.de/doku.php?id=v-team:katalogisierung:statistik:start> (18.04.2015).

4 Ein Beispiel dafür ist die Entscheidung, wann ein körperschaftlicher Urheber als geistiger Schöpfer gilt (d.h. nach alter Terminologie: wann er die Haupteintragung erhält). Gemäß RAK musste dafür nur geprüft werden, ob die Körperschaft im Sachtitel genannt oder dazu zu ergänzen war. Gemäß RDA muss hingegen der Inhalt der Ressource betrachtet und bewertet werden. Vgl. dazu Wiesenmüller, Heidrun; Horny, Silke: Basiswissen RDA. Eine Einführung für deutschsprachige Anwender, Berlin/Boston: De Gruyter Saur, 2015, S. 134-137.

5 So wurden beispielsweise in der Anfangszeit von RAK auch Wörter in Titelzusätzen abgekürzt. Mit dem Wechsel von Zettelkatalogen zu elektronischen Katalogen war dies nicht mehr praktikabel, da nun auch nach Stichwörtern aus dem Titelzusatz gesucht werden konnte.

Mit Blick auf mögliche maschinelle Anpassungen ist außerdem ganz grundsätzlich zu klären, inwieweit man bereit ist, hybride Datensätze zu akzeptieren: Kann man bei den Altdaten eine Vermischung von zwei Regelwerken in Kauf nehmen? Oder ist es besser, bei Datensätzen zu bleiben, die zwar nach veralteten Regeln erstellt wurden, jedoch in sich stringent sind? In jedem Fall muss datentechnisch gekennzeichnet werden, ob es sich um einen reinen RAK-Datensatz, um einen reinen RDA-Datensatz oder um einen Datensatz handelt, der maschinell in einigen Punkten auf RDA angehoben wurde.

Ebenfalls gut überlegt werden muss, an welchen Stellen maschinelle Upgrades vom alten auf das neue Regelwerk zum Einsatz kommen sollen. Denn es versteht sich, dass sie nicht an allen Stellen von gleicher Bedeutung sind. Wichtig sind sie vor allem dort, wo sich Auswirkungen auf die Recherche ergeben – beispielsweise bei Informationen, die in Facetten für einen Drill-down angeboten werden sollen (vgl. Kap. 2.2). Zu bedenken ist außerdem, dass es immer eine gewisse Fehlerquote geben wird, die gegen den zu erwartenden Nutzen abzuwägen ist.

Im Folgenden werden beispielhaft einige Bereiche der Katalogisierung betrachtet und daraufhin geprüft, ob bzw. inwieweit eine maschinelle Umsetzung von RAK nach RDA möglich und sinnvoll erscheint. Dies kann freilich nur eine allererste Annäherung an das Thema sein, der intensivere Untersuchungen folgen müssen.

2. Titeldatensätze

Titeldatensätze werden auch unter RDA nicht grundsätzlich anders aussehen als bisher. Zwar basiert RDA auf dem Modell der „Functional Requirements for Bibliographic Records“ (FRBR),⁶ doch wird es auch künftig keine durchgängige Aufspaltung in Datensätze für Werke, Expressionen und Manifestationen geben. Der Normalfall wird die sogenannte *zusammengesetzte Beschreibung* sein, welche Informationen aus den Ebenen Werk, Expression und Manifestation miteinander kombiniert.⁷ Nichtsdestoweniger gibt es zwischen altem und neuem Regelwerk zahlreiche Unterschiede – beispielsweise Änderungen im Bereich der bibliografischen Beschreibung (vgl. Kap. 2.1) und bei den Beziehungen (frühere Terminologie: Eintragungen) zu Personen (vgl. Kap. 2.3) oder neu eingeführte Informationselemente (vgl. Kap. 2.2).

2.1 Zu übertragende Elemente in der bibliografischen Beschreibung

Bei den Informationen, die direkt aus der Informationsquelle übertragen, d.h. abgeschrieben werden, lässt sich die Grundregel von RDA zusammenfassen mit „Take what you see!“ (Nimm, was du siehst!). Dies entspricht einem Grundprinzip von RDA: „Die Daten, die eine Ressource beschreiben, sollten widerspiegeln, wie sich die Ressource selbst darstellt.“ (RDA 0.4.3.4).⁸ Es wird also – anders als unter RAK – nichts mehr geändert, abgekürzt oder weggelassen. Steht in der Informationsquelle als Ausgabebezeichnung „Zweite, vollständig überarbeitete Auflage“, so wird dies nach RDA genau so übernommen.⁹ Nach RAK hätte man hingegen „2., vollst. überarb. Aufl.“ daraus gemacht.

6 Zu FRBR vgl. Wiesenmüller; Horny: Basiswissen RDA (wie Anm. 4), S. 17-22.

7 Vgl. ebd., S. 120-123.

8 Alle RDA-Stellen werden zitiert nach der deutschen Fassung des RDA Toolkit mit Stand April 2015. Zugang zum RDA Toolkit: <http://www.rdatoolkit.org/> (10.05.2015).

9 Vgl. Wiesenmüller; Horny: Basiswissen RDA (wie Anm. 4), S. 35.

Im Bereich der Ausgabebezeichnungen besteht eine denkbare Umsetzungsregel darin, zumindest die häufig vorkommenden Abkürzungen nach einer festen Liste aufzulösen. Im Beispiel könnte die Angabe dann korrigiert werden zu „2., vollständig überarbeitete Auflage“. Das Wort „Zweite“ wäre allerdings nicht mehr zu rekonstruieren – schließlich kann man der RAK-Aufnahme nicht ansehen, ob im Original eine Ziffer oder ein Zahlwort stand. In solchen Fällen stößt die Methode an ihre Grenzen. Auch würde es immer dann zu einem Fehler kommen, wenn eine Abkürzung wie „Aufl.“ schon in der Ressource stand – denn dann müsste sie nach RDA genau in dieser Form übertragen werden. Nicht auszuschließen sind außerdem Probleme bei Ausgabebezeichnungen, die vom gängigen Schema abweichen. So wäre etwa die Angabe „2. Auflage, ergänzt und bearbeitet“ nach RAK als „2. Aufl., erg. und bearb.“ wiedergegeben worden. Bei der Umsetzung nach der beschriebenen Methode würde sie zu „2. Auflage, ergänzte und bearbeitete“ mutieren.

In Verantwortlichkeitsangaben (frühere Terminologie: Verfasserangabe) wären die Umsetzungsprobleme weitaus größer als bei Ausgabebezeichnungen: Zum einen findet man hier nicht selten Abkürzungen bereits in der Ressource vor, z.B.: „Christoph Lorey / John L. Plews / Caroline L. Rieger (Hrsg.)“. Eine pauschale Auflösung von Abkürzungen wie „Hrsg.“ führt deshalb häufig nicht zu einem RDA-gerechten Ergebnis. Zum anderen wurden nach RAK manche Informationen überhaupt nicht erfasst. Im obigen Beispiel wäre nur der erste Herausgeber übernommen worden: „Christoph Lorey ... (Hrsg.)“. Wollte man die beiden anderen Herausgeber nachträglich ergänzen, so wäre dies nur durch einen Abgleich mit angloamerikanischen Daten möglich. Da gemäß dem früheren Regelwerk „Anglo-American Cataloguing Rules“ (AACR2) bis zu drei Herausgeber übernommen wurden, könnte man beim angeführten Beispiel die fehlenden Namen aus der entsprechenden Titelaufnahme etwa bei der Library of Congress ergänzen. Bei mehr als drei Personen in derselben Verantwortlichkeitsangabe wurde allerdings auch nach AACR2 gekürzt. Diese Methode wäre also nicht nur aufwendig, sondern würde auch nicht in jedem Fall funktionieren.

Insgesamt erscheint es fraglich, ob Aufwand und Nutzen bei derartigen Aktionen noch in einem sinnvollen Verhältnis zueinander stehen. Als Fazit lässt sich deshalb festhalten, dass maschinelle Änderungen bei den zu übertragenden Elementen zwar in gewissen Grenzen möglich sind, aber vermutlich nicht sehr sinnvoll wären.

2.2 Inhaltstyp

Als Beispiel für ein neu eingeführtes Element in RDA kann der Inhaltstyp (content type, RDA 6.9) dienen. Er muss zwingend erfasst werden (sogenanntes „Kernelement“) und gibt den Charakter der vorliegenden Expression an: Wie wird die Ressource wahrgenommen – als Text, als Bild, als Computerprogramm etc.?¹⁰ Dafür gibt es eine Liste von 23 normierten Begriffen (RDA 6.9.1.3). Darunter sind etwa *Text* (Beispiele: Buch, PDF-Dokument), *unbewegtes Bild* (Beispiele: Druckgrafik, Bildband), *zweidimensionales bewegtes Bild* (Beispiele: Film, Videospiel), *gesprochenes Wort* (Beispiel: Hörbuch) und *aufgeführte Musik* (Beispiel: Musik-CD). Es ist naheliegend, den Inhaltstyp im Katalog in Form einer Facette anzubieten, um Treffermengen nach diesem Kriterium einschränken zu können. Dafür

¹⁰ Vgl. ebd., S. 77f.

sollten jedoch auch die Altdaten durchgängig über dieses Merkmal verfügen. Eine maschinelle Ergänzung des Inhaltstyps ist also in jedem Fall anzustreben.¹¹

```

0500 Aau
0600 sf
1100 2014$n[2014]
1130 druck
1140 kart
1500 ger
1700 XA-DE
2000 3-89591-108-9$fin Hülle : EUR 4.90
2113 OCLC: 894732611
2201 9783895911088
2240 BSZ: 415425085
4000 Naturpark Kyffhäuser, Sondershausen und Umgebung$dAusflüge zwischen Nordhausen, der Goldenen
Aue, Kelbra, Tilleda und Bad Frankenhausen; Radtourenvorschläge mit Entfernungsangaben; viele Tipps zu
Sehenswürdigkeiten
4020 2. Aufl., Laufzeit der Kt. bis 2020, neu
4026 1:35 000#1:35000#
4030 Borsdorf bei Leipzig$nBarthel
4060 1 Kt.
4061 mehrfarb.
4062 81 x 59 cm, gefaltet
4170 Schöne Heimat ; 108 : Radwander- und Wanderkarte
4180 #108,2014#!313062021!Schöne Heimat ; 108 : Radwander- und Wanderkarte
4201 Mit Abb. und Texten. - Mit Schummerung und Höhenlinien
    
```

Abb. 1: Beispiel für ein Landkarten-Katalogisat im SWB (mit kart in 1140)

Als Test-Beispiel wurde der Inhaltstyp *kartografisches Bild* ausgewählt, der z.B. für Landkarten und Atlanten zu vergeben ist. Die Ausgangsthese war, dass dieser Inhaltstyp aus den vorhandenen Daten im SWB leicht und zuverlässig zu generieren sein müsste, indem man das Feld 1140 (Veröffentlichungsart und Inhalt) im Pica3-Format auswertet. Für dieses Feld ist u.a. der Code *kart* definiert, der im SWB knapp 250.000mal vergeben wurde. Er ist zu verwenden für „alle Karten und Pläne; unabhängig von der physischen Form. Als Karten gelten auch kartenverwandte Darstellungen wie Atlanten, Blockbilder, Diagrammkarten, Luftbilder, Luftbildkarten, Panoramen, Profile, Satellitenbilder, Vogelschaubilder und dreidimensionale kartographische Materialien wie Globen, Globensegmente und Reliefs.“¹² Diese Definition ist etwas breiter als die des Inhaltstyps *kartografisches Bild*: Es fallen auch Globen und Reliefs darunter, die nach RDA einen anderen Inhaltstyp erhalten, nämlich *kartografische dreidimensionale Form*. Eine solche Unschärfe kann jedoch in Kauf genommen werden, zumal die Zahl der Globen und Reliefs im Datenbestand überschaubar sein dürfte.¹³ In der Theorie sollte also

11 Die maschinelle „Generierung von Inhaltstyp, Medientyp und Datenträgertyp aus vorhandenen Angaben“ wurde auch auf einem Workshop für Systemanbieter am 23.10.2014 in der Deutschen Nationalbibliothek von Renate Polak-Bennemann als „Kernpunkt“ bei der rückwirkenden Anpassung vorhandener Daten bezeichnet. Vgl. ihren Vortrag „Technische Veränderungen“, Folie 10. Präsentationsfolien: https://wiki.dnb.de/download/attachments/99090660/03_Systemanbieterworkshop-Einf%C3%BChrung-technik.pdf (10.05.2015).

12 Feldbeschreibung zu 1140, URL: <http://swbtools.bsz-bw.de/cgi-bin/help.pl?cmd=kat&val=1140®elwerk=RAK> (10.05.2015).

13 Weitere mögliche Unschärfen könnten sich bei digitalen Materialien ergeben, denn in RDA gibt es auch noch den Inhaltstyp *kartografisches bewegtes Bild* und *kartografischer Datensatz*.

ein automatisches Zuspieldes des Inhaltstyps *kartografisches Bild* an alle Datensätze mit dem Code *kart* in 1140 zu einem guten Ergebnis führen und das Problem weitgehend lösen.

Stichproben zeigen, dass *kart* bei Landkarten, welche üblicherweise von Spezialisten erschlossen werden, sehr zuverlässig vergeben wird (Abb. 1). Bei den Atlanten fehlt die Kennzeichnung jedoch meistens (Abb. 2). Katalogisierenden, die nur unregelmäßig mit Kartenmaterial umgehen, ist die Existenz des Codes offenbar nicht ausreichend präsent. Will man auch bei Atlanten den Inhaltstyp zuspielden, müssen folglich weitere Aspekte mit einbezogen werden. Das Titelstichwort „Atlas“ scheidet als Kriterium aus, da es auch in vielen nicht-kartografischen Zusammenhängen auftritt (z.B. medizinische Atlanten, „Wein-Atlas“). Es würde also zu sehr vielen falschen Zuordnungen führen.

Denkbar wäre eine Auswertung der Illustrationsangabe in Feld 4061: Der Inhaltstyp *kartografisches Bild* könnte bei allen Datensätzen ergänzt werden, bei denen an dieser Stelle „zahlr. Kt.“ (zahlreiche Karten) oder „überw. Kt.“ (überwiegend Karten) steht. Aber auch dieses Verfahren führt nicht immer zum Erfolg: Denn bei vielen Atlanten wurde nur „Kt.“ erfasst, was eigentlich für *einige* Karten steht – ein Beispiel dafür zeigt Abb. 2. Würde man *kartografisches Bild* aber auch bei allen Datensätzen zuspielden, die „Kt.“ in der Illustrationsangabe haben, so würde dies zu enormen Ballast führen: Bei einer Einschränkung auf diesen Inhaltstyp würden dann nämlich nicht nur Kartenwerke ausgegeben werden, sondern auch alle Bücher, die nur einige wenige Karten zur Illustration des Texts enthalten.

```
0500 Aau
1100 2010
1130 druck
1500 eng
2000 978-0-00-734583-0$f(hbk.) £90.00
2000 0-00-734583-6$f(hbk.) £90.00
2030 015556238
2113 OCLC: 695933010
2201 9780007345830
2240 BSZ: 330444859
3260 Atlas of Britain
4000 The @Times - Atlas of Britain$d[National atlas of England, Scotland, Wales and Northern Ireland]
4020 1. ed.
4030 London$nTimes Books
4060 448 S.
4061 überw. Ill., graph. Darst., Kt.
4062 37 cm
4212 The @Times atlas of Britain
4700 5090: ddsu/sred; 5550: FRUB11
5010 912.41
5090 1205783139!RN 70229 [Geographie # Regionale Geographie # Europa # Westeuropa # Britische Inseln #
Sachschlüssel zur Geographie # Allgemeine Geographie # Mathematische Geographie und Physiogeographie #
Mathematische Geographie und Kartographie # Kartographie # Teilgebiete und Einzelfragen # Thematische
Karten]
5540 |u|Great Britain; Maps
5540 |u|Great Britain; Civilization
5540 |u|Great Britain; Geography
5540 |u|Great Britain; History
5550 1208942076!Großbritannien
5550 |f|Atlas
```

Abb. 2: Beispiel für ein Atlas-Katalogisat im SWB (ohne *kart* in 1140)

So kommt man nicht umhin, außer dem Code *kart* und der Illustrationsangabe noch ein weiteres Kriterium mit einzubeziehen. Dies könnten Angaben aus der Sacherschließung sein, also geeignete Forms Schlagwörter aus RSWK und LCSH oder einschlägige Systemstellen bzw. Schlüssel in verbreiteten Klassifikationen wie DDC oder RVK. Beim Atlas aus Abb. 2 sind entsprechende Merkmale in den Feldern 5090 (RVK), 5540 (British Library Subject Headings) und 5550 (RSWK) vorhanden. Auch hier ist freilich Vorsicht geboten: Beispielsweise wird das Forms Schlagwort *f. Atlas* der RSWK auch „für Abbildungswerke in der Human- u. Tiermedizin u. der Phytopathologie“¹⁴ vergeben. Solche Fälle müssten über eine Fachzuordnung ausgefiltert werden. Aber selbst bei der aufwendigen Kombination von drei Aspekten – Code in Feld 1140, Illustrationsangabe und Informationen aus der Sacherschließung – dürfte am Ende ein Rest von Titeldatensätzen bleiben, die aus unterschiedlichen Gründen durch das Raster fallen (weil z.B. keine Sacherschließung vorhanden ist) und die deshalb nicht mit dem Inhaltstyp *kartografisches Bild* versehen werden können.

Natürlich müssten entsprechende Verfahren auch für alle anderen Inhaltstypen entwickelt werden. Aber schon das eine Beispiel zeigt, wie komplex maschinelle Upgrades sind. Für die Konzeption der Umsetzungsmethoden sind nicht nur gute Kenntnisse der Regelwerke und Formate nötig, sondern auch umfangreiche und genaue Datenanalysen. Alle Methoden müssen außerdem vorab intensiv getestet und die Ergebnisse anhand einer ausreichend großen Stichprobe geprüft werden. Deutlich wird auch, dass oft mehrere Umsetzungsschritte erforderlich sind, um ein einigermaßen befriedigendes Ergebnis zu erhalten. Die Migration wird überdies durch Inkonsistenzen, Unvollkommenheiten und Katalogisierungsfehler in den Ausgangsdaten erschwert.

2.3 Beziehungen zu Personen

Bei Beziehungen zu Personen (frühere Terminologie: Eintragungen unter Personen) bringt RDA zahlreiche Änderungen mit sich. Vieles, was nach RAK ein „Sachtitelwerk“ war, wird unter RDA zu einem „Verfasserwerk“ (wobei dieser Begriff in der Terminologie von RDA nicht verwendet wird) – beispielsweise Bildbände oder im Team erarbeitete Werke von mehr als drei Verfassern. RDA kennt auch keine Sonderregeln für Nichtbuchmaterialien mehr. So hat ein Roman immer denselben geistigen Schöpfer – egal, ob er als gedrucktes Buch oder als Hörbuch vorliegt. Gemäß dem früheren Regelwerk wurde jedoch bei der Druckausgabe die Haupteintragung unter dem Verfasser gemacht und bei der Hörbuchausgabe unter dem Sachtitel.

Ändert sich nur die Wertigkeit einer Beziehung – wenn z.B. eine Person unter RAK eine Nebeneintragung erhielt, unter RDA aber erster geistiger Schöpfer ist (frühere Terminologie: die Haupteintragung bekommt) – so ist dies für die Recherche an sich unproblematisch. Eine Schwierigkeit ergibt sich jedoch dadurch, dass RDA Werke über ihren geistigen Schöpfer definiert: Der sogenannte normierte Sucheinstieg für ein Werk besteht aus dem ersten geistigen Schöpfer und dem bevorzugten Titel des Werks.¹⁵ Dies kann dazu führen, dass nach RAK katalogisierte ältere Ausgaben gemäß der Logik von RDA nicht zum selben Werk gehören wie jüngere, nach RDA katalogisierte Ausgaben.

14 Verwendungshinweis im GND-Datensatz „Atlas“, URL: <http://d-nb.info/gnd/4143303-8> (10.05.2015).

15 Vgl. Wiesenmüller; Horny: Basiswissen RDA (wie Anm. 4), S. 79f.

Beispielsweise ordnet RDA alle Ausgaben von Agatha Christies „Mord im Pfarrhaus“ – egal, ob sie als Buch oder Hörbuch vorliegen – dem Werk mit dem normierten Sucheinstieg „Christie, Agatha, 1890-1976. The murder at the vicarage“ zu. Die nach RAK katalogisierten Hörbücher würden jedoch aus dieser Gruppe herausfallen, da Agatha Christie bei ihnen nur eine Nebeneintragung hatte und deshalb nicht als geistige Schöpferin im Verständnis von RDA gilt. Datentechnisch liegt der Unterschied nur in einem einzigen Feld: Einmal steht Agatha Christie im Feld für den ersten geistigen Schöpfer (in Pica: 3000, entspricht in MARC: 100), einmal im Feld für weitere Personen (in Pica: 3010; entspricht in MARC: 700).

Für die Benutzerinnen und Benutzer sollte es natürlich in jedem Fall möglich sein, alle Ausgaben eines Werks zu recherchieren – egal, ob diese nach RAK oder nach RDA katalogisiert wurden. Dies lässt sich beispielsweise über ein Werk-Clustering nach der Methode von Magnus Pfeffer erreichen.¹⁶ Titeldatensätze, bei denen der Werktitel und mindestens eine Person (egal, in welchem Feld) übereinstimmen, werden demselben Werk zugeordnet. Der Mechanismus ermittelt deshalb sowohl die gedruckten Ausgaben als auch die Hörbuch-Versionen von „Mord im Pfarrhaus“ korrekt als Manifestationen desselben Werks. Innerhalb eines solchen Werk-Clusters kann man nun bei allen Datensätzen das Feld für den ersten geistigen Schöpfer angleichen. Bei den Hörbüchern wird also die Beziehung zu Agatha Christie vom Pica-Feld 3010 in Feld 3000 verschoben.

Alternativ könnte man das Feld mit der Beziehung zu Agatha Christie unverändert lassen, müsste dann jedoch die Information, zu welchem Werk eine Ausgabe gehört, an anderer Stelle in den Daten verankern. Eine naheliegende Möglichkeit dafür wäre die Verknüpfung aller Datensätze aus dem Cluster mit einem Normdatensatz für das Werk, welcher auch maschinell erstellt werden könnte. Die dritte Möglichkeit besteht darin, überhaupt nicht in die bestehenden Daten einzugreifen, sondern stattdessen das Werk-Clustering bei Bedarf „on the fly“ ablaufen zu lassen. Inzwischen bieten eine ganze Reihe von Katalogsystemen die Möglichkeit, auf Knopfdruck eine Liste anderer Ausgaben desselben Werks anzuzeigen – beispielsweise das Primo-System von Exlibris.¹⁷

Trotz der deutlichen Regelwerksunterschiede zwischen RAK und RDA stellt also die Zusammenführung aller Ausgaben eines Werks in einer modernen technischen Umgebung keine unüberwindliche Schwierigkeit dar. Upgrades der Altdaten in der oben beschriebenen Form sind möglich, aber nicht unbedingt zwingend, da das Ziel auch im Rahmen der Recherche und Trefferaufbereitung erreicht werden kann.

16 Vgl. Pfeffer, Magnus: Using clustering across union catalogues to enrich entries with indexing information. In: Spiropoulou, Myra; Schmidt-Thieme, Lars; Janning, Ruth (Hg.): Data analysis, machine learning and knowledge discovery, Cham: Springer, 2014, S. 437-445 http://dx.doi.org/10.1007/978-3-319-01595-8_47, und Wiesenmüller, Heidrun; Pfeffer, Magnus: Abgleichen, anreichern, verknüpfen. Das Clustering-Verfahren – eine neue Möglichkeit für die Analyse und Verbesserung von Katalogdaten. In: BuB 65 (2013), H. 9, S. 625-629.

17 Als Beispiele können die Primo-Installationen der UB Mannheim (URL: http://primo.bib.uni-mannheim.de/primo_library/libweb/action/search.do?vid=MAN_UB) und des Österreichischen Bibliothekenverbands (URL: http://search.obvsg.at/primo_library/libweb/action/search.do?vid=ACC) dienen (10.05.2015). Zum Werk-Clustering in Primo vgl. auch Schaffner, Verena: FRBR in MAB und Primo – ein kafkaesker Prozess? Möglichkeiten der FRBRisierung von MAB2-Datensätzen in Primo, exemplarisch dargestellt an Datensätzen zu Franz Kafkas „Der Prozess“, Graz-Feldkirch: Neugebauer, 2012 (Schriften der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, Bd. 9), S. 66-72. <http://phaidra.univie.ac.at/o:376819> (10.05.2015).

3. Normdaten

Im Bereich der Normdaten sind wichtige Schritte für den Umstieg auf das neue Regelwerk bereits vor längerer Zeit vollzogen worden: Mit der Umstellung auf die Gemeinsame Normdatei (GND) und der Einführung der sogenannten Übergangsregeln im Jahr 2012 wurden die Ansetzungsregeln in vielerlei Hinsicht an die angloamerikanische Tradition angepasst.¹⁸ Eine Reihe von Abweichungen zwischen den Übergangsregeln und RDA blieb jedoch bestehen, beispielsweise bei den Regeln für untergeordnete Körperschaften. Weitere Angleichungen erfolgten im Rahmen der ersten Stufe des RDA-Umstiegs im Herbst 2014. Mit dem Vollumstieg Ende 2015 werden die Regeln für die Normdaten vollständig RDA-gerecht sein. Die über viele Jahrzehnte angelegten Normdaten entsprechend freilich vielfach noch nicht diesem Standard. An zwei Beispielen sollen nun die Möglichkeiten und Grenzen automatischer Upgrades im Bereich der Normdaten betrachtet werden.

3.1 Universitätsinstitute

Institute von Universitäten gehören zu den untergeordneten Körperschaften. Bei diesen gibt es grundsätzlich zwei Möglichkeiten für die Struktur des bevorzugten Namens: Dieser kann entweder selbständig gebildet sein, d.h. er besteht (nur) aus dem Namen der untergeordneten Körperschaft, oder er wird unselbständig gebildet, d.h. die untergeordnete Körperschaft wird als Abteilung der übergeordneten Körperschaft angegeben.¹⁹

Sowohl nach den RAK (§ 429,1 Anm.) als auch nach den GND-Übergangsregeln²⁰ wurden Universitätsinstitute grundsätzlich selbständig angesetzt, beispielsweise als *Institut für Informatik (Heidelberg)*. Unter RDA gibt es jedoch keine pauschale Regelung mehr, die für alle Universitätsinstitute in gleicher Weise anzuwenden wäre. Je nachdem, wie der Name des Instituts genau aussieht bzw. wie er in der Informationsquelle präsentiert wird, wendet man unterschiedliche Regelwerksstellen an.²¹ Um zu einem RDA-gerechten Ergebnis zu kommen, muss also jedes einzelne Universitätsinstitut geprüft werden. Dabei sind drei Fälle zu unterscheiden.

18 Vgl. Behrens-Neumann, Renate: Die Gemeinsame Normdatei (GND) – ein Projekt kommt zum Abschluss. In: Dialog mit Bibliotheken 24 (2012), H. 1, S. 25-28. <http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/dialog201201volltext.pdf> (10.05.2015), sowie Wiechmann, Brigitte: Die Gemeinsame Normdatei (GND) – Rückblick und Ausblick. In: Dialog mit Bibliotheken 24 (2012), H. 2, S. 20-22. <http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/dialog201202Volltext.pdf> (10.05.2015).

19 Im Fall der Universitätsinstitute gibt es dabei üblicherweise noch eine mittlere Hierarchieebene, nämlich die Fakultät oder den Fachbereich. Gemäß RDA 11.2.2.15 wird die Zwischenstufe übergangen.

20 GND-Übergangsregel K11, URL: https://wiki.dnb.de/download/attachments/51740915/gnd_uebergangsregeln_koerperschaften_11.pdf (10.05.2015).

21 Zur Behandlung untergeordneter Körperschaften nach RDA vgl. Wiesenmüller; Horny: Basiswissen RDA (wie Anm. 4), S. 103-107.


UNIVERSITÄT HEIDELBERG | ZUKUNFT SEIT 1386

Institut für Informatik

Geschäftsführung

Sekretariat

Arbeitsgruppen

MitarbeiterInnen

Anreise



Home

Institut für Informatik

Anschriften und Personen

- Geschäftsführung
- Sekretariat
- Arbeitsgruppen
- Mitarbeiter
- Anreise

Abb. 3: Ausschnitt aus der Website des Instituts für Informatik an der Universität Heidelberg

Der erste Fall sind Namen, die nur aus dem Wort „Institut“ und dem Fach o.ä. bestehen. Beispielsweise findet sich auf der Homepage der Universität Heidelberg das „Institut für Informatik“ (Abb. 3).²² Folgt ein Institutsname diesem Muster, so wird der bevorzugte Name nach RDA unselbständig gebildet – in unserem Beispiel also als *Universität Heidelberg. Institut für Informatik* (RDA 11.2.2.14.5). Im zweiten Fall ist der Name der Universität ein Bestandteil des Institutsnamens, wie z.B. „Geographisches Institut der Universität Bonn“ (Abb. 4).²³ Auch in diesem Fall wird der bevorzugte Name unselbständig gebildet: *Universität Bonn. Geographisches Institut* (RDA 11.2.2.14.6). In der Praxis ist die Abgrenzung zwischen Fall 1 und Fall 2 nicht immer leicht zu treffen: Denn zum einen kommen in den Informationsquellen oft unterschiedliche Namensvarianten vor, und zum anderen lässt sich nicht immer klar feststellen, ob eine Formulierung wie „der Universität XY“ als fester Namensbestandteil zu gelten hat oder ob es sich dabei nur um eine beschreibende Angabe handelt.

22 URL: <http://www.if.uni-heidelberg.de/> (10.05.2015).

23 URL: <http://www.geographie.uni-bonn.de/> (10.05.2015).

Sie sind hier: [Startseite](#)

Geographisches Institut Universität Bonn



Geographisches Institut der Universität Bonn

Meckenheimer Allee 166

D-53115 B o n n

Tel.: +49 (0)228 73-7290; Fax: +49 (0)228 73-5393;

Mail: geographie@uni-bonn.de

Abb. 4: Ausschnitt aus der Website des Geographischen Instituts der Universität Bonn

Der dritte Fall sind Institute mit einem spezifischen Namen, wie das zur Universität Heidelberg gehörige „Kirchhoff-Institut für Physik“. In diesem Fall wird der bevorzugte Name selbständig gebildet: *Kirchhoff-Institut für Physik* (RDA 11.2.2.13). Wäre das Kirchhoff-Institut jedoch auf der Website als „Kirchhoff-Institut für Physik der Universität Heidelberg“ präsentiert worden, so hätte dies gemäß Fall 2 zu einer unselbständigen Namensbildung geführt, also zu *Universität Heidelberg. Kirchhoff-Institut für Physik*.

Die Beispiele zeigen zum einen, wie diffizil die Anwendung der RDA-Regeln ist. Zum anderen wird deutlich, dass trotz aller komplizierten und schwierigen Überlegungen der Name von Universitätsinstituten in der weit überwiegenden Mehrheit der Fälle unselbständig gebildet wird. Eine durchaus bedenkenswerte Möglichkeit wäre deshalb, mit einer automatischen Routine pauschal alle in der GND noch selbständig angesetzten Universitätsinstitute in die unselbständige Form zu bringen. Beim Heidelberger Informatik-Institut würde man dies dadurch erreichen, dass man die Inhalte des Felds 110 (bevorzugter Name der Körperschaft, hier *Institut für Informatik (Heidelberg)*) und des zweiten Felds 410 (abweichender Name, hier *Universität Heidelberg. Institut für Informatik*) miteinander vertauscht (Abb. 5).

```

005 Tbv1
006 http://d-nb.info/gnd/16078976-X
008 kiz
011 f
035 gnd/16078976-X
039 gkd/16078976-X$vzg
043 XA-DE
110 Institut für Informatik$gHeidelberg
410 Universität Heidelberg$bFakultät für Mathematik und Informatik$bInstitut für Informatik
410 Universität Heidelberg$bInstitut für Informatik
510 !255259603!Universität Heidelberg$bFakultät für Mathematik und Informatik$4adue
551 !208952578!Heidelberg$4orta$X1
797 326775919
913 $$gkd$ia$aInstitut für Informatik <Heidelberg>$016078976-X
    
```

Abb. 5: Derzeitiger GND-Datensatz für das Heidelberger Institut für Informatik im SWB

Mit dieser Methode käme man bei relativ geringem Aufwand zu einer hohen Rate an RDA-gerechten Ansetzungen für Universitätsinstitute. Falsch umgesetzt würden allerdings Namen vom Typ „Kirchhoff-Institut für Physik“. Mit entsprechend höherem Arbeitseinsatz ließe sich das Ergebnis sicher noch verbessern – beispielsweise, indem Datensätze, bei denen das Feld 110 mit „...-Institut“ beginnt, von der Routine ausgenommen werden. Aber es fragt sich, ob Kosten und Nutzen dann noch in einem guten Verhältnis zueinander stehen würden. Denn vermutlich gilt auch hier das bekannte Paretoprinzip, demzufolge 80 % der Ergebnisse mit nur 20 % des Aufwands erreicht werden. Will man auch noch die verbleibenden 20 % der Ergebnisse erzielen, so müssten dafür 80 % der Arbeit aufgewendet werden.

3.2 Pseudonyme

Bei den Universitätsinstituten (vgl. Kap. 3.1) muss aufgrund des Regelwerksumstiegs zwar in vielen Fällen der bevorzugte Name geändert werden, doch die Entitäten selbst bleiben konstant: Jedem nach RAK angelegten Normdatensatz für ein Universitätsinstitut entspricht auch unter RDA genau ein Normdatensatz. Weitaus schwieriger liegt der Fall bei Personen, die unter ihrem wirklichen Namen und einem Pseudonym oder unter mehreren Pseudonymen geschrieben haben bzw. schreiben. Denn in solchen Fällen betreffen die Änderungen nicht nur bevorzugte und abweichende Namen, sondern auch die Zahl und den Zuschnitt der Entitäten.

Während es unter RAK für jede reale Person nur einen einzigen Normdatensatz gab, in dem alle Namensformen (auch Pseudonyme) gesammelt wurden, wird gemäß RDA für jede Identität, unter der eine Person auftritt, ein eigener Normdatensatz erstellt. Bei der Katalogisierung einer Ressource muss immer der richtige Normdatensatz verwendet werden – je nachdem, unter welcher Identität der Autor bei diesem Werk auftritt. Katalogisiert man beispielsweise eine Ausgabe von „Alice in Wonderland“, so wird eine Beziehung zur Entität *Carroll, Lewis* (dem Pseudonym von Charles Lutwidge Dodgson) hergestellt. Würde man hingegen ein mathematisches Werk katalogisieren,

das Dodgson unter seinem wirklichen Namen publizierte, so ist der Normdatensatz für die Entität *Dodgson, Charles Lutwidge* zu verwenden.²⁴ Dies ermöglicht dann auch eine gezielte Suche nach den mathematischen Schriften des Autors, ohne dass man sich in der Trefferliste erst durch Hunderte von Katalogisaten für die Kinderbücher arbeiten muss.

```

100 Rowling, J. K.
400 Rowling, Joanne Kathleen$4navo
400 Rowling, Joanne K.
400 Rowling, Joanne
400 Rowling, Joane K.$vFalsche Namensform
400 $PLuolin$VChines.
400 Rauling, Je. Ke.$vUrdu-Form
400 Rolinga, Je. Ke.
400 Rollin, J. K.$vUigur.
500 !420405941!Galbraith, Robert*1968-*$4pseu
500 !422790745!Scamander, Newt*1965-*$4pseu
500 !422791431!Whisp, Kennilworthy*1965-*$4pseu
548 1965$4datI
550 !209103183!Schriftstellerin$4berc
550 !209010754!Lehrerin$4beru
667 Basic Heading
    
```

Abb. 6: Bereits umgearbeiteter GND-Datensatz für J. K. Rowling im SWB (Ausschnitt)

In der Praxis bedeutet die Umstellung erstens, dass die entsprechenden RAK-Normdatensätze jeweils in mehrere RDA-Normdatensätze aufgespalten werden müssen. Im Fall von J. K. Rowling ist dies bereits geschehen: Abb. 6 zeigt den Datensatz für den wirklichen Namen. Dieser ist über Beziehungen im Feld 500 mit den neu angelegten Normdatensätzen für die drei Pseudonyme der Autorin – Robert Galbraith, Newt Scamander und Kennilworthy Whisp – verbunden. Bei Wolfgang Hohlbein steht die Umarbeitung hingegen noch aus: Seine acht Pseudonyme sind derzeit im Normdatensatz noch als abweichende Namen in Feld 400 verankert (Abb. 7). Zweitens müssen alle Titeldatensätze, die mit dem RAK-Normdatensatz verbunden waren, dem richtigen RDA-Normdatensatz zugeordnet werden.

24 Vgl. Wiesenmüller; Horny: Basiswissen RDA (wie Anm. 4), S. 90, 156, 169f. und 272f. Die beschriebene Aufspaltung in verschiedene Identitäten betrifft nur die Formalerschließung, aber nicht die Sacherschließung. Bei allen Werken, die die Person als Thema behandeln, wird derselbe Normdatensatz als Schlagwort verwendet (das sogenannte *basic heading*, also die Identität, unter der die Person am besten bekannt ist). Vgl. GND-Erfassungshilfe EH-P-06 (Pseudonyme), <https://wiki.dnb.de/download/attachments/90411361/EH-P-06.pdf> (10.05.2015).

100 Hohlbein, Wolfgang
 375 m
 400 Hohlbein, Wolfgang E.\$vEymers Pseud.-Lex., Weigand
 400 Hohlbein, Wolfgang\$vfalsche Namensform
 400 Hohlbein, W.
 400 Cholbajn, Volfgang
 400 Chol'bajn, Vol'fgang
 400 Cholmpain, Bolphnkannk
 400 Holbain, Polp'ügang
 400 Holbeini, ...
 400 Holbeins, Volfgangs
 400 Holhbein, Wolfgang
 400 Hörubain, Worufugangu
 400 Huo'erben, Wofugang
 400 \$PWofugang-Huo'erben
 400 Bonnella, Angela\$4pseu
 400 Craven, Robert\$4pseu\$vt. Weigand persönliches Pseud.; teilweise auch benutzt mit Burgdorf, Karl-Ulrich; Titel nicht zuzuordnen, deshalb keine VW bei Burgdorf
 400 Eschenloh, Wolfgang\$4pseu
 400 Heidner, Martin\$4pseu
 400 Marks, Michael\$4pseu
 400 MacCloud, Jason\$4pseu
 400 Mac Cloud, Jason\$4pseu
 400 McCloud, Jason\$4pseu
 400 Mc Cloud, Jason\$4pseu
 400 Smith, Angela\$4pseu
 400 Vernom, Jack\$4pseu
 400 Wolf, Henry\$4pseu\$vt. Weigand persönliches Pseud.; teilw. auch benutzt mit Burgdorf, Karl-Ulrich; Titel nicht zuzuordnen, deshalb keine VW bei Burgdorf

Abb. 7: Noch nicht umgearbeiteter GND-Datensatz für Wolfgang Hohlbein im SWB (Ausschnitt)

Der erste Schritt erfolgt zentral: Die Änderung der Normdatensätze wird nur in der Masterdatei der GND ausgeführt und gelangt dann über automatische Routinen in alle Verbände. Die Zuordnung der Titeldatensätze muss hingegen in jedem Verbundkatalog von neuem durchgeführt werden. Oft sind große Mengen an Titeln betroffen: Mit dem RAK-Normdatensatz für Wolfgang Hohlbein sind beispielsweise im SWB über 500 Titeldatensätze verknüpft. Wenn jeder Fall in jedem einzelnen Verbund manuell bearbeitet werden muss, so ist angesichts der knappen Personalressourcen zu befürchten, dass die Titeldatensätze überhaupt nicht oder erst mit einer erheblichen zeitlichen Verzögerung richtig zugeordnet werden. Auch kann es bei der manuellen Bearbeitung leicht zu Fehlern kommen. Dies hätte Nachteile für die Recherche: Sucht man etwa nach Robert Galbraith als Autor, so würde man einen Teil der relevanten Titel nicht erhalten – nämlich diejenigen, die nicht an den neu entstandenen Normdatensatz umgehängt wurden, sondern noch mit dem ursprünglichen Normdatensatz (der jetzt aber nur noch für den wirklichen Namen der Person steht) verbunden sind. Im SWB sind beispielsweise derzeit zwar die Ausgaben des Romans „Der Seidenspinner“ schon mit dem Normdatensatz für Robert Galbraith verknüpft, nicht aber die Ausgaben des Romans „Der Ruf des Kuckucks“ – diese sind noch mit dem Normdatensatz für J. K. Rowling verbunden.

Es wurde deshalb vorgeschlagen, das Problem über eine erweiterte Indexierung zu lösen: Bei einer Recherche nach einem Personennamen sollen dann nicht nur die bevorzugten und abweichenden Namen berücksichtigt werden, sondern auch Namen von in Beziehung stehenden Personen, sofern sie mit dem GND-Code *pseu* (Pseudonym) oder *nawi* (wirklicher Name) gekennzeichnet sind. Dies würde dazu führen, dass bei der Recherche stets sämtliche Titel ausgegeben werden – egal, ob sie mit einer Pseudonym-Identität oder der realen Identität der Person verbunden sind. Die in den Daten mit nicht unerheblichem Aufwand durchgeführte Differenzierung der verschiedenen Identitäten würde also bei der Recherche wieder außer Kraft gesetzt, sodass man letztlich genau dieselben Treffermengen erhält wie unter RAK.

Weitaus besser wäre freilich eine Lösung, welche die von RDA gewollte Unterscheidung auch in der Praxis berücksichtigt – also primär die Titel anzeigt, bei denen die Person auch wirklich unter der gesuchten Identität auftritt. Natürlich sollte den Recherchierenden zusätzlich eine komfortable Möglichkeit geboten werden, um auch die mit den anderen Identitäten verbundenen Titel zu sehen. Denkbar ist beispielsweise ein expliziter Hinweis und Link auf diese Titel²⁵ oder ein entsprechendes Ranking, bei dem zuerst die Treffer kommen, die mit der tatsächlich gesuchten Identität in Verbindung stehen, und erst danach die Titel, die zu den anderen Identitäten gehören. Die Voraussetzung für derartige Lösungen ist freilich die korrekte Zuordnung der Titeldatensätze zu den RDA-Normdatensätzen. Wie könnten maschinelle Methoden dabei helfen?

Die potenziell betroffenen Datensätze sind anhand der beim Umstieg auf die GND eingeführten Kennung *pip* in Feld 008 identifizierbar. Allerdings muss in jedem Fall zuerst intellektuell geprüft werden, ob wirklich ein Fall von getrennten Identitäten vorliegt. Denn wenn eine Person ausschließlich unter einem Pseudonym publiziert hat und nie unter ihrem wirklichen Namen (wie z.B. George Orwell, der eigentlich Eric Arthur Blair hieß), so gibt es weiterhin nur einen einzigen Normdatensatz.

Die Aufspaltung der betroffenen Normdatensätze könnte durch ein maschinelles Verfahren unterstützt werden, welches automatisch neue Normdatensätze für alle mit den entsprechenden Codes gekennzeichneten Namensformen anlegt und diese miteinander verknüpft. Biografische Informationen würden dabei – mit Ausnahme der Lebensdaten – nur im Datensatz für den wirklichen Namen gehalten und nicht in die Datensätze für die Pseudonym-Identitäten kopiert. Ganz ohne menschliches Eingreifen kann ein solcher Vorgang allerdings nicht ablaufen, wie das Beispiel von Wolfgang Hohlbein zeigt: Hier müsste manuell dafür gesorgt werden, dass aus den vier zusammengehörigen Namensformen (*MacCloud, Jason*; *Mac Cloud, Jason*; *McCloud, Jason*; *Mc Cloud, Jason*) nicht vier neue Normdatensätze entstehen, sondern nur ein einziger neuer Normdatensatz generiert wird. Bei diesem wären drei der vier Namensformen als abweichende Namen zu führen.

Noch wichtiger ist ein maschinelles Verfahren bei der Zuordnung der Titeldatensätze. Zunächst sollten die mit dem ursprünglichen Normdatensatz verbundenen Titel nach Werken geclustert werden (vgl. Kap. 2.3) – denn der geistige Schöpfer ist auf der Ebene des Werks angesiedelt. Alle zu einem

25 Z. B. ein Hinweis wie: „*Rowling, J. K.* verwendet auch die Namen *Galbraith, Robert*; *Scamander, Newt*; *Whisp, Kennilworthy*. Klicken Sie auf die Namen, um die zugehörigen Titel zu sehen!“.

Werk-Cluster gehörigen Titeldatensätze können dann en bloc einem Normdatensatz zugeordnet werden. Für jedes Werk muss nun noch entschieden werden, welche Identität die richtige ist. Dafür bietet sich eine Auswertung der Verantwortlichkeitsangaben an: Es ist zu ermitteln, welcher der in Frage kommenden Namen in den Verantwortlichkeitsangaben eines Werk-Clusters tatsächlich vorkommt. Beispielsweise steht bei allen Ausgaben des Romans „Das Drachenkind“ von Wolfgang Hohlbein, die im SWB vorhanden sind, in der Verantwortlichkeitsangabe der Name „Michael Marks“. Die Zuordnung des Werks zur Pseudonym-Identität *Marks, Michael* ist deshalb eindeutig; die geclusterten Titeldatensätze könnten mit einer entsprechenden Routine maschinell umgehängt werden.

Nicht alle Fälle sind so eindeutig: Mitunter kommen in den Verantwortlichkeitsangaben eines Werk-Clusters die Namen unterschiedlicher Identitäten vor. So wurde der Roman „The running man“ (deutsch „Menschenjagd“) zunächst unter dem Pseudonym Richard Bachmann publiziert. In späteren Ausgaben finden sich Angaben wie „Stephen King schreibt als Richard Bachmann“. Und mittlerweile steht z.T. in der Verantwortlichkeitsangabe nur noch der Name „Stephen King“. In solchen Fällen könnte man entweder über eine Regel festlegen, welche Identität „gewinnt“,²⁶ oder die Entscheidung einem menschlichen Katalogisierer überlassen.

4. Fazit und Ausblick

Wie die beispielhafte Betrachtung verschiedener Aspekte gezeigt hat, sind maschinelle Anpassungen von RAK-Daten an den neuen Katalogisierungsstandard RDA durchaus möglich, aber alles andere als trivial. Insbesondere muss man sich von der Hoffnung verabschieden, dass es eine einzige Lösung für alles geben könnte – vielmehr sind zahlreiche einzelne Schritte nötig, um zu einem zufriedenstellenden Ergebnis zu kommen. Denkbar wäre dies am ehesten als ein iterativer, über einen längeren Zeitraum ablaufender Prozess, bei dem die Daten Schritt für Schritt verbessert werden.

Die hier vorgelegte Betrachtung erfolgte allein aus datentechnischer Sicht. Was derzeit noch völlig fehlt, ist eine Diskussion über die organisatorischen Rahmenbedingungen für solche RDA-Upgrades. Wünschenswert wäre ein kooperativer Ansatz, bei dem die anstehenden Aufgaben auf mehrere Partner verteilt werden. Die an verschiedenen Stellen entwickelten Lösungen müssten dann natürlich auch von anderen nachgenutzt werden können. Von zentraler Bedeutung ist außerdem das Vorhandensein geeigneter Tools für das Processing der Daten. Erfahrungsgemäß stellen die begrenzten Personalressourcen in den IT-Abteilungen einen „Flaschenhals“ dar, an dem viele wünschenswerte Projekte scheitern. Es wäre deshalb zu begrüßen, wenn entsprechende Werkzeuge für das Metadatenmanagement nicht nur von IT-Mitarbeiterinnen und Mitarbeitern eingesetzt werden könnten, sondern auch von Bibliothekspersonal ohne ausgeprägten IT-Hintergrund.

Ein vielversprechende Entwicklung stellt in diesem Zusammenhang das Projekt D:SWARM dar.²⁷ In Kooperation zwischen der SLUB Dresden und der Avantgarde Labs GmbH wird seit 2013 eine

²⁶ Dies könnte entweder diejenige Identität sein, deren Name in der Manifestation mit dem frühesten Erscheinungsdatum vorkommt, oder diejenige, die insgesamt gesehen am häufigsten in den Verantwortlichkeitsangaben vorkommt.

²⁷ Zu D:SWARM vgl. Mittelbach, Jens; Glaß, Robert: A library data management platform based on linked open data. Vortrag gehalten auf der SWIB14 am 2. Dezember 2014 in Bonn. Vortragsfolien: <https://speakerdeck.com/swib14/a-library-data-management-platform-based-on-linked-open-data> (10.05.2015).

Open-Source-Plattform für das Management bibliothekarischer Daten entwickelt, die auf Linked Open Data beruht. Zu den Prinzipien des Projekts gehören zum einen Offenheit und Nachnutzbarkeit: Sowohl Mappings als auch Workflows zur Datentransformation können in der Community geteilt werden. Zum anderen wurde Wert auf ein einfach zu bedienendes, grafisches User-Interface gelegt, das auch ohne vertiefte IT-Kenntnisse bedient werden kann. D:SWARM könnte deshalb eine geeignete technische Umgebung darstellen, um die Herausforderungen zu bewältigen, die sich durch den RDA-Umstieg ergeben.

Literaturverzeichnis

- Behrens-Neumann, Renate: Die Gemeinsame Normdatei (GND) – ein Projekt kommt zum Abschluss. In: Dialog mit Bibliotheken 24 (2012), H. 1, S. 25-28.
<http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/dialog201201volltext.pdf> (10.05.2015).
- Mittelbach, Jens; Glaß, Robert: A library data management platform based on linked open data. Vortrag gehalten auf der SWIB14 am 2. Dezember 2014 in Bonn.
Vortragsfolien: <https://speakerdeck.com/swib14/a-library-data-management-platform-based-on-linked-open-data> (10.05.2015).
- 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, Cham: Springer, 2014, S. 437-445.
http://dx.doi.org/10.1007/978-3-319-01595-8_47.
- Polak-Bennemann, Renate: Technische Veränderungen. Vortrag gehalten auf dem Workshop für Systemanbieter am 23. Oktober 2014 in der Deutschen Nationalbibliothek.
Vortragsfolien: https://wiki.dnb.de/download/attachments/99090660/03_Systemanbieterworkshop-Einf%C3%BChrung-technik.pdf (10.05.2015).
- RDA Toolkit. Resource Description and Access. <http://www.rdatoolkit.org/> (10.05.2015).
- Schaffner, Verena: FRBR in MAB und Primo – ein kafkaesker Prozess? Möglichkeiten der FRBRisierung von MAB2-Datensätzen in Primo, exemplarisch dargestellt an Datensätzen zu Franz Kafkas „Der Prozess“, Graz-Feldkirch: Neugebauer, 2012 (Schriften der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, Bd. 9).
<http://phaidra.univie.ac.at/o:376819> (10.05.2015).
- Vonhof, Cornelia; Stang, Richard; Wiesenmüller, Heidrun: Forschung für die Praxis – Perspektiven für Bibliotheks- und Informationsmanagement. In: o-bib 2 (2015), H. 1, S. 68-74.
<http://dx.doi.org/10.5282/o-bib/2015H1S68-74>.

- Wiechmann, Brigitte: Die Gemeinsame Normdatei (GND) – Rückblick und Ausblick. In: Dialog mit Bibliotheken 24 (2012), H. 2, S. 20-22. <http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/dialog201202Volltext.pdf> (10.05.2015).
- Wiesenmüller, Heidrun; Horny, Silke: Basiswissen RDA. Eine Einführung für deutschsprachige Anwender, Berlin/Boston: De Gruyter Saur, 2015.
- Wiesenmüller, Heidrun; Pfeffer, Magnus: Abgleichen, anreichern, verknüpfen. Das Clustering-Verfahren – eine neue Möglichkeit für die Analyse und Verbesserung von Katalogdaten. In: BuB 65 (2013), H. 9, S. 625-629.