

Authority File

CHRISTEL HENGEL

Normdaten und Metadaten

Die Idee eines Internationalen Authority File



Christel Hengel

Authority files are a special sort of metadata: metadata for metadata. In a set of authority records that is separate from the metadata of a publication, the names or identifiers of entities which differ from those used in the publications are collected and standardized. By using authority records as metadata it is possible to monitor and ensure, for example, that for authors, copyright holders, publishers, key words, etc., a consistent and identical search path and access point is established (authority control). For queries, authority standards ensure that relevant publications can be retrieved via any of the name-forms and identifiers contained in the authority data file. By assigning attributes and relational links to the individual entities, additionally networked access points can be provided. Authority data are hence ideal for forming semantic navigational networks to search and access publications and other entities.

The standardized vocabulary and numbering system of an authority file is, however, only valid in its own scope of application, and the joint usage of title data can be impeded by the use of differing authority files.

The IFLA Working Group FRANAR has developed a model – based on the vision of a joint Virtual Authority File – for establishing the interoperability of national authority files in the library world. In order to prove the feasibility of such a scheme, the Library of Congress, OCLC and the Deutsche Bibliothek have agreed to conduct a joint project in which the construction a VIAF (Virtual International Authority File) will be demonstrated in the case of personal name authority records.

Normdaten sind Metadaten besonderer Qualität: Metadaten zu Metadaten. In einem getrennt vom Metadaten-Set zur einzelnen Publikation gebildeten Normdaten-Satz werden die unterschiedlichen in den Publikationen verwendeten Namensformen bzw. Bezeichnungen für dasselbe Objekt zusammengeführt und normiert. Durch die Verwendung von Normdaten als Metadaten wird kontrolliert und sichergestellt, dass z. B. für Autoren, Urheber, Verleger, Sachinhalte etc. gleich bleibende, identische Sucheinstiege und Zugriffspunkte gebildet werden (Authority Control). Für die Recherche wirkt sich die Normierung so aus, dass die zugehörigen Publikationen über alle im Normdatensatz beinhalteten Namensformen und Bezeichnungen gefunden werden können. Durch die Angabe von Attributen und Relationen zu den einzelnen Objekten werden zudem zusätzliche, vernetzte Zugriffspunkte bereitgestellt. Normdaten sind damit ideal zur Bildung von semantischen Navigationsnetzen für die Suche und den Zugriff auf Publikationen und andere Objekte geeignet.

Das standardisierte Vokabular und das Nummernsystem einer Normdatei hat aber nur im eigenen Anwendungsbereich Gültigkeit, und die gemeinsame Nutzung von Titeldaten kann durch die Nutzung unterschiedlicher Normdateien behindert werden.

Die IFLA-Arbeitsgruppe FRANAR hat mit der Vision eines gemeinsamen Virtuellen Authority File ein Modell entwickelt, wie im Bibliothekswesen Interoperabilität zwischen den nationalen Normdateien hergestellt werden kann.

Um die Realisierbarkeit eines solchen Vorhabens nachzuweisen, haben die Library of Congress, OCLC und Die Deutsche Bibliothek ein gemeinsames Projekt vereinbart, in dem exemplarisch für den Bereich der Personennormdaten ein VIAF (Virtual International Authority File) aufgebaut werden soll.

Bibliographische Metadaten dienen der Suche und dem Auffinden von Netzpublikationen, in denen sie enthalten sind oder denen sie beigegeben sind. Durch die nach Elementtypen differenzierte Erfassung der

einzelnen Metadaten wird es dem Benutzer ermöglicht, seine Suchfragen zu spezifizieren. Der Such-Term kann in seiner speziellen Rolle für die Publikation recherchiert werden, so z. B. ein Personennamen nur in der Funktion als Autor, nicht aber als Bestandteil des Titels oder als Deskriptor. Metadaten erfüllen damit in der Recherche dieselben Funktionen wie die Katalogdaten der Bibliotheken.

Soll in verschiedenen, teilweise heterogen erschlossenen Datenbeständen – auch in einer gemeinsamen Suche in Metadaten und Katalogdaten – recherchiert und sollen gegebenenfalls Daten nachgenutzt werden können, ergeben sich, sieht man von Format- und Schnittstellenfragen ab, drei Bereiche, für die Festlegungen zu treffen sind.

Zunächst die Bestimmung der Entitäten: Welche Publikationen bzw. welche Ebenen von Publikationen werden erschlossen?

Sodann die Festlegung der Elemente-Sets: Welche Datenelement-Typen (Autor, Titel, Urheber, Schlagwort, Verlag ...) sind bei der Erschließung zu berücksichtigen? In welchem Differenzierungsgrad? Wie können unterschiedliche Differenzierungsgrade und Elemente-Sets aufeinander abgebildet oder konvertiert werden?¹

Drittens die Standardisierung der Feldinhalte: Es genügt nicht, die Art der Felder (die Elementtypen) festzulegen; Feldinhalte, insbesondere Namen und Bezeichnungen, die für unterschiedliche Publikationen zutreffen, sollten in gleich bleibender Form suchbar sein. Wie hierbei mit Normdateien gearbeitet werden kann und welche Ansätze es für ihre internationale Zusammenführung gibt, soll im Folgenden kurz beschrieben werden.

Normdaten sind Metadaten besonderer Güte. Sie sind Daten über Objekte aller Art, deren Namen oder Bezeichnungen bei der Erschließung einer Publikation oder eines anderen Objektes benötigt werden. Im Normdatensatz werden, unabhängig von den in der eigenen Institution vorliegenden Publikationen, die relevanten Namensvarianten oder Bezeichnungen für das Objekt zusammengeführt und so eingetragen, dass sie in der Recherche als äquivalent behandelt werden können.

Während die Recherche über die Metadaten zur Publikation im Prinzip eine zweistufige Suche ist, die

Bestimmung der Entitäten

Festlegung der Elemente-Sets

Standardisierung der Feldinhalte

zunächst zum Metadatensatz führt und von dort die Zugriffsmöglichkeit auf die Publikation bietet (auch wenn der Benutzer bei Online-Publikationen häufig direkt zur Publikation weitergeführt wird), verläuft die Suche über Normdaten im Prinzip dreistufig. Der Suchschritt über den Normdatensatz kann allerdings für den Benutzer in der Anzeige übergangen werden.

Zunächst wird der Normdatensatz ermittelt, über den voreingestellten Term im Normdatensatz, der den Zugriff auf den Metadatensatz zur Publikation bietet, der Metadatensatz aufgerufen und von ihm aus der Zugriff auf die Publikation ermöglicht.

Auf internationaler Ebene arbeitet die IFLA-Arbeitsgruppe FRANAR (Functional Requirements and Numbering of Authority Records)² daran, vor dem Hintergrund der »Functional Requirements for Bibliographic Records« funktionale Anforderungen an Normdaten zu definieren.

Die Diskussion darüber, welche Objekte im bibliothekarischen Kontext relevant sind und für welche Typen von Objekten Normdaten gebildet werden sollen, nahm breiten Raum ein.

Als Objekt-Typen schlägt die FRANAR-Gruppe vor:

- Person
- Familie
- Körperschaft (einschließlich Kongressen und Veranstaltungen, einschließlich Gebietskörperschaften)
- Werk, Expression, Manifestation, Exemplar (Einheiten für die unterschiedlichen Stufen im Lebenszyklus von Publikationen)
- Objekt
- Geographikum
- Sachbegriff
- Zeit

Die wichtigsten Datenelemente für Normdaten sind die unterschiedlichen Namensformen bzw. Benennungen, unter denen das Objekt bekannt ist.

Jeder Normdatensatz hat eine oder mehrere Vorzugsbezeichnungen (Ansetzungsformen); mehrere Vorzugsbezeichnungen können z.B. für unterschiedliche Sprach- oder Schriftformen gebildet sein oder nach unterschiedlichen Regelwerken angesetzt sein.

Ein Beispiel aus der PND (Personennamendatei) mit gleichberechtigten Ansetzungsformen nach unterschiedlichen Regelwerken bzw. aus anderen Normdateien:

100 @Hildegardis <Bingensis> [RAK-WB]

120 @Hildegard <von Bingen> [RSWK]

140 @Hildegard <Saint> \$d1098-1179 [LCSH]

Zu den als Namensvarianten und synonyme Bezeichnungen (Äquivalente) erfassten Formen zählen neben abweichenden Benennungen auch sprachli-

che Varianten, unterschiedliche Transliterationsformen und schließlich in Zukunft nach dem Übergang zu Unicode auch Schriftvarianten.

Beispiel aus der PND (Schriftvarianten sind in der PND noch nicht dargestellt):

200 @Hildegard <Heilige, 1098-1179> %SWB-AK

200 @Hildegard <Sankt>

200 @Hildegardis <Sancta>

200 @Hildegardis <Abbatissa>

200 @Hildegard <Sainte>

200 @Hildegarda <di Bingen>

200 @Hildegardis <de Monte Sancti Ruperti>

200 @Hildegard <die Heilige>

200 @Hildegardis <de Bingen>

200 Bingen, Hildegard /von

Eng verwandt mit den Namensformen ist der Identifier, den der Normdatensatz in der Normdatei oder in anderen Normdateien oder Nachweissystemen trägt. Identifier haben gleichzeitig Ähnlichkeiten und Vorteile gegenüber Vorzugsbezeichnungen, da sie jederzeit mit diesen identisch gesetzt werden können und anders als sie keinen terminologischen Veränderungen (wohl aber ggf. administrativen Veränderungen) unterworfen sind. Durch die Angabe der Normdatennummer, die das entsprechende Objekt in einer anderen Normdatei hat, können Verbindungen zu dieser Normdatei hergestellt und gegebenenfalls eine gemeinsame Suche in den damit erschlossenen Beständen organisiert werden.

In PND-Sätzen sind z.B. die PND-Nummer, die SWD-Nummer, die ZKA-Nummer und die LCNA-Nummer für die entsprechende Person enthalten.

Neben Namensformen und Identifiern werden die im jeweiligen Kontext relevanten Attribute des Objektes angegeben, für Personen können dies z.B. sein:

- Lebensdaten, Wirkungsdaten
- Beruf, Funktion
- Wirkungsort
- Sachgebiet
- Geschlecht
- Werke
- Beschreibung

Darüber hinaus sind Relationen zu anderen Objekten aufgeführt. Sie unterscheiden sich von Attributen dadurch, dass in der Normdatei ein Normdatensatz für das entsprechende relationierte Objekt existiert und in der jeweiligen Normdatei-spezifischen Form (über den Identifier, über die Ansetzungsform, über die Ansetzungsform mit Erweiterungen, einseitig oder gegenseitig) unter Kennzeichnung der Relationsart eine Relationierung vorgenommen ist.

die Suche über Normdaten erfolgt dreistufig

Namensformen und Identifier

Attribute von Objekten und Relationen zu anderen Objekten

In Personen-Normdaten könnten z. B. als Relationen erfasst werden:

- Affiliation
- Körperschaft, für die die Person den Forschungs-, Sammlungs- oder Tätigkeitsbereich darstellt
- Werke, Erfindungen, Theorien einer Person
- Wirkliche Person bzw. Sammelpseudonym
- Sachgebiet
- Beruf
- Wirkungsort
- Geburtsort
- Zeit(-epoche)

Relationen zwischen Normdatensätzen

Über die Relationen zwischen Normdatensätzen eröffnen sich weitere Suchmöglichkeiten, die gestatten, die unterschiedlichen Zugriffsmöglichkeiten miteinander zu kombinieren und zu vernetzen. So können z. B. ausgehend vom Körperschaftssatz die Personen gesucht werden, die zu ihr affiliert sind, sowie die Publikationen der entsprechenden Personen ermittelt werden, und zudem kann weiterverfolgt werden, welche Werke, Erfindungen und Theorien von diesen Personen stammen und ob hierzu Nachweise vorliegen.

Über ihre Grundfunktion hinaus, Namensvarianten zusammenzuführen und für die Suche zu standardisieren, nehmen Normdatensätze zunehmend weitergehende Aufgaben wahr. Die Normdaten werden mit zunehmendem Ausbau der Normdateien für ihren Verwendungsbereich zu einem semantischen Netz für die Recherche, in das allerdings jeder neue Normdatensatz bei der Neuerfassung entsprechend eingepasst werden muss. Mit zunehmender Anwendung auch in außerbibliothekarischen Bereichen wie in der Fachinformation, in Archiven, Museen und anderen Anwendungsgebieten wächst dieses semantische Netz in weitere Bereiche hinein.

semantisches Netz

Den Vorteilen von Normdaten wurde auch im Metadatenbereich in den Metadatensets und Application Profiles Rechnung getragen, indem für bestimmte Felder die Möglichkeit eröffnet wird, durch die Angabe von Schemata die Zugehörigkeit einer bestimmten Ansetzungsform zu einer bestimmten Normdatei zu kennzeichnen. Durch die Typisierung, aus welcher Normdatei, welchem Thesaurus bzw. welcher Klassifikation ein Feldinhalt stammt, wird es ermöglicht, die Recherchemöglichkeiten in dieser Normdatei mitzunutzen.

Gleichzeitig wird dabei aber auch die Schwierigkeit deutlich, vor der Anbieter und Anwender von Normdaten angesichts historisch gewachsener, separater Normdateien stehen. Das standardisierte Vokabular und das damit verbundene Nummernsystem der Normdatei hat nur im eigenen Anwendungsbereich Gültigkeit. Die Suche mit dem eigenen Vokabular in ei-

ner fremden Normdatei (oder umgekehrt) wird durch Regelunterschiede und Sprachbarrieren behindert.

Der Gedanke, die bestehenden Normdateien zusammenzuführen und dadurch weltweit die gegenseitige Nutzung von Normdaten zu ermöglichen, wird seit Mitte der 1960er-Jahre in der IFLA verfolgt. Während zunächst die Sicht vorherrschte, Abhilfe durch eine Angleichung der Ansetzungsformen zu schaffen, hat mittlerweile ein Paradigmenwechsel eingesetzt.

Schon mit der Bildung der Arbeitsgruppe FRANAR war die Vorstellung verbunden, nicht über die Standardisierung von verbalen Bezeichnungen, sondern über ein einheitliches Nummernsystem für Normdaten eine Zusammenführung der Vorzugsbezeichnungen und Namensvarianten aus unterschiedlichen Normdateien unter einem gemeinsamen Identifier zu erreichen und damit eine gemeinsame Suche über die angeschlossenen Normdateien zu ermöglichen.

Der Grundgedanke dabei war derselbe, der jeder einzelnen Normdatei zugrunde liegt. Durch die Zusammenführung der als äquivalent zu behandelnden Bezeichnungen unter einer gemeinsamen Vorzugsbezeichnung, in diesem Falle einem Identifier, kann ein gemeinsames Retrieval mit Zugriffsmöglichkeiten auf alle damit verbundenen Daten und Publikationen organisiert werden.

Würde zum Beispiel dieselbe Körperschaft in den LCNA in einer im Englischen üblichen Form, in RAMEAU in einer im Französischen üblichen Form, und in der GKD in einer originalsprachigen Form angesetzt, würden durch die Zusammenführung unter einem gemeinsamen Identifier diese u. U. völlig inkompatiblen Namensformen sowie die dazu in den beteiligten Normdateien aufgeführten Namensvarianten alle in den jeweils zugehörigen Titeldaten gesucht und die Suchergebnisse aufgrund der Zusammenführung unter einem gemeinsamen Identifier zu einem gemeinsamen Suchergebnis zusammengeführt.

Mittlerweile ist eine Weiterentwicklung dieser Sicht eingetreten. Ebenso wie in einem einzelnen Normdatensatz unterschiedliche gleichberechtigte Ansetzungsformen und Identifier als äquivalent vereinbart sein können, können auch im internationalen Bereich die Identifier der bestehenden Normdateien einander als äquivalent zugeordnet werden. Wird eine PND-Nummer einer LCNA-Nummer zugeordnet, so ist diese Beziehung eindeutig. Die dahinter stehenden Objekte sind identisch. Die unterschiedlichen Namensvarianten im PND-Satz und im LCNA-Satz können für die Suche gleichgesetzt werden.

Das eigentliche Problem bei der Zusammenführung der Normdateien besteht also nicht im Fehlen einer Theorie, wie eine Zusammenführung zu bewerk-

stelligen wäre. Es sind die großen Datenbestände sowohl an Normdaten als auch an damit verbundenen Titeldaten, die der Verwirklichung der Idee im Wege stehen.

Um die Realisierbarkeit eines Internationalen Authority File nachzuweisen, haben die Library of Congress, OCLC und Die Deutsche Bibliothek ein gemeinsames Projekt vereinbart, in dem exemplarisch für den Bereich der Personennormdaten die Normdatensätze der LCNA und der PND in einem gemeinsamen VIAF (Virtual International Authority File)³ zusammengeführt werden sollen.

Mit dem Aufbau eines (Virtuellen) Internationalen Authority File für Personennamen, in dem die Personennamensätze der Library of Congress (LoC) und des PND-Verbundes gemeinsam suchbar gemacht werden, werden mehrere Zielsetzungen verfolgt:

- Es wird der Grundstein für eine multinationale Normdatei gelegt, die aus den nationalen Normdateien gespeist wird und damit von vornherein kooperativ und arbeitsteilig angelegt ist.
- Im VIAF werden die unterschiedlichen Ansetzungsformen der nationalen Normdateien in einem Datensatz zusammengeführt und damit gemeinsam suchbar gemacht. Das Projekt soll damit nachweisen, dass über ein gemeinsames Nummernsystem die gemeinsame Datennutzung auch ohne komplette Angleichung der Ansetzungsformen möglich ist.
- Die Zusammenführung und gegenseitige Nutzung der Normdaten, über Regel-, Sprach- und Schriftbarrieren hinweg, ermöglicht zugleich die gegenseitige Nutzung von Titeldaten.
- Das Projekt ist ein Pilotprojekt auch für andere Normdateien. Andere nationale Normdateien für Personennamen sollen folgen, desgleichen Normdateien für Körperschaften, Werktitel und andere Arten von Normdaten.

Der VIAF soll Angaben aus allen Datensätzen des LCNAF und der PND enthalten und soll im Web für Recherche und bibliographische Nachnutzung zur Verfügung gestellt werden. Über die Datenelemente und die Datenstruktur, die der gemeinsame VIAF haben soll, müssen noch im Einzelnen Vereinbarungen getroffen werden. Da im Projekt zunächst vorgesehen ist, den VIAF als einen realen Datenbestand aufzubauen, ist die Bezeichnung »Virtuell« für die Projektzeit nicht zutreffend. Endgültige Entscheidungen über die Datenhaltung – für eine virtuelle Zusammenführung durch die Recherche über verteilt liegende Bestände oder für die Bildung eines realen Datenbestands – werden aber sicher erst in einer späteren Phase dieses oder eines nächsten Projekts zu erwarten sein.

Zusammenführung bestehender Normdateien

Aufbau eines VIAF für Personennamen – Zusammenführung der Personennamensätze der LoC und der PND

Das Projekt wird in drei Phasen ablaufen:

In Phase 1 werden die parallelen Datensätze beider Normdateien zusammengeführt.

Die Normdatensätze der LCNA und der PND werden durch einen Abgleich der damit verbundenen Titeldatensätze mit nachfolgendem Check der Normdatensätze zusammengeführt werden.

Matching der Datensätze

Der Titeldaten- und Normdaten-Abgleich soll bei OCLC erfolgen.

Im Check der Titeldatensätze werden anhand enthaltener identischer bzw. hinreichend ähnlicher Merkmale die als bibliografisch identisch erkannten Sätze zusammengeführt sowie die darin enthaltenen Personen ebenfalls gegeneinander gecheckt und gegebenenfalls zusammengeführt.

OCLC verfügt aus Abgleichen mit den LCNA bereits über Algorithmen zur Identitätserkennung der Titeldaten und Normdaten. Diese Algorithmen sollen verwendet und gegebenenfalls an die andere Datenlage angepasst werden.

Für den Titelabgleich wird der gesamte Titeldatenbestand der LoC und der Gesamtbestand von World-Cat auf der einen Seite sowie der gesamte DDB-Titelbestand und der Titelbestand des Bayerischen Bibliotheksverbundes, der mit PND-Sätzen verknüpft ist, auf der anderen Seite herangezogen. Nach ersten Testlieferungen im Dezember 2002 wurde Anfang Mai 2003 mit der Auslieferung der DDB- und BVB-Daten begonnen. Insgesamt werden rund 5,7 Mio. Titelsätze aus DDB, rund 5 Mio. Titelsätze aus dem BVB sowie ca. 2,3 Mio. PND-Normdatensätze an OCLC geliefert werden.

Bereitstellung des VIAF über das Internet

In Phase 2 wird ein VIAF-Server eingerichtet und der aus den zusammengeführten Normdaten entstandene VIAF für die Suche über das Internet bereitgestellt. Änderungen in den beiden Normdateien LCNA und PND werden durch regelmäßige Updates aus beiden Dateien im VIAF nachvollzogen. Dabei soll für die Datenübernahme das OAI-Protokoll genutzt werden.

Echtbetrieb

In Phase 3 soll schließlich der Server in Betrieb genommen, für die Recherche durch die Bibliotheken und Endnutzer geöffnet und unter Betriebsbedingungen getestet werden. Für diese Phase ist auch die Implementierung von Unicode vorgesehen.

¹ Im bibliothekarischen Bereich setzte sich die IFLA-Studie »Functional requirements for bibliographic records« der IFLA Study Group on the Functional Requirements for Bibliographic Records (München: Saur, 1998) mit diesen Fragestellungen auseinander. Auf der IFLA-Konferenz in Glasgow wurde die Bildung einer FRBR-Arbeitsgruppe vereinbart, die die Arbeit weiterführen wird. Dabei werden die Entwicklungen im Metadaten-Bereich einzubeziehen sein.

² Die Arbeitsgruppe FRANAR wurde 1999 auf der IFLA-Konferenz in Bangkok als UBCIM-Arbeitsgruppe gebildet.

³ Zunächst ist eine physische Zusammenführung der Daten vereinbart. Zu einem späteren Zeitpunkt kann der gemeinsame Authority File virtuell ausgestaltet werden.

DIE VERFASSERIN

Christel Hengel, Die Deutsche Bibliothek, Deutsche Bibliothek Frankfurt am Main, Arbeitsstelle Normdateien, Adickesallee 1, 60322 Frankfurt am Main, hengel@dbf.ddb.de