

Susann Solberg

Der Gang in die Zeitschriftendatenbank

»es könnt alles so einfach sein, isses aber nicht«¹⁾

Mit dem »Geriatric-Report« und den »Speyerer Schriften zu Gesundheitspolitik und Gesundheitsrecht« startete die Deutsche Nationalbibliothek (DNB) am 1. März 2007 die Onlinekatalogisierung von fortlaufenden Sammelwerken in der Zeitschriftendatenbank (ZDB). Die Titel waren zufällig gewählt und nicht Programm.

Außerdem wurde mit der Integration der bis zum 1. März 2007 im PICA-Iltis-System der DNB bereits vorhandenen Zeitschriften- und Schriftenreihendaten in die ZDB begonnen. Dabei handelt es sich um die so genannten Altdaten, die im Zeitraum von 1913 bis Ende Februar 2007 entstanden sind.

Realisierung der aktiven ZDB-Teilnahme lautete sinngemäß der Auftrag des Projektes »Integration der fortlaufenden Sammelwerke der Deutschen Nationalbibliothek in die Zeitschriftendatenbank (ZDB)«, das im Herbst 2002 die Arbeit aufnahm. Die Planung und Umsetzung wurde in enger Zusammenarbeit vor allem mit den Kolleginnen der Zentralredaktion Titel der ZDB durchgeführt.

Die DNB ist einem funktionierenden Katalogisierungsverbund beigetreten, der auf jahrelang entwickelten und erprobten Regelungen basiert. Diese hat sie dementsprechend übernommen.

Im Rahmen der bestehenden Regelungen galt es, für die darüber hinausgehenden Anforderungen der DNB Lösungen und Konzepte sowohl für die Zukunft, d. h. die aktuelle, laufende Katalogisierung in der ZDB als auch für die Vergangenheit, die Integration der Altdaten zu entwickeln.

Dabei mussten neben technischen und fach-

lichen auch mentale Hürden überwunden werden. Anfangs geäußerte Meinungen wie »das haben wir immer schon so gemacht« und eine durchaus ehrenwerte Form von »Datenrettungsmentalität« mussten aufgrund der Komplexität der Anforderungen und aus wirtschaftlichen Erwägungen aufgegeben werden. Für manch einen »Nationalbibliothekar« ein fast schmerzhafter Prozess. Dieser hat sich aber gelohnt, denn dadurch wurde der Weg frei für neue Denkansätze und für im Projekt gerne so titulierte »radikale Lösungen«.

Im Laufe des Projektes stellte sich heraus, dass insbesondere die geplante Integration der Altdaten, die aus einigen »historischen Datenschichten« bestehen, eine sehr komplexe, manchmal fast unlösbar scheinende Aufgabe war.

Durch ein hohes Maß an Kreativität in Bezug auf Datenstrukturen, Datenmodelle und das Datenformat konnten Lösungen gefunden werden. So wurde z. B. ganz modern ein »Daten-Chill-Out« kreiert, die »virtuelle Datenspülmaschine« diskutiert und »Datensumpfgespenster« sowie »rumhängende Daten« dingfest gemacht. Auch die Datensatzfamilien oder -sippen wurden definiert und erfreuen sich heute zwar nicht großer Beliebtheit, aber großer Aufmerksamkeit.

Um welche Größenordnungen handelte es sich bei diesem Vorhaben?

In Pica-Iltis lagen bis zum 1. März 2007 rund 335.000 Datensätze, also Altdaten vor. Darunter sind u. a. die Datensätze zu aktuell circa 60.000 laufenden Zeitschriftentiteln sowie zu rund 63.000 Schriftenreihen.

Außerdem wurden die Angaben von rund 210.000 konventionellen Katalogzetteln aus

Altdatenintegration in die ZDB

Neue Denkansätze

Altdatenintegration als komplexe Aufgabe

**Vorliegender
Altdatenbestand**

den alten Zettelkatalogen in Leipzig und Frankfurt am Main im Rahmen eines Konversionsprojekts elektronisch erfasst.

Welche Voraussetzungen und Ergebnisse mussten vorliegen, um das Ziel zu erreichen und in die Arbeitsroutine zu überführen?

Die Staatsbibliothek zu Berlin – Preußischer Kulturbesitz ist Trägerin der ZDB, für die Systembetreuung einschließlich der Datenhaltung zeichnet seit dem Jahr 2000 die DNB verantwortlich.

**Zusammenführung von
DNB- und ZDB-
Datenbeständen**

Auf technischer Seite wurde deshalb entschieden, die Datenbestände der DNB und der ZDB aus Gründen einer verbesserten und effizienteren Datenhaltung in einen Gesamtbestand zusammenzuführen. Dieser dient nicht nur der ZDB bzw. ihren Teilnehmern, sondern wird auch als Grundlage für alle Be- und Verarbeitungsaktivitäten der DNB genutzt. Diese Bestandssituation, die sich von derjenigen anderer ZDB-Teilnehmer unterscheidet, musste bei den weiteren Überlegungen berücksichtigt werden.

**Analyse unterschiedlicher
Datenstrukturen**

Die Analyse und das Verstehen der bis dato unterschiedlichen Katalogisierungs- und Abgrenzungspraxis zwischen Monografien und fortlaufenden Sammelwerken der ZDB und der DNB, der unterschiedlichen Datenstrukturen bzw. der Datenmodelle sowie der u. a. daraus resultierenden Abweichungen im Datenformat nahmen zu Beginn einen breiten Raum ein und ergaben u. a. einen sehr nützlichen Material- und Erkenntnispool.

Auf dieser Grundlage wurde die weitere Vorgehensweise erarbeitet. Da zum einen die aktuelle Katalogisierung in der ZDB geplant wurde und zum anderen die bereits vorhandenen Altdateien integriert werden sollten, mussten für denselben Sachverhalt für die

Zukunft zum Teil jeweils andere Konzepte erarbeitet werden, als für die Altdateien.

Dabei war die Vorbereitung der Zukunft der laufenden Katalogisierung in der ZDB sehr arbeitsintensiv, aber im Vergleich zur geplanten Altdateienintegration eher unproblematisch: Übernahme und Anwendung des »ZETA-Format und Konventionen für die Zeitschriften-Titelaufnahme in der Zeitschriften-datenbank« und der entsprechenden Verbundarbeitsweise mit allen Konsequenzen für die Arbeitspraxis, die Geschäftsgänge, die Datenerfassung, das Datenformat, die Datendienste, die Abgrenzung Monografie – fortlaufendes Sammelwerk.

Die Altdateien hingegen weisen durch die bis dato abweichende Praxis der DNB z. B. bei der Definition von Unterreihen, Beilagen, Schriftenreihen und Zeitschriften sowie Titeländerungen andere Datenstrukturen auf als die entsprechenden Datensätze in der ZDB. Deshalb musste hier jeweils überlegt werden, wie diese Datensituation in das bestehende Datenformat und die Datenstrukturen eingepasst werden könnte. Dafür wurde ein Datenmodell entwickelt, das den automatischen wie den manuellen Verfahren zur Altdateienintegration zugrunde liegt.

Eine weitere Herausforderung bestand darin, dass vor allem zu den aktuell rund 60.000 laufenden Zeitschriften der DNB nicht nur über 100.000 Titeldatensätze vorliegen, die in die ZDB integriert werden sollten, sondern diese Titeldatensätze jeweils mit weiteren Datensätzen für den Bestandsnachweis in den Frankfurter und Leipziger Lokalsystemen verknüpft sind. Bei zur ZDB abweichenden Bestandsnachweisen etc. ergaben sich dadurch auf der Ebene der Lokalsysteme weite-

**Übernahme
von ZETA**

**Altdateien-
problematik**

**Verknüpfung
der Bestands-
nachweise**

Entwicklung eines gemeinsamen Datenformats

re Datensätze, die infolge der Altdatenintegration »aufgeräumt« werden müssen.

Aufgrund der zu bewältigenden Datenmengen war von Beginn an klar, dass ein rein intellektuelles und manuelles Verfahren für die Integration der Altdaten in die ZDB aus ökonomischen Gründen nicht in Frage kommen würde.

Ein Kernstück für die künftige Katalogisierung wie die retrospektive Datenintegration ist das gemeinsame Datenformat, das zum 1. März 2007 in der ZDB eingeführt wurde.

Sowohl die ZDB als auch die DNB haben bereits vor dem Start der Katalogisierung mit dem PICA-Datenformat gearbeitet. Allerdings bestanden, auch bedingt durch die bis dahin voneinander abweichende Katalogisierungspraxis, einige Unterschiede in der Anwendung des PICA-Datenformats. Dabei handelte es sich z. B. um einzelne Felder, die mit unterschiedlichen Inhalten belegt sind, aber auch um mehrere Felder zu einzelnen Themenkomplexen wie z. B. zu Sekundärformen oder zu Netzpublikationen. Die DNB hat das von der ZDB verwendete PICA-Datenformat und dementsprechend ZETA übernommen. Es bestand dennoch die Notwendigkeit, das ZDB-Datenformat an den Stellen zu erweitern oder zu ändern, wo die DNB aufgrund der Herstellung der verschiedenen bibliografischen Dienstleistungen sowie durch die gesetzliche Aufgabe der Langzeitarchivierung von Netzpublikationen und die geplante Integration ihrer Altdaten in die ZDB besondere, nicht »abschaffbare« Erfordernisse hat. Dafür galt u. a. die Maxime, dass die ZDB-Teilnehmer in ihrem Alltagsgeschäft möglichst wenig gestört werden sollten. Es sollte so wenig wie möglich und nur so viel wie un-

bedingt nötig geändert oder ergänzt werden. Das Ergebnis ist, gemessen an dem Umfang und der Komplexität der Formatharmonisierung, sehr erfreulich ausgefallen. Die Arbeitsgemeinschaft der Datenbankteilnehmer (AGDBT) der ZDB hat in ihrer 34. Sitzung im November 2006 das neue, gemeinsame Datenformat zustimmend zur Kenntnis genommen²⁾.

Die Altdatenintegration kann aufgrund der zu bearbeitenden Datenmengen nur stufenweise realisiert werden.

Altdatenintegration Stufe 1

In den letzten Jahren wurden die alten Zettelkataloge – in Leipzig für den Zeitraum von 1913 bis 1973 und in Frankfurt am Main die beiden Zeitschriftenkataloge für den Zeitraum von 1945 bis 1971 – konvertiert.³⁾ Es wurden 210.000 Katalogzettel bearbeitet. Als Ergebnis wurden im ZDB-Hauptbestand rund 140.000 Exemplardatensätze erfasst. Im DNB-Hauptbestand wurden rund 113.000 Titel- und Exemplardatensätze von Titeln erfasst, die in der ZDB nicht gefunden wurden (Nichttreffer).

Stufe 2 und Stufe 3 betreffen die bis Ende Februar 2007 bereits im Pica-Iltis-System elektronisch verzeichneten Zeitschriften- und Schriftenreihen (Altdaten) der DNB.

Altdatenintegration Stufe 2:

Der Maschinelle Abgleich (MaschA)

Das Ziel bestand darin, so viel automatische Unterstützung wie möglich für die Integration der Altdaten zu entwickeln und nutzen zu können. Die Altdaten sollten ZDB-konform umgearbeitet und bereinigt werden.

Mit dem MaschA wurde ein automatisiertes

Konvertierte
Zeitschriften-
kataloge

Verfahren entwickelt, das aus einem Programmpaket mit einer so genannten Match- und einer Merge-Komponente besteht. Mithilfe des Verfahrens sollte eine möglichst hohe, verlässliche Trefferquote 1 : 1 erzielt werden.

In mehreren Testläufen wurden die Programme gründlich geprüft und verbessert, sodass die Ergebnisse bei einer relativ niedrigen Fehlerquote stabil und verlässlich waren.

Es wurden vier Datengruppen definiert. Die ersten drei Gruppen, – die Treffer, die neuen Titel und die VÜ-Treffer, – wurden im MaschA geprüft und ggf. bearbeitet. Die Sonderkontingente als vierte Gruppe konnten im MaschA nicht berücksichtigt werden.

Vier Datengruppen

Ergebnis des MaschA

- **Treffer:** In der DNB vorhandene Zeitschriften- und Schriftenreihentitel, die bereits in der ZDB nachgewiesen sind. Der ZDB-Datensatz war der Zielsatz. Die Exemplar- und Bestandsdaten sowie ggf. zusätzliche Informationen (z. B. Sacherschließungsdaten) wurden an den ZDB-Satz angehängt und der Datensatz der DNB wurde gelöscht.

- **Neu(er) Titel:** Für die ZDB neue Titel, für die bei der DNB ein oder mehrere Datensätze vorhanden waren. Der oder die Datensätze der DNB wurden soweit möglich an das neue Datenformat angepasst und in die ZDB eingespielt.

- **VÜ-Treffer:** Vermutliche und/oder zu überarbeitende Treffer, die nicht ohne intellektuelle Überprüfung und ggf. manuelle Aufbereitung automatisch bearbeitet werden können. Sie wurden im MaschA gekennzeichnet, müssen aber intellektuell überprüft und je nach Ergebnis entsprechend analog als Treffer

oder als für die ZDB neuer Titel weiterbearbeitet werden.

- **Sonderkontingente (SoKo):** Diese waren nicht Teil des MaschA. Einzelne definierte Gruppen von Publikationen werden beim maschinellen Abgleich aufgrund sehr abweichender Strukturen etc. nicht berücksichtigt und müssen grundsätzlich »per Hand« integriert werden. Das betrifft z. B. Gruppen wie Netzpublikationen und die elektronischen Publikationen auf physischen Datenträgern. Nachdem am 1. März 2007 der Startschuss für die Katalogisierung in der ZDB auf der Grundlage des gemeinsamen Datenformates

Karger Libri

Zeitschriftenagentur

Ihr Partner für

- die Beschaffung aller Formate von Periodika
- die Reduzierung des administrativen Aufwands Ihrer Abonnementsverwaltung
- effiziente Leistungen und persönliche Kundenbetreuung zu fairen Bedingungen

Besuchen Sie uns unter www.libri.ch oder rufen Sie an – wir beraten Sie gerne.

Karger Libri AG
Internationale Zeitschriftenagentur
Allschwilerstrasse 10
Postfach
CH-4009 Basel/Schweiz
Telefon +41 61 306 15 00
Fax +41 61 306 12 34
E-Mail journals@libri.ch

Mehrere Datenkontingente

gefallen war, wurde der MaschA Ende April begonnen. Aufgrund der durch den Abgleich jeweils zu bewegendenden Datenmengen und daraus resultierenden Laufzeiten mussten die Daten in mehrere Kontingente aufgeteilt werden. Diese wurden, um die ZDB-Teilnehmer nicht über Gebühr zu beeinträchtigen, über mehrere Wochenenden verteilt abgearbeitet. Zuerst wurden die Schriftenreihen und die laufenden Zeitschriften abgeglichen, danach die weiteren Zeitschriftendatenkontingente.

Das Ergebnis in Zahlen:

Rund 335.000 Datensätze wurden im MaschA abgeglichen und rund 23 % als Treffer, rund 26 % als Neu und rund 50 % als VÜ-Treffer bearbeitet.⁴⁾

In der Teilmenge der Schriftenreihendatensätze konnte sogar eine maschinelle Bearbeitungsquote von rund 75 % (Treffer und Neu) erzielt werden.

Erfreulich ist, dass die Hälfte der rund 335.000 Datensätze durch den MaschA bearbeitet werden konnte und die Daten jetzt in der ZDB zur Verfügung stehen.

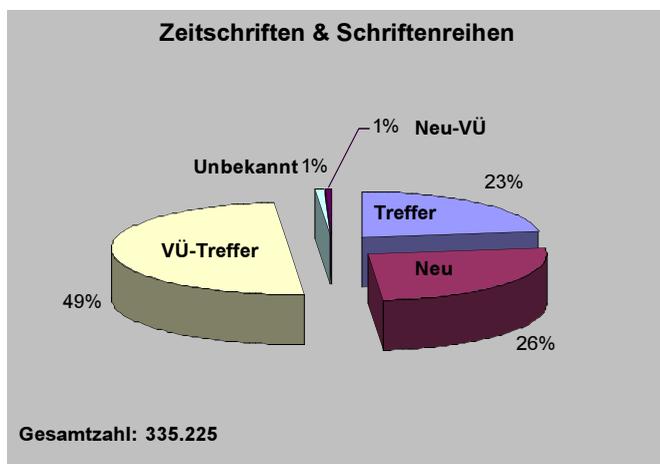
Altdatenintegration Stufe 3: Manuelle Bearbeitung und MaschA-Nachnutzung

Die intellektuelle und manuelle Bearbeitung der restlichen Datensätze, der VÜ-Treffer und der Sonderkontingente, wird trotz des Einsatzes zusätzlichen Personals allerdings einige Jahre in Anspruch nehmen. Mit Priorität werden die laufenden Titel sowie einige Sonderkontingente bearbeitet. Die Integration der Altdatenkontingente ist zum Teil schwierig, dadurch zeitintensiv und erfordert neben Titelaufnahmeerfahrung ein gutes Verständnis von Datenstrukturen gepaart mit Abstraktionsvermögen. Damit die Bearbeitung der VÜ-Treffer und Sonderkontingente nicht zu zeitaufwändig und zu kompliziert wird, wurde der MaschA so konzipiert, dass die Programme auch für die manuelle Bearbeitung nachgenutzt werden können. So muss in vielen Fällen nicht jeder Arbeitsschritt komplett per Hand durchgeführt werden. Zum Beispiel werden Exemplar- und Bestandsdaten oder mit einer Schriftenreihe verknüpfte Stückaufnahmen unter Nachnutzung des MaschA auch hier automatisch umgehängt.

Um die Mitarbeiter in die Lage zu versetzen, zum einen die Katalogisierung in der ZDB sowie zum anderen die Altdatenintegration bewältigen zu können, waren umfangreiche ZETA-Schulungen sowie Schulungen zur Altdatenintegration notwendig. Eine standortübergreifende Arbeitsgruppe erarbeitete ein Schulungskonzept sowie Schulungsunterlagen. Bewusst wurde auf ein Schneeballprinzip für die ZETA-Schulungen verzichtet. Deshalb führte ein kleines Team, bestehend aus Mitgliedern der Arbeitsgruppe, die Schulungen in Leipzig und Frankfurt am Main nach demselben Muster und mit denselben Inhalten.

Zeitintensive Integration restlicher Altdaten

Entwicklung von Schulungskonzepten



ten durch. Da die DNB seit dem 1. März mit dem Gesamtbestand arbeitet und sie Zeitschriften und Schriftenreihen in die ZDB katalogisiert, waren rund 180 Mitarbeiterinnen und Mitarbeiter von dem Wechsel in die ZDB betroffen. Diese sowie weitere Kolleginnen und Kollegen aus anderen Abteilungen wurden in unterschiedlicher Intensität geschult. Die ZDB unterstützte die Schulungsvorbereitung und -durchführung durch das Angebot von zweiwöchigen Hospitationen bei der Zentralredaktion Titel der ZDB sowie durch eintägige Workshops in Leipzig und Frankfurt am Main.

Die jahrelang zum Teil abweichende Katalogisierungspraxis wurde durch die Teilnahme der DNB an der ZDB als gut funktionierendem Katalogisierungsverbund vereinheitlicht. Damit ist die DNB einerseits Nutznießerin der Vorteile der kooperativen Verbundkatalogisierung. Auf der anderen Seite trägt sie zum gegenseitigen Nutzen durch die kooperative Erfassung, (Nach-)Nutzung und Ver-

edelung der in die ZDB eingebrachten Daten für Zeitschriften, zeitschriftenartige Reihen, Zeitungen und Schriftenreihen bei. Sie bringt kontinuierlich eine gleich bleibende bibliografische Qualität in die ZDB ein und reichert die von ihr genutzten oder in die ZDB eingebrachten ZDB-Titelaufnahmen durch zusätzliche Informationen an (z. B. autorisierte ISSN und Key title, RSWK-Ketten, DDC-Notationen, standardmäßige Angabe der Verlagsangabe). Diese Anreicherungen stellen zusammen mit den DNB-Neuaufnahmen für fortlaufende Sammelwerke, vor allem den Titelaufnahmen für die Schriftenreihen, einen Mehrwert für die ZDB und ihre Teilnehmer dar.

Die Mühe hat sich gelohnt! Der herzliche Dank gilt den Kolleginnen und Kollegen der ZDB sowie zahlreichen Kolleginnen und Kollegen der DNB, die in guter Zusammenarbeit, mit viel Engagement und der Bereitschaft, gegen den Strich zu denken, zu dem Gelingen dieses Projektes beigetragen haben.

Synergieeffekte

Anmerkungen

1

Frei nach »Die Fantastischen Vier«
<<http://www.diefantastischenvier.de/>>

2

s. Protokoll der 34. Sitzung der Arbeitsgemeinschaft der Datenbankteilnehmer der Zeitschriftendatenbank (AGDBT)
<http://www.zeitschriftendatenbank.de/downloads/pdf/protokoll_34_agdbt.pdf>

3

Für die Altbestandskonversion wurde mit der ungarischen Firma Medea Services GmbH zusammengearbeitet.

4

Rund 1 % »unbekannt« sind zu überprüfende Titel, zu denen das Programm zu viele mögliche Treffer gefunden hat und keine Zuordnung zu Treffern, Neu oder VÜ durchführen konnte.