

Man versetze sich für einen kurzen Augenblick in ein Wohnzimmer einer Studentin, die eine Hausarbeit in Politologie zum Thema "Der Sozialstaat in der öffentlichen Wahrnehmung der Bundesrepublik Deutschland von 1990 bis 2010" schreibt. Es ist zwei Uhr nachts, die Universitätsbibliothek hat geschlossen und die Studentin muss bis spätestens 14 Uhr am gleichen Tag die Arbeit einreichen. Verzweifelt sucht sie im Online-Katalog nach weiterer Literatur, um die Forschungsdiskussion in den Fußnoten weiter zu differenzieren. Sie findet auch die entsprechenden gedruckten Publikationen, auf die sie leider nicht zugreifen kann, da die Bibliothek geschlossen hat. Dabei hat "ihre" Universitätsbibliothek E-Books lizenziert, die sie allerdings nicht gefunden hat, da die E-Books nicht im Online-Katalog erfasst, sondern nur auf der Homepage der Einrichtung verlinkt sind. Bei der Rückgabe der Arbeit erfährt die Studentin, dass sie eine bessere Note erhalten hätte, wenn sie nur die Forschungsdiskussion differenziert dargestellt hätte.

Eine vollständige Einbindung der lizenzierten E-Books im Online-Katalog ist der Wunsch vieler Bibliotheken. Insbesondere mit der vermehrten Verfügbarkeit konsortial oder national lizenzierter Sammlungen und Pakete ist das Angebot zusehends unübersichtlicher geworden und die bisherige Praxis, elektronische Dokumente parallel zu den gedruckten Exemplaren manuell formal und inhaltlich zu erschließen, stößt an organisatorische und ressourcenbedingte Grenzen. Einer direkten maschinellen Verarbeitung steht das Problem entgegen, dass die bibliografischen Metadaten der E-Books häufig im Format MARC 21 geliefert werden, die lokalen Bibliothekssysteme in der Regel aber mit den Formaten MAB2 und Pica arbeiten und Daten in anderen Formaten nicht verarbeiten können.

Ein Lösungsansatz besteht darin, die bibliografischen Metadaten an zentraler Stelle zu konvertieren und zu harmonisieren und den Bibliotheken im Format MAB2 zur Verfügung zu stellen. Dieser Ansatz ist allerdings nur für statische E-Book-Sammlungen praktikierbar und ist durch den Aufwand der Konversion in „saubere“ MAB2-Daten mit einem deutlichen zeitlichen Verzug verbunden.

Hinzu kommt, dass nicht alle lokalen Bibliothekssysteme diese MAB2-Titeldaten ohne weiteres verarbeiten können: Während bei gedruckten Medien administrative und exemplarbezogene Daten vorliegen und mit dem Titelsatz verknüpft sind, fehlen diese bei den E-Book-Titeldaten und müssen je nach System durch leere Einträge ohne Funktion ergänzt werden. Schließlich bleibt das Problem, wie die Titeldaten von mehreren tausend E-Books und die mit ihnen verknüpften administrativen Schein-Daten z.B. bei Stornierung des Abonnements wieder aus der Bibliothekssoftware entfernt werden. Je nach System ist dies aufwändig oder gar nicht vorgesehen. Lösungen durch direkte Manipulationen der Datenbank bergen aber auch das Risiko, nicht betroffene Daten zu korrumpieren. Dies beinhaltet z.B. das unbeabsichtigte Löschen von Daten und ihren Verknüpfungen.

Es ist also der Einsatz einer Software notwendig, die das Laden von bibliografischen Metadaten in heterogenen Formaten sowie das Löschen von Titeldaten unterstützt. Damit ist auch die Abbildung einer E-Book-Sammlung, die einer ständigen Veränderung unterliegt (z.B. ein „current file“ mit den aktuellsten E-Books aus einer größeren Sammlung), problemlos möglich. Diese Funktionen unterstützen verschiedene lokale Suchmaschinen, deren technische Basis in der Regel eine quelloffene Suchsoftware wie z.B. Lucene¹ ist und

¹ <http://lucene.apache.org>

die mit den lokalen Bibliothekssystemen allein über Webschnittstellen kommunizieren, ansonsten aber eigenständige Softwaresysteme sind.

Die Universitätsbibliothek Mannheim setzt seit November 2009 als erste deutsche Bibliothek das von der Firma Ex Libris entwickelte und vertriebene Produkt Primo produktiv ein. Zentraler Einstiegspunkt für eine Recherche ist die Suche über alle indextierten Felder (der „Google-Schlitz“). Die Suchtreffer werden standardmäßig nach Relevanz sortiert, die sich aus der Häufigkeit der Suchbegriffe im jeweiligen Datensatz errechnet. Weitere Sortiermöglichkeiten sind das Erscheinungsjahr und die Popularität, die sich u.a. aus der Klickhäufigkeit des Treffers errechnet. Auf der rechten Seite des Bildschirms sind Facetten zu sehen, die eine schrittweise Einschränkung der Treffermenge nach konfigurierbaren inhaltlichen und formalen Kategorien erlauben.

The screenshot shows the Primo search interface. At the top, there is a navigation bar with the University of Mannheim logo and the text 'Universitätsbibliothek Mannheim'. Below this, there is a search bar with the text 'the' and a search button 'Los'. To the right of the search bar, there are links for 'Zur Suche', 'Tagging', 'E-Zeitschriften', and 'Hilfe'. Below the search bar, there are several tabs: 'Bücher und mehr', 'Katalog', 'Dawson EBooks', 'Primo Central', 'Safari', and 'FnT + LRC'. Below the tabs, there is a search bar with the text 'the' and a search button 'Los'. To the right of the search bar, there are links for 'Erweiterte Suche' and 'Meine Einstellungen'. Below the search bar, there are search results for 'MAN DAWSON' with 99,832 results. Three results are visible, each with a title, author, and publication information. On the right side, there are facets for 'Meine Ergebnisse verbessern' (Schlagwort and Erscheinungsjahr) and 'Halle Gast' (Anmelden für vollen Zugang).

Abb. 1: Suchfenster von Primo

Neben dem Katalog der UB Mannheim wurden die Titeldaten mehrerer E-Book-Anbieter eingespielt: E-Books aus den Paketen der DFG-geförderten Nationallizenzen, das Angebot des Aggregators DawsonEra sowie das Paket „Safari Tech Books current file“ von Proquest. Die Datenlieferung aus den Nationallizenzen wurde vom GBV aufbereitet und in MAB2 zur Verfügung gestellt. Das Paket „Early English Books“ beispielsweise umfasst ca. 100.000 Datensätze. Das Paket von DawsonEra umfasst ebenfalls ca. 100.000 E-Books namhafter Verlage (u.a. DeGruyter, Elsevier, Springer und Wiley-Blackwell).² Bei DawsonEra gibt es die Besonderheit, dass die E-Books den Studierenden und Wissenschaftlern für einen fünf Minuten langen Preview angeboten und bei Interesse von der Universitätsbibliothek lizenziert und in kurzer Zeit zur Verfügung gestellt werden.

Die Safari Tech Books sind eine vornehmlich dem Bereich der Informatik und Wirtschaftswissenschaften zuzuordnende Sammlung mit E-Books unterschiedlicher Verlage (u.a. O'Reilly Media, John Wiley & Sons, Addison-Wesley).³ Als „current file“ sind in der Lizenz nur die Titel der beiden letzten Jahre enthalten und die Zusammenstellung des Paketes ändert sich ständig.

² <http://www.dawsonera.com/>.

³ <http://www.safaribooksonline.com>.

Die beiden letztgenannten Anbieter bieten Metadaten im Format MARC Exchange an. Alle Daten werden nach dem in Abbildung 2 dargestellten Schema in Primo bearbeitet: Die MAB2- oder MARC-Dateien werden vom Server des Anbieters geladen, in eine Archivdatei gepackt und auf dem Primo-Server abgelegt. Dieses Archiv wird von Primo geharvestet, d.h. im Ursprungsformat geladen und für die weitere Verarbeitung vorbereitet. Anschließend werden die Quelldaten in ein einheitliches Datenformat überführt und angeglichen („normalisiert“). Im letzten Schritt indexiert die Suchsoftware Lucene die entstandenen Daten, die danach für die Recherche in Primo zur Verfügung stehen.

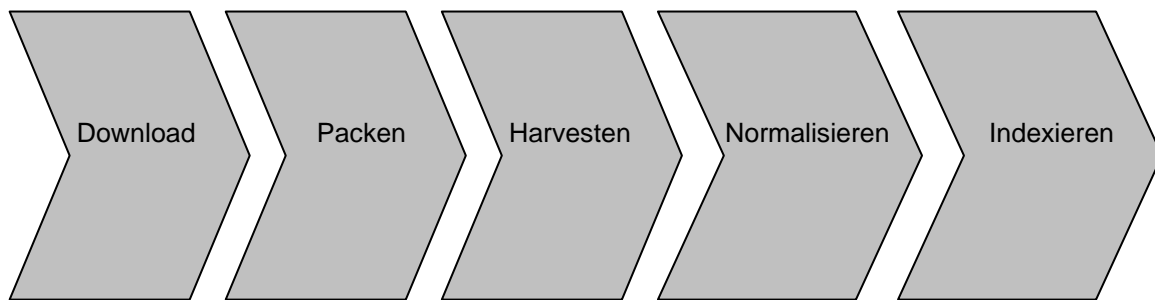


Abb. 2:Workflow für die Einbindung von Metadaten in Primo

Dreh- und Angelpunkt des oben beschriebenen Prozesses sind die Normalisierungsregeln, die die Quelldaten bei der Transformation manipulieren und das dabei zugrunde gelegte Datenformat „Primo Normalized XML“ (PNX). Dieses Format enthält für jeden funktionalen Aspekt, der durch die Suchsoftware realisiert werden soll, eine separate Menge von Feldern. Unterschieden werden u.a. die Bereiche Darstellung, Suche, Sortierung, Verlinkung, Facettierung, Deduplizierung und FRBR. Die Unterfelder sind jeweils an die Anforderungen der Teilfunktion angepasst. So finden sich im Bereich "Darstellung" Felder analog zu der Empfehlung der Dublin Core Metadata Initiative oder im Bereich "Suche" Felder für jeden vom Suchinterface vorgesehenen Schlüssel. Alle Bereiche lassen sich um eigene Felder ergänzen, um lokale Besonderheiten abzubilden und Zusatzfunktionen zu implementieren.

Die Normalisierungsregeln werden mit Hilfe der Primo Publishing Platform festgelegt, die über eine Weboberfläche zugänglich ist. Jeder Datenquelle werden Normalisierungsregeln zugeordnet, die, wenn nötig, individuell für jede Datenquelle gestaltet werden können. Muster für die gängigen bibliografischen Metadatenformate werden vom Anbieter mitgeliefert und dienen als Basis für eigene Anpassungen. Nach den bisherigen Erfahrungen können diese Anpassungen durchaus aufwändig sein, denn die Musterregeln bilden nicht den vollen Umfang der Ausgangsformate ab und können nicht alle Variationen berücksichtigen, die die Auslegung der Katalogisierungsregeln und die Nutzung des Datenformates erlauben. So hat die Universitätsbibliothek Mannheim einen erheblichen Zeitaufwand investiert, um ihre lokalen Titeldaten, die im MAB2-Format vorliegen, vollständig und korrekt nach PNX zu wandeln. Die MARC Exchange Daten von Proquest und DawsonEra hingegen benötigten wesentlich weniger Anpassungen und zur Einbindung der Pakete dieser Anbieter waren nur je zwei Arbeitstage notwendig.

Analog zum Laden der Metadaten gibt es in der Publishing Platform die Möglichkeit, Datenquellen mit einem Mausklick temporär von der Suche auszuschließen oder ganz aus dem Index zu löschen.

Zum heutigen Zeitpunkt können die Nutzerinnen und Nutzer der Universitätsbibliothek Mannheim über das bereits skizzierte Angebot hinaus in ca. 22 Millionen Titeldaten recherchieren, die sich aus den folgenden Datenquellen zusammensetzen:

Datenquelle	Beschreibung	Format	Menge
Katalogdaten	Die im Südwestverbund katalogisierten "klassischen" Titeldaten.	MAB2	1.400.000
Econis	Titeldaten von Aufsätzen und Büchern der Zentralbibliothek für Wirtschaftswissenschaften in Kiel.	MAB2	700.000 (Auswahl)
Online Contents	Aufsatzdaten der Datenbank SwetsWise, die von den deutschen Sondersammelgebietsbibliotheken ergänzt werden.	MAB2	16.000.000 (Auswahl)
Nationallizenzen	Aufsätze und Bücher, die konsortial erworben und von der DFG finanziert wurden.	MAB2	4.000.000
Dawsonera	E-Books des Aggregators Dawson.	MARC 21	100.000
Safari	E-Books des Aggregators Safari.	MARC 21	800
Summe			22.200.800

Die Universitätsbibliothek Mannheim hat die Titeldaten der abonnierten Zeitschriften mit den Aufsatzdaten von Econis und den Online Contents ergänzt. Damit haben ca. 80 % der nachgewiesenen Zeitschriften einen Verweis auf die Aufsätze der letzten 20 Jahre. Diese Quote wurde für die Kernfächer der Universität anhand der von den Fachreferentinnen und Fachreferenten genannten und nach dem Journal Citation Reports gerankten wichtigen Zeitschriften ermittelt.

Die Universitätsbibliothek Mannheim bietet somit einen integrierten Zugriff auf Aufsätze und Bücher in einem Suchsystem an und überwindet damit die künstliche Grenze zwischen den beiden Dokumentarten, die historisch bedingt ist und aus dem Zeitalter der analogen Zettelkataloge stammt. Da alle Titeldaten in lokalen Suchindexen liegen, ist die Antwortzeit von Primo in der Regel unter einer Sekunde. Gerade diese schnellen Antwortzeiten sind ein Vorteil gegenüber der bisher eingesetzten Portaltechnologie, die Suchanfragen über z39/50-Schnittstellen an weitere Datenbanken übergibt und sich hinsichtlich der Antwortgeschwindigkeit des Gesamtsystems immer nach der der abgefragten Datenbanken richten muss.

Dazu kommt, dass die Titeldaten der Aufsätze und Bücher mit den Bestandsinformationen bzw. dem Volltext verknüpft sind und ein unmittelbarer Zugriff auf die Ressourcen möglich ist. Gerade dadurch sind die schwierig recherchierbaren elektronischen Ressourcen für die Nutzerinnen und Nutzer besser als bisher zugänglich. Mit diesem Konzept belebt die Universitätsbibliothek Mannheim den "klassischen" Katalog neu, der einen von außen einsehbaren Bestandsnachweis der von der Bibliothek erworbenen Medien darstellt.

Als Fazit lässt sich festhalten, dass die Software Primo die beschriebenen Probleme bei der Einbindung von E-Books in Bibliothekskataloge überzeugend löst und die Dienstleistung der Universitätsbibliothek erheblich verbessert, da eine Vielzahl von Dokumentarten über eine Suchoberfläche recherchierbar sind. Jetzt kann so auch der Studentin bei ihrer Last-Minute-Recherche geholfen werden, die durch die Online-Ressourcen in der Treffermenge der lokalen Suchmaschine die Forschungsdiskussion detailliert darstellen kann und hoffentlich eine bessere Note für ihre Arbeit erhält.