

Ein schlankes Datenmodell für E-Books Die providerneutrale Beschreibung

Manfred Müller, Bayerische Staatsbibliothek

Zusammenfassung:

Der Artikel zeigt verschiedene Lösungsansätze, wie je nach Prinzipien der Verbund-Datenhaltung eine providerneutrale E-Book-Beschreibung modelliert sein kann. Die skizzierten Datenmodelle bieten einen Ausblick, welche Chancen in der frühzeitigen Adaption des providerneutralen Datenmodells beim Übergang zu den Datenstrukturen in den Cloud-basierten Erschließungsumgebungen bestehen.

Summary:

The article demonstrates current approaches to provider-neutral ebook records in German union catalogs and gives a perspective on how to transfer the outlined data models into cloud-based catalog systems.

Zitierfähiger Link (DOI): <http://dx.doi.org/10.5282/o-bib/2015H4S176-181>

Schlagwörter: E-Book; Metadaten; Datenmodell; Verbundkatalog; Deutschland

Einleitung

Gemäß der in den meisten Verbundkatalogen in Deutschland bisher üblichen Praxis bei der Erfassung oder maschinellen Einspielung von E-Book-Metadaten existieren zu ein- und derselben E-Book-Manifestation mindestens so viele Titelaufnahmen wie es verschiedene Anbieter (Verlage, Aggregatoren etc.) gibt. Diese providerspezifischen Beschreibungen sind damit in gewisser Weise "Dubletten". Sie entsprechen nicht den durch die FRBR und RDA gesetzten Standards und nötigen den Nutzer von regionalen oder überregionalen Nachweis- und Recherchesystemen dazu, aus mehreren identisch aussehenden Treffern bei der Titelrecherche mühsam denjenigen mit dem Besitznachweis und den Zugangsinformationen zu der gewünschten Bibliothek herauszusuchen.

Definition „providerneutral“

Als providerneutrale E-Book-Aufnahme bezeichnet man *eine* umfassende Beschreibung, welche die inhaltlich gleiche E-Book-Ressource als eine Manifestation des Werks abbildet, unabhängig davon, von welchem Anbieter (Verlag, Provider, Aggregator, ...), von wie vielen Anbietern und zu welchem Zeitpunkt die Beschreibung zur Verfügung gestellt wird. Auch unterschiedliche Dateiformate oder abweichende Zugangsbedingungen führen bei einer Anwendung des providerneutralen Prinzips nicht zu separaten Beschreibungen.

Je nach Datenmodell kann man zwei verschiedene Ansätze der providerneutralen Beschreibung feststellen:

1. Entweder werden die providerspezifischen Angaben in der bibliografischen Beschreibung weggelassen und nur in den Lokaldaten abgebildet. Die Aufnahme enthält dann keinerlei auf die Provider bezogene Angaben; sie ist somit im Wortsinne „providerneutral“;
 2. oder die providerspezifischen Angaben werden in der umfassenden Beschreibung addiert. Diese ist dadurch eigentlich „provideraggregiert“.
- Beide Fälle werden in Dokumentationen zusammengefasst und als „providerneutral“ bezeichnet.

Standards

Providerneutrale Beschreibungen von Online-Ressourcen entsprechen den Standards von

- FRBR – Functional Requirements for Bibliographic Records¹ und
- RDA – Resource Description and Access.²

FRBR und in ihrer Folge RDA folgen dem Grundprinzip, dass eine identische E-Book-Ausgabe unabhängig von Anzahl und Art der Provider immer als *eine* Manifestation der Expression eines Werkes anzusehen ist. In den Katalogumgebungen werden primär Beschreibungen auf der Ebene der Manifestation erstellt. Folglich ist eine E-Book-Ausgabe durch *eine* umfassende Beschreibung zu erschließen, auch wenn sie parallel von verschiedenen Providern angeboten wird. Die bibliografische Beschreibung erfolgt demnach je nach Datenmodell providerneutral oder provideraggregiert.

Auf diesem Grundprinzip setzen die Anwendungsrichtlinien großer Erschließungs-Communities für Beschreibungen von Online-Ressourcen auf. Folgende internationale Richtlinien beinhalten Aussagen zur datenmodell- und datenformatgetreuen Umsetzung der providerneutralen Beschreibung:

- Program for Cooperative Cataloging (PCC): Provider-Neutral E-Resource MARC Record Guidelines³
- ExLibris: Alma Community Catalog: Cataloging Standards, Policies, Rights, and Responsibilities⁴
- OCLC: Cataloging Documentation, Practices, and Programs.⁵

Datenelemente

Die bibliografische Beschreibung basiert auf der E-Book-Manifestation der primären Veröffentlichungsinstanz, welche in der Regel der erstveröffentlichende Verlag ist. Datenelemente, die providerspezifische Angaben enthalten, treten hinzu – je nach Datenmodell entweder integriert in die bibliografische Beschreibung und/oder gesondert in den zugeordneten Lokaldatensätzen.

- 1 IFLA (Hg.): Functional Requirements for Bibliographic Records. Final Report, September 1997, as amended and corrected through February 2009. Online-Ausgabe: <http://www.ifa.org/publications/functional-requirements-for-bibliographic-records> (16.10.2015).
- 2 RDA – Resource Description and Access. <http://www.rda-jsc.org> (16.10.2015).
- 3 Program for Cooperative Cataloging (PCC): Provider-Neutral E-Resource MARC Record Guidelines. <http://www.loc.gov/aba/pcc/scs/documents/PCC-PN-guidelines.html> (16.10.2015).
- 4 ExLibris: Alma Community Catalog: Cataloging Standards, Policies, Rights, and Responsibilities. Ex Libris Ltd., 2014. <http://www.exlibrisgroup.com/files/Products/Alma/AlmaCommunityCatalogCatalogingStandardsPolicies.pdf> (16.10.2015).
- 5 OCLC: Cataloging Documentation, Practices, and Programs. <https://www.oclc.org/support/documentation/cataloging.en.html> (16.10.2015).

Providerneutrale/-übergreifende Elemente	Providerspezifische Elemente
Geistige(r) Schöpfer	Provider (Produktsigel)
Titel	Kollektion (Produktsigel)
Verantwortlichkeitsangabe	URL(s)
Ausgabevermerk	Zugangsbedingungen/-beschränkungen
Veröffentlichungsangabe	
Gesamttitelangabe	
Inhalts-/Medien-/Datenträgertyp	
Umfang	
Identifikator/en (ISBN)	Identifikator/en (ISBN)
	Anbieter-IDs
+ Sacherschließung	

Abb.: Datenelemente in den Beschreibungen der E-Book-Manifestation

Datenmodelle in den Verbundkatalogen der Bibliotheksverbände

Im Folgenden werden die vorherrschenden Lösungsansätze skizziert, wie je nach Datenmodell die Beschreibungen für eine E-Book-Ressource bei verschiedenen Providern gestaltet sein können. Die skizzierten Datenmodelle sind so angelegt, dass sie auch im Fall von PDA (Patron Driven Acquisition, nutzergesteuerte Erwerbung) oder beim Vorliegen institutionsspezifischer URLs funktionieren.

Nach Providern getrennt

Separate bibliografische Datensätze für dieselbe E-Book-Ressource für jeden im Verbundkatalog vertretenen Provider. Die Beschreibungen sind jeweils providerspezifisch. Die Zuordnung zu den lizenzierenden Bibliotheken erfolgt über Besitznachweiskennzeichen in der Beschreibung des zur Lizenz passenden Providers. Eventuell vorliegende institutionsspezifische URLs können bei diesem Datenmodell nur im jeweiligen Lokalsystem verankert werden. Anwendender Verbundkatalog: B3Kat.⁶

Nach Providern getrennt plus Lokaldatensätze

Separate bibliografische Datensätze für dieselbe E-Book-Ressource für jeden im Verbund vertretenen Provider. Die Beschreibungen enthalten die auf den Anbieter zutreffenden bibliografischen Elemente inkl. providerbezogene URL. Die Zuordnung der Zugangsinformationen zu den lizenzierenden Bibliotheken erfolgt mithilfe von Lokaldatensätzen. Diese enthalten eine der in der bibliografischen Aufnahme vorhandenen Zugangs-URLs und ggf. weitere bibliotheksspezifische Informationen. Institutionsspezifische URLs können nur über die Lokaldatensätze abgebildet werden. Anwendende Verbände: hbz, HeBIS, SWB.

6 B3Kat: gemeinsame Verbunddatenbank von Bibliotheksverbund Bayern (BVB) und Kooperativem Bibliotheksverbund Berlin-Brandenburg (KOBV).

Providerneutral (provideraggregiert) plus Lokaldatensätze

Zusammengeführter Datensatz für dieselbe E-Book-Ressource aus den Metadaten aller im Verbund durch Lizenzen vertretenen Provider. Die Beschreibung aggregiert die auf alle vertretenen Provider zutreffenden bibliografischen Elemente, unter anderem auch alle URLs der diversen Anbieter. Diese provideraggregierte Form der Datenmodellierung erfordert zusätzlich eine bibliotheks- bzw. lizenzbezogene Darstellung auf der Lokaldatensatzebene. Die bibliotheksbezogenen Lokaldatensätze enthalten die zur Lizenz passende, in der bibliografischen Aufnahme ebenfalls aufgeführte Zugangs-URL und ggf. weitere bibliotheksspezifische Informationen. Anwendende Verbünde: GBV, ZDB (entsprechend dem ZDB-Profil: E-Journals).

Providerneutral (provideraggregiert)

Zusammengeführter Datensatz für dieselbe E-Book-Ressource aus den Metadaten aller im Verbund durch Lizenzen vertretenen Provider. Die Beschreibung aggregiert die auf alle vertretenen Provider zutreffenden bibliografischen Elemente. Im bibliografischen Datensatz wird für jede lizenznehmende Bibliothek ein URL-Feld mit der für die jeweilige Bibliothek gültigen URL plus dem Bibliothekskennzeichen aufgeführt. Bei diesem Modell lassen sich auch institutionsspezifisch gültige URLs im bibliografischen Datensatz abbilden. Das Besitzkennzeichen steuert die Anzeige der auf die eigene Bibliothek bezogenen URL(s) in der lokalen Rechercheoberfläche. Anwendender Verbundkatalog: B3Kat⁶ (Implementierung ist geplant für 2016).

Die Darstellung in den Rechercheoberflächen der Verbünde

Bei einer nach Providern getrennten Präsentation der E-Book-Ressourcen in einem Verbund-Recherchesystem muss der Nutzer eventuell mehrere Beschreibungen in der Vollanzeige konsultieren, um herauszufinden, ob überhaupt eine Beschreibung für die eigene Bibliothek den Zugang zum Volltext bietet und, wenn ja, welche es ist. Dagegen stellt der providerneutrale Ansatz dem Nutzer für eine E-Book-Manifestation lediglich eine Beschreibung zur Verfügung. Die Auswahl unter mehreren Aufnahmen entfällt. Das Erkennen, ob man Zugriff auf den Volltext erhalten kann, erfolgt im Allgemeinen schneller und kann durch visuelle Gestaltung, welche die lizenznehmenden Bibliotheken schneller überblicken und auswählen lässt, unterstützt werden. Speziell bei provideraggregierten Aufnahmen mit ggf. vielen angesammelten URLs und mehreren anbieterspezifischen Datenelementen ist eine optische Aufbereitung zur Unterstützung des Nutzers erforderlich.

Zusammenführung heterogener Metadaten zu providerneutralen Beschreibungen

Die Vorgehensweise, um heterogene Metadaten diverser Datenlieferanten zu providerneutralen bzw. provideraggregierten Aufnahmen zusammenzuführen, gliedert sich in folgende Schritte:

Clustering

Über Elemente wie die Standardnummer ISBN, die persistenten Identifikatoren DOI oder URN, oder die URLs muss festgestellt werden, welche Aufnahmen überhaupt zusammengehören. Wenn zu wenige der vorgenannten Elemente vorhanden sind, muss man eventuell vorhandene Anbieter-Identnummern oder gegebenenfalls auch einen Vergleich von Autor- bzw. Titelement-Zeichenfolgen zusätzlich heranziehen. Zu den weiteren Prüfkriterien gehört das Abgleichen von

Materialcodes, um sicher zu gehen, dass man nur Online-Manifestationen berücksichtigt, und von Erscheinungsjahren, um identische Ausgaben zu identifizieren.

Matching

Wenn das Clustering anhand der oben genannten Kriterien zusammengehörige Aufnahmen festgestellt hat, muss bestimmt werden, welche der Aufnahmen zur Zielaufnahme wird. Das kann beispielsweise erfolgen durch Vergleich der Umfänglichkeit der Aufnahmen oder durch das Auswerten von Zeitangaben zur Erstellung bzw. letzten Aktualisierung der Beschreibung. Die umfassendere, aktuellere Aufnahme gewinnt im Allgemeinen.

Merging

Als letzter Schritt erfolgt dann die Zusammenführung der Aufnahmen, wobei man möglichst dafür sorgen sollte, Qualitätsmerkmale, wie z. B. eine Normdatenverknüpfung und verbale oder klassifikatorische Sacherschließung, in die Zielaufnahme zu übernehmen bzw. in der Zielaufnahme zu erhalten.

Aufwand vs. Nutzen

Die Zusammenführung der Metadaten, welche aus unterschiedlichen Quellen möglicherweise in unterschiedlichen Datenformaten und/oder Formatausprägungen vorliegen sowie gegebenenfalls nach verschiedenen Erschließungsregelwerken (RDA, AACR2, RAK) erzeugt wurden, stellt unzweifelhaft einen administrativen und technischen Aufwand dar. Dieser Aufwand wird aber wieder aufgewogen durch die Tatsache, dass danach nur an *einem* Datensatz Normdatenverknüpfung und Sacherschließung vorgenommen werden müssen, wenn sie nicht ohnehin durch das Merging schon in die Aufnahme transportiert wurden. Eine providerneutrale Aufnahme reduziert zudem dauerhaft den redaktionellen Aufwand in der Katalogumgebung, da weniger Beschreibungen für identische Ressourcen vorhanden und gegebenenfalls zu pflegen sind. Ein weiterer Vorteil ist die schnellere Auswahl der passenden Aufnahme für die Nachnutzung durch weitere Bibliotheken.

Perspektive in Cloud-basierten Katalogisierungsumgebungen

Die Implementierung der skizzierten providerneutralen Datenmodelle bereits in die aktuellen Verbundkatalogsysteme bietet gute Perspektiven, um diese Datenelemente dann beim Umstieg auf Cloud-basierte Katalogumgebungen in deren von Haus aus providerneutral angelegte Datenstrukturen passend zu migrieren. Der providerneutrale bzw. provideraggregierte bibliografische Datensatz trifft in der Cloud auf den *Providerneutral Master Record* in der Community/Network Zone von Alma⁷ oder WMS.⁸ Liegen hingegen im Verbundsystem allgemein noch providergetrennte bibliografische Datensätze vor, so müssten diese vor dem bzw. im Migrationsprozess zusammengeführt werden. Diese Zusammenführung stellt während der Migrationsphase ein weiteres, eigentlich in dieser Phase vermeidbares, Arbeitspaket dar. Die providerspezifischen Angaben und die Zugangsinformationen (URLs) stehen bisher je nach Datenmodell entweder im bibliografischen Datensatz (B3Kat-Konzept) oder in den Lokaldatensätzen. Sie werden bibliotheksbezogen in Alma in den

7 ExLibris Alma. <http://www.exlibrisgroup.com/de/category/Alma> (16.10.2015).

8 OCLC WorldShare Management Services. <https://www.oclc.org/worldshare-management-services.en.html> (16.10.2015).

Inventory Level Record bzw. in WMS in die *Local Bibliographic Data* (providerspezifische Angaben) und in den *Local Holdings Record* (Zugangsinformation, URL) transferiert.

In den Cloud-basierten Systemen ist für alle Arten von Online-Ressourcen, inklusive der E-Book-Produkte, perspektivisch das Bestücken oder sogar das vollständige Generieren der bibliotheks- und providerbezogenen Daten über die Kommunikation mit weiteren Systemmodulen denkbar. Zu diesem Zweck können Tools wie die zugrundeliegenden *Knowledge Bases* oder der *Metadata Collection Manager* benutzt werden. Denn auf deren Grundlage stehen anbieterübergreifend die erforderlichen Metadaten, inkl. deren Updates, zur Verfügung. Als Systemmodul integrierte oder über Schnittstellen angedockte ERM-Systeme⁹ können die Basis bilden, um erforderliche Katalogdatensätze durch den Abgleich mit der Lizenzadministration zu generieren oder auszustatten. Dadurch ist gewährleistet, dass die Metadaten in den Katalogumgebungen passend zum Status und zum inhaltlichen Umfang der Lizenzierung erzeugt und präsentiert werden. Ein weiterer Vorteil ist, dass die Datensätze für das Erreichen dieses Ziels nur einmal, primär im ERMS, administriert werden müssen.

Zusammenfassung

Providerneutrale E-Book-Aufnahmen

- beschreiben den gleichen Inhalt / die identische Manifestation in nur einer Aufnahme,
- entsprechen den aktuellen Standards FRBR und RDA,
- sind Anwendungsfälle für innovative Katalogisierungsumgebungen,
- verursachen einen gewissen Aufwand in der Zusammenführung, reduzieren aber den zukünftigen Pflegeaufwand und
- führen zu nutzerfreundlicherer Präsentation, da eine einfachere Auffindbarkeit des Zugangs zu den E-Books erzielt wird.

9 Electronic Resource Management System (ERMS).