

Erschließung und Datenmanagement

Cloudbasierte Infrastruktur für Bibliotheksdaten (CIB) – Passagen einer Reform

Uwe Risch, HeBIS Verbundzentrale

Zusammenfassung:

Der Beitrag stellt das Projekt „Cloudbasierte Infrastruktur für Bibliotheksdaten“ (CIB) vor und gibt einen Überblick über den bisher erreichten Projektstand.

Summary:

The paper presents the project „Cloud-based infrastructure for library data“ (CIB) and summarizes what has been achieved so far.

Zitierfähiger Link (DOI): <http://dx.doi.org/10.5282/o-bib/2015H4S119-124>

Autorenidentifikation: Risch, Uwe: GND 122592263

Schlagwörter: Cloud Computing; Katalogisierung; Datenmanagement

Das Projekt „Cloudbasierte Infrastruktur für Bibliotheksdaten“ (CIB)¹ zielt auf die Überführung bibliothekarischer Workflows und Dienste in cloudbasierte Arbeitsumgebungen und damit auf die sukzessive Ablösung bestehender Verbunddatenbanken. Ausgangspunkt des am 1.9.2013 gestarteten Projektes ist die heute vorfindliche und von Wissenschaftsrat wie DFG nachdrücklich kritisierte Verfasstheit der bibliothekarischen Infrastruktur in Deutschland. Geprägt von sechs Verbundsystemen und konzipiert auf dem technischen Stand der 1980er Jahre werden hier vor allem regional ausgerichtete Dienstleistungen erbracht, deren technisch-organisatorischer Kern in jedem Verbundsystem letztlich gleich ist.

Der Lösungsansatz des CIB-Projektes macht sich die heute entstehenden Bibliothekssysteme der nächsten Generation zunutze. Angesichts dieser bereits vorfindlichen Systeme sind Eigenentwicklungen, etwa in Form „nationaler“ Katalogisierungssysteme, mit ihren hohen Risiken hinsichtlich Entwicklungsaufwand, institutioneller Verankerung und dauerhafter Pflege nicht zielführend. CIB zielt somit auf eine Modernisierung des traditionellen Kerngeschäfts der Bibliotheken: Katalogisierung, Ausleihe, Erwerbung sind Standardfunktionen jeder Bibliothek und somit (nicht nur prinzipiell) offen für eine Zuarbeit durch externe Dienstleister. „Modernisierung“ bedeutet in diesem Zusammenhang:

1 Gefördert durch die DFG und getragen vom Hessischen Bibliotheksinformationssystem (HeBIS) als Konsortialführer sowie den Bibliotheksverbänden Bayern (BVB) und Berlin-Brandenburg (KOBV).

- die Nutzung des heute im Vergleich zu den Gründungszeiten der Verbände stabilen und schnellen Internets
- ein Aufgreifen der sich in den letzten Jahren entwickelnden Wirtschaftskonzepte bzw. Geschäftsmodelle des so genannten Cloud Computing, in dessen Rahmen mittlerweile auch die großen Bibliothekssystemanbieter standardisierbare Dienste weltweit und damit kostengünstig anbieten können.

Mit der Abkehr von den heute noch dominierenden regionalen und lokalen Systemen eröffnet sich zudem eine Perspektive konsequenter Internationalisierung, die noch befördert und erleichtert wird durch die aktuelle Einführung des internationalen Regelwerks RDA.

All dies wird zu einer Effizienzsteigerung in der Versorgung unserer Nutzerinnen und Nutzer führen. Freiwerdende Mittel können in neue Dienste investiert werden, für die heute weder Personal noch finanzielle Mittel bereitstehen. Sammlung und Bereitstellung neuer Materialarten wie zum Beispiel Forschungsdaten, Langzeitarchivierung, das Vorantreiben der „Konvergenz“ von Bibliotheken, Archiven und Museen oder die Nutzung semantischer Technologien sind Desiderate, die dringend von Bibliotheken angepackt werden müssen.

In Deutschland wird voraussichtlich nicht nur eine einzige bibliothekarische Cloud-Umgebung zum Tragen kommen. Daher strebt das Projekt eine herstellernerneutrale Infrastruktur an, in der unterschiedliche kommerzielle Anbieter eingebunden werden können. Das CIB-Projekt hat von Beginn an eine Zusammenarbeit mit den beiden in Deutschland führenden Herstellern, OCLC (WorldShare Management Services, WMS) und Ex Libris (Alma) gesucht und mit beiden Systemanbietern eine auch die Plattformsynchronisation umfassende Kooperation vereinbart. Angestrebt wird ein möglichst komfortabler Zugriff auf die Daten aller deutschen Bibliotheken unabhängig von der jeweils genutzten Katalogisierungsumgebung. Sowohl die direkt in den WorldCat katalogisierenden Bibliotheken (WMS-Anwender) als auch die Alma-Anwender sollen jederzeit Zugriff auf die bibliografischen Daten der Bibliotheken der jeweils anderen Katalogisierungsumgebung haben. Zwingende Voraussetzung für den Aufbau eines solchen „Einheitlichen deutschen Datenraums“ ist, dass in beiden Umgebungen ein einheitliches Datenmodell implementiert und für die Katalogisierung eingesetzt wird. Die Konzepte für die Synchronisierung der jeweils im WorldCat und in der Alma-Network-Zone verzeichneten Metadaten deutscher Bibliotheken bilden einen Schwerpunkt der Projektarbeiten.

1. Einbindung zentraler deutscher Katalogisierungsressourcen in die internationalen Plattformen

Ohne eine Einbindung zentraler Katalogisierungsressourcen deutscher Bibliotheken können cloud-basierte Plattformen kein ausreichendes Leistungsspektrum bieten.

GND

Sowohl OCLC als auch Ex Libris haben die Integration der GND in ihre Katalogisierungsumgebung zugesagt. In beiden Plattformen sollen identifier-basierte Verknüpfungen zwischen Titel und Normdatensatz möglich sein.

Eine Fachspezifikation der Migration von Normdatensätzen in die Plattformen via MARC-Format ist vom CIB-Konsortium Anfang März 2015 vorgelegt worden. Derzeit wird mit den Anbietern über die konkrete Ausgestaltung der GND-Einbindung verhandelt. Die Integration der GND in die genannten Plattformen soll Ende 2015 fertig gestellt sein.

Katalogisierung von Periodika

Primärkatalogisierung, Datenübernahmen und Datenhaltung aller Materialien und Erscheinungsformen (Monografien und Periodika) erfolgen im CIB-Modell unter einem Dach, also nach einem einheitlichen Regelwerk in den jeweiligen internationalen Katalogisierungsumgebungen. Die heute in Deutschland übliche Aufteilung der Katalogisierung anhand unterschiedlicher Materialtypen erübrigt sich somit. Insbesondere wird die separate Periodikakatalogisierung in der Zeitschriftenbibliothek (ZDB) und, soweit elektronische Ressourcen und deren Lizenzen betroffen sind, in der Elektronischen Zeitschriftenbibliothek (EZB) vermieden. Mit dem Übergang in die internationalen Katalogisierungsumgebungen muss daher die Rolle von ZDB und EZB überdacht werden.

In Gesprächen mit Vertretern der ZDB wurden die unterschiedlichen Positionen festgehalten und vom CIB-Konsortium mit dem Konzept der „Versorgungsschnittstelle“ zur ZDB ein möglicher Weg beschrieben, die ZDB auf qualitativ hohem Niveau auch weiterhin mit allen Titelnenaufnahmen und -änderungen aus dem Bereich der Periodika zu versorgen. Hier werden im nächsten Schritt konkrete Szenarien beschrieben.

2. Datenmanagement und CIB-Datenmodell

Als Grundlage des deutschen Datenraums wird ein gemeinsames Datenmodell benötigt. Bei der im CIB-Projekt angestrebten Vielfalt der cloudbasierten Plattformen besteht ansonsten die Gefahr, dass unterschiedliche Datenmodelle und Formatausprägungen eine einheitliche Katalogisierung durch deutsche Bibliotheken erschweren und zudem für die effiziente, unkomplizierte Übernahme von Daten eine große Hürde darstellen. Das CIB-Projekt hat im März 2015 ein „Datenmodell für Titeldaten“ vorgelegt, in dem zunächst entsprechende Grundanforderungen an die cloudbasierten Katalogisierungsumgebungen beschrieben werden. Dabei wird eine ausgewogene Balance gesucht zwischen einer Anpassung an internationale Gepflogenheiten einerseits und der Sicherung des bisherigen hohen Qualitätsstandards der Metadaten andererseits. Eine zentrale Anforderung ist dabei die Möglichkeit von Datensatzverknüpfungen über eine Identifikationsnummer, womit entsprechende Relationen beispielsweise zwischen Titeldaten- und Normdatensätzen oder auch zwischen verschiedenen Hierarchiestufen einer mehrteiligen Ressource erzeugt werden können.

Das CIB-Datenmodell und die zunächst für die GND vorliegende Feinspezifikation schaffen neben den Voraussetzungen für eine reibungslose Plattformsynchronisation auch die Basis dafür, die in den CIB-Bibliotheken vorhandenen Metadaten ohne Informations- und Serviceverluste in die cloudbasierten Plattformen zu migrieren. Erfreulicherweise konnte sich die AG der Verbundsysteme in ihrer Sitzung vom April 2015 darauf verständigen, dass alle deutschen Verbünde auf der Grundlage des erarbeiteten CIB-Datenmodells ihre Datenlieferungen in den OCLC WorldCat vereinheitlichen und qualitativ zu verbessern suchen.

3. Rechtliche Aspekte des Betriebs cloudbasierter Systeme

Der Betrieb internationaler Katalogisierungsplattformen erfordert keine Verzeichnung personenbezogener Nutzerdaten. Erst cloudbasierte Lokalsystemfunktionen sind, etwa bei der Unterstützung von Ausleihprozessen, auf die Einbindung personenbezogener Daten angewiesen. Innerhalb des CIB-Konsortiums besteht Einigkeit darüber, dass personenbezogene Daten nur dann in cloudbasierten System gehalten werden können, wenn diese Systeme europäischen Datenschutzbestimmungen unterliegen und dies auch von den zuständigen Aufsichtsbehörden abgenommen wurde. Nach intensiven Gesprächen mit Anbietern und den verantwortlichen Stellen in den Landesverwaltungen zeichnen sich hier Lösungen ab. Einen weiteren Schritt konnten die am Projekt beteiligten Berliner Bibliotheken gehen. Deren Verträge mit Ex Libris über die Nutzung von Alma konnten zwischenzeitlich unterzeichnet werden, nachdem von den zuständigen Datenschutzbeauftragten der jeweiligen Hochschule und des Landes ein positives Votum ergangen ist.

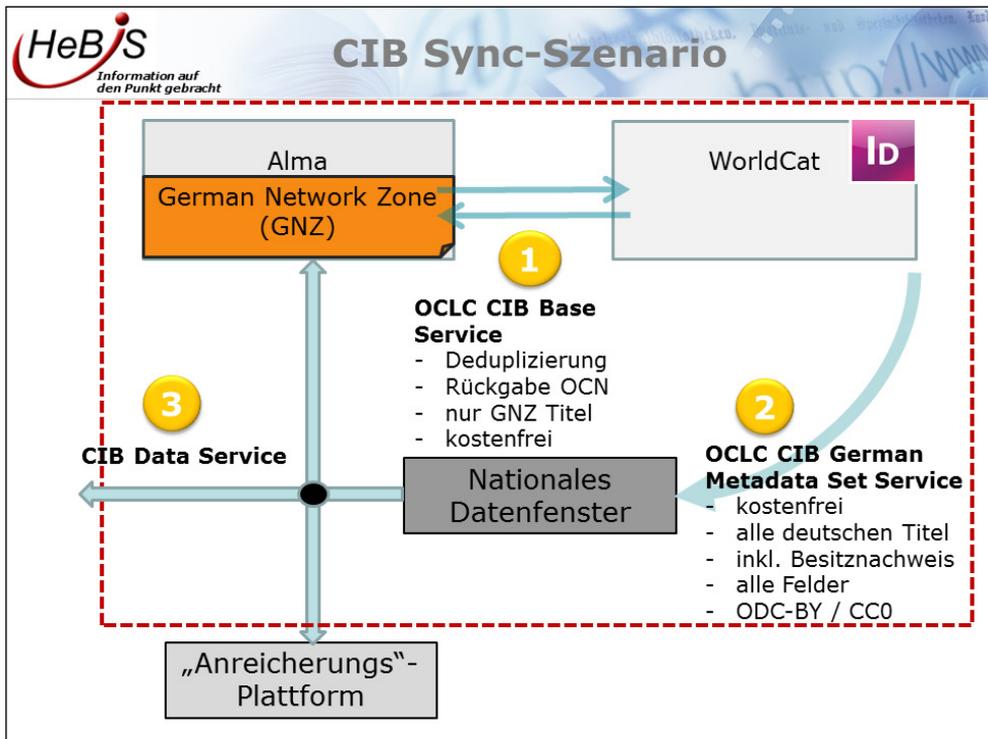
4. Plattform-Synchronisation

Die Pluralität der in Deutschland eingesetzten Katalogisierungsplattformen soll weitgehend erhalten bleiben. Dies impliziert, dass die am Projekt beteiligten Plattformen (derzeit OCLC WorldCat und Ex Libris Alma) miteinander synchronisiert werden, so dass der Titelbestand der teilnehmenden deutschen Bibliotheken in jeder Plattform aktuell und vollständig vorhanden ist. Der damit entstehende einheitliche deutsche Datenraum hat den Vorteil, dass erstmals ein virtueller Nationalkatalog entsteht, ohne jedoch den unzeitgemäßen Weg einer dezidiert nationalen Systemlösung gehen zu müssen. Als weiteres Element dieser Infrastruktur wird ein „Nationales Datenfenster“ aufgebaut, das die Titeldaten mit Bestand in deutschen Bibliotheken aus den Plattformen aufnimmt und als eigenständiges, physikalisches Sicherungssystem dient.

Die Deduplizierung der deutschen Titeldaten erfolgt aus pragmatischen Gründen im WorldCat, da hier der größte Datenbestand gehalten wird. Aus dem gleichen Grund wird als eindeutiger und plattformübergreifender Titelidentifikator die „OCLC Control Number“ (OCN) verwendet.

Einigkeit wurde zwischenzeitlich auch darüber erzielt, dass die Synchronisation der Titeldaten mittelbar über das „Nationale Datenfenster“ stattfinden kann, an das OCLC die deduplizierten und mit einer OCN versehenen „deutschen“ Titeldaten liefert.

Die Synchronisation soll ohne weitere Kosten für die teilnehmenden deutschen Bibliotheken realisiert werden. Die in das „nationale Datenfenster“ gesicherten Daten umfassen alle im OCLC WorldCat verzeichneten Titel mit Besitz in deutschen Bibliotheken, alle Felder/Kategorien der WorldCat-Titelaufnahme und können „by attribution“ (Hinweis auf den Urheber) frei nachgenutzt werden.



Die hier angestrebte Infrastruktur für die Realisierung eines einheitlichen deutschen Datenraums wird von den CIB-Projektpartnern zunächst in Form eines prototypischen Betriebs aufgebaut. Neben der Einbindung des OCLC WorldCat gehört hierzu notwendig auch der Aufbau einer German Network Zone in Ex Libris Alma. Aus der German Network Zone können im Wege der unmittelbaren Synchronisation Titel an den OCLC WorldCat geschickt werden, von wo nach erfolgter Deduplizierung der vergebene Titelidentifikator (OCN) zurückgegeben wird. Die technischen Verfahren der Plattformsynchronisation werden derzeit mit OCLC und Ex Libris abgestimmt, ein Prototyp befindet sich im Aufbau. Die derzeit laufenden Tests einer Ausspielung von Daten aus dem OCLC WorldCat in das „Nationale Datenfenster“ sind vielversprechend.

5. Ausblick

Die grundlegenden Ziele des Projektes – Katalogisierung nach internationalen Standards ohne deutsche Sonderwege, Synchronisation der Plattformen, Herstellerneutralität, datenschutzrechtliche Absicherung – können nach heutiger Kenntnis realisiert werden. Die Verallgemeinerung der innerhalb des CIB-Konsortiums geplanten Infrastruktur ist möglich und wünschenswert, jedoch ohne weitere politische Unterstützung nicht garantiert.

Den Bibliotheken bieten sich derzeit unterschiedliche Möglichkeiten, aus den herkömmlichen Katalogisierungs- und Bibliotheksmanagementsystemen in neue, cloudbasierte Systeme zu wechseln.

Erfolgt dieser Wechsel jeweils individuell, bleiben Synergiepotentiale unberücksichtigt, werden hohe Risiken in der Durchführung von Migrationsprozessen eingegangen und finanzielle Mittel faktisch verschenkt. Statt einer Reform der deutschen Bibliotheksverbände kündigt sich hier womöglich ein schlichter Zerfall und die Entstehung immer neuer Partikularismen an.

Den Bibliotheksverbänden fällt es nach wie vor nicht leicht, angesichts massiver Veränderungen der Nutzererwartungen sowie der technischen Möglichkeiten (Bibliothekssysteme der neuen Generation), eine gemeinsame Reformstrategie zu formulieren. Zu verlockend ist das Beharren auf bestehenden Arrangements.

Derzeit zeichnen sich in Umrissen zwei Entwicklungspfade ab. Die erste Strategie zielt auf eine weitgehend strukturkonservative Beibehaltung der bestehenden Arbeitsteilung zwischen zentralen Verbunddatenbank(en), der Beibehaltung von ZDB und EZB als autonomen Datenquellen zur Erfassung von Periodika bzw. elektronischen Ressourcen und einer bestenfalls selektiven Nachnutzung internationaler Datenquellen auf dem Wege eines Copy-Cataloguings. Die Nutzung von Lokalsystemfunktionen ist in diesem Entwicklungsszenario noch nicht endgültig geklärt. Ex Libris Alma ist für einige (wenige) Bibliotheken bereits fest eingeplant. Jedoch ist die von GBV und HBZ ins Auge gefasste Nutzung von Quali OLE erst im Stadium einer Prüfung. Offenkundige Schwäche dieses Entwicklungspfades ist der Weiterbetrieb des veralteten Bibliothekssystems CBS (ehedem PICA, heute OCLC) und die unklare Nutzungsperspektive für das Lokalsystem Quali OLE. Dass sich letztlich die vielen nationalen Sonderwege des deutschen Bibliothekswesens in diesem Szenario eher zu verfestigen drohen, muss als weiterer Nachteil dieser Strategie gewertet werden. Dennoch: Auch in diesem Szenario lässt sich eine, wenn auch erst in Ankündigungen zu erkennende Abkehr von rein regionalen Verbunddatenbanken erkennen.

Der von den CIB-Partnern eingeschlagene Entwicklungspfad zielt ohne den Umweg einer überregional reformierten Verbunddatenbank auf die Nutzung von Bibliothekssystemen der neuen Generation in einem internationalen Kontext. OCLC WorldCat und eine gemeinsame German Network Zone der deutschen Anwender von Ex Libris Alma sind die beiden Plattformen, die innerhalb des CIB-Modells die Eckpfeiler der Katalogisierung bilden. Eigenentwicklungen oder der Fortbestand hergebrachter Systeme auch über die Abkündigung durch den Systemhersteller hinaus sind in diesem Modell nicht vorgesehen. Synchronisiert über das deutsche Datenfenster bei freiem Zugriff auf alle Titel mit Bestand in deutschen Bibliotheken zeichnet sich hier eine robuste Infrastruktur ab, die noch stärker vereinfacht werden könnte, wenn alle Bibliotheken im WorldCat primär katalogisieren und somit eine Synchronisation der German Network Zone über das deutsche Datenfenster entbehrlich wäre. Nicht verhehlt werden sollte jedoch, dass die Kooperation mit den kommerziellen Systemanbietern nicht immer einfach ist. Auch dort lässt sich insbesondere mit Blick auf die Umsatzzahlen ein vitales Interesse an der Beibehaltung einer Struktur mit möglichst vielen verkauften Instanzen oder Lizenzen beobachten. Die Organisation von Nachfragemacht wäre hilfreich, kommt jedoch derzeit aus politischen Gründen nur schrittweise voran.