

Zum Stand des Projektes »Cloudbasierte Infrastruktur für Bibliotheksdaten« (CIB)

Foto: privat



Mathias Kratzer

Foto: Laura Schulz



Stefan Lohrum

Foto: privat



Uwe Risch

Foto: Simon Wolf



Peter Thiessen

Das Projekt »Cloudbasierte Infrastruktur für Bibliotheksdaten« (CIB) konnte die Entwicklungspfade in eine zukünftig cloudbasierte und international ausgerichtete Infrastruktur eingehend untersuchen und, soweit politisch und technisch möglich, mit konkreten Lösungen versehen. Die erfolgreiche Integration der Gemeinsamen Normdatei (GND) in die Cloudplattformen, die weitgehende Lösung der Datenschutzproblematik, die Erarbeitung eines gemeinsamen Datenmodells und schließlich der Aufbau eines Nationalen Datenfensters waren wichtige Meilensteine. Die Schaffung eines einheitlichen deutschen Datenraumes mithilfe der Plattformsynchronisation konnte nicht gelingen. Die in der Konsequenz vom Projekt empfohlene Primärkatalogisierung im OCLC WorldCat bietet hierzu eine Alternative.

The »Cloudbasierte Infrastruktur für Bibliotheksdaten« (CIB – cloud-based infrastructure for library data) project analysed in detail the development paths of a future cloud-based and internationally-oriented infrastructure and, to the extent politically and technically possible, provided concrete solutions. Important milestones were the successful integration of the Integrated Authority File (GND) into the cloud platforms, the resolution of many data protection problems, the development of a common data model and finally the establishment of a national data window. The creation of a uniform German data area through the use of platform synchronisation proved unsuccessful. Primary cataloguing in the OCLC WorldCat, as recommended by the project, represents an alternative to this.

EINLEITUNG

Im Oktober 2012 hat die Deutsche Forschungsgemeinschaft (DFG) ein Förderprogramm zur Neuausrichtung überregionaler Informationsservices ausgeschrieben, um eine Reform bestehender Infrastrukturen anstoßen, die mit den Empfehlungen des Wissenschaftsrates zur Zukunft der bibliothekarischen Verbundsysteme in Deutschland gefordert wurde. Im Themenfeld »Bibliotheksdateninfrastruktur und Lokale Systeme« der DFG-Ausschreibung wurde das vom Hessischen Bibliotheksinformationssystem (HeBIS), vom Bibliotheksverbund Bayern (BVB) und vom Kooperativen Bibliotheksverbund Berlin-Brandenburg (KOBV) beantragte Vorhaben »Cloudbasierte Infrastruktur für Bibliotheksdaten« (CIB) bewilligt. Das Projekt zielt auf die Überführung bibliothekarischer Workflows und Dienste in cloudbasierte Arbeitsumgebungen und die sukzessive Ablösung traditioneller Verbund- und Lokalsysteme durch internationale Systemplattformen.

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 Entwicklungs-

aufwand, institutioneller Verankerung und dauerhafter Pflege, nicht zielführend. Exemplarisch für die internationalen Cloudsysteme der nächsten Generation stehen Alma von Ex Libris und WorldShare Management Services (WMS) von OCLC. Systeme dieser beiden Anbieter werden bereits heute von der ganz überwiegenden Mehrzahl der deutschen wissenschaftlichen Bibliotheken genutzt. Vor diesem Hintergrund wurden mit beiden Anbietern schriftliche Kooperationsvereinbarungen getroffen, in denen Entwicklungsziele für eine Anpassung dieser Systeme an den deutschen Markt formuliert wurden. Dazu gehörten unter anderem die Einbindung von Normdatenangeboten (insbesondere die Gemeinsame Normdatei, GND) in die neuen Bibliotheksplattformen und die gegenseitige Synchronisation dieser Plattformen.

Mit der angestrebten Synchronisation der internationalen Plattformen ist der Aufbau eines einheitlichen deutschen Datenraumes intendiert. Dieser umfasst die bibliografischen Daten der deutschen Bibliotheken. Als Sicherungskopie wird ein »Nationales Datenfenster«, ein laufend aktualisierter Datenabzug in Form einer physischen Ausspeicherung, aufgebaut.

EINBINDUNG ZENTRALER DEUTSCHER KATALOGISIERUNGSRESSOURCEN IN DIE INTERNATIONALEN PLATTFORMEN

Ohne eine Einbindung zentraler Katalogisierungsressourcen deutscher Bibliotheken können cloudbasierte Plattformen kein ausreichendes Leistungsspektrum bieten. Im deutschen Sprachraum hat sich als kooperativ gepflegtes Instrument die *Gemeinsame Normdatei (GND)* für Ansetzungen von Personen, Körperschaften und Sachbegriffen durchgesetzt. Diese auf den deutschsprachigen Raum ausgerichteten Normdateien müssen auch bei einem Wechsel in internationale ausgerichtete, cloudbasierte Plattformen verfügbar sein.

Sowohl OCLC als auch Ex Libris haben die Integration der GND in ihre Katalogisierungsumgebung zugesagt. In beiden Plattformen sollen Verknüpfungen auf der Basis numerischer Identifikatoren zwischen Titel und Normdatensatz möglich sein. Derzeit (Juni 2016) ist die GND in beiden Plattformen für den lesenden Zugriff eingespielt. Aktualisierungen finden im täglichen Rhythmus statt. Die Nutzung der GND für den

schreibenden Zugriff und damit für Änderungen und Neueinträge ist bisher noch nicht realisiert, steht aber nach Aussage der beiden Anbieter spätestens für 2017 an.

Die Erschließung aller Materialien und Erscheinungsformen (Monografien und Periodika) erfolgt 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 Erscheinungsformen erübrigत sich somit. Insbesondere wird die separate Periodikakatalogisierung in der Zeitschriftendatenbank (ZDB) und, soweit elektronische Ressourcen und deren Lizenzen betroffen sind, in der Elektronischen Zeitschriften Datenbank (EZB) vermieden. Mit dem Übergang in die internationalen Katalogisierungsumgebungen muss deshalb die Rolle von ZDB und EZB überdacht werden.

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. Das CIB-Datenmodell beschreibt das Verhältnis zwischen Norm- und Titeldaten sowie die Ausprägungen des Formats (MARC 21). Ziel ist es, eine plattformunabhängige, einheitliche Katalogisierung zu ermöglichen und eine effiziente, unkomplizierte Übernahme von Daten zu gewährleisten.¹

Das CIB-Projekt hat 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.

Das CIB-Datenmodell und die den Firmen vorgelegten Anforderungen für die Integration der GND 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. Die deutschen Verbünde arbeiten auf der Grundlage des erarbeiteten CIB-Datenmodells aktuell daran, ihre Datenlieferungen in den OCLC WorldCat zu vereinheitlichen und qualitativ zu verbessern.

RECHTLICHE ASPEKTE DES BETRIEBS CLOUDBASIERTER SYSTEME

Innerhalb des CIB-Konsortiums besteht Einigkeit darüber, dass personenbezogene Daten nur dann in cloudbasierten Systemen gehalten werden können, wenn diese Systeme europäischen Datenschutzbestimmungen unterliegen und dies auch von den zuständigen Aufsichtsbehörden insofern abgenommen wurde. Nach intensiven Gesprächen mit Anbietern und den verantwortlichen Stellen in den Landesverwaltungen zeichnen sich hier Lösungen ab. 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 war.

Periodikakatalogisierung

KONZEPT DER PLATTFORM-SYNCHRONISATION

Aus Sicht des Projektes sollte es mit Blick auf den zur Verfügung stehenden Titeldatenbestand für eine deutsche Bibliothek gleichgültig sein, in welcher internationalen Plattform katalogisiert wird – der Titelbestand der teilnehmenden deutschen Bibliotheken sollte in jeder Plattform aktuell und vollständig vorhanden sein. Hierzu wurden im CIB-Projekt Lösungsansätze für eine Synchronisation der Plattformen am Beispiel vom Ex Libris Alma und OCLC WorldCat/WMS gesucht.

gemeinsames Datenmodell

Entwicklung von Lösungsansätzen für die Plattform-Synchronisation

Unmittelbare Synchronisation

Das CIB-Konsortium strebte bei Projektstart eine *unmittelbare Synchronisation* zwischen OCLC WorldCat² und einer (einzigsten) Alma-basierten Katalogisierungsumgebung an. Letztgenannte soll aus Sicht des CIB-Projektes als »German Network Zone« entstehen, in der möglichst alle deutschen Alma-Anwender gemeinsam katalogisieren. Für die Synchronisation zwischen diesen beiden Plattformen sollte neben einer angemessen kurzen Aktualisierungszeit (idealerweise nahezu in Echtzeit) auch die Vollständigkeit der Titeldatensätze (keine Reduzierung der Felder, Besitznachweis zumindest über OCLC-Symbol) und die Einbindung aller Titeldatensätze mit Besitz in deutschen Bibliotheken sichergestellt sein.

Die Projektpartner waren sich sehr schnell darin einig, dass eine zentrale Identifikator-Vergabe im OCLC WorldCat realisiert werden sollte, da hier der weltweit größte Datenbestand für die Deduplizierung zur Verfügung steht. Die Vergabe eines Identifikators garantiert einen einheitlichen Bezugsrahmen für die erfass-

Voraussetzungen für
eine reibungslose
Plattformsynchronisation

kein physikalisches System, sondern eine logische Struktur

ten Titel, der Synchronisationsprozess bestimmt den Umfang der einzubeziehenden Titel (deutsche Bibliotheken) und bildet somit einen »Datenraum«, der kein physikalisches System, sondern eine logische Struktur darstellt. Dieser einheitliche deutsche Datenraum meint eine Qualität der Daten *in den internationalen Plattformen* und nicht etwa ein nationales Katalogisierungssystem. Ein solcher Datenraum hat den Vorteil, dass erstmals ein virtueller Nationalkatalog entsteht, ohne jedoch den unzeitgemäßen Weg einer dediziert nationalen Systemlösung gehen zu müssen. Der einheitliche deutsche Datenraum zielte zudem auf eine Neutralisierung von Anbietermacht, da es für eine (deutsche) Bibliothek mit Blick auf den zur Verfügung stehenden Titeldatenbestand unerheblich sein sollte, welche der synchronisierten Katalogisierungsplattformen lizenziert wird.

Im weiteren Projektverlauf wurde deutlich, dass sich eine unmittelbare Synchronisation nicht realisieren lässt. Ex Libris veränderte zu Projektbeginn die Architektur der Alma-Plattform, so dass sich in den neuen Funktionsbestimmungen von »Community« und »Network Zone« zwar Vorteile für Bibliotheksverbünde, nicht aber für das Synchronisationsanliegen des CIB-Projektes ergaben. OCLC konnte zudem der unmit-

telbaren Lieferung aller Titeldaten mit Besitz in deutschen Bibliotheken an eine deutsche Alma-Katalogisierungsumgebung nicht zustimmen.

Jedoch bot OCLC mit dem »OCLC CIB Base Service« einen Dienst an, um Titeldaten aus einer Alma German Network Zone (GNZ) in den WorldCat zu importieren, diese zu deduplizieren und die vergebene OCN (die »OCLC Control Number« als eindeutigen numerischen Identifikator) an die German Network Zone zurückzuliefern. Damit ist im Titeldatenbestand der German Network Zone jedoch einzig die OCN als Verweis auf einen Titel im WorldCat existent. Die über die bloße OCN hinausgehenden Elemente eines WorldCat-Titeleintrags werden von OCLC nicht geliefert. Zudem ist der »OCLC CIB Base Service« systematisch begrenzt auf Titel, die innerhalb der German Network Zone bereits vorliegen – ausschließlich im WorldCat erfasste Titel werde nicht an die German Network Zone weitergeleitet.

Mittelbare Synchronisation

Nach sehr eingehenden Gesprächen verständigten sich die Projektpartner auf ein Konzept der *mittelbaren* Plattformsynchronisation, welches die dem OCLC Base Service immanenten Funktionslücken für den Aufbau

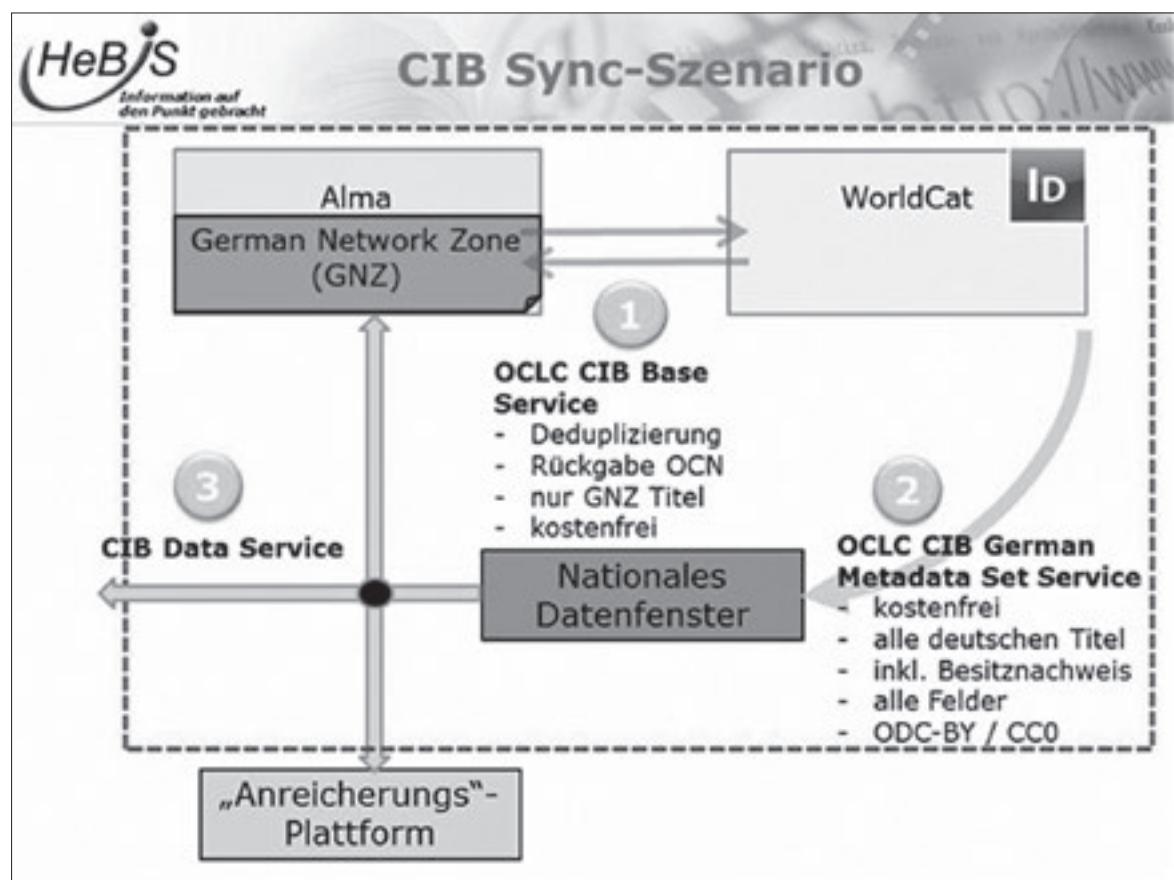


Abb. 1: CIB-Sync-Szenario

eines einheitlichen Deutschen Datenraums komponieren sollte.

Dieses »CIB-Sync-Szenario« (vgl. Abb. 1) sieht eine aktiver Rolle des Nationalen Datenfensters vor. Ursprünglich konzipiert als reines Datensicherungssystem soll es nun zusätzlich eine funktionale Stellung innerhalb der Plattformsynchronisation als Verteilstation für Titeldaten einnehmen. OCLC sicherte zu, das Nationale Datenfenster über einen »OCLC CIB German Metadata Set Service« mit allen Titeldaten deutscher Bibliotheken zu versorgen. Der Aktualisierungszyklus liegt derzeit bei 24 Stunden. Der »OCLC CIB German Metadata Set Service« wird von einer Institution lizenziert und von OCLC ohne weitere Kosten für die in Anspruch nehmende Institution erbracht. Zudem werden keine Beschränkungen des Feldumfangs an den Titeln vorgenommen: Die im WorldCat enthaltenen Titel werden inklusive aller, auch der angereicherten Felder zurückgeliefert (full fieldset). Ebenso wenig erfährt der Umfang der zurückgelieferten Titel eine Begrenzung: Es werden alle Titel deutscher Bibliotheken zurückgeliefert (full recordset). Die Lizenz der zurückgelieferten Titel ist ODC-BY³. Eine Einschränkung in der Nachnutzung dieser Titel existiert daher nur insoweit, als der Ursprung der Daten erkennbar sein muss.

Ausgehend vom Nationalen Datenfenster können die Daten nun an weitere Datenkonsumenten, darunter auch die German Network Zone, weitergereicht werden. Aufgesetzt als kontinuierlicher Lieferprozess entsteht so ein Datenkreislauf, in dem idealerweise alle beteiligten Datenquellen jeweils relevante Änderungen weitergeben bzw. übernehmen können.

Das beschriebene »CIB-Sync-Szenario« entspricht den funktionalen Anforderungen des CIB-Konsortiums. Das Nationale Datenfenster kann seine Sicherungsfunktion erfüllen und steht zudem für eine mittelbare Belieferung der Alma German Network Zone oder weiterer Plattformen mit den deutschen Titeldaten des WorldCat zur Verfügung. Die nun aktive Funktion des Nationalen Datenfensters innerhalb des Synchronisationsprozesses hat zwar Vorteile bei der Skalierung der Synchronisation – eine Einbindung weiterer Plattformen ist verhältnismäßig leicht durchzuführen. Jedoch führt dieses Verfahren im Vergleich zu einer direkten Synchronisation zwischen WorldCat und Alma GNZ zu einer wesentlich höheren Komplexität und ist von daher auch wartungsbedürftiger.

Erprobung des CIB-Sync-Szenarios

Das CIB-Konsortium hat seit März 2015 dieses Konzept der mittelbaren Synchronisation prototypisch umgesetzt. Einer eingehenden Prüfung wurden dabei unterzogen:

1. der OCLC CIB German Metadata Set Service,
2. das prototypisch aufgebaute Nationale Datenfenster,
3. die für den OCLC CIB Base Service wichtigen Einspielungs-, Deduplizierungs- und Merge-Prozesse des OCLC WorldCat.

aktiver Rolle des Nationalen Datenfensters

OCLC CIB German Metadata Set Service

Der Dienst »OCLC CIB German Metadata Set Service« basiert auf dem OCLC-Produkt »Collection Manager«, das für die Lieferung der Titeldaten aus dem WorldCat in das Nationale Datenfenster genutzt wird. Der Collection Manager lieferte während der Tests die Titeldaten im Wesentlichen vollständig. Die Bereitstellung der Daten erfolgt generell 24 Stunden nach Anforderung. Dieser Bereitstellungszyklus wird von OCLC als Standard vorgegeben.

Die Prozesskontrolle war in der vorliegenden Form jedoch nicht ausreichend. Grundsätzlich ist der OCLC Collection Manager ausgelegt für die Lieferungen einzelner Bibliotheksbestände, nicht jedoch für eine Aggregation von Bibliotheksbeständen, die wie im vorliegenden Fall alle deutschen Bibliotheken etwa in Form einer Datenabfrage für die Bibliotheksgruppe »GER« umfasst. Eine solche Möglichkeit zur aggregierenden Datenanforderung müsste von OCLC nachprogrammiert werden. Während der Tests wurde ein Workaround (Anforderung der Daten über eine kombinierte Anfrage über alle Besitzkennzeichen) genutzt. Inkonsistenzen zeigten sich bei Löschungen und Hinzufügungen von Besitzkennzeichen, da der Collection Manager mit seinem Fokus auf die Einzelinstitution Besitzänderungen innerhalb einer Aggregation von Bibliotheken nicht korrekt abbildet. Weitere Probleme ergeben sich aus dem in den Datenlieferungen angegebenen Feld MARC 005 (»Date and Time of Latest Transaction«), in dem das Datum der Ausspeicherung (Übertragungsdatum) und nicht das Datum der letzten Änderung am Masterdatensatz festgehalten ist. Damit wird die Behandlung von Versionskonflikten im Laufe des Synchronisationszyklus erschwert.

Prozesskontrolle nicht ausreichend

Nationales Datenfenster

Der Prototyp des Nationalen Datenfensters als Sicherungs- und Verteilsystem wurde vereinbarungsgemäß bei HeBIS implementiert, wo auch der Dienst OCLC CIB German Metadata Set Service lizenziert wird. Das Nationale Datenfenster ist aktuell in der Lage, Titeldaten aus dem WordCat entgegenzunehmen, zu validieren und abzuspeichern. Es verfügt zudem über eine Standardabfrage- und -lieferschnittstelle (nach OAI-PMH). Die Tests mit rund 12 Mio. Datensätzen verliefen er-

wesentlich höhere Komplexität auch wartungsbedürftiger

folgreich. Der Prototyp des Nationalen Datenfensters ist betriebsbereit.

Einspielungs-, Deduplizierungs- und Merge-Prozesse des OCLC WorldCat

Prozesse innerhalb des OCLC WorldCat nicht zufriedenstellend

Die Einspielungs-, Deduplizierungs- und Merge-Prozesse innerhalb des OCLC WorldCat sind aus der Sicht des Projektes nicht nachvollziehbar und entsprechen in ihren Resultaten nicht den Qualitätserwartungen. Eine Überprüfung des Match&Merge-Verfahrens im WorldCat, die im Rahmen des CIB-Projekts in der Verbundzentrale des BVB durchgeführt wurde, hat u.a. zu folgenden Ergebnissen geführt:

- In signifikanter Höhe konnten bei Lieferungen von E-Books falsche Matches mit Printpublikationen festgestellt werden. In geringer Anzahl wurden auch Datensätze, die unterschiedliche Auflagen beschreiben, als Dubletten eingestuft.
- Trotz dieser Fehler kann das Verfahren allerdings als relativ konservativ und sicher eingestuft werden. Der im WorldCat praktizierte konservative Ansatz der Dublettenerkennung führt infolge jedoch auch zu einer relativ hohen Quote von 35–40 % an Neuaufnahmen trotz vorhandener Matchingkandidaten/Dubletten.
- Wurden gelieferte Datensätze als Neuaufnahmen in den WorldCat eingespielt, wird in der Regel der Großteil des Feldspektrums übernommen. Beobachtete Ausnahmen von dieser Regel waren kausal nicht nachzuvollziehen und bedeuteten in einzelnen Fällen einen Informationsverlust im Datensatz.
- Auch bei der Zusammenführung der gelieferten Datensätze mit vorhandenen Datensätzen im WorldCat waren problematische Fälle zu beobachten, in denen ein Informationsverlust in Titeldaten (einzelne Feldinhalte oder komplett Felder, die nicht übernommen/angereichert wurden) zu konstatieren ist.

Es wurde deutlich, dass die Dokumentation und die angestrebten Parametrisierungsmöglichkeiten dieser Einspielprozesse Verfahrensverbesserungen im OCLC WorldCat voraussetzen, deren Umsetzung derzeit weder zeitlich noch sachlich abgeschätzt werden kann.

Einspielungs- und Match&Merge-Prozesse innerhalb der German Network Zone

Ex Libris hat sich bereit erklärt, in einem ersten Test rund 500.000 Titeldaten aus dem WorldCat in eine German Network Zone einzuspielen. Das Erfolgskriterium wurde mit einer »number of issues < 1%« angegeben. Diese Forderung nach einer Fehlerquote von unter 1% ist angesichts der Komplexität von Ein-

spielprozessen der vorliegenden Art nur ausgesprochen schwer erreichbar und hätte absehbar erhebliche Aufwände und damit deutliche Verzögerungen im Projektablauf impliziert.

Erprobungsergebnisse

Die Tests der mittelbaren Plattformsynchronisation enthielten zwei gravierende Umsetzungsprobleme:

1. Die Einspielungs-, Deduplizierungs- und Merge-Prozesse innerhalb des OCLC WorldCat sind nach derzeitigem Stand weder ausreichend dokumentiert noch können Datenlieferungen an den WorldCat in der notwendigen Form parametrisiert werden. In der Konsequenz stellen sich Einspielvorgänge – und damit insbesondere auch Datenlieferungen im Rahmen der Plattformsynchronisation – als derzeit nicht kontrollierbar dar.

2. Die Datenlieferung aus dem WorldCat via Collection Manager ist mit einem Zeitverzug von 24 Stunden für Synchronisationsprozesse bei Weitem zu langsam. Im Verlauf eines solchen (mindestens) 24-stündigen Synchronisationszyklus können in beiden Datenquellen Änderungen am Ausgangstitel entstehen. Die Anzahl dieser Änderungen wächst mit der Dauer des Synchronisationszyklus. Das Zusammenführen konfliktierender Änderungen am gleichen Titel lässt sich jedoch mit vertretbarem Aufwand in einem solchen Szenario nicht mehr erreichen. Eine pragmatische Minimierung konfliktierender Änderungen etwa durch eine sehr deutliche Verkürzung des Synchronisationszyklus und durch striktere Kontrolle der Deduplizierungs- und Zusammenführungsprozesse (Match&Merge) in den Plattformen ist unter den heutigen Bedingungen leider keine realistische Option.

Als Konsequenz aus den unter 1. und 2. genannten Punkten drohen innerhalb des angestrebten CIB-Sync-Szenarios nicht mehr beherrschbare Konflikte im Bereich der Zusammenführung von Titeldatensätzen. Die technisch zugrundeliegende »Master-Master-Synchronisation«, die Synchronisation zweier autonomer Datenquellen – hier in Form der Ex Libris Alma German Network Zone und des OCLC WorldCat – kann unter den beschriebenen Bedingungen die notwendige Datenqualität nicht garantieren. Nur nebenbei sei erwähnt, dass sich diese Probleme in gleicher Weise ergäben, wenn statt internationaler Plattformen nationale Katalogisierungsdatenbanken synchronisiert werden müssten.

Als Ergebnis der dargelegten intensiven Verhandlungen mit den Systemanbietern, der umfassenden konzeptionellen Arbeiten sowie der detaillierten Tests muss das Konzept der Plattformsynchronisation als nicht realisierbar bewertet werden.

Erprobungsergebnisse

notwendige Datenqualität nicht garantiert

Verlässt man die hier in den Vordergrund gehobene technisch-funktionale Betrachtungsebene mit ihrer Fülle an Test- und Evaluierungsergebnissen ergeben sich weitere Aspekte. Ein Blick auf die Handlungsstrategien der einzelnen Akteure führt zu Einsichten in strukturelle Hindernisse für das Vorhaben der Plattformsynchronisation. Innerhalb der im CIB-Projekt eingegangenen öffentlich-privaten Partnerschaft bestehen die ökonomischen Interessen der Systemanbieter notwendig fort. Aus Sicht von OCLC als einem Anbieter mit einem starken Geschäftsfeldfokus auf Daten ist die kostenlose Weitergabe von Daten im Zuge der Synchronisation nur schwer vorstellbar. Gleiches gilt bei Ex Libris als einem Unternehmen mit Geschäftsfeldfokus auf Anwendungssoftware⁴ für die angestrebte Reduzierung von Alma Network Zonen auf möglichst eine einzige Instanz in Deutschland. Mit der Plattformsynchronisation versuchte das CIB-Konsortium eine herstellerunabhängige, für alle Bibliotheken gleichermaßen nutzbringende Infrastruktur zu entwickeln – ein auch im internationalen Vergleich sehr ambitioniertes Vorhaben. Die beiden beteiligten Systemanbieter hätten im Rahmen dieser Infrastruktur mit verallgemeinerungsfähigen »benefits« für die deutschen Bibliotheken dauerhaft in Lösungen investieren müssen, die ihren ökonomischen Interessen nur begrenzt entsprechen. Es gelang in dieser Konstellation leider nicht, zu einer tragfähigen und nachhaltigen Auslegung der Plattformsynchronisation zu kommen. In diesem Zusammenhang keineswegs hilfreich war die Zersplitterung der Nachfrage- und Verhandlungsmacht deutscher Bibliotheken gegenüber den Systemanbietern. In einer von Partikularinteressen dominierten Verhandlungsumgebung ist eine Begrenzung ökonomischer Interessen der Anbieter nicht realisierbar.

Das CIB-Konsortium nimmt daher Abstand vom Vorhaben der Plattformsynchronisation und empfiehlt die Primärkatalogisierung im OCLC WorldCat. Dieser Lösungsansatz überlässt die Wahl des Bibliotheksmanagementsystems weiterhin den Bibliotheken, würde jedoch die Synchronisationsprobleme deutlich reduzieren. Primärkatalogisierung im WorldCat bezieht sich auf kooperativ nutzbare bibliografische Titelaufnahmen und bedeutet, dass Neuaufnahmen und Änderungen an Datensätzen ausschließlich im WorldCat vorzunehmen sind. Der WorldCat dient somit als *primäre* Datenquelle für die Haltung und Pflege bibliografischer Titeldatensätze. Andere Bibliothekssysteme erhalten ihre bibliografischen Titeldatensätze aus dem WorldCat und erlauben idealerweise keine internen Modifikationen an den gelieferten Daten. Feldinhalte der bibliografischen Titeldatensätze werden

ausgehend von der Primärdatenquelle WorldCat eins zu eins auf den nachgeordnet angeschlossenen Plattformen abgebildet.

Dies bedeutet gleichzeitig, dass für andere Katalogisierungs- und Bibliothekssysteme als das nativ mit dem WorldCat arbeitende WMS-System eine Versorgung mit Titeldatensätzen anhand eines Master-Slave-Synchronisationsmodells erfolgen muss. Entsprechende performante Schnittstellen müssten für den WorldCat zur Verfügung gestellt bzw. entwickelt werden. Auch wäre ein für Bibliotheken bzw. Verbünde attraktives Lizenzierungsmodell OCLCs für eine solche Primärkatalogisierung notwendig. Ein erstes Angebot OCLCs für eine solche »Bundeslizenz« liegt mittlerweile vor, bedarf aber weiterer Ausarbeitungen.

Besondere Beachtung verdient die Nutzung der jeweils plattformspezifischen Knowledge Bases in einem Szenario der Primärkatalogisierung im WorldCat. In allen derzeit am Markt vertretenen cloudbasierten Bibliothekssystemen wird für elektronische Ressourcen konzeptionell sauber zwischen Lizenz- und bibliografischen Metadaten getrennt: Während letztere in einer klassischen Katalogkomponente vorgehalten und bearbeitet werden, verwaltet man erstere in einer vom jeweiligen Hersteller auf Grundlage von Verlags- bzw. Anbieterinformationen gepflegten Knowledge Base. Die bibliografischen Metadatensätze sollten dabei – im Einklang mit den im CIB-Projekt getroffenen Verabredungen – als eindeutigen Identifikator eine OCLC-Nummer (OCN) erhalten. Solange die OCN in jeder Knowledge Base als Metadatenfeld enthalten ist, sind Primärkatalogisierung im WorldCat und effiziente E-Ressourcen-Verwaltung über die Knowledge Base der individuell gewählten Cloudlösung durchaus miteinander vereinbar. Das Aktivieren einer E-Ressource mit gegebener OCN muss dazu lediglich das Setzen des eigenen Besitznachweises an dem durch die OCN identifizierten WorldCat Master Record und ggf. dessen Download in einen lokalen Titeldatenspiegel auslösen – eine Funktion, die in Nicht-OCLC-Systemen derzeit allerdings noch nicht implementiert ist. Mit Realisierung einer solchen Abgleichfunktion stünden dann auch alle kooperativ erfassten Anreicherungen am Titel zur Verfügung.⁵

Gerade in Bezug auf die Nutzung von Knowledge Bases sieht das CIB-Projekt weiteren Klärungsbedarf, der über das Konzept der Primärkatalogisierung im WorldCat und auch über den bisherigen Zuschnitt des CIB-Projektes hinausgeht. Die zunehmende Bedeutung von Knowledge Bases indiziert einen Bedeutungsverlust manueller Verfahren kooperativer Katalogisierung, wie sie heute noch insbesondere im Bereich der analogen Medien dominieren. Automatisier-

strukturelle Hindernisse für das Vorhaben der Plattformsynchronisation

Zersplitterung der Nachfrage- und Verhandlungsmacht deutscher Bibliotheken

Neuaufnahmen und Änderungen an Datensätzen ausschließlich im WorldCat

Wasserzeichen – Schreiber – Provenienzen

Neue Methoden zur Erforschung und Erschließung von Kulturgut im digitalen Zeitalter: zwischen wissenschaftlicher Spezialdisziplin und »catalog enrichment«
Hrsg. von Wolfgang Eckhardt, Julia Neumann, Tobias Schwinger und Alexander Staub
2016. 322 Seiten, gebunden, Fadenheftung
ISBN 978-3-465-04257-0
ZfBB Sonderband 118
Auch als E-Book erhältlich

Die Erforschung der musikalischen Quellen gehört zu den Grundlagen des Faches Musikwissenschaft. Aufschlussreich für alle ihre Gebiete ist eine erweiterte Datenbasis mit tief erschlossenen Musikquellen. Mit dem Pilotprojekt »Kompetenzzentrum Forschung und Information Musik« (KoFIM) an der Staatsbibliothek zu Berlin – Stiftung Preußischer Kulturbesitz soll jetzt die musikalische Quellenforschung durch eine Bibliothek mit großem Musikalienbestand entscheidend vorangebracht werden. Dabei sollen neue Methoden erprobt und entsprechende Geschäftsgänge für die Erschließung entwickelt werden. Das Projekt dient der Tiefenerschließung von bislang noch nicht ausreichend katalogisierten Musikhandschriftenbeständen und der Etablierung eines *catalog enrichment* auf dem Gebiet der digitalen Dokumentation von Schreiberhänden und Wasserzeichen. Im Rahmen dieses von der DFG geförderten Projekts fand im Oktober 2014 an der Staatsbibliothek zu Berlin ein wissenschaftliches Kolloquium statt, das seitens des Fachpublikums auf großes Interesse stieß. Dieser ZfBB-Sonderband dokumentiert die anlässlich des Kolloquiums entstandenen Beiträge.



VITTORIO KLOSTERMANN

te und damit kostengünstigere Erschließungsprozesse können bei digitalen Medien sehr viel leichter greifen. Dies macht Titelanreicherungen für digitale Medien nicht weniger nützlich. Für die Übernahme solcher Anreicherungen werden jedoch plattformübergreifend eindeutige Identifikatoren für die bibliografischen Metadatensätze benötigt, die aus Sicht des CIB-Projektes über die Vergabe einer OCLC Control Number abgebildet werden sollten. Hier besteht weiterhin ein Bedarf für plattformübergreifende Regelungen.

Nicht verschwiegen werden sollte in diesem Zusammenhang, dass im Markt für Bibliothekssysteme zunehmend Anbieter mit einem eigenen Interesse am Vertrieb von Publikationen agieren.⁶ Die bisher als weitgehend anbieterneutral geltenden Knowledge Bases werden mit einer solchen Vermischung von Content- und Technologie-Interessen zumindest der Gefahr ausgesetzt, zukünftig stärker als interessengebundener Vertriebskanal für elektronische Ressourcen zu fungieren. Es lässt sich heute noch nicht abschätzen, welche konkreten Auswirkungen diese Entwicklung zukünftig haben wird.

Schließlich muss festgestellt werden, dass die fortschreitende Dominanz digitaler Medien nicht nur dazu führt, dass Aufwände für Verfahren zur Katalogisierung und zur Plattformsynchronisation zumindest in Teilbereichen kritischer bewertet werden müssen. Wichtig ist auch, dass aus Nutzersicht wünschenswerte Verbesserungen bei Recherche und Informationsaufbereitung zwar effiziente Katalogisierungssysteme voraussetzen, weitergehende und innovative Dienste im Bereich Recherche, Anreicherung oder Kontextualisierung aber durchaus jenseits der großen und notwendigerweise stets um standardisierte Dienste bemühten Plattformen entstehen können. Die im CIB-Projekt erarbeitete Konstellation von cloudbasierten Plattformen und einem Nationalen Datenfenster bietet hierzu einen guten Startpunkt.

A U S B L I C K

Der von den CIB-Partnern eingeschlagene Entwicklungspfad zielt ohne den Umweg einer in welcher Form auch immer 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 Weiterbetrieb hergebrachter Systeme auch über die Abkündigung durch den Systemhersteller hinaus sind in diesem Modell nicht vorgesehen.

Grundlegende Ziele des Projektes – Aufgabe regionaler Katalogisierungssysteme zugunsten cloudbasierter internationaler Systeme, Katalogisierung nach internationalen Standards ohne deutsche Sonderwege, datenschutzrechtliche Absicherung – können nach heutiger Kenntnis realisiert werden. Die ursprünglich vorgesehene Plattformsynchronisation kann jedoch unter den heutigen Bedingungen nicht gelingen.

Vor diesem Hintergrund ist aus Sicht des CIB-Konsortiums eine Primärkatalogisierung im WorldCat zu empfehlen, da neben dem WorldCat keine weitere bibliothekarische Datensammlung mit deziert internationalem Charakter existiert. Ein Angebot über eine »Bundeslizenz« zur Katalogisierung im OCLC WorldCat liegt vor, ist aber in der aktuellen Form bestenfalls ein Startpunkt für weitere Gespräche. Das Konzept der Primärkatalogisierung wird sich in der praktischen Umsetzung bewähren müssen. Aus Sicht von Nicht-OCLC-Systemanwendern ergeben sich Umsetzungsprobleme, die vor allem die Nutzung von Knowledge Bases betreffen.

Für die von Wissenschaftsrat und DFG angestrebte »Neuausrichtung überregionaler Informationsservices« konnte das CIB-Projekt wichtige Beiträge leisten. Die Entwicklungspfade in eine zukünftig cloudbasierte Infrastruktur wurden in ihren Varianten eingehend untersucht und, soweit politisch und technisch möglich, mit konkreten Lösungen versehen. Das Projekt trug mit diesen Ergebnissen und auch durch eine aktive Öffentlichkeitsarbeit dazu bei, die notwendige Diskussion über die zukünftige Bibliotheksinfrastruktur in Deutschland zu strukturieren und zu befördern. Für kommende Entwicklungen kann nun auf eine im Entstehen begriffene Infrastruktur aus globalen, cloudbasierten Plattformangeboten und nationalen Komponenten wie dem Nationalen Datenfenster zurückgegriffen werden. Auf dieser Basis können verbesserte und neue Dienste für Bildung, Wissenschaft und Forschung entwickelt werden.

¹ Der Beitrag von Albrecht/Meßmer/Rusch/Thiessen in diesem Themenheft, S. 258–264 erläutert Details zu diesem Vorhaben.

² Als Titeldatenbasis der WorldShare Management Services (WMS).

³ Nutzung unter Nennung der Herkunftsquelle. Vgl. <http://open datacommons.org/licenses/by/1.0/> oder die Erklärungen OCLCs unter www.oclc.org/data/attribution.en.html [Zugriff am: 28.6.2016].

⁴ Zum mindesten galt die bis zur Übernahme durch ProQuest im Jahr 2015.

⁵ Gewiss wird es stets einige E-Ressourcen geben, denen (noch) keine OCN zugeordnet wurde. Doch sobald entsprechende Zuordnungen im WorldCat erfolgen, könnten diese auch in Alma nachgezogen und nachgenutzt werden.

⁶ Beachtenswert ist hier die Übernahme von Ex Libris durch ProQuest (Cambridge Information Group) und das angekündigte Engagement Ebsco bei einem gegebenenfalls bevorstehenden Neustart der Kuali OLE-Entwicklung.

grundlegende Ziele des Projekts können realisiert werden

DIE VERFASSER

Dr. Mathias Kratzer, Leiter des Referats Verbundnahe Dienste, Verbundzentrale des Bibliotheksverbundes Bayern, Bayerische Staatsbibliothek, Ludwigstraße 16, 80539 München, Tel.: 089-28638-2797, E-Mail: kratzer@bsb-muenchen.de

Dipl. math. Stefan Lohrum, Leiter der Arbeitsgruppe KOBV-Operating, Kooperativer Bibliotheksverbund Berlin-Brandenburg Zuse-Institute Berlin, Takustr. 7, 14195 Berlin-Dahlem, Tel.: 030-84185-197, E-Mail: lohrum@zib.de

Dr. Uwe Risch, Leiter HeBIS Verbundzentrale, Universitätsbibliothek Frankfurt am Main, Bockenheimer Landstr. 134–138, 60325 Frankfurt am Main, Tel.: 069-798-22782, E-Mail: risch@hebis.uni-frankfurt.de

Peter Thiessen, Mitarbeiter des Referats Verbundnahe Dienste in der Verbundzentrale des Bibliotheksverbundes Bayern, Bayerische Staatsbibliothek, Ludwigstraße 16, 80539 München, E-Mail: Peter.Thiessen@bsb-muenchen.de