

Die Common Bibliographic Data Zone (CBDZ) – ein Werkstattbericht

Mit dem Aufbau der »Common Bibliographic Data Zone« (CBDZ) schaffen die Verbundzentralen BSZ, hbz und VZG ein zentrales Dienstleistungsangebot für Bibliotheken, die das cloudbasierte Bibliotheksmanagementsystem Alma der Firma Ex Libris nutzen. Die CBDZ wird auf Basis einer Alma Network Zone realisiert, in der die frei verfügbaren Titeldaten der drei Verbunddatenbanken (GBV, hbz, SWB) dedupliziert enthalten sein werden. Damit werden ca. zwei Drittel der Daten der wissenschaftlichen Bibliotheken in Deutschland über die CBDZ angeboten, die von den Alma-Bibliotheken der drei Verbünde und auch von Alma-Bibliotheken anderer Verbundregionen genutzt werden können. Der Aufsatz beschreibt als Werkstattbericht die Ziele, den aktuellen Stand und die Perspektiven des CBDZ-Projekts, das im November 2015 begonnen hat.

In setting up the »Common Bibliographic Data Zone« (CBDZ) the head offices of the BSZ, hbz and VZG library networks are creating a range of central services for libraries that use the Alma cloud-based library management system of Ex Libris. The CBDZ is based on an Alma Network Zone in which the freely available bibliographic data of the three network databases (GBV, hbz, SWB) will be held in deduplicated form. Thus roughly two-thirds of the data of the academic libraries in Germany will be offered via the CBDZ. These data can be used by the Alma libraries of the three networks and also by Alma libraries of other network regions. The paper is a workshop report which describes the objectives, the current status and the prospects of the CBDZ project which began in November 2015.

HINTERGRUND UND ZIELE

Einführung von Alma in der deutschen Bibliothekslandschaft

Das cloudbasierte Bibliotheksmanagementsystem Alma der Firma Ex Libris hat sich im internationalen Markt etabliert und wird zunehmend auch von deutschen Bibliotheken eingesetzt. Inzwischen sind vier Bibliotheken aus dem SWB mit Alma in Produktion. Weitere Bibliotheken in Deutschland befinden sich in einem Migrationsprojekt zu Alma bzw. interessieren sich für eine Lizenzierung dieser Software.

Alma erlaubt mit seinen drei Zonen – Institution, Network und Community Zone – die Abbildung von lokalen Geschäftsgängen sowie die Nutzung von zentralen und globalen Services, über die hohe Synergieeffekte erzielt werden können. Darüber hinaus deckt Alma bereits jetzt die meisten lokal benötigten Funktionen ab und wird laufend an die Erfordernisse der deutschen Bibliotheksinfrastruktur (insbes. ZDB, GND, Fernleihe) angepasst.

Das CBDZ-Projekt zielt darauf ab, die Alma-Instanzen (Alma Institution Zones) in die deutsche Verbundlandschaft zu integrieren. Damit werden zentrale Services für jetzige und zukünftige Alma-Anwen-

der etabliert, so dass diese Alma optimal nutzen und gleichzeitig die bewährten Verbunddienstleistungen beziehen können. Die CBDZ kann den Umstieg eines Verbundsystems (oder mehrerer Verbundsysteme) auf Alma zu einem späteren Zeitpunkt ermöglichen. Mit dem Aufbau und Produktionsbetrieb der CBDZ können die drei Verbünde Erfahrungen mit Alma sammeln und schrittweise zentrale Services in Alma entwickeln. Die CBDZ könnte eine Grundlage für ein zukünftiges Verbundsystem darstellen: Voraussetzung wäre, dass in Alma alle jetzigen Verbundservices realisiert werden.

Projektstart und Projektumfang

Im Juni 2015 beschlossen die Verbundzentralen der Verbünde GBV, hbz und SWB auf Basis des cloudbasierten Systems Alma der Firma Ex Libris eine gemeinsame Network Zone aufzubauen. Das Konsortium lädt die frei verfügbaren Daten der drei Verbundregionen dedupliziert in die CBDZ und behält die Datenhoheit. Die Verbundzentralen definieren die notwendigen technischen und organisatorischen Rahmenbedingungen für die CBDZ und sorgen für die Synchronisierung der bibliografischen Daten mit den Verbunddatenbanken. Im Vordergrund der Anbindung von Bibliotheken an die CBDZ stehen vorerst die Verbundbibliotheken des CBDZ-Konsortiums, die sich bereits für Alma als Bibliothekssystem entschieden haben oder mittelfristig umsteigen möchten. Es sind jedoch auch Bibliotheken anderer Verbundregionen eingeladen, die CBDZ zu nutzen.

Damit ist es dem CBDZ-Konsortium gelungen, einen wesentlichen Grundstein für eine Alma Network Zone in Deutschland zu legen. Mit dem Aufbau und dem Betrieb der CBDZ werden letztendlich auch wesentliche Ziele von DFG und Wissenschaftsrat zur Neustrukturierung des deutschen Verbundsystems erfüllt: Drei Verbünde kooperieren bei der Einführung der neuen, cloudbasierten Bibliothekssoftware Alma.

MEHRWERTE FÜR INSTITUTIONEN MIT ALMA ALS BIBLIOTHEKS- MANAGEMENTSYSTEM

Die CBDZ – als Alma Network Zone – bietet für die Bibliotheken, die eine Alma Institution Zone lizenziert haben, eine optimale Grundlage für die effektive Durchführung aller lokalen Geschäftsgänge bei gleichzeitiger Nutzung der gewohnten Verbundser-

gemeinsame Network Zone in Deutschland

vices. Folgende Mehrwerte sind bereits zum jetzigen Zeitpunkt benennbar:

- Titeldaten aus dem eigenen Verbund liegen in der CBDZ vor und können direkt nachgenutzt werden (keine Fremddatenübernahme notwendig).
- Titeldaten aus anderen CBDZ-Verbünden werden dedupliziert integriert und erweitern das Fremddatenangebot innerhalb von Alma.
- Neue Titel (Erwerbungskatalogisate), die noch nicht in den Verbünden/der CBDZ vorhanden sind, können in Alma erfasst und in die primären Verbunddatenbanken exportiert werden.
- Vollständige, detaillierte Katalogaufnahmen (Neuzugänge nach Bucheingang, Korrekturen) können zunächst weiterhin über die gewohnten Verbund-Clients (CBS WinIBW bzw. Aleph-Client) durchgeführt werden. Die Komplexität des Umstiegs auf Alma wird dadurch reduziert: Die Bibliotheken können sich bei Alma-Migrationsprojekten auf die Umstellung der lokalen Workflows konzentrieren.
- Die CBDZ wird mit den bewährten Verbundservices des eigenen Verbundes versorgt. Auch Dienstleistungen von »fremden« CBDZ-Partnern können nachgenutzt werden.
- Schrittweise werden die Verbünde und Ex Libris neue Funktionen und Services in Alma implementieren, welche das Zusammenspiel von Alma und Verbunddatenbanken erweitern (insbes. Online-Schnittstelle für die Katalogisierung).
- Die Alma-Anwender in Deutschland schaffen mit der CBDZ eine neue, gemeinsame Community, die eine starke Rolle bei der Entwicklung neuer Alma-Funktionen ermöglichen wird.

Vorteile der CBDZ

aktuelle Verbundsysteme GBV, hbz, SWB

Herausforderungen

Mit dem Aufbau der CBDZ verfolgen die Verbundzentralen BSZ, hbz und VZG das ambitionierte Ziel der Zusammenführung von bibliografischen Metadaten aus drei verschiedenen Verbunddatenbanken. Die Datenqualität nach erfolgter Deduplizierung in Alma muss für den Katalogisierungsbetrieb und den Nachweis von Bibliotheksbeständen geeignet sein. Auch die laufende Deduplizierung in der CBDZ und die Synchronisierung der CBDZ mit den drei Masterdatenbanken stellt eine komplexe Aufgabe dar. Aufgrund der unterschiedlichen Ausgangssituation der drei Verbünde ergeben sich zusätzliche Herausforderungen.

Ausgangssituation

Das BSZ betreibt die Verbunddatenbank auf Basis eines CBS-Systems von PICA/OCLC. Die Verbundteilnehmer katalogisieren im PICA-Format über den WinIBW-Client in der Datenbank. Im hbz-Verbund wird Aleph, auch ein Produkt von Ex Libris, als Verbundsystem eingesetzt; die Katalogisierung findet im Format »Aleph-MAB« über den Aleph-Client statt. Die VZG betreibt ebenfalls ein CBS-System für die Teilnehmer des Gemeinsamen Bibliotheksverbundes (GBV), allerdings unterscheidet sich das PICA-Format von dem des BSZ. Zwar werden die Formate im Zuge der SWB/GBV-Kooperation angeglichen, dies wird aber erst im Jahr 2017 abgeschlossen sein. Da es kein einheitliches Datenformat in den Verbünden gibt, bleibt als einziger gemeinsamer Nenner das Export-Format MARC 21.

In allen Verbünden sind verschiedene Bibliotheksmagementsysteme im Einsatz, deren Daten mit den Verbundsystemen über unterschiedliche standardi-

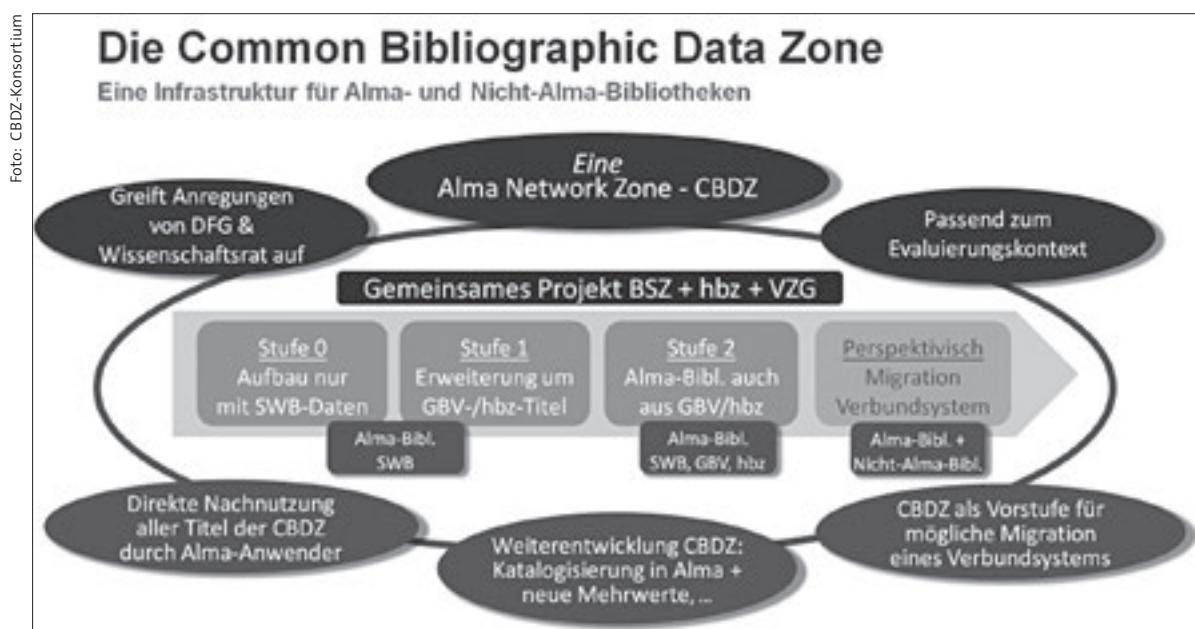


Abb. 1: Die CBDZ-Übersicht

sierte bzw. proprietäre Schnittstellen (offline, online) synchronisiert werden. Diese Faktoren müssen beim Aufbau und Betrieb der CBDZ berücksichtigt werden. Die Common Bibliographic Data Zone kommt als neues System für die Alma-Bibliotheken in der Verbundlandschaft hinzu und muss nahtlos in die existierenden Verbundinfrastrukturen integriert werden, um allen beteiligten Einrichtungen weiterhin einen effizienten Arbeitsprozess zu ermöglichen.

Konzeption

Eine besondere Herausforderung des Projekts CBDZ lag vor allem in den Vorüberlegungen und der anschließenden Konzeption: Wie kann ein deduplizierter Datenpool aus drei Verbünden erstellt und anschließend mit den bestehenden Masterdatenbanken von GBV, hbz und SWB synchronisiert werden? Neben der Klärung der Randbedingungen und Gegebenheiten in den Verbünden und der technischen Voraussetzungen musste auch ein Konsens mit Ex Libris erzielt werden. Dafür wurde eine Arbeitsgruppe, bestehend aus Experten der drei Verbundzentralen eingerichtet, die sich mehrere Monate mit der Konzeption der CBDZ beschäftigte; die Ergebnisse sind inzwischen auch mit Ex Libris abgestimmt.

Deduplizierung

Bei der Zusammenführung der Daten aus GBV, hbz und SWB stellen sich diverse Fragen:

- Mit welchem Tool und in welchem Format soll das initiale Match&Merge-Verfahren stattfinden?
- Wie soll das Ergebnis einer Deduplizierung umgesetzt werden (Gewinner-Titel)?
- Wie können Datenverluste vermieden werden?
- Wie können sinnvolle Anreicherungen zur Verbesserung der Datenqualität stattfinden?
- Wie kann im laufenden Betrieb ein ähnliches Deduplizierungsverfahren wie beim initialen Aufbau der CBDZ implementiert werden?

Die drei Verbundzentralen sind durch die jahrelange Erfahrung mit vielen Migrationsprojekten neuer Teilnehmerbibliotheken sehr versiert und verfügen über gute Tools, um Daten zu deduplizieren. Hierfür stehen die oben genannten Verbundsysteme Aleph und CBS bereits in vollem Umfang zur Verfügung.

Allerdings bedeutet die Nutzung eines der beiden Systeme immer, dass zunächst alle Daten von mindestens einem Verbund in das Zielformat des gewählten Systems erfolgen müssen. So müssten beispielweise die Titeldaten aus dem hbz-Verbund aus dem Aleph-Internformat (MAB-angennähert) über ein standardisiertes Austauschformat (MAB 2 oder MARC 21) zum

PICA-Format konvertiert werden, bevor ein Match&Merge in einem CBS-System stattfinden kann. Umgekehrt müssten sogar die Titeldaten aus zwei Verbünden von PICA über ein Standardformat zum Aleph-Internformat konvertiert werden, damit diese in Aleph weiterverarbeitet und dedupliziert werden könnten.

Alternativ steht Alma als Tool zur Deduplizierung der Verbunddaten im Format MARC 21 zur Verfügung, allerdings kann das Match-Verfahren zum aktuellen Zeitpunkt nur anhand von Identifizieren durchgeführt werden, wie beispielsweise mit ISBN, ISSN oder sonstigen Nummern. Weil diese Identifier in den wenigen Fällen zu einem zufriedenstellenden Deduplizierungsergebnis führen und noch weitere Elemente wie Titel, Verfasser, Veröffentlichungsjahr, usw. mit einbezogen werden müssen, hat sich das CBDZ-Konsortium gegen die Verwendung von Alma für die initiale Deduplizierung entschieden: Darüber hinaus fehlen in Alma noch Tools für eine Deduplizierung sehr großer Mengen an Titeldaten, mit z.B. Gewichtungsmöglichkeiten und der Berücksichtigung der verschiedenen Satztypen (Monografien, Über- und Unterordnungen). Diese langjährig erprobten Deduplizierungsmechanismen existieren in den Verbundzentralen und können die Qualität einer deduplizierten CBDZ für den Katalogisierungsbetrieb sicherstellen.

Es soll nicht verschwiegen werden, dass durchaus auch die Ergebnisse von anderen Projekten wie beispielsweise Culturegraph oder GVI (Gemeinsamer Verbundeindex) im Hinblick auf Nachnutzbarkeit für die Deduplizierung der Verbunddaten betrachtet wurden, allerdings ergeben sich aus Sicht des CBDZ-Konsortiums keine Synergieeffekte, da der Schwerpunkt dieser Projekte auf der Identifikation von Dubletten, nicht aber auf der Zusammenführung, also dem Mergen von Datensätzen für den Katalogisierungsbetrieb liegt.

Grundsätzlich sollen die in der CBDZ vorhandenen Daten eine möglichst hohe Qualität aufweisen, daher sind gute Merge-Mechanismen unerlässlich. Im Idealfall sollte ein gemergerter Datensatz die besten Attribute aus den maximal drei ursprünglichen Quellen erhalten. Diesem Wunschgedanken stehen allerdings technische und organisatorische Grenzen im Weg. So muss einerseits geprüft werden, wie feingranular die Zusammenführung von Daten überhaupt möglich ist, und andererseits müssen Richtlinien getroffen werden, die beispielsweise festlegen, welcher Datensatz angereichert wird und in welchem Umfang. Insbesondere die Anreicherung von Sacherschließungsdaten ist aus Sicht des CBDZ-Konsortiums essentiell und soll in jedem Fall umgesetzt werden.

Konzeptphase

Datenformate und Fragen zur Datenzusammenführung

Datenqualität

Datenflüsse und Schnittstellen

Neben der Deduplizierung ist auch der Datenaustausch der CBDZ mit den Masterdatenbanken ein schwieriges Unterfangen. Dafür müssen verschiedene Fragen geklärt werden:

- Über welche Schnittstelle werden die Updates der bibliografischen Metadaten aus den primären Verbunddatenbanken (Neuzugänge, Korrekturen, Löschungen) in die CBDZ transferiert?
- Über welche Mechanismen werden CBDZ-Titel identifiziert, vollständig aktualisiert bzw. lediglich um einzelne Informationen angereichert?
- Wie können Titelneuzugänge aus der CBDZ in den primären Verbunddatenbanken umgesetzt werden?
- Wie kann es perspektivisch gelingen, Titel-Updates direkt in Alma vorzunehmen (ohne Nutzung des Verbund-Clients) und diese verlustfrei über Online-Schnittstellen in den primären Verbunddatenbanken umzusetzen?
- Über welche Schnittstelle werden die Bestandsdaten aus einer Alma Institution Zone in die primären Verbunddatenbanken transferiert?

Problematisch beim Umgang mit Titel-Updates aus den Verbundkatalogen ist insbesondere die Tatsache, dass in der CBDZ konkurrierende Updates auf zusammengeführte Titel aus unterschiedlichen Quellen umgesetzt werden müssen. Dafür muss es klare Prinzipien für das laufende Match&Merge-Verfahren auf der Seite der CBDZ geben. Hinzu kommt die Schwierigkeit der unterschiedlichen Formate MARC 21 – MAB, beziehungsweise MARC 21 – PICA. Insbesondere bei der Zwei-Wege-Konvertierung von Titeldaten können Datenverluste entstehen; daher müssen die Schnittstellen, bzw. Import- und Exportkonverter der Verbünde, weitestgehend reversibel sein.

GEPLANTE UMSETZUNG – STUFENKONZEPT

Aufgrund der Komplexität des Projekts und der Funktionalitäten, die teilweise noch von Ex Libris oder dem Konsortium entwickelt werden müssen, wurde das Konzept für die konkrete Umsetzung in drei Stufen aufgeteilt:

- **Stufe 0 – Initialer Aufbau der CBDZ mit SWB-Titeln:** Diese Stufe dient dem zügigen Anschluss der produktiven Alma-Bibliotheken im SWB.
- **Stufe 1 – Integration der GBV- und hbz-Titel:** In dieser Stufe werden zusätzliche Titeldaten aus hbz und GBV zur direkten Nachnutzung in Alma importiert.
- **Stufe 2 – Nutzung der CBDZ durch Alma-Bibliotheken mehrerer Verbundsysteme:** Mit Erreichung der Stufe 2 ist die CBDZ ein deduplizierter Datenpool,

bestehend aus Daten der Verbünde GBV, hbz und SWB. Bibliotheken, die Teilnehmer dieser Verbünde sind und Alma einsetzen, können an die CBDZ angeschlossen werden.

Grundannahmen

Nach Projektstart im November 2015 sind folgende zehn Grundannahmen für den Aufbau und den Betrieb der CBDZ definiert worden:

1. *Deduplizierte CBDZ:* Die CBDZ ist möglichst zu jedem Zeitpunkt dedupliziert: Zum Erreichen dieses Ziels ist ein initiales und laufendes Deduplizierungsverfahren notwendig.
2. *Stufenkonzept:* Der initiale Aufbau einer deduplizierten CBDZ ist ein längerer Prozess innerhalb verschiedener Projektphasen, der ein stufenweises Vorgehen bedingt.
3. *Umfang der Stufen:* Mit jeder Stufe nehmen Datenumfang und Funktionalitäten in der CBDZ zu: In früheren Stufen entstehen daher geringere Aufwände, um schnell Ergebnisse zu erzielen. Komplexere Aufgaben können zum Teil erst in späteren Stufen realisiert werden.
4. *Verbunddatenbanken als Master:* Die CBDZ wird parallel zu den drei Verbunddatenbanken aufgebaut, die zunächst weiterhin die primären Katalogisierungssysteme (Master) – insbesondere durch Vergabe eines Verbund-Identifiers – bleiben.
5. *Titel-Updates aus Verbunddatenbanken:* Die CBDZ wird laufend mit den Titel-Updates aus den primären Verbunddatenbanken aktualisiert. Dafür muss innerhalb von Alma ein Deduplizierungsverfahren für neue Verbundtitel implementiert werden. In der CBDZ sind nicht die Bestandsdaten der Nicht-Alma-Bibliotheken aus den Verbundsystemen enthalten.
6. *Bestands-Updates aus Alma:* Die Verbunddatenbanken werden laufend mit den Beständen der Alma-Bibliotheken (Alma Institution Zones) aktualisiert.
7. *Katalogisierung über Alma vs. Nutzung Verbund-Client:* Bibliotheken mit einer Alma Institution Zone sollen so weit wie möglich über die Alma-Anwendung arbeiten können: Dazu gehört auch die Erfassung von Neukatalogisaten in Alma (v.a. für Erwerbungsaufnahmen), die über den Export der Bestandsdaten als Neuzugänge in der jeweiligen Verbunddatenbank importiert werden können. Perspektivisch wird eine vollständige Katalogisierung über die Alma-Anwendung in den primären Verbunddatenbanken über eine Online-Schnittstelle angestrebt. Bis diese Lösung zur Verfügung steht, können die Alma-Bibliotheken weiterhin

- über die bisherigen, gewohnten Verbund-Clients in den primären Verbunddatenbanken arbeiten.
8. *Grundsatzproblem »Konkurrierende Updates auf einen CBDZ-Titel«:* Innerhalb einer deduplizierten CBDZ ist das zentrale Grundsatzproblem zu lösen, wie mit »n Titel-Updates« aus drei Quellsystemen (Verbunddatenbanken) verfahren werden soll.
 9. *Bereich Sacherschließung:* Die Sacherschließung stellt einen sehr wichtigen Aspekt dar, der bei der initialen und laufenden Deduplikation zu berücksichtigen ist. Datenverluste in diesem Bereich sind zu vermeiden, Anreicherungen sind zu ermöglichen.
 10. *Primärkatalogisierung in der CBDZ:* Inwieweit die vollständige Primärerfassung in der CBDZ zukünftig stattfinden kann, wird derzeit noch vom Konsortium untersucht, spätestens mit der Migration eines Verbundsystems wird für dieses die Primärkatalogisierung in Alma stattfinden müssen.

Stufe 0 – Aufbau der CBDZ mit SWB-Daten

Diese Stufe wird als Basis für den gesamten Aufbau der CBDZ betrachtet und kann mit vergleichsweise geringem Aufwand erreicht werden. Zum aktuellen Zeitpunkt (Juni 2016) gibt es in Deutschland vier Bibliotheken, die bereits produktiv mit Alma als Bibliothekssystem arbeiten, diese sind ausschließlich Teilnehmerbibliotheken im Südwestverbund. Der Fokus der Stufe 0 liegt daher auf den Titeldaten des SWB und der Anbindung der Alma-Bibliotheken an die Verbunddatenbank. So sollen die bisherigen Institution Zones sukzessive an die CBDZ (Network Zone) angeschlossen werden und dort alle Titeldaten des Südwestverbundes wiederfinden, um diese für Bestellungen beim

Buchhandel etc. direkt nachnutzen zu können. Abbildung 2 zeigt, wie dies in der Alma-Oberfläche aussieht, die CBDZ verbirgt sich dabei hinter der Registerkarte Netzwerk. Wird diese aktiviert, kann die Bibliothek zusätzlich zum eigenen Bestand die passenden Titel aus dem SWB sehen. Eine zusätzliche Recherche im Metadateneditor über Z39.50, wie sie derzeit von den produktiven Alma-Bibliotheken angewandt wird, entfällt somit. Dies trägt insbesondere zur Beschleunigung im Erwerbsprozess bei.

Übernimmt die Bibliothek einen Titel aus der CBDZ in die eigene Institution wird der Bestandsnachweis in der darauffolgenden Nacht an den SWB gemeldet. Falls keine passende Titelaufnahme in der CBDZ zur Nachnutzung vorhanden ist, kann die Bibliothek über den Metadateneditor in Alma ein neues Katalogisat im Format MARC 21 erfassen, welches in der Nacht gemeinsam mit dem Bestandsnachweis (Holding) an den SWB übergeben wird. Das BSZ fungiert dabei als OAI-Harvester und übernimmt die Nachverarbeitung der aus Alma bereitgestellten Daten, um diese im geeigneten Format und gegebenenfalls gemäß bibliotheksspezifischer Kriterien im Verbund nachzuweisen.

Aufgrund der oben genannten Festlegung, dass keine Master-Master-Synchronisierung zwischen Alma und den Verbunddatenbanken stattfinden soll und es bisher keine reversible Online-Schnittstelle gibt, können Datenänderungen aus Alma derzeit noch nicht in den SWB übernommen werden. Das heißt für die Korrektur vorhandener Titeldaten muss noch der WinIBW-Client verwendet werden. Außerdem gibt es in Alma noch keine deutsche Hilfefunktion für Regelwerks- und Formatfragen, wie dies in den bisherigen Katalogisierungsclients der Verbunddatenbanken

Nutzung der CBDZ

Verbund-Client für Titel-Update

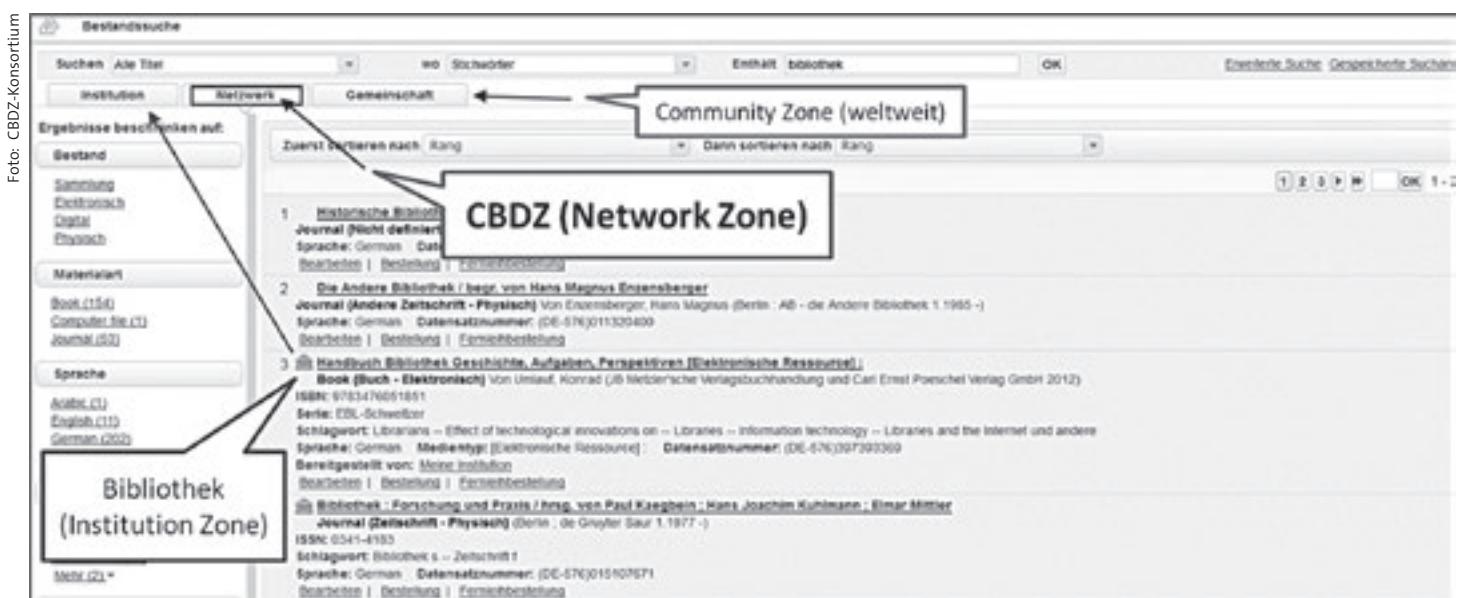


Abb. 2: Die CBDZ in Alma

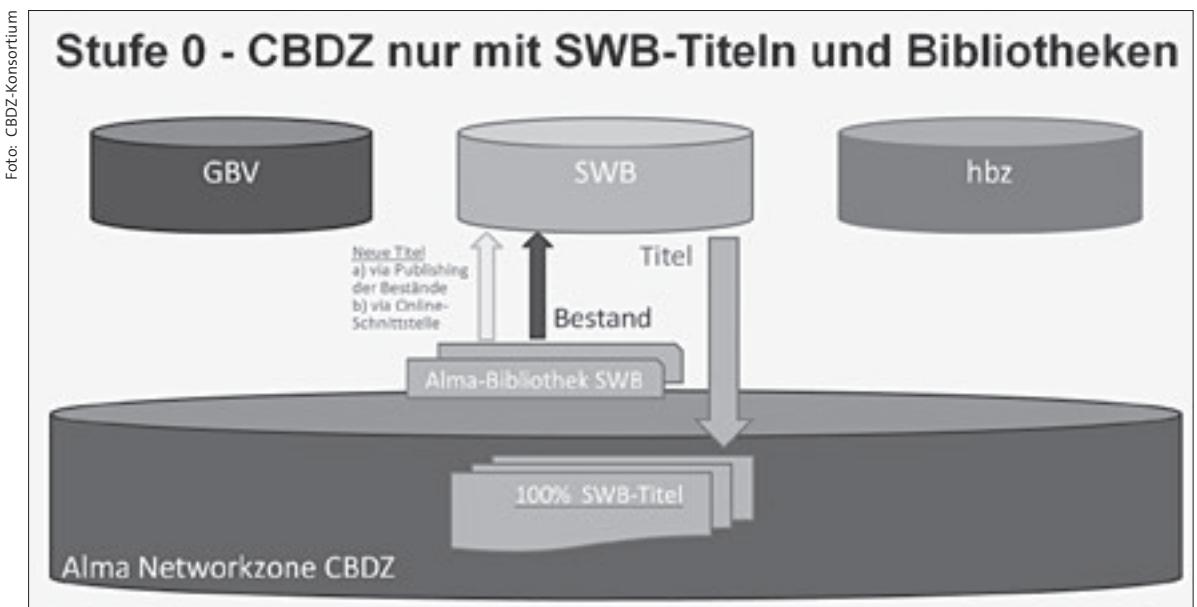


Abb. 3: Stufe 0 – Aufbau der CBDZ

existiert. Somit ist die teilweise Nutzung des WinIBW-Clients einerseits ein Wermutstropfen, andererseits erleichtert dies die langsame Umgewöhnung der Bibliotheksmitarbeiter an das neue Format MARC 21, welches für die Katalogisierung von Erwerbungsaufnahmen zum Einsatz kommt.

Alle neu erfassten Titelaufnahmen und Änderungen in der SWB-Verbunddatenbank werden täglich an die CBDZ weitergegeben, somit werden auch die Alma-Institutionen mit den üblichen Verbunddienstleistungen versorgt. Die Stufe 0 dient als Übergangsphase zur Stufe 1. Abbildung 3 beschreibt den Datenfluss, der bis Ende des Jahres 2016 umgesetzt werden soll. Als Pilotpartner wird die Universitätsbibliothek Mannheim als erste Bibliothek an die Common Bibliographic Data Zone angeschlossen.

Stufe 1 – Integration der hbz- und GBV-Titel in die CBDZ

Ausbau der CBDZ

In dieser Stufe werden erstmals Daten aus den beiden Verbünden hbz und GBV zu den bereits vorhandenen SWB-Titeln zugespielt. Ziel der Stufe 1 ist die Vergrößerung der Menge an Titeldaten zur Nachnutzung, die für bereits angeschlossene Alma-Bibliotheken des SWB in der CBDZ zur Verfügung stehen. Reihenfolge und Umfang der zu importierenden Daten werden in den kommenden Monaten nach eingehender Analyse der Daten festgelegt.

Des Weiteren wird das Konsortium in Stufe 1 Absprachen mit Ex Libris treffen, welche Funktionalitäten in Alma für die laufende Deduplizierung und Zusammenführung von Titeldaten aus unterschiedlichen Quellen benötigt werden. Die initiale Deduplizierung

von hbz-, GBV- und SWB-Daten wird außerhalb von Alma im CBS-System des GBV stattfinden. Dabei werden zunächst nur disjunkte Titel zum SWB-Bestand ermittelt und in die CBDZ geladen, ein Merge findet in dieser Stufe noch nicht statt.

Hintergrund dieser Entscheidung ist die Vermeidung des bereits erwähnten Grundsatzproblems konkurrierender Updates aus verschiedenen Quellen auf gemergte Datensätze in der CBDZ. Somit gibt es in Stufe 1 für jeden Datensatz ausschließlich eine primäre Datenquelle, aus der laufende Updates in der CBDZ nachgeführt werden. Dies wird sichergestellt, indem grundsätzlich nur eine Verbund-ID pro Datensatz gespeichert wird. Diese Art der Implementierung hat den entscheidenden Vorteil, dass bereits vorhandene Funktionalitäten in Alma für die Updates aus den Verbünden genutzt werden können, da Datensätze vollständig überschrieben werden, sobald eine Änderung aus der Masterdatenbank kommt. Ein aufwändiges Merge-Verfahren kann somit umgangen werden, neue Titelaufnahmen müssen jedoch vor dem Import in die CBDZ einen Dublettencheck durchlaufen, der nach den gleichen Regeln wie beim initialen Deduplizierungsverfahren implementiert werden sollte.

Übernimmt eine Alma-Bibliothek Titel aus der CBDZ in den eigenen Bestand, die ursprünglich aus dem hbz oder GBV stammen, wird dieser Datensatz wie ein neues Katalogisat in Alma behandelt und wird gemeinsam mit dem Bestandsnachweis über Nacht im SWB eingespielt. Sobald das Update aus dem Südwestverbund in die CBDZ zurückgespielt wird, wechselt der Alma-Titeldatensatz seinen Replikationspartner und wird ab diesem Moment nur noch aus dem

Stufe 1 - Integration der hbz- und GBV-Titel

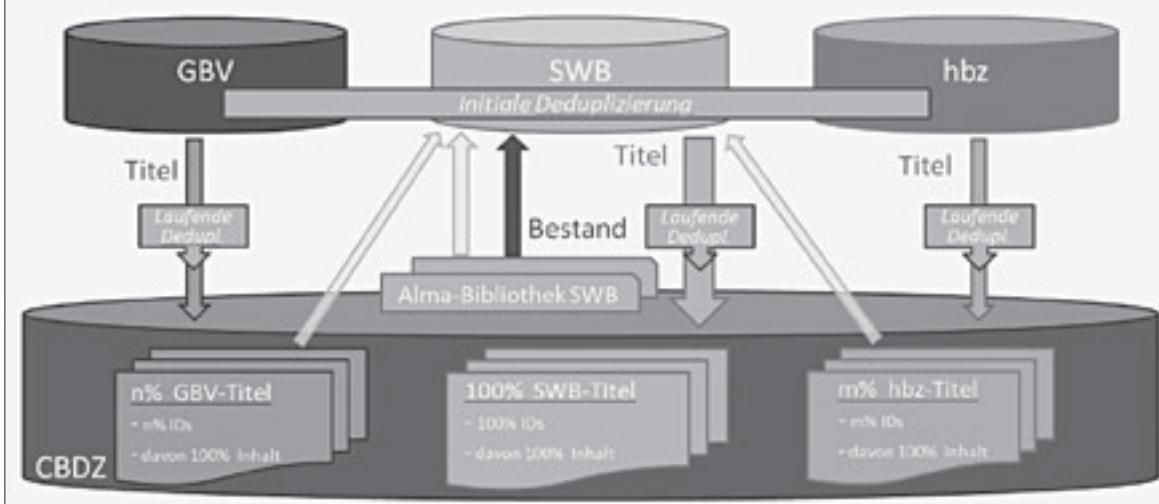


Abb. 4: Stufe 1 – Erweiterung der CBDZ

Stufe 2 - CBDZ mit Alma-Bibliotheken mehrerer Verbünde

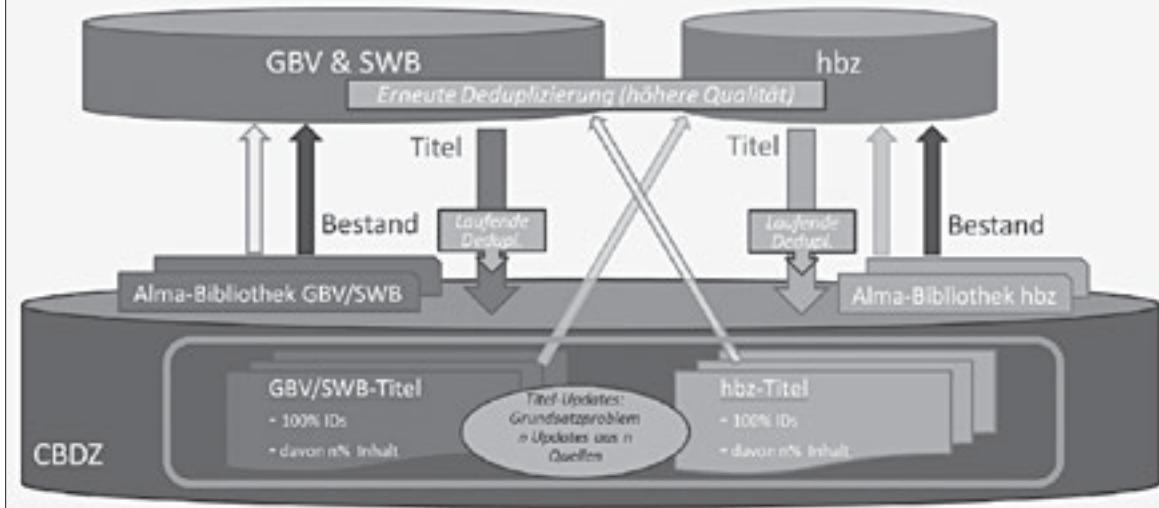


Abb. 5: Stufe 2 – CBDZ zur Nutzung von Bibliotheken mehrerer Verbünde

Südwestverbund aktualisiert. Die in Abbildung 4 dargestellte Stufe 1 ist als Erweiterung der Stufe 0 zu sehen und wird nach derzeitiger Planung ab April 2018 von der Stufe 2 abgelöst.

Stufe 2 – Nutzung der CBDZ durch Alma-Bibliotheken mehrerer CBDZ-Verbundsysteme

Während in den Stufen 0 und 1 lediglich Alma-Bibliotheken, die Teilnehmer am SWB sind, berücksichtigt werden, soll in Stufe 2 die Nutzung der CBDZ für Bibliotheken aus allen drei Verbünden ermöglicht werden. Hierfür ist eine erneute Deduplizierung, inklusive Match&Merge aller Verbunddaten notwendig. Dabei

ist das Ergebnis des Deduplizierungsverfahrens wesentlich relevanter als in Stufe 1, damit Bestände bei Migrationen weiterer Alma-Bibliotheken den richtigen Titel-Entitäten zugeordnet werden können. Im Gegensatz zur Stufe 1 ist es daher nicht mehr ausreichend, nur eine Verbund-ID pro Datensatz zu speichern, sondern jeder existierende Identifier aus GBV, hbz und SWB muss in einem zusammengeführten Datensatz vorhanden sein.

Des Weiteren sollen wichtige regionale Inhalte und Sacherschließungsdaten beim Merge berücksichtigt werden, damit keine Bibliothek – ganz gleich aus welcher Verbundregion – benachteiligt wird, sie muss ihre

CBDZ für Alma-Bibliotheken

aufwändig gepflegten Metadaten auch nach der Zusammenführung der Verbunddaten wiederfinden.

Durch das Vorhandensein mehrerer Verbund-Identifier in den gemergten Titeldaten muss in Stufe 2 zum ersten Mal das Grundsatzproblem konkurrierender Updates aus den primären Systemen gelöst werden. Dies ist sowohl ein technisches als auch ein organisatorisches Problem, da klare Prinzipien für die Festlegungen eines »Gewinner-Titels« nach der initialen Deduplizierung notwendig sind, um ggf. Anreicherungen aus den Verlierersätzen übernehmen zu können. Somit muss auch das Verfahren für das laufende Match&Merge in dieser Stufe vollständig implementiert sein. Das CBDZ-Konsortium hat verschiedene Ansätze evaluiert und eine präferierte Lösung erarbeitet, die jedoch noch nicht endgültig verabschiedet ist. Des Weiteren werden die Verbundzentralen Ex Libris hinsichtlich der Konzeption eines ausgefeilten Match&Merge-Verfahrens beraten, um alle Voraussetzungen in Alma für Stufe 2 zu schaffen.

In Abbildung 5 wird der voraussichtlich ab 2018 angestrebte Datenfluss dargestellt; die bereits in den vorherigen Stufen umgesetzten Funktionalitäten, wie beispielsweise das Melden der Bestandsdaten an die Verbünde, bleiben bestehen.

STAND DER CBDZ UND AKTUELLE ARBEITEN DES KONSORTIUMS

Daten und Importe

Bereits Anfang des Jahres 2016 wurde mit dem Aufbau der Common Bibliographic Data Zone begonnen. Von Ex Libris wurde die benötigte Alma Network Zone aufgesetzt und der initiale Ladevorgang der Daten aus dem Südwestverbund durchgeführt, das BSZ stellte hierfür einen Gesamtabzug der Verbunddatenbank im Format MARC 21 zur Verfügung. Nachdem dieser Vorgang Anfang April abgeschlossen werden konnte, erhielt das CBDZ-Konsortium aufgrund der bereits erfolgten Zertifizierung von Kolleginnen in BSZ und hbz Zugriff auf die Konfiguration. Dies ermöglichte dem BSZ die Einrichtung der Importprofile, so dass die tägliche Aktualisierung der CBDZ mit Daten aus dem SWB ab Mai 2016 in Routine gehen konnte.

Beim Import der SWB-Verbunddaten sind automatisch auch die Daten aus der Zeitschriftendatenbank enthalten, da im BSZ alle Änderungen aus der ZDB im Fünfminutentakt per OAI abgeholt werden. Somit wird auch die CBDZ tagesaktuell mit den Daten aus der ZDB versorgt. Zum Stichtag 1. Juli 2016 befinden sich in der CBDZ etwa 22,3 Millionen Titeldaten.

Metadatenkonfiguration

Unter dem Punkt Metadatenkonfiguration wird in Alma die Konfiguration der Felder verstanden, die zur Katalogisierung bibliografischer Metadaten im Format MARC 21 benötigt werden. Dies hat unter anderem Auswirkungen auf die Validierung der Felder im Metadateneditor. Wird beispielsweise ein Datensatz aufgerufen, der Felder enthält, die nicht im Standard der Library of Congress vorgesehen sind, wird dies als Warnung angezeigt. Im deutschsprachigen Raum gibt es einige Felder, die zusätzlich zu den Standardfeldern erfasst werden, das bekannteste Beispiel ist vermutlich das Feld 689, welches für die Abbildung von Schlagwörtern nach den Regeln für den Schlagwortkatalog (RSWK) genutzt wird. Da dieses und weitere anwenderspezifische Felder auch in den importierten Verbunddaten enthalten sind, muss die Metadatenkonfiguration in Alma entsprechend angepasst werden. In der CBDZ Network Zone wurden bereits alle anwenderspezifischen Felder für den deutschsprachigen Raum konfiguriert, basierend auf dem aktuellen Stand der Expertengruppe Datenformate, welche dem Standardisierungsausschuss der Deutschen Nationalbibliothek untersteht.

Evaluierung und Tests

Ein Großteil der aktuellen und zukünftigen Arbeit des CBDZ-Konsortiums besteht in der Evaluierung der bestehenden Funktionalitäten in Alma und dem konkreten Testen. Aufgrund der monatlichen Software-Release müssen sich die Experten ständig auf den neuesten Stand bringen und hinzukommende Features in die Planungen der CBDZ mit einbeziehen.

Für den Aufbau der Stufe 0 liegt der Fokus insbesondere auf dem Zusammenspiel zwischen Network Zone und Institution Zones, da beispielsweise gewährleistet werden muss, dass lokale Titelfelder oder Verknüpfungen in den Institution Zones nicht durch Updates der Titeldaten in der Network Zone verloren gehen. Hierfür wird eine Arbeitsgruppe gebildet, die sich gemeinsam mit Ex Libris und der Pilotbibliothek der Universität Mannheim mit den Auswirkungen des Anschlusses an die CBDZ beschäftigt.

Aufgrund der zu erwarteten Datenmenge aus den drei Verbunddatenbanken legt das Konsortium besonderen Wert auf eine performante Verarbeitung von Daten, vor allem bei Importen oder Datenkorrekturen im großen Stil. Auch hierbei dürfen sich Änderungen in der CBDZ nicht auf die angeschlossenen Institution Zones auswirken.

Während der Umsetzung der Stufe 1 wird sich der Fokus vermehrt auf die angekündigten Deduplizierungsmechanismen richten, dies betrifft zunächst

die Evaluierung der aktuell zur Verfügung stehenden Möglichkeiten in Alma und den Verbundsystemen, später auch die Konzeption der Funktionalitäten, die das Konsortium für die Umsetzung der Stufe 2 benötigt.

Des Weiteren wird sich das Konsortium mit der Bereitstellung von E-Book-Paketen in der CBDZ befassen. Zwar liefern die Verbünde bereits bibliografische Metadaten zu elektronischen Ressourcen, allerdings sind noch einige konzeptionelle Fragen zur effizienten Umsetzung der Alma-Paketstrukturen ohne manuelle Nacharbeiten in der CBDZ offen.

FAZIT

Mit der CBDZ schaffen die drei Verbünde GBV, hbz und SWB eine auf der Alma-Software basierende Plattform (Network Zone), mit der die Alma-Bibliotheken in Deutschland optimal unterstützt und mit den bewährten Verbundservices versorgt werden können.

Die CBDZ wird laufend mit den drei Verbundkatalogen (GBV, hbz, SWB) aktualisiert, die zunächst weiterhin die primären Katalogisierungsdatenbanken darstellen. Perspektivisch soll diese Network Zone auch die mögliche zukünftige Migration eines Verbundsystems auf Alma ermöglichen. Das Projekt zum Aufbau und Betrieb einer deduplizierten CBDZ ist sehr komplex und fordert viele Ressourcen und Erfahrungen der drei Verbundzentralen.

Die CBDZ vereint die zentralen Bibliotheksmanagementsysteme CBS (OCLC), Aleph und Alma (beide Ex Libris) in einer gemeinsamen Architektur und bietet eine Ergänzung/Alternative zur Katalogisierung in WorldCat. Damit schafft sie großes Potential für die deutsche Bibliotheksinfrastruktur. Mit dem Kooperationsvorhaben der CBDZ greifen die drei Verbundzentralen die Empfehlungen von Wissenschaftsrat und Deutscher Forschungsgemeinschaft auf. Das Kooperationsprojekt stellt einen wichtigen Meilenstein zur Neuausrichtung der überregionalen Informationsstrukturen in Deutschland dar.

CBDZ konform mit Empfehlungen von Wissenschaftsrat und DFG

DIE VERFASSER

Michaela Sieber, Alma-Projektleiterin und Mitglied der CBDZ-Geschäftsstelle, Bibliotheksservice-Zentrum Baden-Württemberg (BSZ), Felix-Wankel-Str. 4, 78467 Konstanz, Tel.: 07531-88-4282, E-Mail: michaela.sieber@bsz-bw.de

Günter Hupfer, Systembibliothekar und Betreuung Projekte hbz-Verbunddatenbank, Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen (hbz), Jülicher Straße 6, 50674 Köln, E-Mail: hupfer@hbz-nrw.de