

Denise Baumgart, Francesca Schulze

Die Servicestelle der Deutschen Digitalen Bibliothek

Hilfe bei der Datenablieferung



DEUTSCHE DIGITALE BIBLIOTHEK
Kultur und Wissen online

Problem:
Vereinheitlichung
heterogener
Datenbestände

Nach und nach sollen die Metadaten aus allen deutschen Kultur- und Wissenschaftseinrichtungen in die Deutsche Digitale Bibliothek (DDB) integriert werden und über das Internetportal Nutzerinnen und Nutzern aus Deutschland und der ganzen Welt zur Verfügung stehen. Das klingt genauso kompliziert wie es ist: Die Einrichtungen aus den unterschiedlichen Kultur- und Wissenschaftsbereichen verfügen über viele verschiedene Datenformate und erschließen ihre Objekte nach unterschiedlichen Regeln. Die Herausforderung besteht darin, all diese heterogenen digitalen Datenbestände auf ein gemeinsames Format zu vereinheitlichen. - Dieser Aufgabe nehmen sich die 13 Einrichtungen des Kompetenznetzwerks¹⁾ der DDB an und haben gemeinsam mit dem technischen Betreiber, FIZ Karlsruhe - Leibniz-Institut für Informationsinfrastruktur²⁾ einen Workflow von der ersten Kontaktaufnahme der potenziellen Datenlieferanten bis zur Integration und Präsentation der Daten innerhalb der DDB erstellt. Der Workflow wurde an der Deutschen Nationalbibliothek (DNB) erprobt und soll zukünftig als standardisiertes Verfahren für die Integration aller Datenquellen angewendet werden.

Eine Schlüsselrolle übernimmt dabei die Servicestelle, die ein organisatorisches Bindeglied zwischen den abliefernden Institutionen und dem technischen Betreiber bildet. Diese leistet fachliche und organisatorische Unterstützung bei der Vorbereitung der Lieferung von Daten an die DDB und bietet neuen Partnern als zentrale Anlaufstelle einen »Service aus einer Hand«. Die Servicestelle übernimmt dabei nicht die Rolle eines Aggregators. Der eigentliche Datenfluss nimmt den direkten Weg zwischen der kulturwissenschaftlichen Einrichtung oder dem Aggregator und dem technischen Betreiber.

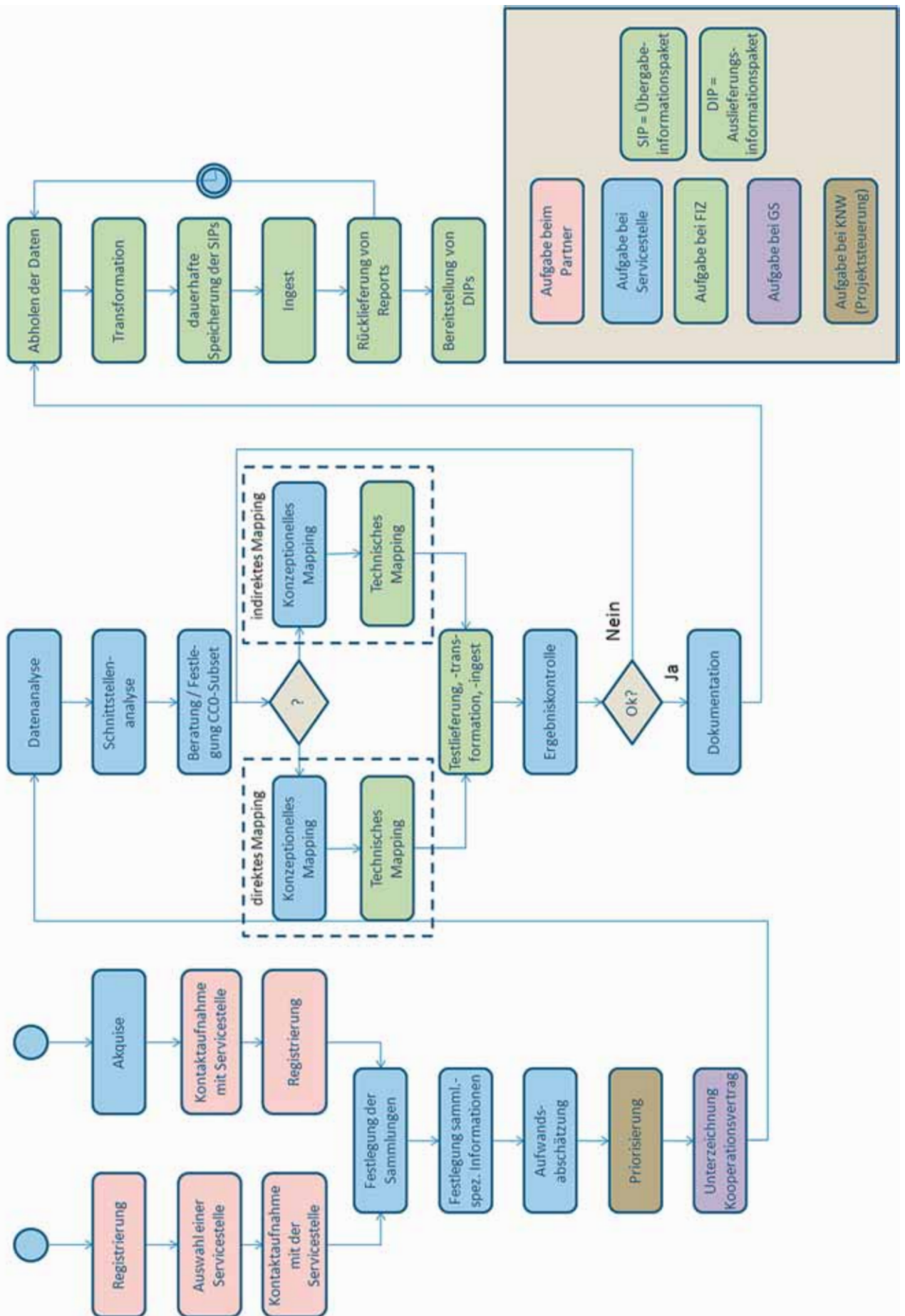
Servicestelle
übernimmt eine
Schlüsselrolle

Die DDB-Servicestelle leistet folgende Unterstützung

Unterstützung des Erstkontakts potenzieller Partner für die DDB:

- Mitarbeit bei der Akquise neuer Partner in Zusammenarbeit mit der Geschäftsstelle (Stiftung Preußischer Kulturbesitz), dem Koordinator (DNB) und dem technischen Betreiber (FIZ Karlsruhe) des Kompetenznetzwerks,

Datenworkflow der Deutschen Digitalen Bibliothek



- Bereitstellung von (spartenspezifischen) Erstinformationen an interessierte Einrichtungen, Inhaltliche und rechtliche Klärungen:
 - Inhaltliche Abstimmung hinsichtlich der zu liefernden Sammlungen anhand allgemeiner Kriterien und individueller Erfordernisse,
 - Festlegung und Festschreibung der Lieferbedingungen, auch im Hinblick auf die Weitergabe der Daten an Europeana,
- Organisatorische und fachliche Arbeiten gemeinsam mit dem Partner und dem FIZ Karlsruhe:
- Verfahrensfestlegung für die Lieferung von Datenformaten und digitalen Objekten,
 - Erstellung von Regeln für die Repräsentation der Lieferdaten im gemeinsamen Metadatenschema der DDB (konzeptionelles Mapping),
 - Überprüfung der Datenkonsistenz im gesamten Workflow der Datenintegration,
 - Dokumentation der Testergebnisse und Übergabe der Ergebnisse an den technischen Betreiber und den Datenlieferanten.

Schritt eins: Kontaktaufnahme und Registrierung

Die Registrierung³⁾ ist der erste Schritt, der Kultur- und Wissenschaftseinrichtungen den Weg in die DDB ebnet. Mit dem Ziel, alle deutschen Kultur- und Wissenschaftseinrichtungen zu erfassen, hat das Kompetenznetzwerk Einrichtungen aus allen Sparten (Archiv, Bibliothek, Denkmalpflege, Mediathek/AV-Medien, Museum und Wissenschaft) eingeladen. Alle Einrichtungen werden nach der Registrierung auf einer virtuellen Kultur- und Wissenschaftslandkarte präsentiert und so als Teil der DDB an prominenter Stelle sichtbar. Hinsichtlich der Datenzulieferung dient die Registrierung dazu, qualifizierte Ansprechpartner für die Kommunikation mit der DDB-Servicestelle zu gewinnen, die die Themen Metadaten, Datenakquise, Ladeprozesse und Digitalisierung bei den jeweiligen Einrichtungen betreuen.

Virtuelle Kultur- und Wissenschaftslandkarte präsentiert die teilnehmenden Einrichtungen

Für die Kontaktaufnahme zur DDB gibt es zwei Möglichkeiten:

- Die DDB-Servicestelle nimmt direkt Kontakt zu einem potenziellen Partner auf. Die Einrichtung registriert sich im nächsten Schritt.
- Die Einrichtung registriert sich aufgrund der Einladung durch das Kompetenznetzwerk, welches wiederum den Kontakt zur Servicestelle für die Einrichtung vermittelt.

Schritt zwei: Inhaltliche und rechtliche Klärungen

Im nächsten Schritt legt die Servicestelle gemeinsam mit der Einrichtung fest, welche Sammlungen und spartenspezifische Informationen an die DDB geliefert werden sollen. Für sammlungsbezogene Informationen (Metadaten) und digitale Objekte, die einer Sammlung angehören, gelten bestimmte Qualitätskriterien, die das Kompetenznetzwerk definiert hat und die sich an gängigen Metadaten- und Digitalisierungsstandards orientieren. Die Auswahl der benötigten Informationen erfolgt mithilfe von Metadaten-Kernsets. Um eine optimale Vernetzung, Suche und Darstellung der Objekte und Institutionen in der DDB zu ermöglichen, berücksichtigen die Metadaten-Kernsets auch sparten- und werkspezifische Besonderheiten und die damit verbundenen Metadatenangaben. Desweiteren werden die Bedingungen der Weitergabe der Daten an das europäische Kulturportal Europeana⁴⁾ geklärt und ein entsprechendes Set aus den Lieferdaten für den Transfer nach Europeana gebildet. Daten, die an die Europeana weitergegeben werden, stehen prinzipiell aufgrund der entsprechenden Anforderung der europäischen Kollegen unter der Lizenz CC0 1.0 Universal Public Domain Dedication⁵⁾. Die Kooperation mit der DDB wird durch einen Vertrag zwischen der Einrichtung und der DDB-Geschäftsstelle formalisiert. Spezifische Informationen zu Qualitätskriterien für Metadaten und digitale Objekte sind in der dem Vertrag beigelegten Anlage zu finden.

Metadaten-Kernsets

Schritt drei: Organisatorische und fachliche Arbeiten

Im Zuge der Datenanalyse klärt die Servicestelle mit dem Partner, in welchem Format und über welche Schnittstelle die Daten der DDB zur Verfügung gestellt werden können. Die Metadaten sollen nach Möglichkeit in einem von der DDB bevorzugten Metadatenstandard (z. B. MARCXML, EAD, METS/MODS, Dublin Core, LIDO) oder in einem proprietären, aber klar definierten Metadatenformat vorliegen, welches innerhalb einer Sparte als einheitliches Eingangsformat für die DDB definiert wurde. Die Liste der präferierten Metadatenformate findet sich auf der DDB Website. Zum Dienstleistungsangebot der Servicestelle gehört es auch, Partner bei der Spezifikation eines Datenformates für die DDB zu unterstützen, sofern die Daten nicht in einem der DDB-Eingangsformate lieferbar sind. Als Übertragungswege für Metadaten und digitale Objekte stehen den Partnern die gängigen Schnittstellen OAI-PMH und ftp-Upload zur Verfügung. Neben Metadaten ist auch die Übertragung von Derivaten in Form von Thumbnails sowie Bild-, Video-, oder Audiodateien an die DDB möglich und erwünscht.

Das Kompetenznetzwerk der DDB hat ein Modell entworfen, das sowohl allen spartenspezifischen Eigenheiten als auch gleichen Elementen innerhalb der Objektbeschreibungen und deren Verlinkung zu anderen Datenbeständen Rechnung trägt. Mithilfe dieses Modells werden schließlich die Kulturobjekte der verschiedenen Einrichtungen im Onlineportal der DDB nachgewiesen. Das DDB-Metadatenmodell für die Verlinkung und facettenbasierte Suche der Datenbestände basiert auf dem CIDOC (Comité international pour la documentation) Conceptual Reference Model (CRM)⁹⁾, das von dem Fachkomitee für Dokumentation im Internationalen Museumsbund CIDOC (ICOM) entwickelt wurde und seit 2006 ein offizieller ISO-Standard ist (ISO 21127:2006). CIDOC-CRM ist ein formalisiertes Begriffsmodell, das die Integration, Zugriffsvermittlung und den Austausch verschiedenartig strukturierter Information aus dem Bereich des kulturellen Erbes unterstützt und daher als geeignetes Referenzmodell für die DDB ermittelt wurde. Das Modell für die verschiedenen

Objektansichten innerhalb der DDB basiert auf den spartenspezifischen Metadaten-Kernsets der Kultur- und Wissenschaftseinrichtungen.

Damit die heterogenen Daten aus den Eingangsformaten der verschiedenen Einrichtungen in das gemeinsame Metadatenformat der DDB überführt werden können, ist die Erstellung von Regeln erforderlich (konzeptionelles Mapping), anhand derer die Transformation der Daten erfolgt (technisches Mapping). Diese Arbeiten erfolgen in enger organisatorischer Abstimmung zwischen dem Datenlieferanten, der Servicestelle und dem Betreiber, wobei zunächst die Servicestelle gemeinsam mit dem Datenlieferanten das konzeptionelle Mapping erstellt. Das konzeptionelle Mapping definiert die Transformationsregeln für die:

- Abbildung des Eingangsformats auf das DDB-Metadatenformat,
- Zuweisung von Metadaten zu den DDB-Facetten,
- Anzeige der Metadaten in der Trefferliste und der Anzeige der Volltreffer,
- Verwendung der Metadaten in Suche und erweiterter Suche.

Das technische Mapping erstellt das FIZ Karlsruhe in Absprache mit der Servicestelle. Während dieses Vorgangs werden die Regeln in anwendbare Transformationsskripte (xslt-Skripte) umgewandelt. Die gelieferten Testdaten werden abhängig vom Format, in dem die Einrichtung ihre Daten zur Verfügung gestellt hat, auf einem der folgenden Wege transformiert: Beim direkten Mapping werden die Daten aus dem Eingangsformat unmittelbar auf das Metadatenformat der DDB abgebildet. Für Formate, die keinem der DDB-Eingangsformate entsprechen, erfolgt zunächst eine Transformation auf ein passendes sparten- oder sammlungsspezifisches Eingangsformat, für das bereits ein Mapping auf das Metadatenformat der DDB existiert (indirektes Mapping). Das Kompetenznetzwerk hat bereits Mappings für alle aktuell bekannten Eingangsformate erstellt, die für die Integration weiterer Datenquellen in die DDB nachnutzbar sind und somit zukünftigen Datenlieferanten den Weg in die DDB erleichtern.

Nach erfolgreichem Ingest stellt das FIZ Karlsruhe die Testdaten in einem Metadaten-Testsystem zur Verfügung. Die Qualitätssicherung der Daten und Mappings erfolgt in einem iterativen Prozess, in

Einhaltung
von Metadaten-
standards

Das DDB-
Metadatenmodell

Konzeptionelles
Mapping

Technisches
Mapping

dem die Einrichtung und die Servicestelle die Daten und Mappings im Metadaten-Testsystem überprüfen und die Ergebnisse aus dem jeweiligem Abnahmeschritt dem FIZ zur Verfügung stellen. Nachdem eine Institution ihre Daten nach eingehender Prüfung final abgenommen und freigegeben hat, werden sie in das Produktivsystem der DDB geladen. Am Ende des Prozesses sind die Kulturobjekte einer Einrichtung über das digitale Schaufenster der DDB für alle interessierten Nutzerinnen und Nutzer zugänglich.

Die Realisierung des Datenexports ist ein wichtiger Meilenstein, an dem das Kompetenznetzwerk kontinuierlich arbeitet, um die in der DDB zusammengeführten Sammlungen auch an Europeana liefern und somit den deutschen Beitrag für das europäische Kulturportal leisten zu können. Der Grundstein für eine erfolgreiche Kooperation mit Europeana wurde bereits gelegt: Informationsexperten aus dem DDB-Kompetenznetzwerk haben ein initiales Mapping eines DDB-Eingangsformats zum Austauschformat der Europeana (Europeana Data Model⁷⁾) erstellt. Das Mapping- und Liefer-

verfahren wird gegenwärtig prototypisch anhand von Testdatenlieferungen an Europeana erprobt und gemeinsam mit Informationsexperten der Europeana begutachtet. Die Ergebnisse aus diesem Testscenario fließen in den Datenworkflow der DDB ein. Auch im Verfahren der Datenablieferung an Europeana wird die Servicestelle der DDB künftig eine zentrale Funktion übernehmen.

Die DDB geht im Spätherbst dieses Jahres mit einer Betaversion online. Innerhalb des Betabetriebs steht den Nutzerinnen und Nutzern ein fester Datenbestand mit rund 6,4 Mio. Datensätzen aus 360 deutschen Kultur- und Wissenschaftseinrichtungen zur Verfügung. Um zu Beginn des öffentlichen Betriebs eine möglichst hinreichende Stabilität des Systems zu gewährleisten, geht die DDB mit einem reduzierten Funktionsumfang an den Start. Nach und nach werden dann weitere Funktionalitäten hinzu geschaltet und die Daten der registrierten Einrichtungen hinzu geladen. Das Portal soll auf diesem Weg kontinuierlich weiter ausgebaut werden.

Anmerkungen

1 Über das Kompetenznetzwerk der Deutschen Digitalen Bibliothek informiert die Website:

<<http://www.deutsche-digitale-bibliothek.de>>

2 <<http://www.fiz-karlsruhe.de/>>

3 Informationen zur Registrierung finden sich auf der Website der Deutschen Digitalen Bibliothek:

<<http://www.deutsche-digitale-bibliothek.de>>

4 <<http://www.europeana.eu>>

5 <<http://creativecommons.org/publicdomain/zero/1.0>>

6 <<http://www.cidoc-crm.org/>>

7 <<http://pro.europeana.eu/web/guest/aggregators-and-providers>>