

Errichtung eines nationalen Zeitungsportals auf der Basis der organisatorischen und technischen Infrastruktur der Deutschen Digitalen Bibliothek (DDB) – „DDB-Zeitungsportal“

Beschreibung des Vorhabens – Projektanträge im Bereich „Wissenschaftliche Literaturversorgungs- und Informationssysteme“ (LIS)

**LIS-Förderprogramm: e-Research-Technologien
(Vorhaben zur Implementierung von e-Research-Technologien)**

Dr. Elisabeth Niggemann (Frankfurt am Main), Matthias Razum (Eggenstein-Leopoldshafen), Prof. Dr. Thomas Bürger (Dresden), Barbara Schneider-Kempf (Berlin)

Beschreibung des Vorhabens (Öffentliche Kurzfassung)

1 Ausgangslage und eigene Vorarbeiten

Bedarf für ein nationales Zeitungsportal / DFG-Projekt „Digitalisierung historischer Zeitungen“

In den Jahren 2013 bis 2016 wurde unter Leitung der Sächsischen Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB) und unter Beteiligung weiterer Bibliotheken¹ das DFG-geförderte Pilotprojekt „Digitalisierung historischer Zeitungen“ durchgeführt, in dessen Rahmen auch ein neuer ZDB-Katalog mit erweiterten Suchmöglichkeiten für die Zeitungssuche sowie mit spezifischen Features zur Unterstützung von Digitalisierungsprojekten entwickelt² und der DFG-Viewer optimiert wurde.³ Als weiteres Ergebnis wurde ein Masterplan⁴ vorgelegt, der auf die Einrichtung einer DFG-Förderlinie zur großflächigen Unterstützung von Digitalisierungsvorhaben zielt, die historische Zeitungen zum Gegenstand haben, und dafür ein entsprechendes Vorgehensmodell einschließlich heranzuziehender Auswahlkriterien beinhaltet.

Andererseits adressiert der Masterplan auch infrastrukturelle Desiderate im Kontext der Zeitungsdigitalisierung und benennt in diesem Zusammenhang vor allem den aus der Wissenschaft heraus geäußerten dringenden Bedarf für einen zentralen Zugang zu den digitalisierten Zeitungen – etwa in Form eines nationalen Zeitungsportals. Ein solcher zentraler Zugang ergibt insbesondere für die wissenschaftliche Nutzung einen deutlichen Mehrwert.

Diese Erkenntnis stützt sich auch auf ein entsprechendes Meinungsbild, das im Rahmen dieses DFG-Projekts bei einem Workshop mit Wissenschaftlerinnen und Wissenschaftlern im Herbst 2014 in Bremen erhoben wurde. Zentrale Anforderungen waren dabei vor allem (1) eine übergreifende Volltextsuche in den digitalisierten Zeitungsbeständen, (2) browsingbasierte Einstiegspunkte (etwa über Kalender und Zeitungstitel), (3) eine integrierte Anzeigekomponente (Viewer) mit Funktionen unter anderem für Treffer-Highlighting, Seitenrotation und -zoom und dem nahtlosen Kopieren gefundener Volltextstellen (Copy&Paste) sowie (4) eine konsistenten Möglichkeit, auf die enthaltenen Zeitungen bzw. Einzelausgaben persistent zu referenzieren und sie somit zitierfähig zu machen.

Einen zentralen Zugang zu digitalisierten Zeitungen gibt es in einigen anderen Ländern bereits – etwa mit ANNO⁵ in Österreich oder Trove⁶ in Australien – und in Deutschland auch mit regionalen Abdeckungen

¹ Die Konsortialpartner des Projekts waren die Staatsbibliothek zu Berlin (SBB), die Staats- und Universitätsbibliothek Bremen (SuUB), die Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB), die Deutsche Nationalbibliothek (DNB), die Universitäts- und Landesbibliothek Halle (ULB) sowie die Bayerische Staatsbibliothek (BSB).

² Siehe <http://zdb-katalog.de/index.xhtml>.

³ Siehe <http://gepris.dfg.de/gepris/projekt/271857751>.

⁴ Empfehlungen zur Digitalisierung historischer Zeitungen in Deutschland (Masterplan Zeitungsdigitalisierung) – Ergebnisse des DFG-Projektes „Digitalisierung historischer Zeitungen“ Pilotphase 2013-2015, 29.01.2016 (Partner: SBB (Berlin), SuUB (Bremen), SLUB (Dresden), DNB (Frankfurt), ULB (Halle), BSB (München)). S. http://www.zeitschriftendatenbank.de/fileadmin/user_upload/ZDB/z/Masterplan.pdf.

⁵ Siehe <http://anno.onb.ac.at/>.

⁶ Siehe <http://trove.nla.gov.au/>.

(digiPress⁷, ZEFYS⁸). Ihnen ist gemein, dass sie zur Bereitstellung der Funktionalitäten auf eine vollständig zentrale Datenhaltung zurückgreifen können – das heißt, insbesondere auch alle Images liegen zentral vor.

Im Rahmen des Pilotprojekts stellte sich nun die Frage nach konkreten Realisierungsoptionen für ein nationales Zeitungsportal für Deutschland. Dabei konzentrierten sich die Überlegungen relativ schnell auf die Deutsche Digitale Bibliothek (DDB)⁹ – und zwar vor allem angesichts der positiven Entwicklungen, die diese genommen hatte: Nach einer etwa anderthalbjährigen Betaphase ging die DDB Ende März 2014 mit der öffentlichen Präsentation einer ersten Vollversion in den Regelbetrieb und etabliert sich seither als nationales Nachweis- und Zugangsportal für das digitalisierte Kultur- und Wissenserbe in Deutschland.

Für die DDB als technische und organisatorische Basis eines nationalen Zeitungsportals sprechen insbesondere folgende Erwägungen:

- die bestehende technische und organisatorische Infrastruktur mit erprobten und etablierten Technologien und Prozessen im Bereich der Zusammenführung, Verarbeitung und Darstellung verteilter digitaler Bestände und Sammlungen und einer vertrauenswürdigen und stabilen Betriebssituation beim technischen Betreiber der DDB (FIZ Karlsruhe),
- die langfristige Perspektive der DDB als gesamtstaatliches Vorhaben, dessen dauerhafte Weiterführung seitens des Bundes und der Länder als Unterhaltsträger als gesichert gelten darf,
- die erklärte Bereitschaft der Verantwortlichen innerhalb der DDB, eine um zeitungsspezifische Komponenten erweiterte technische und organisatorische Infrastruktur dauerhaft zu betreiben,
- die Tatsache, dass für durch öffentliche Förderung digitalisierte Sammlungen und Bestände ohnehin Verpflichtungen bzw. starke Empfehlungen bestehen, diese an die DDB zu liefern, und die DDB bereits jetzt zahlreiche digitalisierte historische Zeitungen aus mehreren teilnehmenden Bibliotheken nachweist.

In dem auf verteilte Datenbestände ausgerichteten Betriebsmodell liegt ein wesentliches Alleinstellungsmerkmal der DDB etwa gegenüber vollständig zentralen Lösungen wie ANNO in Österreich oder auch digiPress der BSB. Denn während in diesen Fällen alle Daten (auch die hochauflösenden Bilddateien) zentral gehalten werden, ist dies im Kontext der DDB-Architektur nicht erforderlich. Zum einen können die an einem nationalen Zeitungsportal teilnehmenden Einrichtungen nicht dazu verpflichtet werden, ihre Digitalisate an zentraler Stelle zu speichern. Zum anderen würde ein solcher vollständig zentraler Ansatz auch Daten duplizieren, was für Speicherung und Betrieb erhebliche Zusatzkosten mit sich bringt.

Die Realisierung des Zeitungsportals auf der Grundlage der DDB als spartenübergreifend angelegtes Gesamtvorhaben hat zudem den Vorteil, dass darin auch die Integration von Zeitungsbeständen problemlos möglich ist, die nicht in Bibliotheken, sondern beispielsweise in Archiven und anderen Einrichtungen lagern.

Deutsche Digitale Bibliothek (DDB)

Die Deutsche Digitale Bibliothek (DDB) verfolgt das Ziel, das Kultur- und Wissenserbe aus Deutschland in digitaler Form zusammenzuführen und über ihre Angebote an unterschiedliche Zielgruppen zu vermitteln. Dazu gehören neben dem DDB-Portal vor allem Anwendungen, die auf dem API der DDB basieren. Derzeit weist die DDB ca. 23 Mio. Objekte aus ca. 360 Einrichtungen aller Kultursparten nach. Alle vier Antragsteller sind aktive Mitglieder im Kompetenznetzwerk „Deutsche Digitale Bibliothek“ mit derzeit 14 Mitgliedern.

- Der DDB-Geschäftsbereich Technik, Entwicklung, Service mit der übergreifenden Projektkoordination und -steuerung sowie der Servicestelle ist an der **Deutschen Nationalbibliothek** angesiedelt. Hier werden die konzeptionelle Weiterentwicklung und die Betreuung der Datenpartner verantwortet.
- **FIZ Karlsruhe** ist der technische Betreiber der DDB und darüber hinaus verantwortlich für die Softwareentwicklung eines Großteils der DDB-Komponenten.
- Die **Stiftung Preußischer Kulturbesitz**, zu der die **Staatsbibliothek zu Berlin (SBB)** gehört, ist gemäß Verwaltungs- und Finanzabkommen als Trägerin der DDB-Geschäftsstelle bestellt und mit dem Geschäftsbereich Finanzen, Recht, Kommunikation betraut. Ihr Präsident, Prof. Dr. Hermann Parzinger, ist Vorstandssprecher der DDB.

⁷ Siehe <https://digiPress.digitale-sammlungen.de/>.

⁸ Siehe <http://zefys.staatsbibliothek-berlin.de/>.

⁹ Siehe <https://www.deutsche-digitale-bibliothek.de/>.

- Die **SLUB Dresden** betreibt die DDB-Fachstelle Mediathek-Bild/Ton und ist im Rahmen eines Pilotprojekts bereits mit der prototypischen Realisierung einer auf dem DFG-Viewer basierenden Anzeige Komponente für die DDB betraut.

Die DDB soll das organisatorische und technische Rückgrat für das DDB-Zeitungsportal bilden: Sie stellt dafür nicht nur die zentrale Datenhaltung (Metadaten) bereit, sondern übernimmt auch die Prozesse zur Zusammenführung und Integration der Datenbestände, die im Zeitungsportal sichtbar werden (insbesondere Akquise, Datenclearing, Transformation, Review, Ingest und Aktualisierung). Durch die enge Verzahnung mit den Strukturen der DDB und ihrer technischen Infrastruktur werden erhebliche Synergieeffekte für das Zeitungsportal erzielt und dessen langfristiger Betrieb sichergestellt.

Die Basis-Architektur der DDB wurde in den Jahren 2016-17 grundlegend erneuert. Mit der verteilten NoSQL-Datenbank Apache Cassandra¹⁰ und dem verteilten Compute-Framework Apache Spark¹¹ – der Erweiterung von Hadoop¹² und dem deutlich verbesserten Nachfolger von MapReduce – sowie der auch bisher schon verwendeten Suchmaschine Apache Solr¹³ basiert das DDB-Backend nun durchgängig auf modernen und skalierbaren Technologien, die sie auch für funktionale Erweiterungen sowie erhebliche mengenmäßige Zuwächse zukunftsfähig macht.

Im Rahmen eines DDB-internen Vorprojekts wurden zudem zwei wesentliche Aspekte bezüglich der technischen Machbarkeit eines auf der DDB basierenden Zeitungsportals untersucht – nämlich

1. die Entwicklung eines Objekt-Viewers zur integrierten Darstellung mehrseitiger Druckwerke, für die die erforderlichen Digitalisate nicht zentral aufseiten der DDB vorgehalten werden, sondern verteilt bei den einzelnen Datenpartnern liegen,
2. die prototypische Umsetzung einer Volltextsuche.

In beiden Bereichen konnte die grundsätzliche technische Realisierbarkeit nachgewiesen werden.

Für den Objekt-Viewer wurde auf die durch den DFG-Viewer¹⁴ genutzte Basis-Technologie Kitodo.Presentation¹⁵ zurückgegriffen, für die im Wesentlichen ein angepasstes Layout mit einem auf das DDB-CI abgestimmten Design entwickelt wurde. Dabei wird der Viewer als Gesamtseite an den Browser ausgeliefert, die im Kontext der DDB als Lightbox eingeblendet wird. Um die Integration des Viewers in die Objektseiten für Zeitungen zu verbessern und damit insbesondere zu vermeiden, dass kleinste nutzerseitige Interaktionen ein vollständiges Neuladen der Seite erfordern, muss die Kommunikation zwischen Portal und Kitodo.Presentation auf die Ebene einzelner Seitenbestandteile (Widgets) heruntergebrochen werden. Diese Weiterentwicklung steht für die Realisierung des Zeitungsportals noch aus (siehe AP 7).

Für die Realisierung der Volltextsuche wurde der vorhandene Solr-basierte Index aufseiten der DDB, der bisher nur Metadaten enthält, entsprechend erweitert. Die notwendigen Prozesse zur routinemäßigen Verarbeitung von Volltexten müssen allerdings noch spezifiziert und implementiert werden. Auch die eigentliche Suchfunktion muss noch erheblich ausspezifiziert und erweitert werden (beispielsweise um eine unscharfe Suche sowie wörterbuchbasierte Ansätze zur Synonymsuche und Berücksichtigung alter Wortformen) – siehe AP 8.

Zeitschriftendatenbank (ZDB)

Die Zeitschriftendatenbank (ZDB) ist das zentrale Nachweisinstrument für Zeitschriften, Zeitungen, Schriftenreihen und andere periodisch erscheinende Veröffentlichungen aus allen Ländern, in allen Sprachen, ohne zeitliche Einschränkungen, in gedruckter, elektronischer oder anderer Form.

Im Oktober 2017 wies die ZDB folgende Mengengerüste nach:

- ZDB-Titelaufnahmen	1.850.000
- davon Serienaufnahmen	271.000
- davon elektronische Ressourcen	198.000
- davon Zeitungen	61.000
- Einrichtungen mit ZDB-Beständen	4.317

Die Daten der ZDB zeichnen sich insbesondere durch disambiguierte Titelenitäten, denen eindeutige Identifikatoren (ZDB-IDs) zugewiesen sind, aus. Die hohe Qualität der ZDB-Daten ist seit vielen Jahren

¹⁰ Siehe <http://cassandra.apache.org/>.

¹¹ Siehe <https://spark.apache.org/>.

¹² Siehe <https://hadoop.apache.org/>.

¹³ Siehe <http://lucene.apache.org/solr/>.

¹⁴ Siehe <https://dfg-viewer.de/>.

¹⁵ Siehe <https://www.kitodo.org/software/kitodopresentation/>.

durch qualifizierte Katalogisierer/-innen in der SBB gewährleistet. ZDB-IDs werden in Deutschland und im deutschsprachigen Ausland in den verschiedensten Nachweis- und Präsentationsumgebungen verwendet. Die ZDB ist fest in die Strukturen der deutschen und österreichischen Bibliotheksverbände integriert und versorgt über diese die teilnehmenden Bibliotheken mit gewünschten spezifischen Ausschnitten des ZDB-Datenbestandes.

Im Ergebnis des Pilotprojekts „Digitalisierung historischer Zeitungen“ haben die dort kooperierenden Bibliotheken einstimmig den Einsatz der ZDB als einrichtungsübergreifendes Steuerungsinstrument für Zeitungs-Digitalisierungsprojekte empfohlen. Gleichzeitig haben die Bibliotheken auf bestehenden Optimierungsbedarf des ZDB-OPACs¹⁶ hingewiesen und insbesondere die Implementierung spezifischer Such- und Präsentationsmöglichkeiten für die Materialart Zeitungen gewünscht.

Die Betreiber der ZDB, die Deutsche Nationalbibliothek (DNB) und die Staatsbibliothek zu Berlin (SBB), haben diesem Wunsch entsprochen und im Rahmen des einjährigen DFG-geförderten Projekts „Weiterentwicklung der Zeitschriftendatenbank (ZDB) zum Steuerungs- und Nachweisinstrument für Zeitungsdigitalisierungsprojekte und für Zeitungen“ (Laufzeit: 2014-15) den Webauftritt der ZDB grundsätzlich neu gestaltet. Der neue Webkatalog der ZDB¹⁷ zeichnet sich mit Blick auf Zeitungen insbesondere durch die folgenden Eigenschaften und Funktionalitäten aus:

- grundsätzliche Überarbeitung aller Oberflächen in Richtung vereinfachte, intuitive Bedienbarkeit; Zielgruppen sind vor allem Wissenschaftlerinnen und Wissenschaftler sowie nicht-bibliothekarische Einrichtungen, denen der Informationszugang erheblich erleichtert wurde,
- umfangliche Visualisierungen komplexer Sachverhalte:
 - einrichtungsübergreifender Nachweis und Abgleich von Beständen, um die oft komplexe Überlieferungslage übersichtlich begreifbar zu machen,
 - Zuordnungen von Zeitungen und bestandshaltenden Einrichtungen zu Orten auf geographischen Karten,
 - Dynamisch generierte Visualisierungen von Titelhistorien und Titelrelationen (Vorgänger- und Nachfolger-Zeitungen, Zeitungsbeilagen),
 - Ermöglichung spezifischer Zeitungssuchen durch Bereitstellung entsprechender Selektionsfacetten,
 - Ermöglichung der Eingabe von für Zeitungsdigitalisierungsprojekte notwendigen Informationen (mittels vom Webkatalog unabhängiger Redaktionsschnittstellen) und Einrichtung entsprechender Suchmöglichkeiten (Ankündigung von Digitalisierungsvorhaben, Art der Digitalisierung).

Die ZDB soll für das geplante nationale Zeitungsportal grundsätzlich als wichtige Datengrundlage fungieren. In einer zweiten Projektphase soll das Zeitungsportal dann auch auf der Nutzeroberfläche mit dem ZDB-Katalog verknüpft werden.

Deutsche Nationalbibliothek (DNB)

Die DNB hat die Aufgabe der lückenlosen Sammlung, Verzeichnung, Archivierung und Bereitstellung aller deutschen und deutschsprachigen Publikationen ab 1913 sowie im Ausland erscheinender Germanica und Übersetzungen deutschsprachiger Werke. Daneben pflegt sie im nationalen und internationalen Rahmen kooperative Beziehungen und ist federführender Partner in der deutschen Regelwerks- und Normierungsarbeit sowie maßgeblich an der Entwicklung internationaler Standards beteiligt. Mit vielen ihrer Angebote und Aktivitäten zielt die DNB auch auf die spartenübergreifende Vernetzung der Kultursparten ab – etwa mit der durch sie betriebenen Gemeinsamen Normdatei (GND), dem gegenwärtig zur dauerhaften Identifizierung von Kulturobjekten erweiterten URN-Service und nicht zuletzt mit ihrer zentralen Rolle innerhalb der DDB als Koordinatorin des Kompetenznetzwerks und Betreiberin der Servicestelle. Die DNB ist in einer Vielzahl von kooperativen Projektvorhaben in diesen Feldern aktiv. Gemeinsam mit der SBB ist sie zudem für die Pflege und Weiterentwicklung der ZDB verantwortlich.

FIZ Karlsruhe

FIZ Karlsruhe arbeitet seit mehreren Jahren intensiv und erfolgreich an e-Research-Lösungen und Dienstleistungen für unterschiedliche Disziplinen im Rahmen von Forschungsprojekten und kommerziellen Aufträgen. Schwerpunkte bilden dabei kooperative Projekte mit einer Vielzahl von Hochschulen, Forschungs- und wissenschaftsnahen Einrichtungen. Mit Informationsservices, virtuellen Forschungsumgebungen, Diensten zum Forschungsdatenmanagement und zur digitalen Langzeitarchivierung deckt FIZ

¹⁶ Siehe <http://dispatch.opac.dnb.de/>.

¹⁷ Siehe <http://zdb-katalog.de/index.xhtml>.

Karlsruhe weite Teile des Lebenszyklus von Daten ab. Zentral ist bei allen Vorhaben die enge Verknüpfung von Entwicklung und Betrieb. Hierbei kann FIZ Karlsruhe auf sein eigenes Rechenzentrum zurückgreifen.

Im Bereich der Informationsservices sind neben der DDB das Archivportal-D zu nennen, das im Rahmen zweier DFG-geförderter Projekte (2013-2017) auf Grundlage der DDB-Infrastruktur aufgebaut wurde und seit 2014 im produktiven Betrieb bei FIZ Karlsruhe läuft. Seit 2016 verantwortet FIZ Karlsruhe Entwicklung und Betrieb für das DFG-geförderte Portal für Technikfolgenabschätzung OpenTA. Darüber hinaus betreibt FIZ Karlsruhe die IT-Infrastruktur für das Kompetenzzentrum Bibliometrie.

Im EU-Projekt SCAPE (Scalable Preservation Environments, 2011-2014) befasste sich FIZ Karlsruhe mit den Herausforderungen der digitalen Langzeitarchivierung sehr großer Repositorien. Im DFG-geförderten Projekt "Nationales Hosting elektronischer Ressourcen" beschäftigte sich FIZ Karlsruhe mit Strategien für den dauerhaften Zugriff auf e-Books und e-Journals. Aus dem DFG-geförderten Projekt RADAR (2013-2016) ging Anfang 2017 ein gleichnamiger Dienst für die Archivierung und Publikation von Forschungsdaten hervor.

Im DFG-geförderten Projekt e-Kinematix baute FIZ Karlsruhe eine virtuelle Forschungsumgebung (VFU) für die Getriebetechnik auf. Seit 2015 entsteht die im Rahmen des Leibniz-Wettbewerbs geförderte VFU TOPORAZ für die objekt- und raumbezogene geisteswissenschaftliche Forschung.

Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB)

Als Konsortialführer der DFG-geförderten Pilotprojekte zur Zeitungsdigitalisierung der Staats-, Landes- und Universitätsbibliotheken Berlin, Bremen, Dresden, Halle, München und der Deutschen Nationalbibliothek hat die SLUB die Massendigitalisierung von rund 450.000 verfilmten Zeitungsseiten erprobt. Schwerpunkte waren dabei die Entwicklung von Format- und Präsentationsstandards¹⁸ für digitalisierte historische Zeitungen sowie die Weiterentwicklung frei nachnutzbarer, medienspezifischer technischer Module für den DFG-Viewer¹⁹, für die Produktions- und Präsentationssoftware Kitodo²⁰ und für eine automatisierte Integration der Volltexterkennung in diesen Digitalisierungsworkflow. Diese Initiative wurde anschließend im Rahmen des Sächsischen Landesdigitalisierungsprogramms²¹ fortgesetzt und um eine (noch experimentelle) Strukturerkennung zur automatischen Ausgaben- und Artikelseparierung ergänzt.

Im Ergebnis dieser Projekte wurden etliche Zeitungen sowohl nach Filmen wie auch nach Originalvorlagen digitalisiert. Die Digitalisate werden mit unterschiedlichen Erschließungstiefen auf der Basis der vereinbarten Mindeststandards frei zugänglich und nachnutzbar präsentiert.²²

Auch über die Zeitungsdigitalisierung hinaus engagiert sich die SLUB aktiv in der Entwicklung und dem Release Management der quelloffenen Software Kitodo und dem darauf aufbauenden DFG-Viewer. Insbesondere treibt sie die medientypologische Erweiterung dieser freien Werkzeuge voran und pflegt die zugrundeliegenden DFG-Formatstandards (METS, MODS, ALTO, TEI). Im Rahmen der technischen Erneuerung der Deutschen Digitalen Bibliothek hat die SLUB die prototypische Integration von Kitodo.Presentation als Viewer in das DDB-Portal vorgenommen und damit die grundsätzliche Eignung der Technologie erwiesen. Diese Vorarbeiten sollen nun Basis für die im Zeitungsportal zu realisierenden Präsentationskomponenten sein (siehe AP 7).

Staatsbibliothek zu Berlin (SBB)

Die SBB verfügt über den größten und umfangreichsten Zeitungsbestand deutscher Bibliotheken. In den letzten Jahren wurden signifikante Bestandssegmente historischer Zeitungen digitalisiert die über das Zeitungsinformationssystem ZEFYS der SBB angeboten werden. Im Oktober 2017 wurden 281.990 Ausgaben von 192 historischen Zeitungen in ZEFYS nachgewiesen. Darüber hinaus wurden wichtige Sonderprojekte zur Amtspresse Preußens, zur DDR-Presse und zu historischen Berliner Zeitungen durchgeführt, deren Ergebnisse ebenfalls zugänglich sind.

Zeitungsdigitalisierungen in der SBB werden sowohl durch externe Dienstleister als auch im eigenen Digitalisierungszentrum durchgeführt, das seit 2010 betrieben wird. In den vergangenen Jahren wurden dort zahlreiche Projekte durchgeführt, die von Massendigitalisierungsvorhaben – z.B. im Rahmen der nationalen VD-Kampagnen oder von Europeana – bis hin zur konservatorisch begleiteten Boutique-Digitalisierung von mittelalterlichen Handschriften und fragilen Sondermaterialien reicht. Durchgeführt werden alle

¹⁸ Siehe <http://dfg-viewer.de/profil-der-metadaten/>.

¹⁹ Siehe <http://dfg-viewer.de/>.

²⁰ Siehe <http://www.kitodo.org/>.

²¹ Siehe <https://www.slub-dresden.de/sammlungen/landesdigitalisierungsprogramm/>.

²² Siehe <https://digital.slub-dresden.de/kollektionen/143/>.

Digitalisierungsvorhaben auf Basis der Workflow-Software Kitodo.Production, an deren Weiterentwicklung und Funktionsanpassung an hinzukommende Materialgattungen – bislang vor allem Zeitungen und Archivalien – die SBB beteiligt ist, während für die Präsentation der Digitalisate ein im Haus realisierter, technisch ambitionierter Viewer zum Einsatz kommt.

Im Rahmen des Pilotprojekts zur Zeitungsdigitalisierung in Deutschland war die SBB neben der Neuentwicklung des ZDB-Webkatalogs auch für die zeitungsspezifische Funktionsanpassung von Kitodo verantwortlich.

Im Zuge einiger von der EU geförderter Projekte im Zeitungsbereich – zu nennen ist das EFRE-Projekt „Digitalisierung historischer Berliner Tageszeitungen“ und insbesondere das groß angelegte Projekt „Europeana Newspapers“ – wurden die Kompetenzen der SBB nachhaltig erweitert und vertieft. So hat die SBB in ihrer Funktion als Projektkoordinatorin von Europeana Newspapers an der Schaffung internationaler Standards und Empfehlungen für die Digitalisierung, Präsentation und Aggregation digitalisierter Zeitungen mitgewirkt und betreut auch weiterhin die thematische Sammlung „Europeana Newspapers“, die im Rahmen von EU-Projekten (bspw. Europeana DSI, Europeana Rise of Literacy) sowohl in Umfang als auch hinsichtlich der bereitgestellten technischen Funktionalitäten weiter ausgebaut wird.

Aktuell ist die SBB auch Kooperationspartner im Digging Into Data-Projekt „Oceanic Exchanges: Tracing Global Information Networks In Historical Newspaper Repositories 1840-1914“ und kooperiert mit dem Projekt „impresso - Media Monitoring of the Past: Mining 200 years of historical newspapers“.

1.1 Projektbezogene Publikationen

1.1.1 Veröffentlichte Arbeiten aus Publikationsorganen mit wissenschaftlicher Qualitätssicherung, Buchveröffentlichungen sowie bereits zur Veröffentlichung angenommene, aber noch nicht veröffentlichte Arbeiten

Altenhöner, Reinhard; Beinert, Tobias; Brantl, Markus; Luckfiel, Robert; Müller, Uwe (2015): Digitalisierung von Kulturgut. In: Praxis-Handbuch Bibliotheksmanagement / hrsg. von Rolf Griebel; Hildegard Schäffler; Konstanze Söllner. – Berlin; München; Boston: De Gruyter Saur, 2015, Bd. 2, S. 763–811.

Hubrich, Jessica: Visualisierung von Titelzusammenhängen. Titelhistorie und Netzwerkgraph im neuen ZDB-Katalog. In: Offen(siv)e Bibliotheken: Neue Zugänge, neue Strukturen, neue Chancen. 32. Österreichischer Bibliothekartag Wien, 15. - 18. September 2015. Hrsg. v. Bruno Bauer, Andreas Ferus und Josef Pauser. Graz-Feldkirch: Neugebauer, 2016. S. 269 - 282.

Hubrich, Jessica (2016): [Visual Representations of Newspaper Information](#). Proceedings of the IFLA 2016 News Media Conference, 20-22 April 2016, Hamburg, Germany.

Hubrich, Jessica und Hans-Jörg Lieder: Zeitungssuche interaktiv: der neue ZDB-Webkatalog. Verschriftlichung des Bibliothekarvortrags für Kongressband. Erschienen im Dezember 2014 in [o-bib 2014/1, S. 305 – 311](#).

Neudecker, Clemens (2016): [Who cares about yesterday's news? Use cases and requirements for newspaper digitization](#). Proceedings of the IFLA 2016 News Media Conference, 20-22 April 2016, Hamburg, Germany.

Neudecker, Clemens; Antonacopoulos, Apostolos (2016): [Making Europe's Historical Newspapers Searchable](#). Proceedings of the 12th IAPR International Workshop on Document Analysis Systems (DAS2016), 11-14 April 2016, Santorini, Greece, pp. 405-410. [DOI:10.1109/DAS.2016.83](#).

Neudecker, Clemens (2016a): [An Open Corpus for Named Entity Recognition in Historic Newspapers](#). Proceedings of the 10th edition of the Language Resources and Evaluation Conference, 23-28 May 2016, Portorož, Slovenia, pp. 4348-4352.

1.1.2 Andere Veröffentlichungen

Hubrich, Jessica; Lieder, Hans-Jörg (2016): Die Zeitschriftendatenbank und die Digitalisierung historischer Zeitungen in Deutschland. In: Dialog mit Bibliotheken 28 (2016) 2, S. 22–28.

Lieder, Hans-Jörg (2016): Coordinating Newspaper Digitisation: Some Facts and Figures. Paper for the IFLA International News Media Conference, Hamburg 2016. <https://blogs.sub.uni-hamburg.de/ifla-newsmedia/wp-content/uploads/2016/04/Lieder-Coordinating-Newspaper-Digitisation-%E2%80%93-Some-Facts-and-Figures.pdf>.

1.1.3 Patente

Entfällt.

2 Ziele und Arbeitsprogramm

2.1 Voraussichtliche Gesamtdauer des Projekts

Für das Projekt ist eine Gesamtlaufzeit von 48 Monaten vorgesehen. Eine Förderung wird zunächst für die Dauer von 24 Monaten beantragt. Die Förderung für eine anschließende zweite Projektphase wird vor dem Abschluss der ersten Projektphase beantragt. Gewünschter Beginn der Projektförderung ist der 1. April 2018.

2.2 Ziele

Das Hauptziel des Projekts „DDB-Zeitungsportal“ besteht darin, auf der technologischen und organisatorischen Basis der DDB ein Portal zu entwickeln und in den Produktivbetrieb zu überführen, das den zentralen und nutzerfreundlichen Zugang zu den digitalisierten historischen Zeitungen aus Deutschland ermöglicht. Durch den unmittelbaren Zugang zu den Zeitungsbeständen mit ihren Volltexten bildet das Zeitungsportal eine Ergänzung zu dem auf bibliographische Daten fokussierten ZDB-Katalog. Dieses nationale Zeitungsportal soll folgende Anforderungen und Rahmenbedingungen erfüllen:

1. Funktionale Anforderungen

Das Zeitungsportal soll sich hinsichtlich seines Funktionsumfangs an den Bedürfnissen wissenschaftlicher Nutzerinnen und Nutzer orientieren und (in der ersten Ausbaustufe) zumindest folgende Funktionalitäten bieten:

- eine übergreifende Volltextsuche in den digitalisierten Zeitungsbeständen (siehe AP 6 und AP 8),
- unterschiedliche browsende Einstiegspunkte (zumindest über einen Kalender und über Zeitungstitel sowie ggf. über Erscheinungsorte und Verbreitungsgebiete) (siehe AP 6),
- eine unmittelbar in die Portalumgebung integrierte Anzeigekomponente (Viewer), mit der die Zeitungen sowie einzelne Seiten (a) im Überblick und per stufenloser Zoom-Funktion im Detail betrachtet sowie gedreht und gedruckt, (b) die im Text gefundenen Suchtreffer per Highlighting auf Wortebene markiert sowie (c) Textabschnitte im Volltext per Copy&Paste direkt kopiert und weiterverwendet werden können (siehe AP 7).
- Eine konsistente Möglichkeit, auf die enthaltenen Zeitungen bzw. Einzelausgaben persistent zu referenzieren und sie somit zitierfähig zu machen, und damit auch eine klare Schnittstelle zu virtuellen Forschungsumgebungen bereitzustellen (siehe AP 7).

2. Content-Abdeckung und technische Voraussetzungen aufseiten der Content-Lieferanten

Das Zeitungsportal soll mittelfristig alle Bestände digitalisierter Zeitungen in deutschen Kultur- und Wissensinstitutionen nachweisen und in dem beschriebenen Funktionsumfang zugänglich und nutzbar machen. Dabei liegt der Fokus bis auf weiteres auf Beständen aus Bibliotheken und anderen Kulturerbeerichtungen und auf historischem, das heißt rechtfreiem Material²³. Dies entspricht auch der intendierten DFG-Förderlinie. Zum Abschluss der ersten Förderphase sollen

- prioritär Ergebnisse DFG-finanzierter Digitalisierungsprojekte (hier in erster Linie im Rahmen des Pilotprojekts durchgeführte bzw. mit Hilfe der intendierten Förderlinie geplante Digitalisierungsvorhaben) in das Zeitungsportal integriert werden, bei denen davon auszugehen ist, dass die DFG-Praxisregeln und insbesondere das METS/MODS-Anwendungsprofil Zeitungen berücksichtigt werden (siehe AP 5 und AP 9),
- weiterhin Bestände, die zwar in Form von METS/MODS geliefert werden können, aber nicht über eine Volltexterschließung (OCR) verfügen.

Insbesondere wird auch die SBB ihre bisher über ZEFYS bereitgestellten Zeitungsdigitalisate in Form von METS/MODS bereitstellen, so dass diese im Verlauf der ersten Förderphase des Projekts in das Zeitungsportal integriert werden können.

In der ersten Förderphase des Projekts sollen außerdem weitere Bestände identifiziert werden, die zukünftig – beispielsweise in einer zweiten Förderphase – in das Zeitungsportal integriert werden sollen, auch wenn sie über abweichende Formate bereitgestellt werden. Hier soll insbesondere untersucht werden, inwieweit die IIIF-Presentation-API ausreicht, um die funktional notwendigen Daten zu übermitteln bzw. wie sie aufseiten datengebender Einrichtungen ausgestaltet sein muss. Zu den zu betrachtenden Datenbeständen zählen beispielsweise diejenigen der Bayerischen Staatsbibliothek (digiPress). Darüber hinaus sollen weitere Bestände erhoben, dokumentiert und hinsichtlich ihrer technischen und formatseitigen Beschaffenheit analysiert werden (siehe AP 10).

Die Bereitstellung der Bilddaten seitens der teilnehmenden Bibliotheken auf der Grundlage IIIF-konformer Schnittstellen (IIIF-Image-API) ist wünschenswert und soll seitens des Zeitungsportals funktional bereits während der ersten Phase unterstützt werden. IIIF kann zum gegenwärtigen Zeitpunkt allerdings nicht zur Voraussetzung dafür gemacht werden, dass Zeitungsbestände in das Zeitungsportal integriert werden. Daher werden durch den Zeitungsviewer sowohl Bilder von herkömmlichen Webservern als auch die IIIF-Image-API unterstützt (siehe AP 7).

²³ Für weitere Ausbaustufen des Zeitungsportals ist die Implementierung eines granularen Rechtemanagements vorstellbar, mit dem beispielsweise aus rechtlichen Gründen zu sperrende Ausgaben oder Teile davon geschwärzt werden und im Index entsprechend gekennzeichnet werden können.

3. Unterstützung einer verteilten Datenhaltung

Das Zeitungsportal soll als zentraler Dienst realisiert werden, der nutzerseitig als bruchlos wahrgenommen wird – der also die vollständige Rezeption von Zeitungsausgaben ermöglicht, ohne dabei das Portal verlassen zu müssen – allerdings auf einer teilweise verteilten Datenhaltung basiert. Dies betrifft vor allem die Volltexte und die Digitalisate (hochauflösende Image-Dateien), die technisch durch die teilnehmenden Einrichtungen bereitgestellt und im Zuge der Datenintegration in die DDB / Zeitungsportal nicht vollständig dupliziert werden. Das heißt, die Image-Dateien werden im Zuge der Integration der betreffenden Datenbestände in das Zeitungsportal nicht an die DDB übertragen, sondern erst im Moment der portalseitigen Nutzung dynamisch angezogen und im Browser dargestellt.

Diese Bedingung hat technische Gründe (Bereitstellung des Speicherplatzes), bildet aber vor allem die föderale und auf einer verteilten Verantwortung basierende Kulturlandschaft in Deutschland ab. Herausforderungen ergeben sich hierbei auch hinsichtlich der Nutzererfahrung (Latenzzeiten beim Seitenaufbau, Größe von zu übertragenden Datenpaketen an das Endnutzengerät). Der verteilte Ansatz ist neben den weitreichenden Synergien ein wesentliches Alleinstellungsmerkmal für die DDB als Basis-Infrastruktur für das Zeitungsportal.

4. Zukunftsfähigkeit und Anschlussfähigkeit

Das Zeitungsportal muss so konzipiert und umgesetzt werden, dass es für aktuelle und zukünftige Entwicklungen möglichst offen und anschlussfähig ist und dabei zukunftsorientierte Konzepte und Technologien einsetzt bzw. deren Einsatz in zukünftigen Ausbaustufen ermöglicht. Ein Beispiel hierfür ist IIIF, ein in Entwicklung befindlicher Standard zur Herstellung von Interoperabilität von und mit Bild-Repositoryn²⁴, dem gegenwärtig großes Potenzial im Hinblick auf die Umsetzung von Nutzungsszenarien zuerkannt wird, die auf der gemeinsamen Nutzung verteilter Bild-Bestände basieren. IIIF soll bereits in der ersten Entwicklungsphase des Zeitungsportals eine wichtige Rolle spielen (insbesondere die Nutzung der Image-API als präferierte Option zum Einbinden von Bilddateien). Allerdings ist nicht davon auszugehen, dass sich IIIF kurzfristig so weit durchsetzt, dass die relevanten Datenbestände ausschließlich darüber eingebunden werden. Außerdem sind viele Aspekte von IIIF selbst noch in der Entwicklung, sodass weitergehende IIIF-basierte Funktionen (Presentation-API, Search-API) erst später realisiert werden können. Die Architektur des Zeitungsportals wird aber so beschaffen sein, dass derartige Erweiterungen später möglich sind.

Ein weiterer Aspekt betrifft die Anschlussfähigkeit zu vergleichbaren Vorhaben. Hier ist insbesondere *Europeana Newspapers*²⁵ zu nennen – ein Teilvorhaben von *Europeana*, das gegenwärtig wegen der Schließung des TEL²⁶-Portals auf eine neue technische Basis gestellt wird. Aus Sicht von *Europeana* soll es seitens großer Aggregatoren wie der DDB eine harmonisierte Sicht auf die Zeitungsdaten geben, die dann in *Europeana Newspapers* eingebettet werden. Hier ist eine koordinierte Zusammenarbeit geplant, die sich im Rahmen der nächsten Förderphase von *Europeana* „DSI-3“ vor allem auf dem Feld von IIIF sowie in der Erarbeitung optimaler Präsentationsformen und Dienste für die Datenaggregation und Nutzung manifestieren wird (siehe AP 11).

2.3 Arbeitsprogramm und Umsetzung

Das Zeitungsportal soll mit Blick auf die oben formulierten Zielstellungen als Teil der DDB realisiert werden und möglichst viele Komponenten der bestehenden organisatorischen und technischen Infrastruktur nachnutzen. Das betrifft folgende Aspekte:

- **Backend.** Es soll keine eigene Datenhaltung aufgebaut werden. Stattdessen werden für das Zeitungsportal relevante Bestände in das DDB-Backend geladen und dort auch als DDB-Objekte behandelt.
- **Prozesse.** Auch die Prozesse, die im Rahmen der Integration von Datenbeständen in das Zeitungsportal erforderlich sind (Akquise, Mapping, Datenclearing, Transformation, Ingest, Indexierung, Datenupdate usw.) werden einschließlich der DDB-seitig beteiligten Organisationseinheiten übernommen (Servicestelle, Fachstelle Bibliothek, FIZ Karlsruhe) und – wo notwendig – entsprechend erweitert bzw. modifiziert (siehe Abbildung 2).
- **Frontend.** Die Benutzungsoberfläche des Zeitungsportals wird als materialspezifische Sicht auf den DDB-Bestand realisiert – das heißt hier werden aus dem in der DDB nachgewiesenen Ge-

²⁴ Siehe <http://iiif.io/about/>.

²⁵ Siehe <http://www.europeana-newspapers.eu/>.

²⁶ TEL = The European Library. Das Portal, das bisher Zugang zu den Beständen von 48 Nationalbibliotheken und führenden Forschungsbibliotheken bot, ist unter <http://www.theeuropeanlibrary.org/> zu erreichen, wird aber seit 2017 nicht mehr aktualisiert und mit neuen Daten versorgt.

sambestand nur die Zeitungen angezeigt bzw. in die Such- und Navigationseinstiege einbezogen, und es werden zeitungsspezifische Funktionen umgesetzt – etwa ein kalenderbasierter Einstieg sowie die Möglichkeit zum Einschränken von Suchergebnissen über die Kalenderfunktion. Softwareseitig wird das Zeitungsportal allerdings mit der bestehenden DDB-Frontend-Software integriert. Es wird also keine völlig losgelöste Software entstehen – wie etwa im Fall des Archivportal-D²⁷, einer ebenfalls material- bzw. spartenspezifischen Sicht auf die DDB (siehe Abbildung 1).

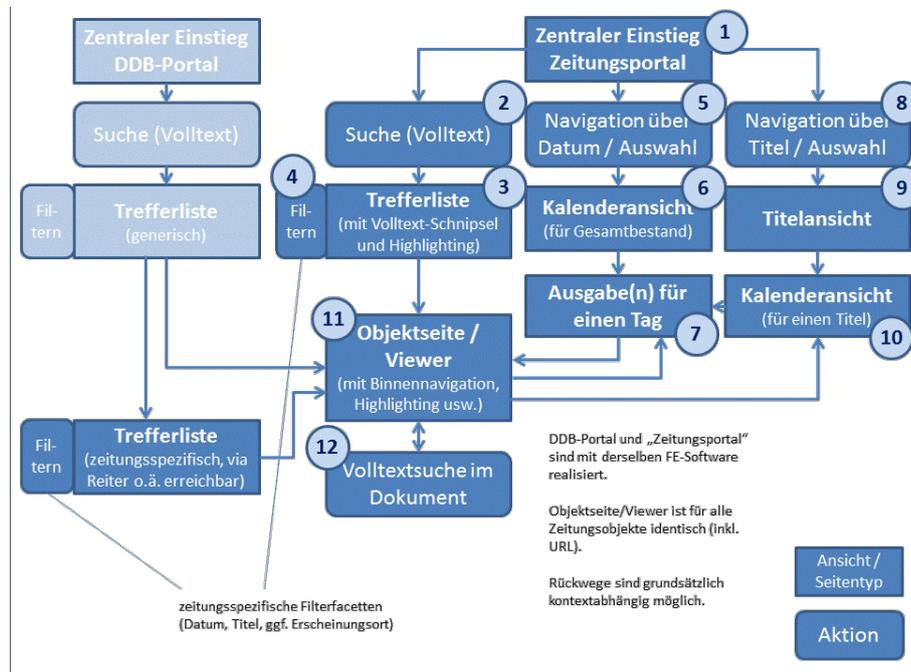


Abbildung 1: Überblick Use Cases für das Zeitungsportal (1. Ausbaustufe) und Verhältnis zum DDB-Portal

Durch dieses Vorgehen können einerseits erhebliche Entwicklungsressourcen für das Zeitungsportal gespart werden. Andererseits wird dadurch die Perspektive eines langfristigen Betriebs des Zeitungsportals gewährleistet, dessen Aufwände nicht allein in der technischen Bereitstellung von Hard- und Software sowie dessen regelmäßiger Pflege und Wartung liegen, sondern wesentlich auch durch die laufenden Prozesse im Zusammenhang mit der Integration und Pflege von Datenbeständen bestimmt werden.

Auf der anderen Seite müssen im Zusammenhang mit der Errichtung des Zeitungsportals eine Reihe von Erweiterungen der DDB-Systeme und -Prozesse vorgenommen werden, deren Notwendigkeit sich aus den zeitungsspezifischen Anforderungen ergibt. Dazu zählen insbesondere

- ein **eigenständiger Gesamteinstieg**, der das Zeitungsportal als gesonderten Bereich erkennbar macht und der auch unabhängig von der DDB angesteuert werden kann (eigene URL / Domain, Branding / Anmutung, besondere Funktionen) (siehe AP 6),
- **zeitungsspezifische Einstiege und Navigationsmöglichkeiten**, die – ausgehend vom Einstieg in das Zeitungsportal – den browsingbasierten Zugang über zeitungstypische Eigenschaften (Datum, Titel, Erscheinungsort) ebenso erlaubt wie das facettenbasierte Filtern von Suchergebnissen (Drill down) über diese Eigenschaften (siehe AP 6),
- die **Volltextsuche** und die damit verbundenen Aspekte (Indexanreicherung für die Nutzung wörterbuchbasierter Funktionen, Darstellung von Suchergebnislisten) (siehe AP 6, AP 8)
- ein auf die Darstellung von Zeitungen optimierter **Viewer**, der unter der Randbedingung des teilweise verteilten Betriebsszenarios eine hinreichende Performanz und damit eine gute Nutzererfahrung ermöglicht (u.a. durch Nutzung der Image-API von IIIF sowie durch die Unterteilung des Viewers in Einzel-Widgets zur Vermeidung unnötigen Datenaustauschs bei Nutzerinteraktionen) (siehe AP 7),
- **Anpassungen am Datenmodell**, durch die zeitungsspezifische Metadaten abgebildet werden (Anwendungsprofil für Zeitungen, kalendergenaue Datumsangaben, Mehrfachausgaben, Datenanreicherungen durch Übernahme von Titeldaten aus der ZDB) (siehe AP 4),

²⁷ Siehe <https://www.archivportal-d.de/>.

- **Anpassungen am Verarbeitungsprozess** für die Daten im Zuge der Transformation (Umsetzung des METS/MODS-Anwendungsprofils für Zeitungen, Anziehen von in der METS-Datei verlinkten Volltextdaten aus den lokalen Repositorien der Datenpartner, Anziehen von Titeldaten aus der ZDB auf der Basis der in den Daten mitgelieferten ZDB-ID) (siehe AP 5).

Abbildung 1 zeigt einen Überblick über die in der ersten Projektphase zu realisierenden Interaktionsstrukturen im Zeitungportal (dunkelblaue Kästchen) sowie das Verhältnis zum generischen DDB-Portal (detaillierte Beschreibung siehe AP 6 und AP 7).

2.3.1 AP 1 – Projektmanagement

Die Projektpartner werden eine Projektgruppe zur Steuerung des Vorhabens einrichten. Das Projektmanagement ist aufgrund der vielen Projektbeteiligten und einzubeziehenden Gruppen besonders erfolgskritisch.

Die Gesamtprojektleitung liegt bei der Deutschen Nationalbibliothek. Sie wird in enger Abstimmung mit den beteiligten Partnern für die Erstellung und Überwachung von Termin- und Meilensteinplänen und die Planung und Steuerung des Ressourceneinsatzes sorgen. Zu ihren Aufgaben zählen außerdem das Risikomanagement, mit dem auf unvorhergesehene Entwicklungen reagiert wird, sowie das Konfliktmanagement, mit dem etwaige Meinungsverschiedenheiten zwischen den Partnern angegangen und gelöst werden. Außerdem sichert die Projektleitung in ihren Rollen als technische Koordinatorin der DDB und Betreiberin der DDB-Servicestelle gemeinsam mit FIZ Karlsruhe, dem technischen Betreiber der DDB und verantwortlichem Softwareentwickler, die Koordination mit den sonstigen Entwicklungen und operativen Belangen der DDB zu.

Zur Koordination des Projektteams und der unterschiedlichen Arbeitspakete werden vierzehntägliche Video- bzw. Telefonkonferenzen und halbjährliche Projekttreffen durchgeführt. Hier werden auch regelmäßig mögliche Projektrisiken adressiert. Konflikte, die auf Arbeitsebene nicht gelöst werden können, werden durch ein bei Bedarf zusammentretendes Steuerungsgremium entschieden, in denen alle Projektpartner auf Leitungsebene vertreten sind.

Ein Projektwiki, eine interne Mailingliste und ein Ticket-System ermöglichen der Projektgruppe das kollaborative Arbeiten. Der vorliegende Antrag bildet den Rahmen für das Projektmanagement. Insbesondere die angegebenen Verantwortlichkeiten für die Arbeitspakete als auch die darin definierten Aktivitäten mit den jeweiligen Zeitangaben sind die Grundlage für die weitere Projektplanung.

Die folgende Übersicht zeigt den zeitlichen Verlauf der einzelnen Arbeitspakete im Gesamtprojekt.

		Projektmonate																									
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
Arbeitspakete	AP 1																										
	AP 2																										
	AP 3																										
	AP 4																										
	AP 5																										
	AP 6																										
	AP 7																										
	AP 8																										
	AP 9																										
	AP 10																										
	AP 11																										

Arbeitsschritte und Aktivitäten²⁸

- Gesamtplanung und regelmäßiger Review (M1 und fortlaufend)
- Projektleitung und Projektsteuerung inkl. Risiko- und Konfliktmanagement (M1-M24)
- Koordination mit den anderen DDB-Entwicklungen (M1-M24)
- Evaluation des Projektfortschritts, Berichtswesen (M1-M24)
- Interne Kommunikation: Wiki, JIRA, Mailing-Listen (M1-M24)

2.3.2 AP 2 – Bedarfserhebung und wissenschaftliche Begleitung

Der grundsätzliche Bedarf für ein übergreifendes Zeitungportal wurde bereits im Rahmen des Pilotprojekts „Digitalisierung historischer Zeitungen“ festgestellt. Dazu fand im Herbst 2014 ein zweitägiger Work-

²⁸ Alle Zeitangaben beziehen sich auf Projektmonate (M1 = 1. Projektmonat, M24 = 24. (letzter) Projektmonat).

shop mit Wissenschaftlerinnen und Wissenschaftlern statt, in dem dieses infrastrukturelle Desiderat benannt wurde und in dem auch die zentralen funktionalen Anforderungen an ein solches Zeitungsportal erfasst wurden. Dieser wissenschaftlich begründete Bedarf wurde auch im Masterplan beschrieben, der als Ergebnis des Pilotprojekts erstellt wurde und in diesem Punkt eine positive und zustimmende Begutachtung erfahren hat (siehe Abschnitt 1 in diesem Antrag). Er bildet die Grundlage für den vorliegenden Antrag und die in der ersten Projektphase umzusetzende Basisfunktionalität.

Gleichwohl soll das Vorhaben zur Errichtung eines nationalen Zeitungsportals eng mit den wissenschaftlichen Communities rückgekoppelt werden. Dazu wird eine wissenschaftliche Begleitgruppe etabliert, die in das Review von Konzepten und Prototypen eingebunden wird und damit in einer frühen Projektphase eine zusätzliche steuernde Rolle erhält.

Außerdem soll zum Ende der ersten Projektphase eine weitere Bedarfsanalyse durchgeführt werden, um den in einer möglichen zweiten Förderphase umzusetzenden Funktionsumfang zu definieren. Zu diesen später umzusetzenden Funktionen zählen nach aktuellem Stand unter anderem:

- Berücksichtigung der Articlebene innerhalb von Zeitungen (Tiefenerschließung) für Suche, hierarchische Darstellung und Metadatenanzeige (abhängig vom betrachteten Strukturelement),
- nachgelagerte, das heißt aufseiten des Zeitungsportals durchgeführte OCR und Layoutanalyse/Artikelsegmentierung (hier sollen auch die bis dahin erreichten Ergebnisse des Projekts OCR-D²⁹ berücksichtigt werden, in dem sich ein Projektmodul ab 2018 mit dem Komplex Layoutanalyse beschäftigen wird),
- frontseitige Integration/Verzahnung zwischen Zeitungsportal und ZDB-Katalog (siehe AP 11),
- volltextbasierte Analysen (für Trend-Themen usw.),
- User Generated Content (z. B. für Verbesserungen im dargestellten Volltext),
- Umsetzung rechtlicher Anforderungen / Rechtmanagement (insbesondere für die Integration noch urheberrechtsbehalteter Zeitungsbestände).

Arbeitsschritte und Aktivitäten

- Initiierung einer wissenschaftlichen Begleitgruppe (M1)
- Workshop zum Konzeptreview mit Projektgruppe und Begleitgruppe (Vorbereitung und Durchführung) (M4-M5)
- Workshop zum Review eines Prototypen und zur Bedarfsanalyse für Phase II mit Projektgruppe und Begleitgruppe (Vorbereitung und Durchführung) (M14-16)
- Erstellung eines priorisierten Anforderungsdokuments als Basis für eine zweite Projektphase (M17-M18)

2.3.3 AP 3 – Anforderungsanalyse

Zu Beginn des Projekts soll auf der Basis der oben beschriebenen Funktionalitäten, die das Zeitungsportal abbilden soll und für die der Bedarf bereits grundsätzlich festgestellt ist (siehe AP 2 und Abschnitt 1), eine detaillierte Anforderungsanalyse durchgeführt werden. Dabei werden die im Pilotprojekt ermittelten Bedarfe in konkrete Nutzungsszenarien übersetzt, die portalseitig realisiert werden sollen:

- Volltextsuche mit Trefferliste und Highlighting sowie anschließender Drill-Down-Möglichkeit (Zeitangaben, Zeitungstitel, Erscheinungsort, ggf. weitere),
- Browsingeinstiege (Kalender, Zeitungstitel, Erscheinungsort),
- Viewerfunktion für einzelne Zeitungsausgaben (unterschiedliche Ansichten, Seiten- und Strukturübersicht mit Binnennavigation, Suchwort-Highlighting, Zoomfunktion, Rotationsfunktion, Suchfunktion in der Ausgabe, Druckfunktion, Exportfunktion, ...),
- Persistente Zitierfähigkeit einzelner Zeitungsausgaben,
- Rückkehr zum Such-/Browsingkontext sowie Navigationsmöglichkeiten über weitere Pfade.

Daneben werden aber auch Anforderungen erfasst, die an früheren Stellen der Prozess- und Verarbeitungskette entstehen und sich beispielsweise daraus ergeben, welche Formate und Anwendungsprofile durch das Zeitungsportal unterstützt werden müssen.

Arbeitsschritte und Aktivitäten

- initialer Anforderungsworkshop (Projekttreffen) für Phase I (M1)
- Formulierung der Anforderungen in den unterschiedlichen Funktionsbereichen (M1)
- Konsolidierung und Dokumentation der Anforderungen (M2)
- abschließender Anforderungsworkshop (Projekttreffen) (M3)

²⁹ Siehe <http://ocr-d.de/>.

- Finalisierung des Anforderungsdokuments für die Phase I (M4)

2.3.4 AP 4 – Feinkonzept

Auf der Grundlage der Anforderungsanalyse und unter Einbeziehung der bisherigen Ergebnisse wird ein Feinkonzept erstellt, das als Grundlage für die anschließende technische Entwicklung dient. Das DDB-Datenmodell muss auf der Basis des METS/MODS-Anwendungsprofils für Zeitungen angepasst und um die zeitungsspezifischen Titelinformationen aus der ZDB erweitert werden. Zudem wird ein UI-Konzept erstellt, aus dem Interaktionspfade und das Layout der unterschiedlichen Seitentypen hervorgehen und auf dem der Designentwurf für die Gesamtseite basiert. Schließlich wird in diesem Arbeitspaket auch ein Integrationskonzept entwickelt, in dem das Zusammenspiel der unterschiedlichen Backend- und Frontendkomponenten im Zeitungsportal beschrieben ist.

Das Feinkonzept wird im Rahmen der wissenschaftlichen Begleitung zurückgekoppelt (betrifft vor allem UI-Konzept, siehe AP 2).

Arbeitsschritte und Aktivitäten

- Anpassung des Datenmodells (M3-M4)
- Entwicklung eines UI-Konzept mit Mockups, Klickpfaden usw. (M4-M5)
- Entwicklung eines Integrationskonzepts (M4-M5)
- Designentwurf (M6-M8)

2.3.5 AP 5 – Liefer- und Transformationsprozesse

Die Liefer- und Transformationsprozesse, die für die Integration von Datenbeständen in die DDB genutzt werden, sind auch die Grundlage für das Laden von Daten in das Zeitungsportal: Die etablierten Workflows, Prozessschritte und Werkzeuge können nachgenutzt werden. Allerdings müssen diese Prozesse zur Realisierung des Zeitungsportals mit den zeitungsspezifischen Funktionen und Datenformaten modifiziert und erweitert werden. Dies betrifft insbesondere das METS/MODS-Anwendungsprofil für Zeitungen, die Übernahme von Titelinformationen aus der ZDB sowie die Behandlung und Verarbeitung von Volltextinformationen zusätzlich zu den Metadaten. In Abbildung 2 ist der Transformations- und Ladeprozess für das Zeitungsportal dargestellt. Die dunkelblau eingefärbten Teile müssen gegenüber den vorhandenen Standardprozessen bei der DDB neu entwickelt bzw. erweitert werden (Nummerierungen aus der Grafik; die Punkte 1-3 werden vollständig in Eigenleistung der DDB erbracht):

- (1) Der für die DDB bereits etablierte Transformationsprozess für METS/MODS-Daten muss für Zeitungen angepasst werden. Zunächst muss eine Selektion der Datensätze vorgenommen werden, die sich auf Zeitungen beziehen, da die entsprechenden Schnittstellen der Einrichtungen i.d.R. für weitere Materialtypen genutzt werden. Die Selektionskriterien werden sehr wahrscheinlich pro Datengeber voneinander abweichen.
- (2) Daten bereits jetzt digitalisierter Zeitungen liegen zwar häufig im Format METS/MODS vor, entsprechen aber nur bedingt den Anforderungen, die in dem METS/MODS-Anforderungsprofil für Zeitungen erarbeitet werden. Daher müssen diese Daten teilweise bereinigt und angereichert werden (Bereinigung von Datensätzen zu Ausgaben, Anreicherung mit ZDB-Identifizier).
- (3) Die vorhandenen Transformationskripte für METS/MODS-Daten müssen auf das Anwendungsprofil für Zeitungen angepasst werden.
- (4) Die für die frontendseitigen Funktionalitäten erforderlichen Titeldaten für die Zeitungen werden auf der Basis der mitgelieferten (bzw. noch ergänzten, siehe (2)) ZDB-Identifizier durch eine schnittstellenbasierte Abfrage in der ZDB angereichert und den gelieferten Daten hinzugefügt. Hierbei werden auch Vorgänger und Nachfolger berücksichtigt. Dieser Anreicherungsprozess muss vollständig neu implementiert werden.
- (5) Die in den METS-Dateien referenzierten Volltextdateien (in der Regel ALTO) werden während des Verarbeitungsprozesses per HTTP-Anfrage bei den datenliefernden Einrichtungen nachgeladen. Dieser Prozess muss neu implementiert und in den Gesamtworkflow integriert werden.
- (6) Die nachgeladenen Volltextdaten müssen anschließend für die Indexierung vorbereitet und indiziert werden.

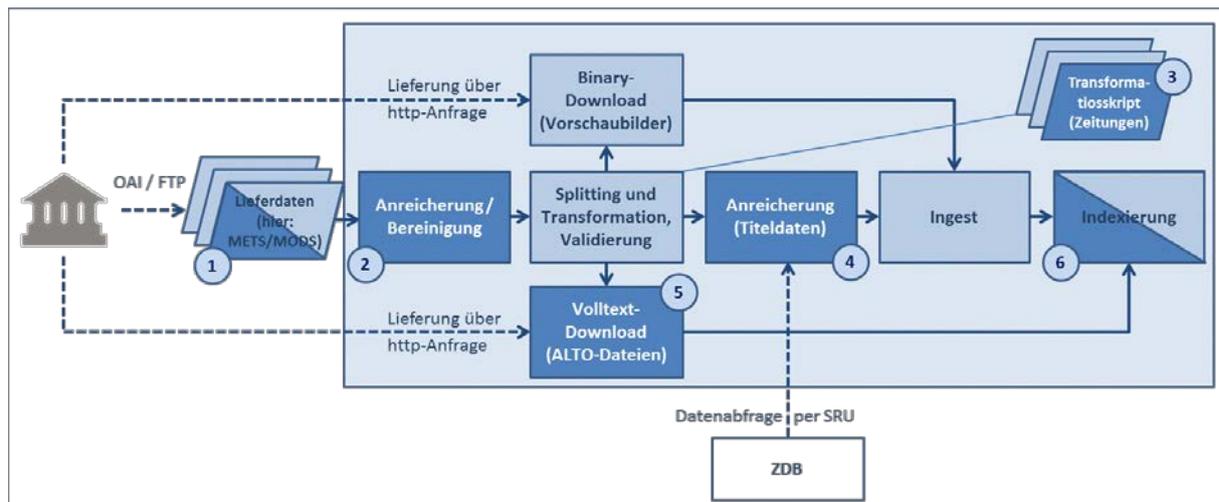


Abbildung 2: Transformations- und Ladeprozesse DDB-Zeitungsportal

Arbeitsschritte und Aktivitäten

- Anpassung der Lieferprozesse zur Selektion von Zeitungsdaten, siehe (1) (M4-M5)
- Mapping-Anpassungen für METS/MODS-Anwendungsprofil Zeitungen, (3) (M6-M8)
- Prozessenerweiterung zur Integration von ZDB-Daten, (4) (M7-M8)
- Prozessenerweiterung für Download und Indexierung von Volltexten, (5) und (6) (M7-M9)
- Evaluierung der Umsetzungsergebnisse und Qualitätssicherung (M10-M11)
- Bereinigung und Anreicherung von Beständen, die nicht dem Anwendungsprofil entsprechen, (2) (M10-M14)

2.3.6 AP 6 – Nutzeroberfläche

Für das Zeitungsportal wird eine neue Nutzeroberfläche entwickelt, die den im Rahmen der Anforderungsanalyse (siehe AP 3) erhobenen Anforderungen und dem darauf basierenden Feinkonzept (AP 4) entspricht. Dabei geht es im Wesentlichen um einen eigenen Portaleinstieg, über den zeitungsspezifische Navigations- und Recherchemöglichkeiten erreichbar sind (siehe dazu auch Abbildung 1).

Neben dem Haupteinstieg für das Zeitungsportal (1), das auch unabhängig vom Einstieg in das generische DDB-Portal erreichbar ist, sollen vor allem folgende Recherche- und Navigationsfunktionalitäten entwickelt werden (Nummerierung siehe Abbildung 1):

- Übergreifende Volltextsuche über den gesamten Zeitungsbestand (2), deren Trefferliste Snippets und das Highlighting übereinstimmender Suchterme (3) beinhaltet, anschließend über zeitungsspezifische Filterfacetten wie Titel, Datum und Erscheinungsort eingeschränkt werden kann (4) und auf die Objektansicht einer Zeitungsausgabe (11) führt,
- kalenderbasierter Einstieg (5), der über mehrere Hierarchiestufen bis zur Auswahl eines Tages (6) führt, für den alle an diesem Tag erschienen Ausgaben mit Titelangabe und verkleinerter Titelseite (7) angezeigt werden und schließlich zur Objektansicht (11) führt.
- alphabetischer und ggf. mit Suchfunktion ausgestalteter Einstieg über Zeitungstitel (8), der zu einer Einstiegsseite für eine Zeitung (9) führt, von der über eine Kalenderansicht (10) wiederum die Ausgabe(n)³⁰ für den gewählten Tag (7) und eine Einzelausgabe im Viewer (11) erreicht werden können.

Für die Softwareentwicklung kommt dabei das Webframework Grails³¹ mit Bootstrap³² als CSS-Framework zum Einsatz – moderne Technologien zur Entwicklung responsiver Webapplikationen, die für die Entwicklung des DDB-Portals bereits im Einsatz und vielfach erprobt sind.

Arbeitsschritte und Aktivitäten

- Implementierung der Einstiegsseite und alternativer Navigationspfade (M6-M10)
- Erweiterung der bisherigen Suche um Volltextsuche (M11-M14)

³⁰ Im Falle einiger historischer Zeitungen sind täglich mehrere Ausgaben erschienen.

³¹ Siehe <https://grails.org/>.

³² Siehe <http://getbootstrap.com/>.

- Integration in bestehendes System (inklusive Ergebnissen aus AP 7) und dazu notwendige Anpassungen im Backend (M6-M19)
- Evaluierung der Umsetzungsergebnisse und Qualitätssicherung (M20-M22)

2.3.7 AP 7 – Viewer und Image-Integration

Zentraler Bestandteil des Zeitungsportals wird eine Viewer-Komponente sein (siehe Abbildung 1 (11)), mit der einzelne Zeitungsausgaben betrachtet werden können und für die eine Vielzahl von Funktionen zur Verfügung steht (Binnennavigation, Zoom, Rotation, Helligkeits- und Kontrastverbesserung, Highlighting von Suchbegriffen, Suche innerhalb der angezeigten Ausgabe (12), Verknüpfung mit anderen Ausgaben desselben Tages (7) bzw. desselben Zeitungstitels über Kalender (10) usw.). Außerdem wird der Objekt-Viewer eine Möglichkeit enthalten, zitierfähige Links auf einzelne Ausgaben zu generieren.

Abbildung 3 zeigt schematisch die technische Realisierung des Objektviewers für die Darstellung einer Zeitungsausgabe als Seitentyp innerhalb des Zeitungsportals (a) und das Zusammenspiel mit den unterschiedlichen Backend-Komponenten (d), (e), (f).

Zur Einbindung der Zeitungs-Images, die nicht zentral gehalten werden, sondern on-the-fly von den besitzenden Einrichtungen angezogen werden, wird die herkömmliche Referenzierung der ganzen Bilddateien (jpg) (h), (i) ebenso unterstützt wie die Image-API von IIIF (g). Um die Verarbeitung der Bildinformationen aufseiten der Viewer-Komponente (a) zu vereinheitlichen und außerdem das mixed-content-Problem³³ zu lösen sowie die vielfach nicht vorhandene CORS³⁴-Unterstützung seitens der Datengeber auszugleichen³⁵, wird aufseiten des Zeitungsportals ein Image-Proxy (f) implementiert. Diese Komponente ruft dann die Images beim Datengeber ab und leitet diese an den Viewer weiter. Aus Sicht des Viewers erscheinen die Images dadurch als lägen sie lokal im Zeitungsportal vor, so dass CORS-Restriktionen und mixed-content-Probleme abgefangen werden können, ohne dass jeder einzelne Datengeber entsprechende Maßnahmen ergreifen muss.

Der Viewer soll nach aktuellem Planungsstand auf der Basis der DFG-Viewer-Technologie entwickelt werden und dafür die Open-Source-Komponente Kitodo.Presentation (e) nutzen. Dabei soll die Kommunikation zwischen Viewer-Backend (e) und Frontend/Portal (a) auf der Basis einzelner Widgets (c) erfolgen. Gegenüber der Auslieferung einer jeweils vollständigen Seite durch das Viewer-Backend werden durch dieses Vorgehen einerseits die Möglichkeiten erweitert, die einzelnen Funktionselemente individuell anzuordnen und außerdem mit Standard-Bestandteilen aus der DDB zu kombinieren³⁶ (b), (d). Andererseits wird dadurch vermieden, dass bei jeder einzelnen Nutzerinteraktion die gesamte Seite neu geladen werden muss. Dadurch wird die notwendige Datenübertragung zum Browser verringert und das Antwortzeit-Verhalten verbessert.

³³ Unter „mixed content“ wird eine per https aufgerufene Website verstanden, in die unter anderem über eine unverschlüsselte http-Verbindung aufgerufene Inhalte eingebunden sind. Wegen Sicherheitsrisiken werden derartige Seiten nicht in allen Browsern korrekt dargestellt bzw. verursachen Warnungen und Fehlermeldungen. Direkt eingebundene Bilddateien, die auf Webservern von Liefereinrichtungen liegen, führen zu diesem Phänomen, wenn die Übertragung per http stattfindet, da das Zeitungsportal ebenso wie das DDB-Portal über https ausgeliefert wird. IIIF schreibt dagegen die Verwendung von https zwingend vor.

³⁴ Mit CORS (= Cross-Origin Resource Sharing) können Regeln definiert werden, unter denen Webbrowser in Ergänzung zur eigentlich gültigen Same-Origin-Policy die Einbindung von Inhalten unterschiedlicher Quellen ermöglicht wird. Diese Regeln müssen auf Seiten des Datenanbieters definiert werden.

³⁵ Ohne CORS-Unterstützung kann kein WebGL verwendet werden, was für performante browserseitige Funktionen wie stufenlose Rotation, Helligkeits- und Kontrastanpassungen usw. aber erforderlich ist.

³⁶ Das betrifft vor allem die beschreibenden Metadaten, die in der DDB in einem speziellen View-Format gespeichert sind.

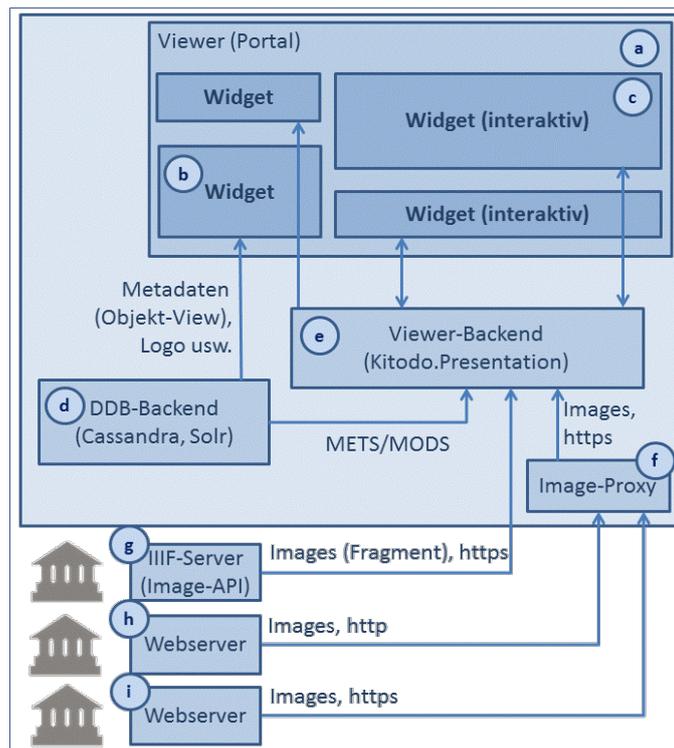


Abbildung 3: Zusammenspiel Frontend, backendseitige Viewer-Komponente, DDB-Backend und Image-Proxy

Arbeitsschritte und Aktivitäten

- Integration von Kitodo.Presentation auf Widget-Basis (SLUB)
 - a. Isolierung der Kitodo.Presentation-Funktionsmodule in Form von Widgets (M5-M6)
 - b. Konfiguration und Bereitstellung des Viewer-Widgets mit stufenlosem Zoom, freier Rotation, Helligkeit-/Kontrastregler, Koordinaten-basierendem Highlighting, Volltext-Anzeige mit Kopierfunktion und Bildsynchronisation (M7-M8)
 - c. Konfiguration und Bereitstellung des Navigations-Widgets mit Inhaltsverzeichnis zur Binnennavigation anhand der METS-Strukturdaten mit Seitenzahlen (M9)
 - d. Konfiguration und Bereitstellung des Paging-Widgets zum Blättern (vorwärts/rückwärts, +/- 5 Seiten, Anfang/Ende) und Anspringen von Einzelseiten (M9)
 - e. Konfiguration und Bereitstellung des Thumbnail-Widgets mit Übersicht über eine Zeitungsausgabe inkl. Navigationsmöglichkeit (M10)
 - f. Umsetzung einer Druckausgabe für alle Widgets (M10)
 - g. Konfiguration und Bereitstellung eines Volltext-Widgets zur Suche innerhalb des Volltexts der aktuellen Ausgabe inkl. Highlighting der Suchtreffer (M11)
 - h. Integration mit DDB-Komponenten (M11-M12)
 - i. Evaluierung der Umsetzungsergebnisse und Qualitätssicherung (M13-M14)

- Entwicklung und Integration eines Image-Proxy mit HTTPS- und CORS-Unterstützung (FIZ)
 - a. Technisches Feinkonzept für Proxy-Architektur (M5-M7)
 - b. Evaluierung von http/2 für das gebündelte Ausliefern von Image-Kacheln (M8)
 - c. Implementierung des Proxies (M9-M12)
 - d. Evaluierung der Umsetzungsergebnisse und Qualitätssicherung (M13-M14)

2.3.8 AP 8 – Volltextsuche

Der vermutlich wichtigste Zugang zum Zeitungsportal wird die Freitextsuche in den Volltexten der Zeitungen sein. Dabei gibt es zwei unterschiedliche Nutzungsszenarien:

1. die Suche über dem gesamten Zeitungsbestand (bzw. zuvor durch nutzerseitige Selektion definierte Teilbestände), die zu einer Trefferliste mit gefundenen Zeitungsausgaben führt (siehe Abbildung 1 (2), (3)), sowie
2. die Suche in einer einzelnen Zeitungsausgabe, die Fundstellen innerhalb dieser Ausgabe erbringen kann (siehe Abbildung 1 (12)).

In beiden Fällen müssen die Fundstellen in den jeweiligen Dokumenten per Highlighting hervorgehoben werden – im Szenario 1 in einem oder mehreren Snippets pro Treffer sowie anschließend in der Seitendarstellung des Viewers und im Szenario 2 unmittelbar im Viewer.

Die frontendseitige Realisierung der beiden Szenarien erfolgt in den Arbeitspaketen 6 bzw. 7. Die technische Integration der Volltexte in den Solr-basierten Suchindex als Teil des Ingest-Vorgangs erfolgt in Arbeitspaket 5.

Dieses Arbeitspaket umfasst

- die Umsetzung fachlicher Anforderungen an die Suche, die sich aus der Generierung der Volltexte per OCR sowie aus dem historischen Alter der Texte ergeben (z.B. unscharfe Suche, Teilwortsuche, Lemmatisierung, POS-Tagging),
- das Testen, Konfigurieren und Optimieren der Such- und Ranking-Parameter im Zusammenspiel mit den anderen Navigationselementen und auf der Grundlage repräsentativer Datenbestände aus dem Zeitungsbereich,
- Maßnahmen zur nutzerfreundlichen Indexierung der Inhalte des Zeitungsportals durch externe Suchmaschinen zur Verbesserung der Auffindbarkeit.

Arbeitsschritte und Aktivitäten

- Formulierung von fachlichen Anforderungen für die Suche auf der Basis von Testfällen (M7-M8)
- Evaluierung bestehender Verfahren und Ansätze zur Suche in Volltexten und zum Natural Language Processing (M9-M10)
- Entwicklung und technische Realisierung eines Konzepts für die Volltextindexierung und -suche (M11-M20)
- Evaluierung der Umsetzungsergebnisse (M15, M21-M22)
- Optimierung und Bugfixing (M16-M18, M23-M24)

2.3.9 AP 9 – Integration von Zeitungsbeständen

Bereits im Verlauf der ersten Projektphase sollen Zeitungsbestände in das Zeitungsportal integriert und mit den implementierten Funktionen nutzbar sein. Das betrifft prioritär Ergebnisse DFG-finanzierter Digitalisierungsprojekte (hier in erster Linie im Rahmen des Pilotprojekts durchgeführte bzw. mithilfe der intendierten Förderlinie geplante Digitalisierungsvorhaben), bei denen davon auszugehen ist, dass die DFG-Praxisregeln und insbesondere das METS/MODS-Anwendungsprofil Zeitungen berücksichtigt wird (siehe AP 5) sowie außerdem Bestände, die zwar in Form von METS/MODS geliefert werden können, aber nicht über eine Volltexterschließung (OCR) verfügen.

In dieses Arbeitspaket fällt die operative Übernahme von Zeitungsbeständen ausgewählter Pilotpartner in den Datenbestand der DDB auf der Basis der erweiterten Liefer- und Transformationsprozesse sowie die Übergabe an den regulären Datenlieferungs- und -verarbeitungsdienst bei den beteiligten Organisationseinheiten (Servicestelle, Fachstelle Bibliothek, FIZ Karlsruhe). Diese Aktivitäten erfolgen weitgehend in Eigenleistungen der DDB.

Arbeitsschritte und Aktivitäten

- Operative Integration von Beständen von Pilotpartnern inkl. Qualitätssicherung (M9-M24)
- Übergabe an den regulären Lieferbetrieb (M22-M24)

2.3.10 AP 10 – Vorbereitung der Integration weiterer Zeitungsbestände

Das DDB-Zeitungsportal soll als Nachweis und zentraler Zugang für alle digitalisierten historischen Zeitungen in Deutschland dienen. Dazu zählen die Bestände, die durch DFG-Förderungen digitalisiert wurden und werden – insbesondere im Rahmen der einzurichtenden Förderlinie. Diese stehen im Mittelpunkt der Betrachtung für die erste Förderphase.

Im Rahmen des Projekts sollen aber auch andere Bestände integriert werden, die nicht mit Mitteln der DFG digitalisiert wurden und deren Bereitstellungsformat daher nicht unbedingt den DFG-Praxisregeln entspricht, die aber für die wissenschaftliche Nutzung ebenso relevant sind und allein aufgrund ihres Umfangs einen erheblichen Mehrwert für das Zeitungsportal darstellen. Dies ist nach aktuellem Planungsstand für eine zweite Förderphase vorgesehen. In der ersten Förderphase sollen entsprechende Bestände identifiziert und Vorbereitungen zu deren späterer Integration getroffen werden.

Dies betrifft insbesondere solche Bestände, die in Form von IIIF (Presentation-API) exponiert werden wie etwa im Fall der Bayerischen Staatsbibliothek (digiPress), aber auch Zeitungen, die in anderen Formaten vorgehalten werden und ggf. auf Umwegen nach METS/MODS transformiert werden müssen.

Mit diesem Arbeitspaket sollen relevante Bestände und Formate identifiziert werden sowie Schnittstellen und ggf. zusätzliche Austauschformate definiert oder bestimmt werden, mit denen eine Übernahme der Bestände in das DDB-Zeitungsportal möglich ist. Ein wesentlicher Schwerpunkt liegt dabei auf der IIIF-Presentation-API. Hier sind Kriterien für deren konkrete Ausgestaltung erforderlich, damit die bereitgestellten IIIF-Manifeste denselben Informationsgehalt haben wie vergleichbare METS/MODS-Daten gemäß Anwendungsprofil für Zeitungen.

Neben alternativen Formaten für die Beschreibung von Zeitungsstrukturen (METS/MODS vs. IIIF-Manifest) müssen hier auch Alternativen zu ALTO als Austauschformat für die OCR-Ergebnisse untersucht werden (z. B. TEI), um später weitere Datenbestände in das Zeitungsportal zu integrieren.

Arbeitsschritte und Aktivitäten

- Identifikation relevanter Zeitungsbestände, die nicht über den Standard-Lieferweg integriert werden können (M8-M10)
- Konzeption eines Lieferweges auf der Basis der Presentation-API von IIIF inkl. Anforderungen an die IIIF-Manifeste, Transformationsschritte u. ä. (M11-M13)
- Untersuchung alternativer Volltextformate und Konzeption einer Verarbeitungskette zur Integration in das Zeitungsportal (M14-M16)

2.3.11 AP 11 – Verzahnung mit Schwestervorhaben

Die Projektpartner, die das DDB-Zeitungsportal realisieren wollen, stehen in enger Kooperation mit Europeana Newspapers; so ist die SBB innerhalb der Europeana für die Weiterentwicklung und das Management der thematischen Zeitungssammlung verantwortlich. SBB und DNB betreiben darüber hinaus gemeinsam die ZDB, in deren Webkatalog im Rahmen des Pilotprojekts zur Digitalisierung historischer Zeitungen die zeitungsspezifischen Funktionen auf Titel- und Bestandsebene deutlich erweitert wurden. Das DDB-Zeitungsportal wird mit beiden Vorhaben kooperieren und vorhandene Entwicklungen soweit wie möglich nachnutzen bzw. anbinden.

Die in der ZDB erfassten Titeldaten für Zeitungen bilden bereits in der ersten Projektphase die Grundlage für den Hierarchieaufbau und die Navigations- und Suchmöglichkeiten. Sie werden im Rahmen des Transformations- und Ingestprozesses schnittstellenbasiert abgefragt und dienen als Grundlage für die Anreicherung der Lieferdaten (siehe AP 5). Neben dieser datenseitigen Integration soll bereits in der ersten Projektphase eine leichtgewichtige Anbindung des ZDB-Katalogs an das Zeitungsportal realisiert werden – und zwar in Form von Links aus dem Objektviewer und unterschiedlichen Navigationsknoten des Zeitungsportals (siehe Abbildung 1) auf entsprechende Seiten des ZDB-Katalogs. Insbesondere soll für jeden Zeitungstitel (siehe u.a. Abbildung 1 (9)) ein Link auf den entsprechenden Eintrag im ZDB-Katalog angeboten werden, wo weiterführende Informationen und Interaktionsmöglichkeiten verfügbar sind. Dafür werden im Rahmen dieses Arbeitspakets entsprechende Vorgaben gemacht.

Darüber hinaus sollen Zeitungsportal und ZDB-Katalog frontendseitig so miteinander verzahnt werden, dass die jeweiligen Funktionsmodule (ZDB: vollständiger Titelnachweis, Besitznachweis, Titelhistorie und -relationen, Verbreitungsorte; Zeitungsportal: Recherche und direkter Zugang zu den digitalisierten Zeitungen) aufeinander verweisen und die Nutzer entsprechend leiten. Dazu soll in der ersten Projektphase ein Konzept erarbeitet werden, das die unterschiedlichen Nutzungsszenarien der beiden Portale in den Blick nimmt und mögliche Mehrwerte für die Nutzer des je anderen Portals berücksichtigt. Ziel ist, in der zweiten Projektphase eine frontendseitige Verzahnung von Zeitungsportal und ZDB zu realisieren.

Mit Europeana Newspapers besteht ebenfalls eine enge Kooperation: Die SBB ist Projektpartner für die Neu-/Weiterentwicklung dieser zeitungsspezifischen Europeana-Sicht. Die DDB/DNB kooperiert in der aktuellen Europeana-Projektphase DSI-3 im Themenfeld IIF und Zeitungen mit Europeana. Diese Kooperation soll auch im Rahmen des DDB-Zeitungsportals genutzt und ausgebaut werden. Mittelfristig ist insbesondere die enge technische Zusammenarbeit im Bereich Image-Bereitstellung (IIF) vorgesehen, die in eine harmonisierte Lieferung von Zeitungsdaten in standardisierten Datenformaten durch die DDB an Europeana mündet.

Arbeitsschritte und Aktivitäten

- Formulierung von Vorgaben zur Verlinkung aus Seiten des Zeitungsportals in den ZDB-Katalog (M2-M3)
- Erstellung eines Konzepts zur frontendseitigen Verzahnung zwischen ZDB-Katalog und Zeitungsportal (M14-M15)
- Erstellung eines Konzepts für die Verzahnung mit Europeana Newspapers (Schwerpunkt: Infrastruktur) (M16-M17)

2.4 Maßnahmen zur Erfüllung der Förderbedingungen und Umgang mit den Projektergebnissen

Der vorliegende Projektantrag beschreibt ein Vorhaben zur Implementierung von e-Research-Technologien im Sinne des DFG-Förderprogramms „e-Research-Technologien“. Dem Vorhaben ging eine Untersuchung im Rahmen des Pilotprojekts „Digitalisierung historischer Zeitungen“ voraus, aus der die Bereitstellung und Zugänglichmachung eines übergreifenden Zeitungsportals für die digitalisierten Zeitungsbestände in Deutschland als wesentliches Desiderat hervorgeht. Aufgrund dieses grundsätzlich formulierten Bedarfs, der im Rahmen der gutachterlichen Evaluierung des Masterplans zur Digitalisierung historischer Zeitungen vonseiten der DFG bestätigt wurde, soll das Zeitungsportal nun entwickelt werden.

Das Projekt zur Errichtung eines Zeitungsportals für Deutschland ist so angelegt, dass dafür die organisatorische und technische Infrastruktur, die die Deutsche Digitale Bibliothek aufgebaut hat und betreibt, soweit wie möglich nachgenutzt und integriert wird. Dadurch werden nicht nur während der Konzeptions- und Aufbauphase des Zeitungsportals erhebliche Synergien genutzt, mit denen der Entwicklungsaufwand deutlich gesenkt wird. Durch die starke Integration des Zeitungsportals mit den technischen Komponenten der DDB und den erforderlichen Prozessen und Organisationseinheiten wird auch der langfristige Betrieb des Zeitungsportals gesichert. Insbesondere die an der DNB ansässige DDB-Projektkoordination und der technische Betreiber der DDB (FIZ Karlsruhe) werden in der Konzeptions- und Entwicklungsphase des Zeitungsportals dafür Sorge tragen, dass der operative Betriebsaufwand des Zeitungsportals, der zusätzlich zum DDB-Betrieb entsteht, möglichst gering ist. Dies betrifft auch die Wartung und Pflege der Softwarekomponenten.

Die DDB verpflichtet sich, das entwickelte Zeitungsportal nach Fertigstellung langfristig zu betreiben und auch für dessen funktionale Erweiterung und Aktualisierung – dies ggf. mit zusätzlichen Fördermitteln – Sorge zu tragen. Bei einem ähnlichen Vorhaben, dem Archivportal-D, ist dies bereits erfolgreich umgesetzt worden: Nach Abschluss der durch die DFG geförderten Entwicklungsphase ist die Betriebsverantwortung einschließlich der Übernahme entsprechender Aufwände auf die DDB übergegangen. Durch die angestrebte softwareseitige Integration des DDB-Zeitungsportals in das DDB-Portal werden die spezifischen Betriebsaufwände für das Zeitungsportal vermutlich sogar geringer sein als beim Archivportal-D, für das eine eigene Frontend-Software entwickelt wurde.

Sämtliche im Projekt entwickelte Software wird unter eine Open-Source-Lizenz gestellt und auf GitHub³⁷ veröffentlicht.

2.5 Erläuterungen zur inhaltlichen und finanziellen Projektbeteiligung von Kooperationspartnerinnen und Kooperationspartnern im Ausland

Entfällt.

3 Literaturverzeichnis

Empfehlungen zur Digitalisierung historischer Zeitungen in Deutschland (Masterplan Zeitungsdigitalisierung) – Ergebnisse des DFG-Projektes „Digitalisierung historischer Zeitungen“ Pilotphase 2013-2015, 29.01.2016 (Partner: SBB (Berlin), SuUB (Bremen), SLUB (Dresden), DNB (Frankfurt), ULB (Halle), BSB (München)).

³⁷ Die DDB veröffentlicht alle Softwarekomponenten auf GitHub, siehe <https://github.com/Deutsche-Digitale-Bibliothek>.

Mühlberger, Günter: Digitalisierung historischer Zeitungen aus dem Blickwinkel der automatisierten Text- und Strukturerkennung (OCR). *Zeitschrift für Bibliothekswesen und Bibliographie*, Jahrgang 58 (2011), Heft 1. [doi:10.3196/186429501158135](https://doi.org/10.3196/186429501158135).

Müller, Christa (2016). ANNO – Der digitale Zeitungslesesaal der Österreichischen Nationalbibliothek. *Bibliothek Forschung und Praxis*, 40(1), pp. 83-89. [doi:10.1515/bfp-2016-0012](https://doi.org/10.1515/bfp-2016-0012).