

Lisa Landes

Deutsches Zeitungsportal – Abschlussbericht 2026

<https://doi.org/10.25969/mediarep/24467>

Veröffentlichungsversion / published version
Buch / book

Empfohlene Zitierung / Suggested Citation:

Landes, Lisa: *Deutsches Zeitungsportal – Abschlussbericht*. Frankfurt am Main: Deutsche Nationalbibliothek 2026. DOI: <https://doi.org/10.25969/mediarep/24467>.

Nutzungsbedingungen:

Dieser Text wird unter einer Creative Commons - Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0/ Lizenz zur Verfügung gestellt. Nähere Auskünfte zu dieser Lizenz finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Terms of use:

This document is made available under a creative commons - Attribution - Non Commercial - No Derivatives 4.0/ License. For more information see:

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Deutsches Zeitungsportal – Abschlussbericht

1 Allgemeine Angaben

DFG-Geschäftszeichen: BU 2228/43-1, BO 3288/16-1, NI 522/16-1, SCHO 1192/20-1, RA 1965/4-1, RA 1965/7-1, SCHN 743/59-1, SCHN 743/73-1

Projektnummern: 404633104 sowie 454852652

Titel des Projekts: „Errichtung eines nationalen Zeitungsportals auf der Basis der organisatorischen und technischen Infrastruktur der Deutschen Digitalen Bibliothek (DDB)“ sowie „Ausbau und Optimierung des DDB-Zeitungsportals – („DDB-Zeitungsportal V. 2.0“)“

Namen der Antragstellenden: SLUB Dresden: Thomas Bürger (bis 31.07.2018), Achim Bonte (bis 31.08.2021), Katrin Stump (ab 01.05.2022); DNB: Elisabeth Niggemann (bis 31.12.2019), Frank Scholze; FIZ Karlsruhe: Matthias Razum; SBB Berlin: Barbara Schneider-Kempff (bis 30.06.2021), Achim Bonte

Dienstanschriften: Katrin Stump, Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB), 01054 Dresden; Frank Scholze, Deutsche Nationalbibliothek, Adickesallee 1, 60322 Frankfurt; Matthias Razum, Leibniz-Institut für Informationsinfrastruktur GmbH (FIZ KA), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen; Achim Bonte, Stiftung Preußischer Kulturbesitz, Staatsbibliothek zu Berlin, Unter den Linden 8, 10117 Berlin

Namen der Mitverantwortlichen, Namen der Kooperationspartnerinnen und -partner: --

Berichtszeitraum (gesamte Förderdauer): 01.01.2019–31.07.2025

2 Zusammenfassung

Ziel des Projekts war der Aufbau eines zentralen, bestands- und jahrhunderteübergreifenden Portals für digitalisierte historische Zeitungen, das attraktive Nutzungsoptionen für Wissenschaft und die interessierte Öffentlichkeit bietet. Das [Deutsche Zeitungsportal](#) wurde im Oktober 2021 als Subportal der [Deutschen Digitalen Bibliothek](#) (DDB) freigeschaltet und kann kostenfrei und ohne Registrierung genutzt werden. Aktuell sind 1.991 Zeitungen von 23 Einrichtungen (insgesamt 26 Mio. Seiten) aus den Jahren 1671–1994 online. Der Bestand wird kontinuierlich weiter ausgebaut.

Über eine Volltextsuche sind ca. 90% der Zeitungen erschlossen. Außerdem kann über Titel, Ort und Datum auf Zeitungen zugegriffen werden. Übersichtsseiten der einzelnen Zeitungen bieten weitere Informationen und Links zur weiterführenden Recherche in der Zeitschriftendatenbank. Nutzer*innen können eigene Korpora anlegen und, z. B. für die Weiterverwendung in Forschungsumgebungen, herunterladen. Über die DDB-API sind die Inhalte für Big-Data-Analysen im Bereich der Digital Humanities zugänglich. Durch die Anbindung des Deutschen Zeitungsportals an die Deutsche Digitale Bibliothek ist der weitere Betrieb und Ausbau gesichert.

Summary

The aim of the project was to build an internet portal that gives central access to digitised historic newspapers from several centuries, bringing together separate collections and offering valuable services for the sciences and the interested public alike. The [German Newspaper Portal](#) was launched in October 2021 as a subportal of the [German Digital Library](#) and can be used free of charge and without registration. At the moment, it contains 1,991 historic newspapers from 23 institutions (a total of 26 mio pages) from the years 1671 to 1994. The collection is continuously being expanded.

Ca. 90% of the newspapers are available via full text search. Newspapers can also be accessed via title, place, and year of publication. Each newspaper has a separate overview page that provides further information and links to the German Union Catalogue of Serials. Users can create their own corpora and download them, e.g. for further use in research environments. The portal also features an API which gives access for big data studies by digital humanists. The connection between the German Newspaper Portal and the German Digital Library ensures its continued operation and expansion.

3 Arbeits- und Ergebnisbericht

3.1 Ausgangslage und Zielsetzung

Die Digitalisierung und Vervolltextung historischer Zeitungen wird seit den 2010er Jahren verstärkt vorangetrieben und seitens der DFG seit 2018 durch eigene Förderlinien unterstützt. Ziele und Rahmenbedingungen der Digitalisierungsaktivitäten wurden 2016/2017 in einem Masterplan¹ veröffentlicht. Darin wird auch der aus der Wissenschaft heraus

¹ 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

geäußerte dringende Bedarf für einen zentralen Zugang zu digitalisierten Zeitungen deutlich gemacht – etwa in Form eines nationalen Zeitungsportals, das die bislang verteilten Bestände von Bibliotheken, Archiven etc. auf einer Webseite zusammenführt. Als Träger des nationalen Zeitungsportals wird im Masterplan die Deutsche Digitale Bibliothek² empfohlen, ein Portal, das den zentralen Nachweis von digitalisiertem deutschem Kulturerbe zur Aufgabe hat und nun um ein weiteres Portal zu historischen Zeitungen ergänzt werden sollte.³

3.2 Ergebnisse

Das neue Portal wurde unter dem Namen „Deutsches Zeitungsportal“ am 28. Oktober 2021 freigeschaltet und ist seitdem unter <https://www.deutsche-digitale-bibliothek.de/newspaper> erreichbar. Es kann kostenfrei und ohne Registrierung genutzt werden, über die DDB-API⁴ können die Inhalte (Bilder, Volltexte und Metadaten) im großen Stil heruntergeladen werden. Zum Start des Portals wurden von neun Datenpartnern⁵ knapp 600.000 Zeitungsausgaben bereitgestellt, die sich auf 247 Zeitungen aus dem Zeitraum 1671–1950 verteilten. Mit 3,8 Millionen Seiten verfügten ca. 84% aller Seiten über eine Volltexterschließung.

Das Deutsche Zeitungsportal sollte laut Masterplan mit vier Kernfunktionen online gehen: einer übergreifenden Volltextsuche, Einstiegspunkten über Kalender, Zeitungstitel und -ort, einem integrierten Viewer sowie einer Möglichkeit, die enthaltenen Ausgaben persistent zu referenzieren. Alle vier Kernfunktionen wurden wie geplant mit dem Launch bereitgestellt.

Ein großer Teil der Arbeit in der zweiten Projektphase widmete sich dem Zuwachs der bereitgestellten Inhalte. Hier wurden große Erfolge erzielt, besonders auch durch die Zusammenarbeit mit regionalen Zeitungsaggregatoren wie zeit.punktNRW.⁶ Aktuell werden von 23 Datenpartnern 4.170.066 Zeitungsausgaben bereitgestellt, die sich auf 1.991 Zeitungen aus dem Zeitraum 1671–1994 verteilen. Mit 26.035.735 Seiten verfügen ca. 90% aller Seiten über eine Volltexterschließung.

(Dresden), DNB (Frankfurt), ULB (Halle), BSB (München)).
https://www.zeitschriftendatenbank.de/fileadmin/user_upload/ZDB/z/Masterplan.pdf [10.10.2025].

² S. <https://www.deutsche-digitale-bibliothek.de/> [10.10.2025].

³ S. Masterplan Zeitungsdigitalisierung https://www.zeitschriftendatenbank.de/fileadmin/user_upload/ZDB/z/Masterplan.pdf, S. 60. [10.10.2025].

⁴ <https://labs.deutsche-digitale-bibliothek.de/app/ddbapi/> [10.10.2025].

⁵ Badische Landesbibliothek, Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden, Staatsbibliothek zu Berlin - Preußischer Kulturbesitz, Staats- und Universitätsbibliothek Hamburg Carl von Ossietzky, Universitäts- und Landesbibliothek Düsseldorf, Landesbibliothek Oldenburg, Universitäts- und Landesbibliothek Sachsen-Anhalt, Bibliothek der Friedrich-Ebert-Stiftung, Deutsches Exilarchiv.

⁶ <https://zeitpunkt.nrw/> [10.10.2025].

Die Freischaltung des Zeitungsportals stieß auf großes öffentliches Interesse, allein am Tag des Launchs verzeichnete das Portal mehr als 44.000 Besuche. Nach einem sich anschließenden Rückgang der Nutzung (Sommer 2022 ca. 7.000 Besuche monatlich) stiegen die Nutzungszahlen sukzessive an. Aktuell verzeichnet das Portal rund 20.000 Besuche im Monat (s. Abbildung unten). Die durchschnittliche Dauer eines Besuchs beläuft sich auf ca. 15 Minuten, was zeigt, dass das Zeitungsportal gut angenommen und intensiv genutzt wird. Im November/Dezember 2021 sowie von Dezember 2023 bis Mai 2024 wurde das Zeitungsportal über Google-Ads-Anzeigen beworben.

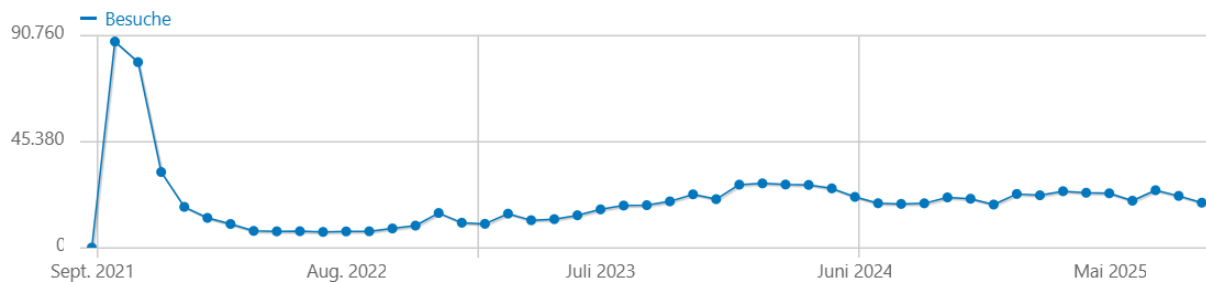


Abbildung 1: Besuche pro Monat von September 2021 bis September 2025

3.3 Arbeitsschritte im Berichtszeitraum

Die Projektarbeit wurde im Wesentlichen so umgesetzt wie geplant. Die wichtigsten Ziele, die Freischaltung eines zentralen Zeitungsportals für Deutschland sowie der Aufbau einer nachhaltigen Infrastruktur für den dauerhaften Betrieb und fortlaufende Weiterentwicklungen, wurden erreicht. Gerade was die Anzahl der verfügbaren Ausgaben angeht, wurden die eigenen Erwartungen übertroffen.

Wie bei einem solch umfangreichen Projekt erwartbar, gab es auch Abweichungen von ursprünglichen Konzepten und Planungen. Einzelheiten hierzu sowie zu den Arbeitsschritten im Berichtszeitraum finden sich in den Abschnitten zu den einzelnen Arbeitspaketen (s. S. 6–12 für die zweite Projektphase und s. den angehängten Zwischenbericht, S. 4–10, für die erste Projektphase).

Ein Problem während der gesamten Projektlaufzeit waren Vakanzen bei den vorgesehenen Personalstellen. Teilweise konnten Stellen, v. a. im IT-Bereich, über Monate hinweg nicht besetzt werden. Teilweise verließen Projektbeschäftigte ihre (befristeten) Projektstellen vorzeitig. Dies führte zu Verzögerungen, die kostenneutrale Laufzeitverlängerungen erforderten, und letzten Endes zu unterschiedlichen Laufzeiten bei den Projektpartnern, weil neue Beschäftigte gefunden und eingearbeitet werden mussten.

Auf technischer Ebene war die enge Verbundenheit des Deutschen Zeitungsportals mit der DDB ein großer Vorteil: Bestehende Software, Infrastruktur und Workflows konnten zu großen Teilen nachgenutzt werden, auch ist der langfristige Betrieb des Zeitungsportals durch die Einbindung in die DDB gesichert. Dass die Einbindung in die DDB auch Nachteile hatte (z. B. durch Abhängigkeiten bei den Release-Zyklen), wurde damit mehr als aufgewogen.

Datenseitig haben sich Bestände als schwierig erwiesen, die nicht mit dem im Projekt aufgebauten und den Empfehlungen des Masterplans folgenden Standardworkflow (Metadaten in METS/MODS, Volltexte in ALTO) bearbeitet werden konnten. Zwar war das Problem vom Standard abweichender Daten vorher bekannt und in mehreren Arbeitspaketen adressiert, die dafür vorgesehenen Lösungsansätze konnten allerdings nicht immer oder nicht wie vorgesehen umgesetzt werden (s. v.a. AP 4 und AP 5). Andere Projektplanungen dagegen haben sich als nicht sinnvoll herausgestellt, weil z. B. die nötigen Voraussetzungen auf Seiten der Datenpartner nicht vorlagen (z. B. artikelseparierte Daten, s. AP 6) oder sich die DDB-internen Überlegungen in der Zwischenzeit verändert hatten (SQL-Schnittstelle, IIIF, s. AP 8).

Da kein Konzept für den Umgang mit Zeitungen aus der NS-Zeit existierte, entstand ungeplant Aufwand für eine Klärung diverser Fragen (Recherche zu den gängigen Praktiken, Einholung der Expertise unterschiedlicher Stakeholder, Formulierung eines Konzepts für das Deutsche Zeitungsportal, Abstimmung mit den DDB-Gremien sowie BKM). Im Ergebnis wurde eine Lösung gefunden: NS-Zeitungen werden mit einem Disclaimer versehen und über eine virtuelle Ausstellung zur Presse in der NS-Zeit kontextualisiert.⁷

3.4 Erfahrungen über die angewandten Methoden und Nachnutzungsmöglichkeiten —

3.5 Ausführungen zur Anschlussfähigkeit der Entwicklungen an nationale und internationale Infrastrukturen

Das Deutsche Zeitungsportal und seine Bestände sind auf vielfältige Weise mit nationalen und internationalen Infrastrukturen verbunden bzw. streben dies an, s. AP 3.

3.6 Öffentlichkeitsarbeit

S. hierzu AP 1 „Projektmanagement und Öffentlichkeitsarbeit“ sowie die Abschnitte 4–4.2.

⁷ <https://ausstellungen.deutsche-digitale-bibliothek.de/ns-presse/> [10.10.2025].

3.7 Arbeitsschritte im Berichtszeitraum

Über die Arbeitspakete der ersten Projektphase wurde im Zwischenbericht informiert. Hier folgen nun die Ergebnisse in den einzelnen Arbeitspaketen der zweiten Projektphase.

AP 1 – Projektmanagement und Öffentlichkeitsarbeit

Das Projektmanagement wurde wie geplant durchgeführt.

Der Fokus der Öffentlichkeitsarbeit lag auf den wissenschaftlichen Nutzenden: Es wurden Artikel und Blog-Beiträge geschrieben, Vorträge auf Tagungen gehalten und ein Podcast veröffentlicht. Eine Liste der Beiträge zur Öffentlichkeitsarbeit findet sich auf S. 11 dieses Berichts. Besonders hervorzuheben ist die zweitägige Fachtagung „Zeitungsportal meets DH: Was braucht die Forschung, was können Zeitungsportale anbieten? Ein Austausch zwischen Wissenschaft, Archiven und Bibliotheken“, die zusammen mit zeit.punktNRW und dem Institut für Zeitungsforschung Ende Oktober 2023 in Dortmund veranstaltet wurde und einen intensiven Austausch zwischen Betreibern von Zeitungsportalen und deren wissenschaftlichen Nutzenden ermöglichte.⁸ Sie war außerdem der Auftakt einer Reihe von Tagungen, bei der Zeitungsportale mit der Wissenschaft in Austausch treten. Die zweite Tagung mit dem Titel „Analyse digitalisierter Zeitungen mit KI: Projekte, Probleme, Perspektiven“ wird am 12. November 2025 online stattfinden.⁹

Daneben gab es auch Formate, die sich an die allgemeine Öffentlichkeit richteten, wie ein Vortrag im Rahmen des Deutschen Digitaltags 2022 und die Präsentation des Zeitungsportals auf den Genealogentagen 2024 und 2025. Darüber hinaus wurde das Projekt in den Social-Media-Kanälen der Projektpartner thematisiert und auf Veranstaltungen wie der Leipziger Buchmesse und der BiblioCon präsentiert.

AP 2 – Wissenschaftliche Begleitgruppe und Nutzerforschung

Die Zusammenarbeit mit der wissenschaftlichen Begleitgruppe wurde in der zweiten Förderphase mit dem dritten Workshop im Oktober 2022 fortgesetzt.¹⁰ Dabei wurden der aktuelle Entwicklungsstand des Deutschen Zeitungsportals vorgestellt und anstehende Themen wie Korpusbildung, Datenanreicherung, Schnittstellen sowie die

⁸ Zeitungsportal meets DH. Was braucht die Forschung, was können Zeitungsportale anbieten? Ein Austausch zwischen Wissenschaft, Archiven und Bibliotheken, in: H-Soz-Kult, 22.09.2023, <https://www.hsozkult.de/event/id/event-138844> [10.10.2025].

⁹ <https://www.dortmund.de/themen/studium-wissenschaft-und-forschung/institut-fuer-zeitungsforschung/forschung-und-vermittlung/fachtagungen/analyse-digitalisierter-zeitungen-mit-ki/> [10.10.2025].

¹⁰ Mitglieder der wiss. Begleitgruppe waren in der 2. Phase: PD Dr. Astrid Blome (Institut für Zeitungsforschung Dortmund), Dr. Estelle Bunout (Universität Luxemburg), Dr. Marten Düring (Universität Luxemburg), Dr. Pim Huijnen (Universität Utrecht), Jana Keck (DHI Washington), Iris Klawatsch (Österreichische Nationalbibliothek), Dr. Jörg Lehmann (Universität Tübingen), Dr. Günter Mühlberger (Universität Innsbruck), Maria Elisabeth Müller (Staats- und Universitätsbibliothek Bremen), PD Dr. Claudia Resch (Österreichische Akademie der Wissenschaften), Dr. Thomas Werneke (FU Berlin).

Öffentlichkeitsarbeit besprochen.¹¹ Feedback und Wünsche der Begleitgruppe flossen, soweit möglich, in die Weiterentwicklung der einzelnen Funktionen ein. Diese Rückkopplung mit Vertretungen der Wissenschaftscommunity hat sich für das Projekt als äußerst hilfreich erwiesen und soll auch nach dem Ende der Projektförderung fortgesetzt werden.

Neben der wissenschaftlichen Nutzung ist für das Deutsche Zeitungsportal auch das Nutzungsverhalten der interessierten Öffentlichkeit relevant. So wurde im Herbst 2023 der erreichte Stand des Zeitungsportals durch eine externe Agentur in einem Usability-Test mit 12 Proband*innen evaluiert. Die hierbei gewonnenen Erkenntnisse über Use Cases und dafür nötige Weiterentwicklungen des Zeitungsportals sind in ein neues Release des Zeitungsportals im Herbst 2024 eingeflossen.¹²

AP 3: Verzahnung mit Schwestervorhaben

Das Zeitungsportal steht in Kontakt mit vielen Akteuren, die sich mit dem Themenfeld Zeitungen und Zeitungsdigitalisierung beschäftigen. Eine der wichtigsten Kooperationen besteht mit der Zeitschriftendatenbank (ZDB), deren standardisierte Daten zu einzelnen Zeitungstiteln eine tragende Rolle in der Datenarchitektur des Zeitungsportals spielen. In der zweiten Projektphase wurde u.a. geprüft, inwieweit der Datenfluss von der ZDB zum Deutschen Zeitungsportal durch die Nutzung von Schnittstellen optimiert werden kann. Da Datendumps gerade für den Aufbau von großen Indices gut geeignet sind, erfolgt halbjährlich eine Neuindizierung auf Basis des von der ZDB bereitgestellten RDF-Dumps. Zusätzlich wird über die OAI-Schnittstelle der ZDB monatlich geprüft, ob sich Änderungen ergeben haben, um die dann die DDB-interne Datenbasis entsprechend angereichert wird. Die Aktualität der ZDB-Daten im Deutschen Zeitungsportal konnte auf diese Weise deutlich verbessert werden.

Dem Wunsch der wissenschaftlichen Begleitgruppe nach mehr „Kontextinformationen“ entsprechend wurde im Zeitungsportal auf den Übersichtsseiten von Zeitungen ein aus der ZDB übernommener Link ergänzt, der auf externe, einzelne Zeitungstitel beschreibende Informationsressourcen führt (falls diese in der ZDB hinterlegt sind). In der ZDB wurde hierfür zuvor ein geeignetes PICA-Feld identifiziert und angepasst, damit die entsprechende Information auch katalogisiert werden kann. Die Anzeige der Links im Deutschen Zeitungsportal erfolgt aktuell noch nicht, da hierfür noch ein Software-Release nötig ist, das aber kurz bevorsteht.

¹¹ S. Anlage 2, Protokoll des Workshops mit der wiss. Begleitgruppe.

¹² S. Anlage 3, Management Summary der Zeitungsportal-Nutzerstudie 2023.

Für den Nachweis von Zeitungsportal-Beständen in der ZDB wurde eine Handreichung formuliert und veröffentlicht.¹³ Datenpartner, die sich an der ZDB beteiligen, können damit die ZDB-Einträge ihrer ans Zeitungsportal gelieferten Zeitungen anreichern. Bei Datenpartnern, die nicht an der ZDB teilnehmen, steht die ZDB-Redaktion bereit, die entsprechenden Einträge zu übernehmen. Seitens der ZDB wurden in einem Initiallauf Links zum DDB-Zeitungsportal ergänzt.

Auch mit weiteren wichtigen Stakeholdern (anderen Zeitungsportalen, einschlägigen Arbeitsgruppen, NFDI-Konsortien) ist die Projektgruppe weiter in Kontakt geblieben. Besonders hervorzuheben ist hier das nordrhein-westfälische Zeitungsportal zeit.punktNRW. Neben der Lieferung der zeit.punktNRW-Bestände gibt es viele weitere gemeinsame Aktivitäten, v.a. im Bereich Öffentlichkeitsarbeit. Im Rahmen des NFDI-Konsortiums Text+ werden die Zeitungsportal-Inhalte weiterverwendet: Seit November 2023 können sie über die Federated Content Search von Text+ durchsucht werden.¹⁴

Hingegen war es leider noch nicht möglich, die Inhalte des Zeitungsportals an die Europeana weiterzureichen. Zwar hat die Europeana ihr Zeitungsportal in der Projektlaufzeit einem Re-Design unterzogen,¹⁵ allerdings können nur die darin bereits enthaltenen Bestände im Volltext angezeigt werden. Für das Einspielen von weiteren Volltext-Beständen stehen seitens der Europeana noch keine Workflows zur Verfügung.

AP 4: Evaluierung, Erweiterung und Optimierung der Verarbeitung von Volltexten

Der Ausbau der im Zeitungsportal angebotenen Inhalte, v.a. der Volltext-Zeitungen, ist zentral für die Nutzerzufriedenheit. Hier wurden in der zweiten Projektphase große Zuwächse erzielt: Von 3,8 Mio. Seiten mit Volltexten beim Launch zu aktuell 26 Mio. Seiten. Beteiligt an dem Launch waren neun Datenpartner mit Volltext-Ausgaben, sind es nun 23. Wie geplant, wurden die für das Einspielen von Volltext-Zeitungen nötigen Workflows mit den schon bestehenden Einspiel-Workflows der DDB verbunden und in die grundständigen Aufgaben der DDB übernommen. Damit ist sichergestellt, dass auch nach dem Ende des Projekts regelmäßig neue Zeitungsbestände ins Zeitungsportal eingespielt werden. Durch eine Kooperation mit der Universitätsbibliothek Mannheim konnten für einige Bestände, die ohne Volltexte eingespielt worden waren, nachträglich Volltexte erstellt werden.

Im Kontext des DFG-Vorhabens OCR-D hat sich zwischenzeitlich ALTO als bevorzugtes Ausgabeformat für maschinell erzeugte Volltexte etabliert, während PAGE-XML vor allem

¹³ S. <https://pro.deutsche-digitale-bibliothek.de/node/1809> [10.10.2025].

¹⁴ <https://blog.dnb.de/deutsches-zeitungsportal-goes-text/> [10.10.2025].

¹⁵ <https://www.europeana.eu/de/themes/newspapers> [10.10.2025].

als Zwischenformat innerhalb der OCR-Prozesse Verwendung findet. Die ursprünglich im Projektantrag angedachte Transformation von PAGE-XML nach ALTO wurde somit obsolet. Das vor allem im Kontext der Digital Humanities für manuell erzeugte, kuratierte oder annotierte Volltexte verwendete TEI-Format ist dagegen weiterhin hoch relevant. Die zunächst geplante Transformation von TEI in ALTO im Rahmen der DDB-Ingestprozesse wurde verworfen. Statt proprietäre Transformationen für die DDB-Prozesse zu entwickeln, wurde die TEI-Unterstützung in die dem Viewer zugrundeliegende quelloffene Technologie Kitodo.Presentation¹⁶ integriert.

AP 5: Integration weiterer Zeitungsbestände

Durch eine Aktualisierung der im Viewer verwendeten OpenLayers-Komponente von Version 3 auf Version 7 wurde die bisher Kitodo-spezifische Implementierung der IIIF Image API durch eine generische und leistungsfähigere ersetzt, die in einer sehr viel größeren Community etabliert ist. Somit kann in Zukunft ein deutlich breiteres Spektrum von IIIF-Image-Servern der Datenpartner unterstützt werden, und Pflege sowie Weiterentwicklung dieser zentralen Komponente ist künftig nicht mehr allein Aufgabe der DDB-Entwicklung, sondern Gemeinschaftsaufgabe einer internationalen Community, was der Zukunftsfähigkeit dient.

Die Recherche zu Zeitungsbeständen in Archiven hat ergeben, dass in den überwiegenden Fällen die DFG-Vorgaben zu Datei- und Metadatenformaten nicht angewendet werden und die Bestände darum nicht über den Standardworkflow ins Zeitungsportal eingespielt werden können. Selbst wenn standardisierte Metadaten vorliegen (dann meist im EAD-Format¹⁷), fehlen in den meisten Fällen unverzichtbare Datenfelder wie z. B. das Tagesdatum einer Ausgabe. Ein weiteres Hindernis für die Übernahme von Zeitungen aus Archiven im großen Stil stellen die fehlenden personellen und technischen Ressourcen, v. a. in kommunalen und städtischen Archiven dar.

Um auch kleinen Einrichtungen die Teilnahme zu ermöglichen, wurde von der Fachstelle Bibliothek Structmeta,¹⁸ eine Open-Source-Software, entwickelt, mit der sich METS/MODS-XML-Dateien herstellen lassen, die zeitungsportalkonform sind. Zwar waren die Pilotpartner mit der leichten Bedienbarkeit der Software sehr zufrieden, doch es wurden auch Nachteile dieses Vorgehens sichtbar: Es entstanden (zu) große Datenpakete, für die die DDB-Ingestarchitektur nicht ausgelegt ist. Außerdem war es extrem aufwändig, nachträglich

¹⁶ <https://www.kitodo.org/software/kitodopresentation> [10.10.2025].

¹⁷ https://de.wikipedia.org/wiki/Encoded_Archival_Description [10.10.2025].

¹⁸ S. <https://github.com/Deutsche-Digitale-Bibliothek/ddblabs-structmeta/blob/main/README.md> [10.10.2025].

Änderungen an den einzelnen Datenpaketen vorzunehmen (z. B. für Updates oder Löschungen). Dieser Lösungsansatz konnte darum nicht im großen Stil weiterverfolgt werden. Stattdessen wird an einer Kooperation mit der Verbundzentrale (VZG) des Gemeinsamen Bibliotheksverbundes (GBV) gearbeitet. Dort entsteht eine Dienstleistung für kleinere Einrichtungen, die ihre Kulturdaten online präsentieren möchten. Dieses Angebot kann perspektivisch auch für das Einspielen von Zeitungsbeständen in das Deutsche Zeitungsportal genutzt werden.

AP 6: Erweiterung des Zeitungsportals für die Darstellung von Zeitungsartikeln

Die in diesem Arbeitspaket geplanten Vorhaben zur Darstellung der Articlebene wurden nicht umgesetzt, da sich herausgestellt hat, dass es nur eine sehr kleine Menge an Zeitungsbeständen gibt, bei denen die Artikel separiert sind. Der Aufwand, der für den Umbau an der Daten- und Softwarearchitektur angefallen wäre, war daher nicht verhältnismäßig. Auch ist nicht abzusehen, dass es kurz- bis mittelfristig deutlich mehr artikelseparierte Bestände geben wird – auf einen technischen Durchbruch bei der automatischen Artikelseparierung wird derzeit noch gewartet.

AP 7: Korpusbildung

Die Anforderung, im Zeitungsportal eigene Korpora für die wissenschaftliche Forschung bilden zu können, kam aus der wissenschaftlichen Begleitgruppe und wurde mit ihr im Workshop im Oktober 2022 konkretisiert. Das darauf basierende Konzept wurde mit dem Release im Herbst 2024 umgesetzt: Angemeldete Nutzende können seitdem bis zu 100 Zeitungsseiten in Listen speichern. Für jede Liste wird eine persistente ID vergeben, die das Korpus definiert und referenzierbar macht. Über ein Textfeld wird die intellektuelle Beschreibung des Korpus ermöglicht. Für die eigene Analyse können die benötigten Informationen (Volltexte, Digitalisate, Metadaten der Zeitungsseiten) selektiert und heruntergeladen werden. Die Listen können außerdem veröffentlicht und somit mit anderen Personen geteilt werden.

AP 8: Datenweitergabe

In diesem Arbeitspaket war vorgesehen, sowohl eine SRU/CQL-Schnittstelle als auch eine IIF Presentation API zu evaluieren, entsprechende Schnittstellen zu konzipieren und diese ggf. zu entwickeln und ins Zeitungsportal zu integrieren. Eine SRU/CQL-Schnittstelle wurde nicht umgesetzt, da die wissenschaftliche Begleitgruppe diese Technologie in der Zwischenzeit für überholt und in der Wissenschaft als nicht etabliert ansah.¹⁹ Die

¹⁹ S. Anlage 2, S. 4.

Transformation der Lieferformate in IIIF-Manifeste zur Bereitstellung über eine IIIF Presentation API konnte in der Projektlaufzeit nicht umgesetzt werden, da das Thema DDB weit angegangen werden soll und keine Insellösung nur für das Zeitungsportal entwickelt werden sollte. Es handelt sich dabei aber für die DDB um eine hoch relevante Technologie, die weiterverfolgt wird und bei der die Vorarbeiten schon weit gediehen sind.

AP 9: Konzeption von Datenanreicherungs-Diensten für OCR & NER

Um den Anteil der Zeitungen mit Volltexten im Zeitungsportal zu erhöhen, wurde untersucht, ob eine OCR-Komponente als zentraler Dienst im Zeitungsportal verankert werden sollte. Um die möglichen Verbesserungen gegenüber den zu erwartenden Aufwänden zu bewerten, wurde von der SBB eine Evaluation auf Basis der OCR-D-Software mit digitalisierten historischen Zeitungen durchgeführt. Ausgehend von einem Testkorpus im Zeitungsportal enthaltener Seiten wurde Ground Truth erstellt und die Qualität verschiedener in OCR-D verfügbarer Verfahren evaluiert. Hierbei zeigte sich, dass die OCR-D-Software, die vor allem für die Volltextdigitalisierung von historischen Drucken, wie sie in den VD-Verzeichnissen enthalten sind,²⁰ entwickelt wird, bei historischen Zeitungen insbesondere im Bereich der Layouterkennung noch Schwächen aufweist. Daher wurde die Option, einen solchen Dienst beim Zeitungsportal zu verankern, als nicht sinnvoll und effizient bewertet. Stattdessen wird darauf gesetzt, dass Weiterentwicklungen von OCR-D (wie etwa im OCR-D-Layout-Projekt²¹) zukünftig Qualitätsverbesserungen bewirken und Volltexte mit akkurater Layouterkennung direkt von den Datenpartnern für das Zeitungsportal bereitgestellt werden können.

Eine weitere Form von Datenanreicherung sind die sog. Named Entities, also Namen von Personen, Orten und Körperschaften. Hierfür wurde analog zur OCR seitens der SBB ein Test- und Evaluationskorpus auf Grundlage digitalisierter Zeitungen erstellt. Hierauf wurden neun KI-basierte Sprachmodelle im Hinblick auf die Performanz der Named Entity Recognition (NER) evaluiert. Eine zwingende Voraussetzung für NER ist eine hohe Qualität der Layouterkennung und Volltexte, denn wenn hierbei etwa Spalten verschiedener Artikel als ein Abschnitt erkannt werden, führt dies zu inhaltlichen Vermischungen, welche sprachbasierte KI-Modelle schnell an ihre Grenzen bringen. Damit ist eine NER-Verarbeitung der Volltexte im Zeitungsportal immer nur dann sinnvoll, wenn eine gute Layouterkennung bzw. sogar einzeln separierte Artikel als Ausgangsbasis bereits vorliegen – beides ist zum aktuellen Zeitpunkt noch nicht im benötigten Umfang der Fall (s. auch AP

²⁰ https://www.aq-sdd.de/Webs/agsdd/DE/Home/home_node.html [10.10.2025].

²¹ <https://gepris.dfg.de/gepris/projekt/517459941> [10.10.2025].

6). Die Evaluationsergebnisse und erstellten Test- und Evaluationsdaten werden von der SBB als Open Source und Open Data im Zusammenhang mit einem wissenschaftlichen Artikel für die KONVENS 2025 Konferenz veröffentlicht.

4 Öffentlich zugängliche Projektergebnisse

Im Folgenden werden die Publikationen aus der zweiten Projektphase aufgeführt. Die Publikationen der ersten Projektphase sind im Zwischenbericht aufgelistet.

4.1 Publikationen mit wissenschaftlicher Qualitätssicherung

Dinger, Patrick/Landes, Lisa: „The German Newspaper Portal: A National Aggregator for Digitised Historical Newspapers“, in: Europeana Tech 16 (Newspapers), Oktober 2020, <https://pro.europeana.eu/page/europeanatech-insight> [10.10.2025].

Landes, Lisa: „Deutsches Zeitungsportal: Deutsche Digitale Bibliothek (Germany)“, in: Beals, M. H. and Emily Bell, with contributions by Ryan Cordell, Paul Fyfe, Isabel Galina Russell, Tessa Hauswedell, Clemens Neudecker, Julianne Nyhan, Sebastian Padó, Miriam Peña Pimentel, Mila Oiva, Lara Rose, Hannu Salmi, Melissa Terras, and Lorella Viola. The Atlas of Digitised Newspapers and Metadata: Reports from Oceanic Exchanges. Loughborough: 2020. DOI: 10.6084/m9.figshare.11560059. <https://www.digitisednewspapers.net/histories/ddb/> [10.10.2025].

Landes, Lisa: Beitrag zum Deutschen Zeitungsportal in „Vergangenheitsformen“, dem Redaktionspodcast von H-Soz-Kult. Staffel 1 - Episode 4: Ethische Fragen im Umgang mit digitalen Quellen, 26.05.2023, <https://www.hsozkult.de/podcast/staffel-1/ethische-fragen> [10.10.2025].

4.2 Weitere Publikationen und öffentlich gemachte Ergebnisse

28.09.2025: „Das Deutsche Zeitungsportal als Quelle für Familienforschung“, Vortrag beim 75. Genealogentag in Frankfurt am Main.

27.09.2025: „Historische Quellen nutzen mit dem Deutschen Zeitungsportal“, Workshop beim 75. Genealogentag in Frankfurt am Main.

23.09.2025: „Mitmachen beim Deutschen Zeitungsportal: Der Aufbau einer neuen zentralen Erschließungsinfrastruktur zur Unterstützung von kleineren Kultureinrichtungen“, Vortrag bei den Goobi-Tagen 2025 in Göttingen.

16.04.2025: „Deutsches Zeitungsportal: Wie könnt ihr hier suchen?“, DDB-Blog-Beitrag (für das Dossier der WDR-APP „Zeitzeugen 1949“, <https://www.planet-schule.de/schwerpunkt/app-zeitzeugen-1945/app-zeitzeugen-1945-unterricht-100.html>), <https://www.deutsche-digitale-bibliothek.de/content/blog/deutsches-zeitungsportal-wie-koennt-ihr-hier-suchen> [10.10.2025].

16.04.2025: „Deutsches Zeitungsportal: Welche Quellen findet man hier?“, DDB-Blog-Beitrag (für das Dossier der WDR-APP „Zeitzeugen 1949“, <https://www.planet-schule.de/schwerpunkt/app-zeitzeugen-1945/app-zeitzeugen-1945-unterricht-100.html>), <https://www.deutsche-digitale-bibliothek.de/content/blog/deutsches-zeitungsportal-welche-quellen-findet-man-hier> [10.10.2025].

21.03.2025: „Eines für Alle, Alle für Eines? Viele Zeitungssammlungen, eine Webseite: Das Deutsche Zeitungsportal der DDB“, Vortrag beim 7. Workshop Retrodigitalisierung in der ZB Med, Köln.

10.02.2025: „Familienforschung mit der Deutschen Digitalen Bibliothek, Archivportal-D und Deutsches Zeitungsportal: Frei zugänglich, kostenlos und ohne Anmeldung“, Online-Vortrag auf der Genealogica 2025.

15.11.2024: „Das Deutsche Zeitungsportal: Einführung, API-Nutzung und Data LAB“, Vortrag im Seminar des NLP course for Digital Humanities and Cultural Studies im Master Program der DMGK Mainz 2024/25, IEG Mainz/Universität Mainz.

28.10.2024: „Neues Gewand für alte Zeitungen: Das Deutsche Zeitungsportal präsentiert sich mit neuem Design, mehr Funktionalitäten und über 4 Mio. Zeitungsausgaben“, Beitrag im DDB-Journal, <https://www.deutsche-digitale->

bibliothek.de/content/blog/das-deutsche-zeitungsportal-praesentiert-sich-mit-neuem-design-mehr-funktionalitaeten-und-ueber-4-mio-zeitungsausgaben/ [10.10.2025].

26.10.2024: „Familienforschung mit der Deutschen Digitalen Bibliothek – Archivportal-D und Deutsches Zeitungsportal: Frei zugänglich, kostenlos und ohne Anmeldung“, Vortrag (und Infostand) auf dem DAGV - Deutsche Arbeitsgemeinschaft genealogischer Verbände e. V.: 74. Deutscher Genealogentag, Berlin.

18.09.2024: „NS-Zeitungen im Deutschen Zeitungsportal“, Online-Vortrag beim Workshop: „Datenethik. Erfahrungsberichte aus der Praxis“ im Rahmen des zweiten NFDI4Memory Community Forums, Halle/online.

12.09.2024 „Historische Forschung digital: Ein Workshop zum Deutschen Zeitungsportal“, Workshop auf der Summer School „Digitale Methoden der Zeitungsanalyse“ der Zentralbibliothek Zürich und der AG Zeitungen und Zeitschriften (Zürich), <https://www.zb.uzh.ch/de/events/summer-school-digitale-methoden-der-zeitungsanalyse> [10.10.2025].

20.12.2023: „Deutsches Zeitungsportal goes Text+“, Beitrag zum DNB-Blog, <https://blog.dnb.de/deutsches-zeitungsportal-goes-text/> [10.10.2025].

12.12.2022: „Ein Jahr Deutsches Zeitungsportal: Über eine Million Ausgaben verfügbar“, Beitrag im DDB-Journal, <https://www.deutsche-digitale-bibliothek.de/content/journal/aktuell/ein-jahr-deutsches-zeitungsportal-ueber-eine-million-ausgaben-verfuegbar> [10.10.2025].

24.06.2022: „Neues von gestern: Online Lesen und Staunen in historischen Zeitungen. Live-Präsentation“, virtueller Workshop im Rahmen des Digitaltags 2022.

22.06.2022: „Das Deutsche Zeitungsportal: Ein neuer Zugang zu historischen Zeitungen“, Vortrag auf der Jahrestagung des Mikrofilmarchivs der deutschsprachigen Presse e.V., Paderborn. Verschriftlicht in: Zeitungs-Mikrofilm-Nachrichten (ZMN), 25. Jahrgang 2023, Nummer 26, S. 7–8.

01.06.2022: „Das Deutsche Zeitungsportal – Historische Zeitungen neu präsentiert und neu entdeckt“, Vortrag auf dem 8. Bibliothekskongress Leipzig 2022 (110. Deutscher Bibliothekartag).

16.03.2022: „Das Deutsche Zeitungsportal – Der neue zentrale Zugang zu digitalisierten historischen Zeitungen“, in: Dialog mit Bibliotheken 2022/1, S. 18–20.

14.02.2022: „Diskretion Ehrensache – (fast) druckfrische Kontaktanzeigen aus dem Zeitungsportal“, Beitrag im DDB-Journal, <https://www.deutsche-digitale-bibliothek.de/content/blog/diskretion-ehrensache-fast-druckfrische-kontaktanzeigen-aus-dem-zeitungsportal> [10.10.2025].

27.10.2021: „Presse in der Zeit des Nationalsozialismus“, virtuelle Ausstellung der Deutschen Digitalen Bibliothek, kuratiert von Johanna Niedbalski, <https://ausstellungen.deutsche-digitale-bibliothek.de/ns-presse/> [10.10.2025].

04.10.2021: „Das Deutsche Zeitungsportal – Aktueller Stand und Herausforderungen“, Vortrag bei der Fachtagung „Die Kölnische Zeitung in ihrem Kontext: Forschung – Kontroversen – Digitalisierung“ in Köln.

19.07.2021: „Das Deutsche Zeitungsportal: Ein neuer Zugang zu historischen Zeitungen (auch) für die Digital Humanities“, Vortrag bei der VDB-Fortbildung: „Wissenschaftliche Bibliotheken und Digital Humanities: Potentiale für die Fachreferatsarbeit“, online.

25.11.2020: „Das Deutsche Zeitungsportal: Chancen und Herausforderungen der Zusammenführung digitalisierter historischer Zeitungsbestände“, Forschungskolloquium „Digital History“ der HU Berlin, virtuell, Abstract/Blog-Post: <https://dhistory.hypotheses.org/265> [10.10.2025], Vortrag auf Youtube: <https://youtu.be/sKhi-65I5CM> [10.10.2025].

22.09.2020: „Ein Portal, viele Partner – Herausforderungen beim Aufbau eines bestandsübergreifenden Zeitungsportals“, Vortrag auf der Tagung „Informationen – digital verpackt“ der Zentralbibliothek, virtuell.

Der für das Zeitungsportal erstellte Source Code wird gemäß den FAIR4RS Principles als Open-Source-Projekt im öffentlich zugänglichen Code-Repository²² von FIZ Karlsruhe abgelegt und ist unter der Apache Software Licence (ASL 2.0) nachnutzbar.

²² <https://dev.fiz-karlsruhe.de/stash/projects/DDB> [10.10.2025].