Empfehlungen
zur Digitalisierung historischer Zeitungen
in Deutschland
(Masterplan Zeitungsdigitalisierung)

Ergebnisse des DFG-Projektes
„Digitalisierung historischer Zeitungen“
Pilotphase 2013-2015

Partner:

Berlin: Staatsbibliothek zu Berlin – Preußischer Kulturbesitz (SBB)
Bremen: Staats-und Universitätsbibliothek Bremen (SuUB)
Dresden: Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB)
Frankfurt am Main: Deutsche Nationalbibliothek Frankfurt am Main (DNB)
Halle/S.: Universität- und Landesbibliothek Sachsen-Anhalt Halle/S. (ULB)
München: Bayerische Staatsbibliothek München (BSB)

Dresden, 29. Januar 2016 / Berlin, 12. Juni 2017

[Hinweis: Es handelt sich um eine revidierte Version des Dokuments, das aufgrund von Hinweisen aus der Begutachtung durch die DFG an einigen Stellen überarbeitet und aktualisiert wurde. Für diese Überarbeitung hat im Zusammenwirken aller ehemaliger Projektpartner die Staatsbibliothek zu Berlin von der SLUB Dresden die Federführung übernommen. Die durchgeführten Änderungen an diesem Dokument beziehen sich inhaltlich im Wesentlichen auf

- eine Aktualisierung und Präzisierung von Informationen zur ZDB und ihrem inhaltlichen Umfeld, Formulierung eines eigenen Kapitels dazu (Nr. 5)
- eine neu formulierte Passage zu Bestandslückenergänzungen
- ein neu formuliertes Kapitel zu Kostenkorridoren
- die Erweiterung des Anhangs
- einige Zahlen sowie die Gesamtberechnung wurde an neuere Erkenntnisse angepasst

Über diese Themen hinaus wurde der Masterplan redaktionell durchgearbeitet, aber nicht durchgehend aktualisiert, zitierte Quellen (Links), Angaben zu Projekten und Literatur sind also ganz überwiegend auf dem Stand der ersten Version 2016 verblieben.

Kennzeichnungen zur Aktualisierung erfolgten zum Erhalt der Lesefreundlichkeit nur bei längeren Passagen.

Darüber hinaus wurde der Masterplan um eine Zusammenstellung praktisch orientierter Handreichungen und ein Unterstützungstool ergänzt, die digitalisierungswillige Einrichtungen dabei unterstützen sollen, ein Projekt zur Digitalisierung von Zeitungen möglichst erfolgreich vorzubereiten. Die Handreichungen und der Textgehalt des Unterstützungstools finden sich im Anhang zum Masterplan. Um einerseits die Handhabbarkeit zu erhöhen und andererseits auch den erforderlichen laufenden Aktualisierungsbedarf zu bedienen, sind diese Hilfsmittel zusätzlich in das entsprechende, einfach gehaltene Unterstützungstool eingebunden. Dieses findet sich unter

http://www.zeitschriftendatenbank.de/zeitungsdigitalisierung

Die Staatsbibliothek zu Berlin wird die laufende Pflege der dort enthaltenen und verlinkten Informationen übernehmen. Hinweise dazu bitte an zdb-hotline@sbb.spk-berlin.de]

Inhaltsverzeichnis

Einleitung: Warum Zeitungen?

Die Digitalisierung der Zeitungen des 17. Jahrhunderts mit Förderung der Deutschen Forschungsgemeinschaft (DFG) eröffne die Chance, „ein ganzes Jahrhundert neu zu entdecken", so Volker Hagedorn in einem ganzseitigen ZEIT-Beitrag „Die *Breaking News* von damals kann man jetzt online lesen" (DIE ZEIT, 22.10.2015, S. 18)[1]. Schon vor der Epoche der Aufklärung habe sich die Zeitung zum Medium einer kritischen Öffentlichkeit entwickelt: „Wer die Zeitungen nicht achtet", schrieb Kaspar Stieler im Jahr 1695 in „Zeitungs Lust und Nutz", „bleibet immer und ewig ein elender Prülker und Stümper in der Wissenschaft der Welt".
Zwischen 1605 und 1945 fand die Zeitung mit rund 21.000 Zeitungsunternehmungen im deutschen Sprachgebiet ihre früheste und weiteste Verbreitung. Aus dem Flugblatt der Frühen Neuzeit erwachsen, etablierte sie sich als wichtigstes Nachrichten- und Unterhaltungsmedium mit den typischen Merkmalen Periodizität, Publizität, Aktualität und Universalität. Mit ungezählten Originalbeiträgen in Text und Bild zu allen gesellschaftlichen Bereichen, Ressorts und Disziplinen sind Zeitungen reiche Quellen, „Sekundenzeiger der Geschichte" (Schopenhauer).
In Deutschland werden Zeitungen in Archiven und Bibliotheken zum Teil nach gesetzlichen Vorgaben gesammelt und in Magazinen aufbewahrt, sind jedoch oftmals schlecht zugänglich: Aufgrund der Massen, ihres großen Formats und ihrer Fragilität sowie der aufwendigen und deshalb überwiegend ungenügenden Erschließung haben sie über Jahrzehnte ein Schattendasein geführt. Damit blieb für die Fachwissenschaften und ein breites interessiertes Publikum ein reiches Quellenreservoir nur schwer zugänglich.
Mit der Sicherheitsverfilmung verbesserte sich die Situation, doch erst mit den Möglichkeiten der Digitalisierung können große Corpora an Texten und Bildern für wissenschaftliches Arbeiten technisch aufbereitet und zugleich für eine breite kulturelle und bildungspolitische Nutzung bereitgestellt werden. Bereits bei einem DFG-Rundgespräch 2009 in der SuUB Bremen sprachen sich Wissenschaftlerinnen und Wissenschaftler unterschiedlicher Disziplinen für eine möglichst umfassende digitale Bereitstellung historischer Zeitungen aus.

Um die bis 2010 in Deutschland weitgehend unkoordinierten Aktivitäten der Zeitungsdigitalisierung in stimmige und effektivere Strukturen zu lenken und an den Bedarfen der Wissenschaften auszurichten, hatte die DFG dem Gemeinschaftsantrag von sechs Bibliotheken in Bremen, Berlin, Dresden, Halle, Frankfurt/M. und München für eine Pilotphase zugestimmt.

Damit sollten

- verschiedene Werkzeuge, Verfahren und Erschließungstiefen in Projekten erprobt
- Nachweis- und Präsentationsstrukturen konkret verbessert
- Empfehlungen zur Auswahl unter inhaltlichen, rechtlichen, konservatorischen und organisatorischen Kriterien erarbeitet
- sowie ein Masterplan für einen systematischen Ausbau der Zeitungsdigitalisierung insbesondere für die wissenschaftliche Nutzung erstellt werden.

[1] http://www.zeit.de/2015/43/historische-zeitungen-digitalisierung/komplettansicht

Der vorliegende Masterplan fasst die Ergebnisse der Pilotphase 2013-2015 zusammen, beschreibt den aktuell erreichten Stand der Zeitungsdigitalisierung in Deutschland und gibt auf der Grundlage eigener Erfahrungen und unter Einbeziehung der Erkenntnisse anderer nationaler und internationaler Zeitungsdigitalisierungsprojekte abgestimmte Empfehlungen für eine Hauptphase zur „Digitalisierung historischer Zeitungen in Deutschland“.

1. Ausgangslage

1.1 Situation bis Projektbeginn der Pilotphase April 2013

Ausgangspunkt des hier beschriebenen Projekts war der – im internationalen Vergleich – offenkundige Nachholbedarf Deutschlands bei der Zeitungsdigitalisierung. Während große Zeitungsdigitalisierungsprojekte z.B. in Australien, England, USA, Österreich oder in den Niederlanden durch Nationalbibliotheken und Verlage ins Leben gerufen wurden, mussten in Deutschland mit seiner ausgeprägten regionalen Zeitungslandschaft und einer sehr hohen Zeitungsdichte andere Wege gefunden werden. Im Unterschied zu Nationalbibliotheken anderer Länder besitzt die vergleichsweise junge, 1913 gegründete Deutsche Nationalbibliothek (DNB) an ihrem Standort Leipzig nur Teile der historischen Zeitungsüberlieferung. Die laufenden Zeitungen sammelt die DNB in Form von Mikrofilmen und seit einigen Jahren als digitale Kopie.

Die DNB betreut zusammen mit der Staatsbibliothek zu Berlin die Zeitschriftendatenbank, die umfangreichste Nachweisplattform für Zeitschriften- und Zeitungs-Bestände im deutschen Sprachgebiet. Die Staatsbibliothek zu Berlin gründete mit ihrem herausragenden internationalen Zeitungsbestand ein Zeitungsinformationssystem ZEFYS, das keinen nationalen Anspruch erhebt, jedoch umfangreiche Bestände, etwa die preußische Amtspresse, historische Berliner Tageszeitungen und Zeitungen der DDR digital präsentiert. Die Bayerische Staatsbibliothek München digitalisierte bayerische Zeitungen im Rahmen der Bayerischen Landesbibliothek Online und konnte seit der 2007 mit Google vereinbarten Digitalisierungskooperation große Teile ihres Zeitungsbestandes digitalisieren, die schrittweise in die Zeitungsplattform digiPress integriert werden. Neben weiteren regionalen Ansätzen in mehreren Bundesländern gab es vergleichsweise wenige fachlich getriebene Projekte im deutschsprachigen Raum wie etwa die zur Exilpresse (Deutsche Nationalbibliothek) oder zur jüdischen Publizistik (Compact Memory, Stadt- und Universitätsbibliothek Frankfurt am Main).
Im Rahmen des ersten DFG-Rundgesprächs an der SuUB Bremen im November 2009 mit Wissenschaftlerinnen und Wissenschaftlern und engagierten Einrichtungen (Bibliotheken, Archive, Presseforschung, Mikrofilmarchiv Dortmund) wurde der Bedarf an einer leicht zugänglichen digitalen Präsentation historischer Zeitungen für Forschung und Lehre in Deutschland festgestellt. Angesichts des immensen Aufwands, in Archiven und Bibliotheken Zeitungen im Original oder als Mikrofilm durchzusehen, sei die digitale Transformation und Bereitstellung von Zeitungen dringend notwendig.
Vor diesem Hintergrund stellten sechs Bibliotheken nach mehreren koordinierenden Arbeitstreffen aufeinander abgestimmte Anträge für eine Pilotphase:

- die Staatsbibliothek zu Berlin und die Deutsche Nationalbibliothek einen Antrag zur funktionalen Verbesserung der Zeitschriftendatenbank (ZDB) als Nachweis- und Steuerungsinstrument für die Zeitungsdigitalisierung,
- die SLUB Dresden einen Antrag zur medientypologischen Erweiterung des DFG-Viewers als Open Source-Präsentationsplattform nach vereinbarten Mindeststandards für eine homogene digitale Präsentation von Zeitungen
- die Staatsbibliotheken Berlin und München sowie die Staats-, Landes- und Universitätsbibliotheken Bremen, Dresden und Halle Pilotanträge zur Digitalisierung ausgewählter Zeitungen mit unterschiedlichen Verfahren und Werkzeugen sowie in verschiedenen Erschließungstiefen, um die Ergebnisse vergleichend auszuwerten und auf dieser Grundlage in abgestimmter Form weiterführende Maßnahmen zu empfehlen.

Während der Vorbereitungsphase eröffnete sich den Staatsbibliotheken Berlin und Hamburg die Möglichkeit, im Rahmen des von der Staatsbibliothek zu Berlin koordinierten Europeana Newspapers-Projekts digitalisierte Berliner und Hamburger Zeitungen als Volltexte aufzubereiten und zu präsentieren. Die Staatsbibliothek zu Berlin hat im Rahmen des vorliegenden DFG-Projekts außerdem die Entwicklung des Workflows „Goobi-Presentation" betreut, die Massendigitalisierung von vier Berliner Zeitungen dann im Rahmen zweier anderer EU-Förderungen realisiert. Diese Projektergebnisse sind in die mit diesem Masterplan vorgelegten Bewertungen mit eingeflossen.

1.2 Zeitungsportale anderer Länder

Während in Deutschland mit ZEFYS (Berlin) oder digiPress (München) Ansätze für Zeitungsportale, jedoch nicht mit nationalem Anspruch, entstanden, bestehen in europäischen Ländern, in Australien und den USA maßgeblich von Nationalbibliotheken initiierte nationale Portale. Diese unterscheiden sich stark von der deutschen Situation aufgrund der

- pressegeschichtlichen Entwicklungen des jeweiligen Landes,
- der Finanzierungsformen zur Digitalisierung der Zeitungen,
- und hinsichtlich der Auswahl der Zeitungen und der verfügbaren Zeiträume.

Als Rechercheinstrumente und reiches Quellenreservoir sind sie auch für die deutsche Forschung relevant und zugleich als Benchmark für ein forschungsfreundliches deutsches Angebot zu beachten.

In ANNO – AustriaN Newspapers Online[2], dem im deutschen Sprachraum stark genutzten Angebot der Österreichischen Nationalbibliothek, stehen aktuell über 840 Zeitungen und zeitungsähnliche Journale mit 15 Mio. Zeitungsseiten aus dem Zeitraum 1568-1944 online[3], – davon wurden im Rahmen des EU-Projekts Europeana Newspapers für 4,6 Mio. Seiten Volltexte erzeugt (zu dem Projekt ausführlicher S. 8f.). ANNO ist modular aufgebaut und innerhalb von dreizehn Jahren in Zusammenarbeit mit zahlreichen regionalen

[2] http://anno.onb.ac.at/
[3] Vgl.: Weber, Albert: Internationaler Workshop: Digitizing German-Language Cultural Heritage from Eastern Europe (Institut für Ost- und Südosteuropaforschung, Regensburg, 27./28. April 2015), in: Spiegelungen. Zeitschrift für deutsche Kultur und Geschichte Südosteuropas Jg. 10 (2015), Heft 1, S. 249-254, hier S. 254.

Partnereinrichtungen zu einem erfolgreichen digitalen Lesesaal für Zeitungen entwickelt worden.
Delpher[4], das nationale Portal der Niederlande, bietet Zugang zu mehr als 9 Mio. Zeitungsseiten des Zeitraums 1618-1993, alle sind mit Volltexten und Artikelseparierung erschlossen. Durch die Einbindung eines historischen Wörterbuchs in die Suche können hier als Besonderheit auch historische Schreibvarianten gefunden werden. Zudem wurden sämtliche Artikelüberschriften manuell korrigiert, um die Genauigkeit der Suche zu erhöhen. Durch Visualisierungstools ist das Portal sehr nutzer- und forschungsfreundlich. Mit zahlreichen Zeitungen der NS-Zeit ist dieses Portal eine wichtige Online-Quelle für die Zeit zwischen 1933 und 1945.
Das British Newspaper Archive[5] mit 12 Mio. Zeitungsseiten ist das einzige kostenpflichtige nationale Portal. Die Digitalisierung erfolgte im Rahmen einer Private-Public-Partnership zwischen der British Library und Microsoft. Im Portal der Nationalbibliothek Wales[6] sind aktuell 1,1 Mio. Seiten bzw. 15 Mio. Artikel verzeichnet, darunter auch solche, die an der British Library nur gegen Bezahlung zugänglich sind.
Das Portal DIGI[7] der Nationalbibliothek Finnlands enthält 3,5 Mio. Seiten und wird ständig erweitert. Mit 'Digitalkoot'[8] ist ein Onlinespiel mit dem Portal verknüpft, über welches Spieler Volltexte und Segmentierungen korrigieren. In Dänemark begann die Königliche Bibliothek Kopenhagen 2014 mit der Digitalisierung aller dänischen Zeitungen seit 1668 – mit einem Dienstleister sollen innerhalb von zwei Jahren 32 Mio. Seiten digitalisiert, mit OCR und Artikelseparierung bearbeitet und in einem nationalen Portal bereit gestellt werden.[9]
Mit knapp 20 Mio. Seiten und 187 Mio. Artikeln setzt international das australische Portal TROVE[10] nach wie vor Standards – hier wurde auch erstmals in größerem Umfang mit einer Korrektur der Volltexte durch Benutzer gearbeitet. Das Portal zeigt nebeneinander die wichtigsten Sucheinstiege über die Kalenderfunktion, die Volltextsuche und eine geographische Suche.
Im Oktober 2015 überschritt das US-Zeitungsportal Chronicling America[11] die Zahl von 10 Mio. Seiten historischer Zeitungen aus dem Zeitraum 1836-1922. Eine Einbindung in die Digital Public Library of America (DPLA) analog zur Europeana ist für 2016-17 vorgesehen.

Im April 2015 hat die International Coalition on Newspapers (ICON) eine vorläufige Auswertung der weltweiten Zeitungsdigitalisierung auf Basis der ICON-Datenbank mit 171.000 Zeitungstiteln aus 160 Ländern vorgelegt: „A Comparative Analysis of Newspaper Digitization to Date"[12]. Danach sind mindestens 30.000 Titel weltweit digitalisiert worden, die meisten in Europa und in den USA, überwiegend auf der Grundlage von Mikrofilmen, und mit deutlichem Schwerpunkt auf Zeitungen des 19. und 20. Jahrhunderts (insbes. 1880-1920). Die Studie des Center for Research Libraries (CRL) berechnet die amerikanischen Investitionen 2004-2016 auf 28.5 Mio. U.S.Dollar für 9,5 Mio. Seiten und schätzt den Bedarf

[4] http://www.delpher.nl/nl/kranten
[5] http://www.britishnewspaperarchive.co.uk/
[6] http://newspapers.library.wales/
[7] http://digi.kansalliskirjasto.fi/sanomalehti
[8] http://www.digitalkoot.fi/
[9] http://en.statsbiblioteket.dk/national-library-division/the-danish-newspaper-collection
[10] http://trove.nla.gov.au/newspaper
[11] http://chroniclingamerica.loc.gov/
[12] http://www.crl.edu/sites/default/files/d6/attachments/events/ICON_Report-State_of_Digitization_final.pdf. Und Umfrage Europeana, URL: http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D4.1-Europeana-newspapers-survey-report.pdf.

bis 2020 auf rund 46 Mio. U.S. Dollar. Die 21-seitige Übersicht zieht diese Schlussfolgerungen:

1. zu viele schutzbedürftige, vom Verfall bedrohte Zeitungen blieben noch unberücksichtigt und unzugänglich,
2. die Digitalisierung folge selektiven und nicht strategischen Zielen aufgrund einer fehlenden steuernden Datenbasis,
3. Zeitungsdigitalisierung sei überwiegend lokal getrieben, forschungsrelevante Zeitungen blieben dadurch oft unberücksichtigt,
4. Digitalisierung und uneingeschränkte Verfügbarkeit seien nicht synonym (als Beispiel werden die digitalisierten DDR-Zeitungen in ZEFYS genannt, die jedoch nach Authentifizierung kostenfrei zugänglich sind),
5. ICON plädiert für eine leistungsstarke internationale Datenbasis für historische Zeitungen.

Vor dem Hintergrund dieser amerikanischen Zwischenbilanz kommt den strukturbildenden Maßnahmen in Europa und insbesondere in Deutschland mit dem umfangreichsten Zeitungsbestand seit dem 17. Jahrhundert eine besondere Bedeutung zu.

Im Rahmen des ICT Policy Support Programme (ICT-PSP) hat die EU-Kommission das Projekt Europeana Newspapers – A Gateway to European Newspapers Online[13] als „Best-Practice-Network" mit über 4 Mio. EUR gefördert. An dem von der Staatsbibliothek zu Berlin koordinierten Projekt waren 18 Projektpartner und 11 assoziierte Partner sowie 35 Netzwerkpartner vom Februar 2012 bis zum März 2015 beteiligt. Die Hauptziele waren:

1. Aggregation von Metadaten digitalisierter Zeitungstitel, deren Überführung in das Europeana Data Model (EDM) sowie die Integration der Titeldaten in die Europeana und in die Zeitschriftendatenbank (ZDB)
2. Verarbeitung von 10 Millionen Zeitungsseiten mit technischen Verfahren zur Erstellung von Volltexten, davon 8 Millionen Seiten mit OCR (Optical Character Recognition) und weitere 2 Millionen Seiten mit OLR (Optical Layout Recognition) sowie experimentelle Bearbeitung mit NERD (Named Entity Recognition & Disambiguation) in drei Sprachen (Deutsch, Französisch, Niederländisch)
3. Entwicklung eines europäischen Portals für Online-Recherchen in digitalisierten Zeitungen
4. Entwurf eines auf gängigen Standards wie METS und ALTO aufbauenden Metadatenprofils speziell für digitalisierte Zeitungen
5. Entwicklung von (Software-)Werkzeugen für die Qualitätskontrolle sowie von experimentellen Verfahren, die eine Einschätzung der zu erwartenden Qualität von Volltexten ermöglichen
6. Durchführung von Informationsveranstaltungen, Verbreitung der Projektergebnisse.

In die Europeana[14] konnten mehr als 20 Mio. Seiten importiert werden, davon sind im TEL-Portal (The European Library)[15] aktuell ca. 12 Mio. Seiten vollständig durchsuchbar. Die deutschsprachigen Zeitungen haben die Staatsbibliothek zu Berlin (4 Berliner Zeitungen, 1,5 Mio. Seiten), die Staats- und Universitätsbibliothek Hamburg (7 Hamburger Zeitungen, 1,7 Mio. Seiten), die Österreichische Nationalbibliothek (333 Zeitungen, 4,6 Mio. Seiten) und die

[13] http://www.europeana-newspapers.eu/
[14] http://europeana.eu/portal/search.html?query=europeana_collectionName%3A92*ewspapers*
[15] http://www.theeuropeanlibrary.org/tel4/newspapers

Landesbibliothek Dr. Friedrich Teßmann Südtirol beigesteuert (47 Zeitungen, 1 Mio. Seiten).[16]

1.3 Parallele Entwicklungen während der Pilotphase 2013-2015 in Deutschland

Die parallel zum DFG-Pilotprojekt und zum Europeana Newspapers-Projekt laufenden und begonnenen weiteren Vorhaben zeigen die Dynamik und den hohen Bedarf an einer digitalen Transformation des Mediums Zeitung. Zu nennen sind exemplarisch Landesdigitalisierungsprogramme in Baden-Württemberg, Berlin, Nordrhein-Westfalen, Rheinland-Pfalz oder Sachsen. Die Landesbibliothek Karlsruhe hat 29 badische Amtsblätter und Zeitungen, die Universitäts- und Landesbibliothek Bonn 145 regionale Amtsblätter und Zeitungen im Umfang von 1,5 Mio. Seiten digitalisiert. In Zusammenarbeit mit dem LVR-Archivberatungs- und Fortbildungszentrum plant die ULB Bonn weitere rheinische Zeitungen in einem Großprojekt zu digitalisieren. Die SLUB Dresden will 2017 mit regionalen Fördermitteln ein seit den 90er Jahren aufgebautes Mikrofilmarchiv sächsischer Zeitungen digitalisieren.

Im Rahmen des mit Förderung des Europäischen Fonds für regionale Entwicklung (EFRE) an der Staatsbibliothek zu Berlin durchgeführten Projekts Digitalisierung historischer Berliner Zeitungen (DAHLIE) wurden vier historische Berliner Tageszeitungen aus den Jahren 1870 bis 1932 mit ca. 1,5 Mio. Seiten mittels Hochleistungs-Rollfilmscanner inhouse digitalisiert, manuell nachbearbeitet und mit Metadaten angereichert. Die Weiterverarbeitung erfolgte im beschriebenen Europeana Newspapers-Projekt mit dem Ergebnis, dass 1,5 Mio. Seiten in der Europeana nachgewiesen und die Volltexte im TEL-Portal durchsuchbar sind. In das Berliner Zeitungsinformationssystem (ZEFYS) wurden die Imagedateien eingebunden, im Zuge eines Umbaus des Portals 2016 ist auch die Integration der elektronischen Volltexte vorgesehen. Die Digitalisierungsarbeiten wurden mit einem Dienstleister und noch nicht mit der Open Source-Software Goobi durchgeführt, die inzwischen im vollen Funktionsumfang zur Verfügung steht und künftig eingesetzt wird. Die Arbeiten wurden so durchgeführt, dass die erzeugten Daten in Goobi integriert werden können und damit für die Weiterverarbeitung bereitstehen. So entstand eine nachhaltige Infrastruktur für die Zeitungsdigitalisierung, die auch für künftige Vorhaben genutzt werden kann.

Das Digitale Forum Mittel- und Osteuropa (Difmoe)[17], ein Konsortium von 14 internationalen Einrichtungen, will rund 34 Zeitschriften und Zeitungen mit Volltextrecherche besonders aus dem Bereich der Südosteuropaforschung zugänglich machen. Im Blog „Minorities Records – Transfering culture diversity into digital"[18] sollen auch Digitalisierungsstrategien diskutiert werden. Das Vorhaben des Instituts für Ost- und Südosteuropaforschung Regensburg wird von der Staatsministerin für Kultur und Medien (BKM) gefördert.

Google hatte die Zeitungsdigitalisierung (in Zusammenarbeit mit vorwiegend amerikanischen Zeitungsverlagen) Anfang des Jahres 2011 offiziell beendet. Davon zu trennen ist allerdings das Google Books-Programm mit Bibliotheken, an dem u.a. die Bayerische Staatsbibliothek und die Österreichische Nationalbibliothek teilnehmen. In diesem Kontext werden auch

[16] http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D4.5_Report_on_newspapers_aggregated_by_TEL_3.0.pdf
[17] http://www.difmoe.eu/?content=Periodika
[18] http://minorecs.hypotheses.org/

gebundene, urheberrechtsfreie Zeitungsbände berücksichtigt, da die Digitalisierung – konservatorische Eignung vorausgesetzt – unabhängig vom Medientyp erfolgt. Auf diese Weise wurde im Rahmen der Public-Private-Partnership nahezu der gesamte urheberrechtsfreie Zeitungsbestand der Bayerischen Staatsbibliothek digitalisiert. Mittlerweile liegen rund 1050 vorwiegend bayerische Zeitungstitel mit ca. 10 Mio. Seiten (inkl. Volltext) vor. Zeitungen werden von Google selbst aber ohne zeitungstypische Erschließung im Normalprogramm Google Books integriert. Eine titelübergreifende Suche, die sich auf Zeitungen beschränkt, oder eine Suche innerhalb eines Zeitungstitels über mehrere Bände hinweg ist so nicht möglich. Um die von Google digitalisierten Zeitungen für die Wissenschaft optimal zugänglich zu machen und differenzierte Recherchen anbieten zu können, ist ihre Erschließung und eine dem Medientyp entsprechende Präsentation erforderlich. Ab 2016 plant die Bayerische Staatsbibliothek daher sukzessive alle von Google bereits digitalisierten Zeitungstitel zu erschließen und in die neue Zeitungsplattform digiPress[19] zu integrieren.

Für den Zeitraum nach 1945 haben inzwischen einzelne Verlage Zeitungsarchive digitalisiert und ermöglichen Volltextrecherchen sowohl kostenfrei *(*z.B. DIE ZEIT[20], das Hamburger Abendblatt[21], die Neue Zürcher Zeitung[22]) als auch kostenpflichtig (FAZ, Mindener Tageblatt[23] u.v.a.) in unterschiedlicher Qualität. In der Regel stehen diese Bestände nicht oder nur auf der Basis von Sondervereinbarungen als Volltextkorpora für wissenschaftliche Datenanalysen zur Verfügung.

1.4 Fazit und Herausforderungen an die Pilotphase

Die zahlreichen parallelen Aktivitäten verdeutlichen das große Interesse an der Digitalisierung historischer Zeitungen, sie zeigen aber auch, wie notwendig eine Koordinierung mit dem Ziel verbesserter Transparenz und überzeugender ggf. virtueller Aggregierung für nutzerfreundliche Sucheinstiege ist.[24] Der Befund der International Coalition on Newspapers (ICON), dass die Zeitungsdigitalisierung bislang selektiven und nicht strategischen Kriterien und Zielen folgt, gilt insbesondere für Deutschland, wo eine für Zeitungsdigitalisierung zuständige Instanz fehlt. Erst wenn die zahlreichen Projekte und zum Teil kleinteiligen Angebote nach einheitlichen Mindeststandards und in kritischer Masse zusammengeführt sind, werden Reichtum und Qualität deutscher Zeitungsüberlieferung für die nationale und internationale Forschung sichtbar und effektiv nutzbar.
Die Pilotprojekte im Rahmen der DFG-Förderung konzentrierten sich deshalb auf die Entwicklung strukturbildender Maßnahmen, die Erprobung unterschiedlicher Digitalisierungsverfahren und deren Auswertung in Form eines Masterplans.

[19] http://digipress.digitale-sammlungen.de
[20] http://www.zeit.de/2015/index
[21] http://www.abendblatt.de/archiv/
[22] http://zeitungsarchiv.nzz.ch/search/
[23] http://www.mt.de/archiv/
[24] Die Verwirrung über den aktuellen Stand aus Sicht universitärer Nutzer belegt exemplarisch die Webseite der Leuphana Universität Lüneburg. Die hier aufgeführten Links verdeutlichen den Mangel an einem überzeugenden deutschen Zeitungsportal: http://www.leuphana.de/universitaet/personen/dagmar-bussiek/lehrangebot/zeitungsarchiv/digitalisierte-zeitungen.html

Im Einzelnen wurden:

1. die Zeitschriftendatenbank (ZDB) in Richtung Nutzerorientierung und Funktionalität so verbessert, dass sie den besitzenden Institutionen als Datenplattform für Zeitungsdigitalisierungsprojekte dient und gleichzeitig Wissenschaftlerinnen und Wissenschaftlern transparenter als zuvor differenzierte Quelleninformationen und Bestandsnachweise für den Materialtyp Zeitung bietet,
2. mit dem DFG-Viewer einheitliche Mindeststandards für die Präsentation von Zeitungen erarbeitet und etabliert, um Aggregierungen heterogener Projektpräsentationen praktisch zu unterstützen und nicht zuletzt auch den zahlreichen Initiativen in Städten und Gemeinden ein kostenfrei nachnutzbares Präsentationstool an die Hand zu geben,
3. in den Pilotprojekten unterschiedliche Verfahren erprobt, die Anzahl der digitalen Zeitungsseiten um 1,5 Millionen erhöht und auf dieser Basis der vorliegende Masterplan mit Empfehlungen für eine Hauptphase „Digitalisierung historischer Zeitungen in Deutschland“ erstellt.

2. Die Projektteile zu Infrastrukturverbesserungen

2.1 Weiterentwicklung der Zeitschriftendatendank zum wissenschaftsfreundlichen Nachweis- und Steuerungsinstrument

Die Zeitschriftendatenbank (ZDB) ist das wichtigste nationale Nachweisinstrument für Zeitschriften und Zeitungen und verzeichnet ca. 1,8 Mio. Titeldaten mit rund 14,1 Mio. Bestandsnachweisen von Periodika in mehr als 4.000 Wissenschafts- und Kultureinrichtungen aus Deutschland und Österreich. 10,1 Mio. Verknüpfungen mit der Gemeinsamen Normdatei (GND) belegen eine hohe Datenqualität und den ausgereiften Standardisierungsgrad der ZDB. Enthalten sind rund 60.000 internationale Zeitungstitel, darunter 39.525 deutsche Zeitungen bzw. zeitungsähnliche Blätter. Von diesen stammen wiederum 21.583 Zeitungstitel aus dem Zeitraum 1600-1945, die in deutscher Sprache gedruckt oder im deutschen Sprachraum erschienen sind.

Die am häufigsten vorkommenden Erscheinungsorte historischer deutscher Zeitungen sind Berlin (1777 Titel), München (740), Wien (526), Köln (451), Leipzig (406) und Hamburg (377). Dies sind rund 20% der rund 21.000 Zeitungen, was andererseits bedeutet, dass rd. 80%, also der weitaus größte Teil, über das ganze deutsche Sprachgebiet verteilt gedruckt wurde. Erst mit der Reichsgründung findet eine starke Konzentration in Berlin statt. Unter den Zeitungen des deutschen Sprachgebiets befinden sich mehr als 300 Titel in rund 40 Sprachen, insbesondere Französisch, Polnisch, Russisch und Englisch, ein weiteres Indiz für die Vielfalt der Zeitungslandschaft.
Innerhalb der letzten Jahre ist dank der zahlreichen parallelen Aktivitäten im europäischen Raum einschl. der Google-Aktivitäten in München und Wien, regionaler Programme und nicht zuletzt des vorliegenden DFG-Pilotprojekts die Zahl digitalisierter Zeitungen deutlich angestiegen. Allerdings müssen viele der bislang rund 4.000 von Bibliotheken in der ZDB nachgewiesenen Digitalisate noch zeitungsspezifisch erschlossen und in Präsentationen eingebunden werden. Auch zeigt die ZDB, dass viele digitalisierte Zeitungen kleine Bestandssegmente umfassen, nicht selten auch nur eine Ausgabe, die auf Benutzerwunsch hin digitalisiert wurde.

Der Zuwachs an Aktivitäten verstärkt den Zeitdruck, die im Projekt entwickelten Verfahren und Methoden möglichst bald strukturbildend nicht nur für ein koordiniertes Gesamtvorgehen einzusetzen, sondern auch zu abgestimmten Präsentationsformaten und einer überzeugenden nationalen Portallösung mit übergreifender medienspezifischer Recherchemöglichkeit zu kommen.
Die 21.583 Zeitungstitel des Zeitraums 1600 bis 1945 verteilen sich auf 1.654 besitzende Einrichtungen, vorwiegend Bibliotheken, da Archivbestände (mit Ausnahme Bayerns) noch nicht systematisch in die ZDB aufgenommen wurden. Rund 40 Bibliotheken weisen 500 und mehr Titel nach, auf über 2.000 Bibliotheken verteilen sich geringere Nachweiszahlen mit einer deutlichen Spitze im Bereich von 50 – 150 Titeln. Diese Bibliotheken und Archive sind mit den umfangreichsten Bestandsnachweisen in der ZDB vertreten:

Bibliothek	Anzahl Titel
Staatsbibliothek zu Berlin	6178
Bayerische Staatsbibliothek München	5250
Deutsche Nationalbibliothek Frankfurt/M. und Leipzig	4462
Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden	2149
Universitäts- und Landesbibliothek Sachsen-Anhalt Halle/S.	1362
Bibliothek des Bundesarchivs	1346
Bibliothek der Friedrich-Ebert-Stiftung Bonn	1290
Württembergische Landesbibliothek Stuttgart	1164
Universitätsbibliothek München	1102
Universitäts- und Stadtbibliothek Köln	1088

Tabelle 1: Einrichtungen mit den umfangreichsten Zeitungsbeständen

Es zeigt sich also einerseits, dass sich umfangreiche Bestände in einigen Einrichtungen konzentrieren. Andererseits belegt der extrem hohe Anteil an nur in einer Einrichtung nachgewiesenen Zeitungen, dass ein nationales Digitalisierungsprogramm nicht ohne die Berücksichtigung vieler, auch kleinerer Einrichtungen durchgeführt werden kann.

Da sich die Bestände einer einzigen Zeitung oftmals auf mehrere besitzende Einrichtungen verteilen, ist eine die gesamte Laufzeit eines Titels berücksichtigende Digitalisierung häufig nur als Kooperationsprojekt mehrerer Einrichtungen möglich. Die Daten der ZDB können – trotz noch zu beseitigender Unschärfen – den Ausgangspunkt für eine maschinelle Zuordnung von Teilen des gesamten Erscheinungsverlaufs zu bestandsführenden Bibliotheken bilden. Deshalb ist die ZDB als Steuerungsinstrument für eine verteilte Digitalisierungsinitiative gut geeignet. Die ZDB ist das substantiellste internationale und nationale Nachweisinstrument für Periodika, insbesondere für Zeitschriften und Zeitungen, allerdings ist sie WissenschaftlerInnen noch zu wenig bekannt. Auch wird sie von Archiven für den Nachweis ihrer reichen Bestände noch zu wenig genutzt. Die im Pilotprojekt

erreichte deutliche verbesserte Präsentation des ZDB-OPAC soll die Attraktivität dieser Datenbasis sichtbarer machen. Seit dem 7.7.2015 liegt eine Betaversion mit Visualisierungstools und zugleich mit Steuerungsfunktionen für Zeitungs- und Zeitschriftendigitalisierungen vor.[25] Diese ZDB-Verbesserung kommt auch bereits der laufenden Erschließung und Digitalisierung der Periodika des 18. Jahrhunderts im Rahmen des VD 18 zugute.

Verbesserung der Zeitungssuche und -präsentation

Die neue, intuitiv bedienbare Suchoberfläche des neuen ZDB-Katalogs integriert einen Zeitstrahl für Suchanfragen nach Zeiträumen. Auf der Suchergebnisseite sind nachträgliche Sucheinschränkungen über Facetten möglich. Auf diese Weise soll auch die Einschränkung auf die Zeitungssuche erfolgen. Von einer gesonderten Zeitungen-Sicht innerhalb der ZDB wurde abgesehen, weil es 1. komplexe Misch- und Übergangsformen zwischen den Medientypen Zeitschrift, Zeitung und verwandten Periodika gibt und weil 2. die neuen Funktionalitäten des ZDB-OPAC für alle Materialarten der ZDB nutzbar sein sollen. Diese weiteren Facetten unterstützen die spezielle Zeitungssuche:

- Frequenz (täglich, drei- bis fünfmal wöchentlich, zweimal wöchentlich, wöchentlich, vierzehntäglich, halbmonatlich, monatlich)
- Verbreitungsorte und damit assoziierte Karte für Verbreitungsorte.

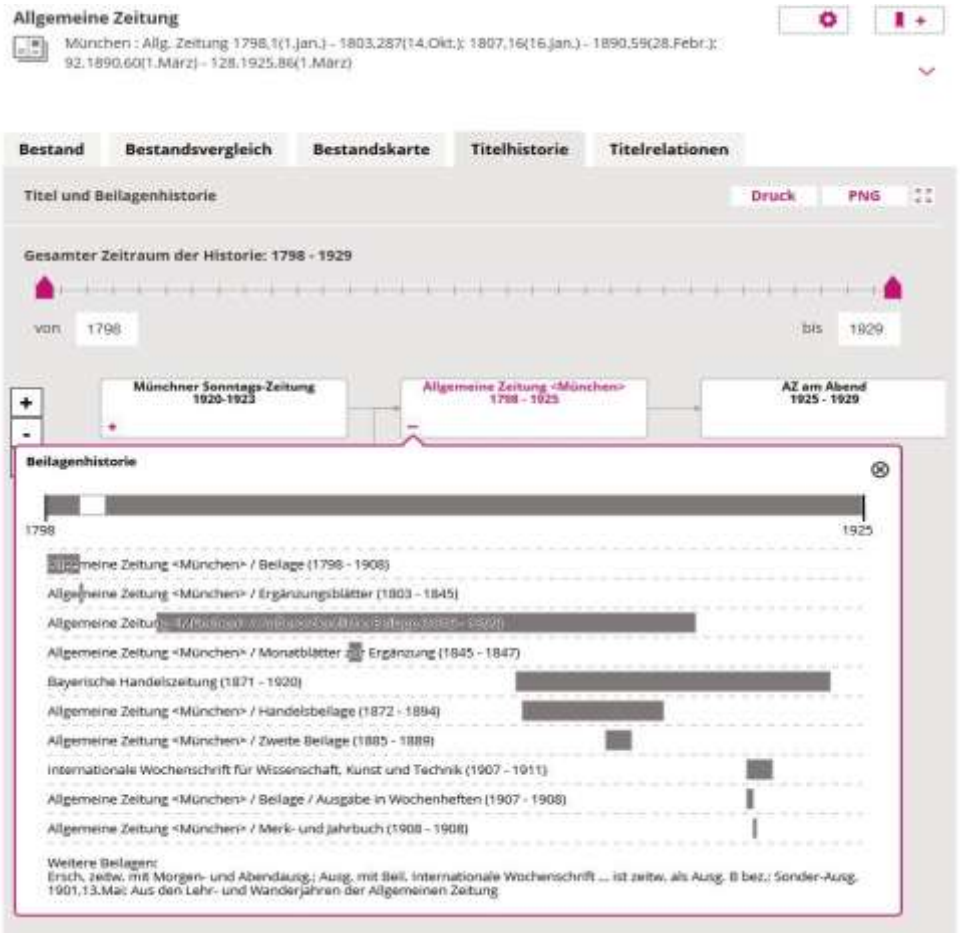

Abbildung 1: Titel- und Beilagenhistorie am Beispiel der Münchner Allgemeinen Zeitung

[25] Vgl. http://beta.zdb-opac.de/zdb/index.xhtml

(Cotta´sche Zeitung) in der ZDB-Beta-Version

In der „Titelhistorie" werden zeitliche Vorläufer und Nachfolger eines Zeitungstitels mit den Namensänderungen sowie entsprechenden Beilagen in chronologischer Reihung angezeigt.

Auf der Grundlage der „Titelrelationen" kann das gesamte relationale Umfeld eines ausgewählten Titels visualisiert werden (vgl. Abbildung 2). Durch diese navigierbare Sicht (mit Zoom, „Knoten" können expandiert oder minimiert, einzelne Titel in Vollansicht angesteuert werden) sind auch komplexe Titelzusammenhänge jetzt besser überschaubar. Neben Visualisierungen von Verlaufsformen und Titelzusammenhängen bietet die Betaversion des ZDB-Katalogs folgende grundsätzliche Neuerungen: Autosuggest zur Unterstützung von Suchanfragen, Breadcrumbtrail zur Anzeige und Abwahl von ausgewählten Filterbegriffen, Suche und Anzeige von originalschriftlichen Titeln, Bestandskarte (zur Beantwortung der Frage: welche Bibliotheken besitzen welche Anteile der Zeitung), Standortkarte (mit zusätzlichen Bibliotheksinformationen).

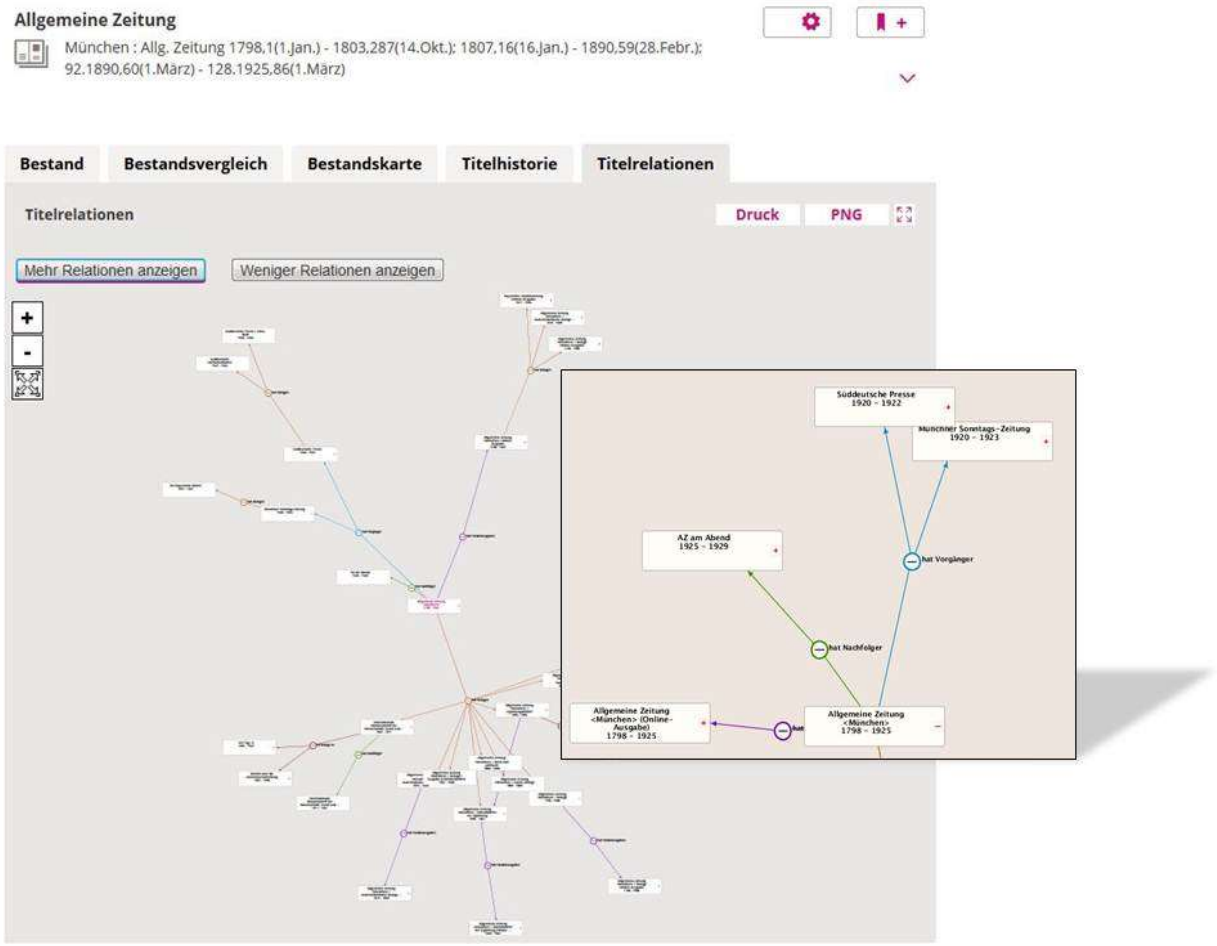

Abbildung 2: Auszug Titelrelationen am Beispiel der Münchner Allgemeinen Zeitung (Cotta´sche Zeitung) in der ZDB-Beta-Version

Funktionalitäten zur Unterstützung von Digitalisierungsprojekten

Zur Unterstützung von Digitalisierungsprojekten wird in der erweiterten Suche eine Checkbox „geplante Digitalisierung“ bereitgestellt. In der Detailsicht eines Titels wird ein „Bestandsvergleich" angeboten, der erstmals die verfügbaren Jahrgänge eines bestimmten

Titels in verschiedenen auswählbaren Einrichtungen übersichtlich visualisiert und damit die Steuerung von Digitalisierungsprojekten, insbesondere Vergleich und Auswahl geeigneter Originale als Digitalisierungsvorlagen, systematisch unterstützt. In Abbildung 3 wird die Verfügbarkeit der Cotta´schen Allgemeinen Zeitung in Bibliotheken angezeigt:

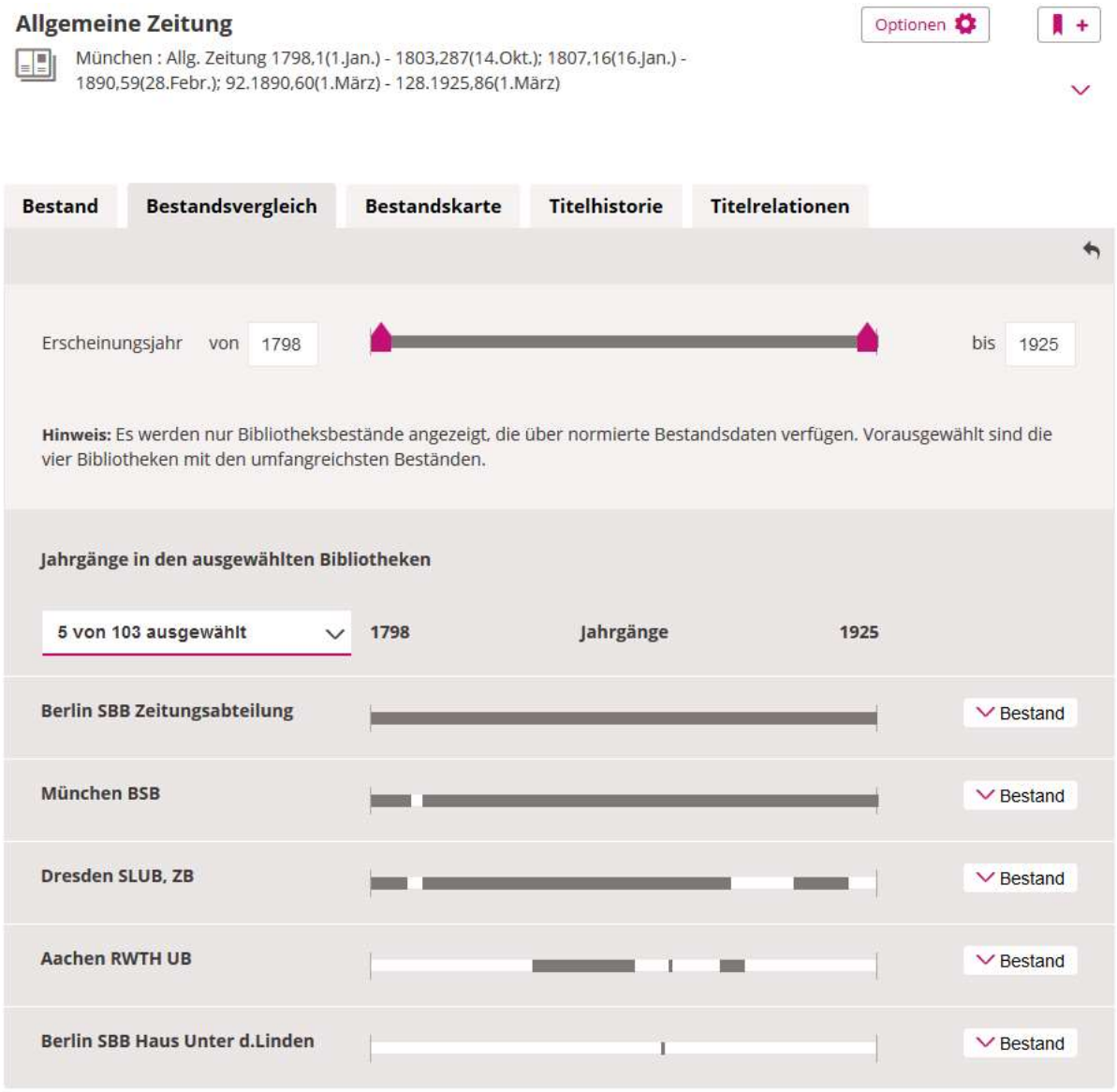

Abbildung 3: Bestandsvergleich am Beispiel der Münchner Allgemeinen Zeitung (Cotta´sche Zeitung) in der ZDB-Beta-Version

Geplante Aktivitäten nach Projektabschluss

Nach Projektabschluss waren noch diverse Optimierungen der Suchoberflächen geplant; der aktuelle Stand kann Kapitel 5 entnommen werden.

Alle genannten Arbeiten nach Projektabschluss wurden bzw. werden in Eigenleistung der Deutschen Nationalbibliothek in Kooperation mit der Staatsbibliothek zu Berlin durchgeführt.

Empfehlungen

Mit den dargestellten Entwicklungen der ZDB ist bereits vieles erreicht, das Potential für eine neue Qualität als Forschungsinstrument jedoch noch nicht ausgeschöpft. Vor allem sollte die

Datenbasis weiter ausgebaut und verbessert werden, um noch bessere Bestandskenntnisse zu gewinnen und daraus wiederum strategische und wissenschaftliche Erkenntnisse ableiten zu können. Deshalb sind diese weiteren Schritte zu empfehlen:

1. In der ZDB fehlen Nachweise insbesondere kommunaler Archive. Beispielhaft wurden mit Förderung der DFG Zeitungsbestände aus bayerischen Archiven und Bibliotheken in der ZDB nachgewiesen.[26] In anderen Ländern (z.B. Sachsen-Anhalt,[27] Sachsen) gibt es koordinierende Vorarbeiten. Es sollten strukturbildend Anreize gesetzt werden, um fehlende Nachweise insbesondere aus kommunalen Archiven und Bibliotheken in der ZDB zu ergänzen.
2. Die Verbreitungsorte von Zeitungen sind in der ZDB nicht vollständig erfasst. Da diese Informationen als Datenbasis für Visualisierungen benötigt werden, wird empfohlen, entsprechende Datenverbesserungen projektbasiert zu fördern.
3. Für die Zeitungsdigitalisierung ist die Nachnutzung von Vorarbeiten (Mikroverfilmung, Katalogisierung, Komplettierung) unerlässlich. Deshalb streben ZDB und das Mikrofilmarchiv der deutschsprachigen Presse e.V. (MFA) an, die ca. 12.000 Nachweise zu den etwa 42.000 Masterfilmrollen von Zeitungen und Zeitschriften aus der allegro-Datenbank des MFA in die ZDB zu überführen. Damit werden Parallelstrukturen zusammengeführt, die Bestandsnachweise in der ZDB nochmals deutlich verbessert, und das Mikrofilmarchiv kann die ZDB künftig als primäres Katalogisierungswerkzeug nutzen.
4. Nach einem zügigen Ausbau ist die ZDB die beste nationale Datenbasis für Periodika und speziell für die Zeitungsdigitalisierung und sollte deshalb mit einem künftigen nationalen Zeitungsportal funktional eng verbunden werden.

2.2 Medientypologische Weiterentwicklung des DFG-Viewers

Die kooperative Digitalisierung deutscher Drucke (VD17, VD18) erforderte die Entwicklung eines gemeinsamen Viewers zur Umsetzung eines gemeinsamen Mindeststandards und zur Einhaltung der Praxisregeln der DFG. Der sogenannte DFG-Viewer hat sich in der Praxis bewährt und wurde bzw. wird in Zusammenarbeit mit den Fachcommunities medientypologisch für die spezifischen Anforderungen an digitale Präsentationen von Handschriften, Nachlässen, Zeitungen und aktuell auch für Archivalien erweitert. Diese mit der Zeitungs-Fachcommunity abgestimmten 6 Arbeitspakete wurden innerhalb von 18 Projektmonaten an der SLUB Dresden (mit eigenem Abschlussbericht)[28] abgeschlossen:

1. generische Umsetzung der dreistufigen Kalendernavigation (Titel, Jahrgang, Ausgabe)

[26] In diesem DFG-Projekt wurden auch nachgewiesene Titel, zu denen bislang keine Exemplare mehr gefunden wurden, mit aufgenommen. Vgl. Kurzaufsatz zum Projekt (S. 67 ff.): http://staatsbibliothek-berlin.de/fileadmin/user_upload/zentrale_Seiten/ueber_uns/pdf/Bibliotheksmagazin/bibliotheksmagazin_0903.pdf und http://www.bayerische-landesbibliothek-online.de/zeitungen-amtsblaetter

[27] Vgl. Dorothea Sommer u. a., „Zeitungsdigitalisierung: Eine neue Herausforderung für die ULB Sachsen-Anhalt. Werkstattbericht aus der Pilotphase des DFG-Projekts ‚Digitalisierung historischer Zeitungen'", ABI-Technik: Zeitschrift für Automation, Bau und Technik im Archiv-, Bibliotheks- und Informationswesen 34, Nr. 2 (2014): 75–85. Vgl. Manfred Pankratz, Hans Bursian: Zeitungen in Sachsen-Anhalt, ein Nachweis. Halle (Saale): Universitäts- und Landesbibliothek Sachsen-Anhalt, 2008 (Schriften zum Bibliotheks- und Büchereiwesen in Sachsen-Anhalt, 91).

[28] Vgl. https://intranet.slub-dresden.de/display/DRIT/Medienspezifische+Weiterentwicklung+des+DFG-Viewers?preview=/29656740/50562790/BU%202228-16-1_Abschlussbericht.pdf

2. stufenloser Zoom für alle Zeitungsformate über OpenLayers
3. freie Bildpositionierung (Panning) über OpenLayers
4. verteilte Volltextsuche (SRU-Schnittstelle, ALTO-Format)
5. Überarbeitung der Formatdokumentationen für METS und MODS
6. Erstellung von Beispielen für den Demonstrator.

Der von der SLUB Dresden betriebene DFG-Viewer ist ein freier Webdienst, der ohne lokale Installation von jedem Interessenten verwendet werden kann. Der Quellcode wurde auf der Entwicklungsplattform GitHub[29] unter der Open-Source-Lizenz GPL3 veröffentlicht und kann ebenfalls frei nachgenutzt werden. Die kostenfreie Nachnutzung soll allen Einrichtungen, deren Bestände forschungsrelevant sind, Kooperationen mit Digitalisierungsvorhaben erleichtern. Die medientypologische Weiterentwicklung folgt dem Wunsch vieler Wissenschaftlerinnen und Wissenschaftler nach einer Option, über die sehr unterschiedlichen lokalen Präsentationsformen hinaus auch mit einem Medien und Institutionen übergreifenden, einheitlichen Viewer arbeiten zu können, um nicht ständig die Anzeigeform wechseln zu müssen.

Abbildung 4: DFG-Viewer mit Volltextsuche

[29] https://github.com/slub/dfg-viewer

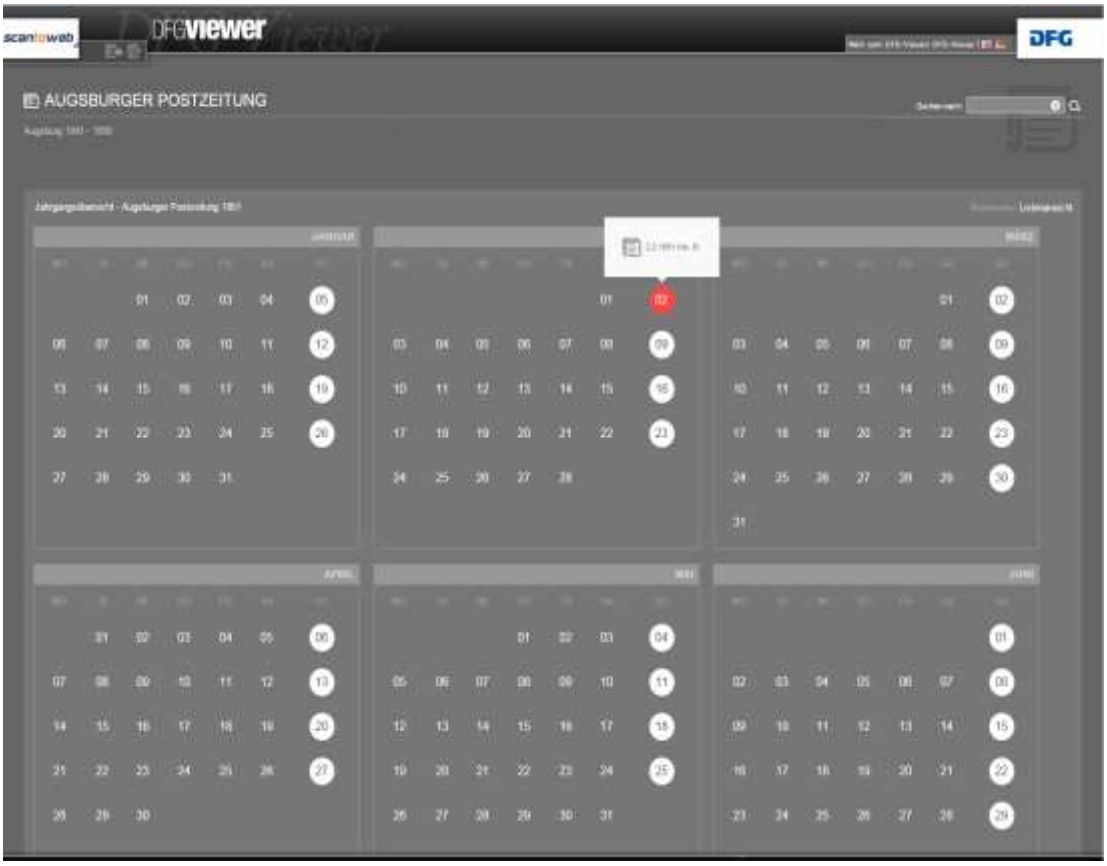

Abbildung 5: DFG-Viewer mit Kalendernavigation

Empfehlung

Der DFG-Viewer steht einem nationalen Zeitungsportal als einheitliche Präsentationsoption zur Verfügung und sollte mit dem Ziel einer nahtlosen Integration in die künftige Portalarchitektur weiter entwickelt werden.

3. Die Pilotprojekte: Erprobung von Verfahren und Werkzeugen, Qualitäts- und Kostenfaktoren im Vergleich

3.1 Arbeitsschwerpunkte der Partnerbibliotheken im Projekt, Mengengerüste

Die historischen Zeitungen stellten aufgrund des Umfangs, der Formate und der Vielfältigkeit neue Herausforderungen an bereits etablierte Digitalisierungsprozesse. Die Pilotpartner erprobten deshalb anhand einzelner Zeitungen exemplarisch unterschiedliche Verfahren und Werkzeuge. Im Folgenden werden zunächst die Anteile der Partner beschrieben, dann die Ergebnisse ausgewertet und untereinander abgestimmte Empfehlungen formuliert.

Überblick Kostenfaktoren Digitalisierung und Erschließung					
	SuUB Bremen	SLUB Dresden		ULB Sachsen-Anhalt	BSB München
Anzahl der Seiten	375.000	411.362	52.800	145.000	295.051 (601.051 mit Google)
Anzahl der Zeitungs-unternehmen[30]	500	6	1	1	2
Speicher	10 TB	14 TB		4,6 TB	19,21 TB
Gesamtkosten	453.007,88 €	224.538,44 €		209.226,00 €	462.755 €[31]
DFG-Finanzierung (ohne Programmpauschale)	252.171,00 €	100.149,00 €		187.684,00 €	286.902,00 €
Workflow					
Goobi		x	x		
ZEND					x
Visual Library	x			x	
Scannen					
Inhouse vom Original			x	x	
Inhouse von der Sekundärform/Film	x	x			
Dienstleister vom Original					x
Dienstleister von der Sekundarform/Film					x
Metadaten/ Grunderschließung					
bibliographische Beschreibung	x	x	x	x	x
Ausgabensegmentierung	x	x	x	x	x
kalendarische Indexierung	x	x	x	x	x
Tiefenerschließung					
Artikelebene			x	x	x
Normdatenverknüpfung			x		
Volltext					
OCR		x		x	x
OCR+Artikelsegmentierung					x

[30] Zeitungen ändern im Laufe ihres Erscheinungszeitraumes oftmals die Titel. Wir zählen hier die Zeitungsunternehmungen, nicht Zeitungstitel mit den jeweiligen Änderungen.
[31] Die Angaben beziehen sich auf die 295.051 neu produzierten Seiten.

OCR-Evaluierung				x	x
Schnittstellen					
Schnittstelle mindest: OAI (METS/MODS für Kalender)	x	x	x	x	x
Schnittstelle erweitert: SRU (ALTO für Volltexte)		x		x	x
Speicherung					
Speicherungskapazitäten	x	x	x	x	x
Langzeitarchivierung	x	x	x	x	x

Tabelle 2: Überblick Kostenfaktoren Digitalisierung und Erschließung

Die Staats- und Universitätsbibliothek Bremen (SuUB) digitalisierte ihren über Jahrzehnte gesammelten Bestand von Reproduktionen deutschsprachiger Zeitungen des 17. Jahrhunderts (ca. 500 Zeitungsunternehmen mit 750 Titeln und 375.000 Seiten). Auf dieser Basis übernahm sie die vollständige digitale Transformation des Mediums Zeitung für das gesamte 17. Jahrhundert: Ca. 80.000 Ausgaben wurden manuell mit Erscheinungsdaten und Berichtszeiträumen in beiden historischen Kalendersystemen erschlossen. Im ersten Quartal 2016 erfolgt die Integration der Nachweise in das Verzeichnis der im deutschen Sprachgebiet erschienenen Drucke des 17. Jahrhunderts (VD17), das damit systematisch um das Medium Zeitung ergänzt wird. Nach Vorabsprachen mit der Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG/GBV) in Göttingen erfolgt eine Schnittstellenanpassung, die Arbeiten werden in Eigenleistung durchgeführt.

Die SuUB Bremen hat im Pilotprojekt ferner die Entwicklung effizienter Verfahren zur Bestandslückenergänzung bei der Zeitungsdigitalisierung übernommen und in Zusammenarbeit mit dem Institut für Deutsche Presseforschung an der Universität Bremen die Workshops mit den Wissenschaftlerinnen und Wissenschaftlern organisiert, bei denen die Bedarfe der wissenschaftlichen Disziplinen erhoben und priorisiert wurden.

Abbildung 6: Zeitungen des Bremer Zeitungsprojekts

Die Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden (SLUB) hat neben der Federführung des Pilotprojekts, der medientypologischen Weiterentwicklung des DFG-Viewers, der Mitwirkung beim Test des Zeitungs-Moduls der Open-Source-Software Goobi-Production (separates Teilprojekt Berlin) die Erprobung kostengünstiger Verfahren bei der Massendigitalisierung anhand sechs verfilmter Zeitungen und die Digitalisierung einer Zeitung vom Original mit exemplarischer Erprobung bibliothekarischer Artikelerschließung einschließlich GND-Verknüpfung der Artikel-Autoren übernommen. Da die ausgewählte "Illustrirte Zeitung" (Leipzig, New York) in Teilen im Rahmen des BSB-Google-Projekts digitalisiert wurde[32], hat sich die SLUB, um Doppelungen zu vermeiden, bei dieser Zeitung auf Digitalisierungstests vom Original beschränkt und stattdessen die ähnlich umfangreiche „Leipziger Volkszeitung" als wichtiges Medium der Arbeiterbewegung des 20. Jahrhunderts ins Portfolio aufgenommen. Damit wurde der geplante Gesamtumfang von ca. 450.000 Seiten beibehalten. Hauptziel war die Ermittlung kostengünstiger Mengenverfahren. Qualitätsvergleiche zur Digitalisierung vom Original und Film, die Bewertung einer vertieften bibliographischen Erschließung am Beispiel der „Dresdner Abend-Zeitung" und rechtliche Klärungen zur Digitalisierung einer NS-Zeitung waren zusätzlich übernommene Aufgaben.

Abbildung 7: Drei Zeitungen digitalisiert vom Film, ein Farbdigitalisat vom Original (SLUB)

Die Staatsbibliothek zu Berlin (SBB) beteiligte sich mit der Entwicklung der Open-Source-Software Goobi zur Zeitungspräsentation in Zusammenarbeit mit der SLUB Dresden. Eine Dokumentation dieses Teilprojektes findet sich im Anhang. Die ursprünglich geplante Massendigitalisierung von vier Berliner Zeitungen nach Mikrofilmen konnte im Rahmen eines EFRE-Projekts der EU, die OCR-Umsetzung im Rahmen des Europeana Newspapers-Projekt realisiert werden. Die Erkenntnisse aus beiden Projekten sind in diesem Bericht berücksichtigt.

Die Universitäts- und Landesbibliothek Halle/S. (ULB) digitalisierte mit dem „Hallischen Tageblatt" eine überregional bedeutsame Zeitung des Pietismus, der im 17. und 18. Jahrhundert von Halle aus in die Welt ausstrahlte. Die in verschiedenen Formaten vorliegende Zeitung wurde vom Original digitalisiert und mit Texterkennungssoftware auf der Grundlage einer OCR-Testreihe bearbeitet. Es erfolgte eine vertiefte Strukturdatenerfassung

[32] Vgl. https://de.wikipedia.org/wiki/Illustrirte_Zeitung und Übersicht der Digitalisate unter: https://de.wikisource.org/wiki/Illustrirte_Zeitung.

bis auf die Artikelebene. Ein Hauptgewicht lag auf der Weiterentwicklung der Standards persistenter Fragment-basierter Adressierung und Referenzierung (URN Granular 2.0) in Kooperation mit der Deutschen Nationalbibliothek mit dem Ziel, künftig Artikel und Seiten persistent zu adressieren.


Wochenblatt
Beym Jahreswechsel.

Abbildung 8: Hallesches Tageblatt/Hallisches Wochenblatt

Die Bayerische Staatsbibliothek München (BSB) hat im Pilotprojekt zwei Zeitungen mit unterschiedlichen Schwerpunkten bearbeitet. Zum einen wurde die „Allgemeine Zeitung/Cotta'sche Zeitung" als wichtiges Leitmedium des 19. Jahrhunderts mit rund 598.000 Seiten bereitgestellt. Von dieser Gesamtmenge wurden mit DFG-Finanzierung 291.906 Seiten durch Dienstleister neu digitalisiert und OCR-erfasst. 306.000 Seiten wurden aus dem Google-Projekt durch die BSB in Eigenleistung strukturiert und bereitgestellt. Die OCR-Erfassung der neu zu produzierenden Seiten erfolgte hierbei vom Original und wurde ausführlich evaluiert. Zum zweiten wurde exemplarisch eine halb-automatische Artikelseparierung im Rahmen der OCR-Erfassung einer Frakturschrift des 20. Jahrhunderts durch einen Dienstleister erprobt, konkret bei der kurzlebigen, aber bedeutenden Wochenschrift „Illustrierter Sonntag/Der gerade Weg" (1929-1933) mit 3.145 Seiten.


Allgemeine Zeitung
Der gerade Weg
1932: Glück oder Unglück?
Sonntag
Riesenbau auf der Theresienwiese

Abbildung 9: Allgemeine Zeitung/Cotta´sche Zeitung und Illustrierter Sonntag/Der gerade Weg

Bibliothek	Titel	Jahre	Ausgaben	Seiten
SuUB Bremen	ca. 500 deutschsprachige Zeitungsunternehmen des 17. Jhdts.	1605-1700	ca. 80.000	ca. 375.000
SLUB Dresden	Dresdner Abend-Zeitung Leipziger Volkszeitung Sächsische Arbeiterzeitung / Dresdner Volkszeitung The Dresden Daily Leipziger Jüdische Zeitung/ Allgemeine Jüdische Familienblatt Leipziger Jüdische Wochenschau Der Freiheitskampf	1817-1857 1894-1933 1890-1933 1906-1910 1922-1933 1928-1933 1930-1945	8.127 11.526 12.730 1.341 584 187 4.866 39.361	52.800 155.575 167.200 5.660 4.701 1.226 77.000 Summe 464.162
ULB Sachsen-Anhalt	Hallesches Tageblatt	1799-1892	14.081	145.454
BSB München	Allgemeine Zeitung / Cotta'sche Zeitung Illustrierter Sonntag / Der Gerade Weg	1798-1929 1929-1933	65.229 216 65.445	597.906 3.145 Summe 601.051

Tabelle 3: Titel und Mengengerüst

3.2 Workflow

Die in diesem Abschnitt benannten Schritte wurden im Rahmen der Gutachterempfehlungen zur Dokumentation der Projektergebnisse in einem web-basierten Auftritt strukturiert zusammengetragen und als Arbeitshilfe ebenso publiziert, wie weitere Handreichungen, die von dieser Webadresse aus erreichbar sind:

http://www.zeitschriftendatenbank.de/zeitungsdigitalisierung

Am Anfang jeder Vorbereitung steht – oder sollte stehen – die Recherche in der Zeitschriftendatenbank (ZDB): Wie verteilen sich die Bestände auf die Häuser, wo ist der umfangreichste Bestand vorhanden, ist ein Bestand zur Digitalisierung vorgemerkt, gibt es Verfilmungen? Es folgt – ggf. nach weiteren Recherchen, solange die ZDB noch nicht alle Bestände erfasst – die Prüfung und Entscheidung für die Vorlagen im Original oder als Film (s.a. die entsprechende Handreichung). Damit verbunden werden sollte nach Möglichkeit die Prüfung auf Vollständigkeit, das Sortieren in der richtigen Reihenfolge, die Kollationierung und ggf. Bestandslückenergänzungen. Der Nachweis der beabsichtigten Digitalisierung in der ZDB erfolgt je nach Software an unterschiedlichen Stellen im Prozess, jedoch so rechtzeitig, dass Doppeldigitalisierungen künftig leichter vermieden werden (s.a. die entsprechende Handreichung). Bei der Zusammenarbeit mit einem Dienstleister sind die Vorbereitungen besonders aufwändig und müssen bei den Kosten berücksichtigt werden (auch hier liegt eine gesonderte Handreichung vor). Der nächste Schritt ist das Scannen. Dabei werden die Images erstellt, begleitet von der Qualitätssicherung mit möglichem Nachscannen bzw. der Weiterverarbeitung der Images. Dem Scanprozess schließt sich die Strukturierung und in einem weiteren Schritt die Erschließung an. Dies kann automatisch, teilautomatisch oder manuell erfolgen. Eine Volltexterkennung mittels OCR setzt darauf auf; OCR und Erschließung können jedoch auch in einem Schritt vorgenommen werden. Die Bereitstellung in den Digitalen Sammlungen und damit verbunden die Erzeugung einer URN

ist der vorletzte Schritt, bevor die Daten dann einerseits zur Langzeitarchivierung und andererseits zur Einbindung in nationale und internationale Portale übergeben werden. Qualitätskontrollen finden an verschiedenen Stellen im Digitalisierungsprozess statt. Die im Pilotprojekt erprobten Verfahren verdeutlichten, dass es bei der Zeitungsdigitalisierung in allen Schritten des Workflows spezifische Qualitäts- und Kostenfaktoren gibt, die das jeweilige Ziel und Ergebnis des Projekts maßgeblich beeinflussen.

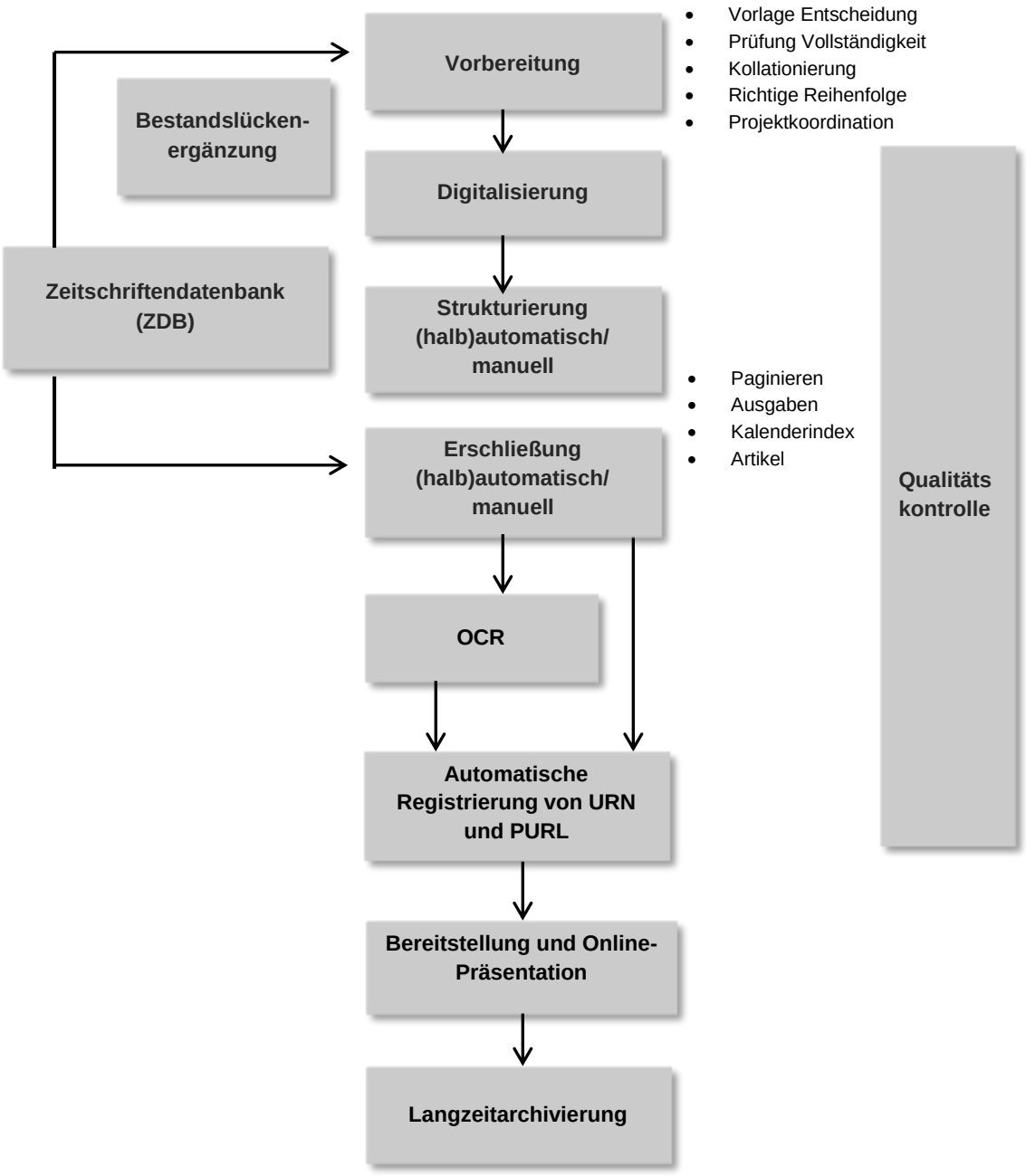

Abbildung 10: Workflow Zeitungsdigitalisierung

Als Projektergebnis kann zusammenfassend festgestellt werden, dass die in den verschiedenen Teilprojekten eingesetzten Workflows und die sie unterstützenden

Softwarelösungen auf eine jahrelange Entwicklungs- und Umsetzungspraxis zurückblicken können und einen hohen Reifegrad erreicht haben. Sie sind in der Praxis der Häuser fest verankert. Ein näherer Vergleich zeigt grundsätzlich ähnliche Abläufe. Vor diesem Hintergrund empfiehlt das Projekt, die Festlegungen zum Workflow und der jeweiligen unterstützenden Software dem Projektpartner zu überlassen, solange die Anforderungen der fortgeschriebenen DFG-Praxisregeln eingehalten werden. Auf die Nutzung des entsprechenden Unterstützungswerkzeugs wird verwiesen.[33]

3.3. Auswahl, Vorbereitung der Vorlagen, Scanverfahren

	SUuB Bremen	SLUB Dresden	ULB Sachsen-Anhalt	BSB München
Original		x	x	x
Mikrofilm		x		
Reproform	x			
Scanner	Scamax Durchlaufscanner 403cd color duplex Aufsichtsscanner i2s Copybook ONYX RGB	Auflichtscanner Proserv Scanntech 602i, anteilig Qidenus Scan Roboter Zeutschel OM 1600	Zeutschel OS 8000 A2-Auflichtscanner mit Canon-Camera EOS 5D Mark II sowie Glasplatte und Buchwippe A0-Auflicht-Flachbettscanner Scann TECH 400i der Firma ProServ für Folio- und Quartformate	Zeutschel OS 14000; Buchwippe OT180
Bildqualität	Graustufe, max. 300ppi	Farbe, max. 300ppi[34] bitonal, max. 300 ppi	Farbe, max. 300ppi	Graustufe, max. 300 ppi
Inhouse/Dienstleister	Inhouse	Inhouse	Inhouse	Dienstleister

Tabelle 4: Scanverfahren im Überblick

Die Pilotpartner erprobten Scanverfahren von drei verschiedenen Vorlagen. Die SuUB Bremen wählte die über Jahrzehnte gesammelten und bibliographisch erschlossenen Papierreproduktionen der Mikrofilme von den weltweit verstreuten Zeitungsoriginalen des 17. Jahrhunderts als Vorlagen. Dies gründete auf der langen Genese der im Zuge der Erforschung der Zeitungen des 17. Jahrhunderts zusammengetragenen Sammlung. Die Papierreproduktionen waren einerseits im Vergleich zu den älteren Mikrofilmen hinsichtlich der Qualität, aber auch des für das Scannen und die Nachbearbeitung benötigten Zeitaufwands wirtschaftlicher. Andererseits war die alphanumerische Sortierung der Reproduktionen das Ergebnis einer intensiven Forschung, infolge derer die Einzelzeitungen titelspezifisch geordnet wurden. Mit diesen Reproduktionen ließen sich alle Zeitungen des 17. Jahrhunderts vergleichsweise schnell scannen.
Die SLUB Dresden, die ULB Sachsen-Anhalt und die BSB München scannten Zeitungen vom Original. In der ULB Sachsen-Anhalt und SLUB Dresden erfolgte die Digitalisierung der Originale in Farbe. Die BSB München wählte Graustufen. Bei sechs Zeitungen nutzte die

[33] http://www.zeitschriftendatenbank.de/zeitungsdigitalisierung
[34] Zur Verwendung von ppi (Pixel-per-Inch) vgl. http://www.andrewdaceyphotography.com/articles/dpi/

SLUB Dresden Mikrofilme als Vorlagen, die von Originalzeitungen aus Gründen der Bestandserhaltung angefertigt wurden. Da bei diesen Filmen über Graustufen keine Verbesserung der Lesbarkeit und der OCR-Ergebnisse erreichbar waren, wurden die Digitalisate der Filme mit Rücksicht auf den Speicherplatz bitonal gespeichert.
Im Ergebnis stellen die Kooperationspartner fest: Die Wiedergabe in Farbe bietet ein authentisches Bild der originalen Zeitungsseite, also hohe Faksimilequalität. Farbe und Graustufe verbessern in der Regel die Lesbarkeit, wobei jedoch letztlich allein die Qualität der Originalvorlage selbst über die Güte der Reproduktion und Lesbarkeit entscheidet. Der Scanaufwand bei Farbe oder Graustufe ist jeweils gleich. Bei Farbscans erhöht sich der Speicherplatzbedarf um ca. den Faktor 3.
Im Hinblick auf eine OCR-Erschließung hatte die ULB Sachsen-Anhalt zu Beginn der Pilotphase umfangreiche Tests zur Auflösung der Scans durchgeführt mit dem Ergebnis, dass 300 (statt 400) ppi als sehr gut und ausreichend anzusehen sind, was zudem zu einer Ersparnis beim Speicherbedarf führt.

Empfehlungen

Vorlagenwahl

Vom Original zu scannen ermöglicht eine Reproduktion in bester Faksimilequalität mit einem optimalen Gesamteindruck des Originals und bietet zudem sehr gute Voraussetzungen für eine spätere OCR. Es ist die zeitaufwändigste Variante, sowohl inhouse als auch durch einen Dienstleister. Insbesondere unikale und seltene Zeitungen sowie Zeitungen von besonderem kulturhistorischem Wert, z.B. mit wichtigen Bildanteilen, sollten vom Original in Farbe oder in Graustufe gescannt werden.
Von Mikrofilmen zu scannen ermöglicht es, bereits sicherungsverfilmte fragile Originale zu schonen und die mit einer Verfilmung durchgeführten Vorleistungen (Verzeichnung, ggf. Lückenergänzung) zu nutzen. Von Filmen kann schneller und preisgünstiger als von Originalen eine Massendigitalisierung erfolgen. Die Qualität der Filme ist allerdings vor einer Digitalisierung zu prüfen (s.a. die entsprechende Handreichung zu diesem Thema).
Digitalisierungen von Filmen sind dann uneingeschränkt zu empfehlen, wenn die Filme in guter Qualität, d.h. mit Qualitätskontrollen entstanden sind, und insbesondere dann, wenn bereits Lückenergänzungen vorgenommen wurden.
Von Reproformen sollte nur in begründeten Ausnahmen digitalisiert werden, weil es zu Qualitätsverlusten kommen kann, da die bereits reproduzierte Form die Auflösung begrenzt. Eine gut begründete Ausnahme ist die Bremer Sammlung, die über Jahrzehnte von weltweit verstreuten Originalen angelegt worden ist. Eine andere Ausnahme wäre dann gegeben, wenn ein Original nicht mehr existiert oder sowohl Original als auch Mikrofilm zum Zeitpunkt der Digitalisierung eine schlechtere Qualität aufweisen als eine zu einem früheren Zeitpunkt entstandene Reproform.

Scanparameter

Bei der Massenverarbeitung von Zeitungsfilmen wurden im Rahmen des Pilotprojekts und des Berliner EFRE-Projekts Graustufenscans angefertigt, für den Arbeitsschritt OCR wurden die Scans binarisiert, da die OCR-Qualität nur rund 1% geringer ausfiel als bei Graustufe

und erheblich Speicherplatz eingespart werden konnte. Die Binarisierung erfolgte dabei mit der adaptiven „Gatos"-Methode, die aktuell die besten Ergebnisse für die OCR ermöglicht.[35]
Für die Anzeige im Web sollten gleichwohl die qualitativ besten Scans, überwiegend Graustufen, mit der maximalsten Auflösung verwendet werden.
Wann immer aus Kostengründen sinnvoll und möglich empfehlen die Partner als Scanparameter eine Digitalisierung vom Original in Farbe (RGB), mindestens jedoch in Graustufen und diese jeweils mit 300 ppi.[36]

Dienstleister oder Inhouse?

Die Zusammenarbeit mit einem Dienstleister, d.h. das Outsourcing der Imageproduktion, empfiehlt sich, wenn ein zuständiges öffentliches Digitalisierungszentrum nicht zur Verfügung steht, die Anschaffung eines technischen Equipments nicht nachhaltig genutzt und wirtschaftlich betrieben werden kann bzw. projektspezifische Anforderungen (z.B. Imageproduktion, OCR-Erfassung, OCR mit Artikelseparierung) inhouse nicht umsetzbar sind oder sehr aufwendig werden. Allerdings dürfen keine konservatorischen Gründe gegen eine externe Digitalisierung sprechen. Auch ist zu berücksichtigen, dass genügend personelle Ressourcen für die Durchführung des erforderlichen Vergabeverfahrens und für die Dienstleisterbetreuung zur Verfügung stehen müssen. Insbesondere die Kontrolle laufender Datenlieferungen (d.h. Umsetzung der Arbeiten gemäß Leistungsbeschreibung, Zeitpläne, Validität und Qualitätskontrolle der Daten), das Monitoring des Reklamationsworkflows sowie die präzise zu formulierenden Freigaben zur Datenlöschung sind hier zu nennen. Werden im Rahmen des Ausschreibungsverfahrens Testdaten erbeten, müssen auch hierfür, neben den organisatorischen und rechtlichen Aufgaben, genügend Ressourcen für vergleichende Auswertungen zur Verfügung stehen. Bei der Bayerischen Staatsbibliothek erfolgt bei der Lieferung größerer Datenmengen der Datentransfer durch skiptgesteuerten Zugriff auf den „Datenbereitstellungsserver" des Dienstleisters über eine Netzverbindung. Hierzu sind bei Projektbeginn entsprechende technische Arbeiten sowohl beim Auftraggeber als auch beim Dienstleister erforderlich (s. hierzu auch die entsprechende Handreichung).

3.4. Bestandslückenergänzung

Die SuUB Bremen erarbeitete im Pilotprojekt effiziente Verfahren zur Bestandslückenergänzung bei der Zeitungsdigitalisierung. Mit der verbesserten Nachweisqualität von Zeitungen in der Zeitschriftendatenbank (ZDB) wird die Ergänzung von Lücken zukünftig einfacher werden. Zu empfehlen ist, vor Beginn der Digitalisierung die jeweiligen Zeitungsausgaben auf Vollständigkeit zu prüfen, so dass Bestandslücken identifiziert werden und eine Aufwandsschätzung für eine digitale Ergänzung erfolgen kann.
Das Verfahren zur Beschaffung/Ergänzung von Bestandslücken sollte sich an etablierten Geschäftsgängen bspw. aus der Fernleihe orientieren. Der Aufwand und die jeweilige Ausgestaltung der Bestandslückenergänzung (Originalausgaben, Verfilmungen, Papierreproduktionen, Digitalisate) sind abhängig von den Zielen des Digitalisierungsprojekts, insbesondere hinsichtlich der automatisierten Weiterverarbeitung. Die konkreten Nutzerinteressen, die physische Beschaffenheit, Seltenheitswert und

[35] Vgl. http://utopia.duth.gr/~ipratika/DIBCO2013/
[36] Zur wissenschaftlichen Diskussion siehe auch: Alexander Rindfleisch, „Stand und Perspektive der Zeitungsdigitalisierung im internationalen Vergleich", 5. Mai 2010, http://edoc.hu-berlin.de/docviews/abstract.php?id=30694, S. 38 ff.

Beschaffbarkeit der Vorlagen sind zu berücksichtigen. Grundsätzlich sollte die Vollständigkeit des digitalen Exemplars auf Jahrgangs-, Ausgaben- und Seitenebene das Ziel der Bestandslückenergänzung sein. Im Idealfall enthält das Digitalisat im Ergebnis jede Seite, deren Existenz bspw. durch durchgehende Seiten- oder Ausgabenzählung bekannt ist oder angenommen werden darf. Doch ist insbesondere beim Vergleich mehrerer vorhandener Exemplare der Aufwand zu berücksichtigen, der nötig wäre, das vollständigste aller Exemplare zu ermitteln. Zu empfehlen ist deshalb ein kontrolliert pragmatisches Vorgehen im Sinne einer stichprobenhaften Überprüfung.

Aus Gründen der Nutzerfreundlichkeit sollte eine über mehrere Institutionen verteilte Darstellung von Digitalisaten einer Zeitung vermieden werden. Empfohlen wird, dass eine Einrichtung alle Ausgaben einer Zeitung (also den gesamten Zeitungsverlauf einschl. aller zugehörigen Titeländerungen) digital zusammenführt und hostet und den Nutzern so den vollständigen Zugriff auf die jeweilige Zeitungsunternehmung ermöglicht. Die Aufgabe der digitalen Präsentation und Speicherung sollte also die Institution mit dem jeweils umfangreichsten Bestand zu einer Zeitungsunternehmung bzw. ein dafür ggf. zuständiges Digitalisierungszentrum übernehmen. Bei der Zusammenarbeit mehrerer Institutionen ist der Gebrauch standardisierter, offener Schnittstellen wichtig.[37] Zudem sollten einheitliche rechtliche Nutzungsbestimmungen gelten und möglichst keine Einschränkungen bei der Nutzung der Digitalisate erklärt werden.

Die Ergänzung von Bestandslücken kann an verschiedenen Stellen des Workflows integriert werden. Bei der Vorbereitung des Projekts sollte nach Möglichkeit eine stichprobenhafte Prüfung auf Vollständigkeit (Jahrgänge/Ausgaben) erfolgen. Mithilfe der Bestandsnachweise in der ZDB sollte der vollständigste Bestand einer Zeitung ermittelt werden, um ggf. fehlende Jahrgänge/Ausgaben zu ergänzen. Zu berücksichtigen sind teils hohe Aufwände bei der Erfassung von Streubeständen. Verzichtet man auf eine Vollständigkeitsprüfung in der Vorbereitungsphase, können während der Bearbeitungsprozesse der Qualitätssicherung des Scanprozesses, der Strukturierung und der Erschließung Bestandslücken erfasst und dokumentiert werden. Die Veröffentlichung der Digitalisate bei der Bereitstellung in den Digitalen Sammlungen bietet zusätzlich die Möglichkeit, Nutzer (Bibliotheken) aktiv in die Bestandslückenergänzung miteinzubeziehen.

Zusammenfassend ist festzustellen, dass die Bestandslückenergänzung im Kontext der Zeitungsdigitalisierung ein zeit- und arbeitsintensives Vorhaben ist. „Vollständigkeit“ im Sinne des Vorhandenseins eines lesbaren Bestands möglichst ohne Lücken auf Jahrgangs-, Ausgaben- und Seitenebene ist nur bedingt erreichbar. Insbesondere beim Vergleich mehrerer vorhandener Exemplare ist stets der Aufwand zu berücksichtigen. Es erscheint angezeigt, gerade bei einer größeren Anzahl zu digitalisierender Zeitungstitel ein pragmatisches Verfahren der Vervollständigung jedes Titels anzustreben.

3.5 Grunderschließung, Tiefenerschließung, OCR, URN-Granular

Das Pilotprojekt hat verschiedene Erschließungstiefen erprobt, um zum einen strukturbildend Grundstandards durchzusetzen (Strukturdatenset des DFG-Viewers,

[37] So bietet z.B. das (auch im EU-Projekt Europeana Newspapers verwendete) International Image Interoperability Framework (IIIF, http://iiif.io/) technische Schnittstellen, um an verschiedenen Orten gehostete Daten in einer einheitlichen gemeinsamen Präsentation zusammenzuführen.

persistente Adressierung jeder Einzelseite einer Zeitung) und zum anderen best practice-Modelle einer modularen Erschließung von Zeitungen für wissenschaftliche Anwendungen mit den jeweiligen Kosten zu beschreiben. Aufgrund der Erfahrungen in der Projektphase und angesichts der entstehenden großen Datenmengen sehen die Partner die Durchführung einer OCR als entscheidendes Modul bereits für den Grundstandard an - sofern die Qualität der Vorlagedaten dies zulässt -, auch wenn der Empfehlungsstand der DFG hier derzeit nur auf die Bilderstellung abhebt.

		Vorgehen	Anwendungen
Grund-Standard	Stufe 1	Imagedigitalisierung mit Strukturdatengenerierung gem. DFG-Viewer- Strukturdatensets mit manueller oder halbautomatischer Ausgabentrennung	chronologische Anzeige und Suchfunktion (Jahre, Monate, Tage)
	Stufe 2	OCR-Volltextgenerierung	Stichwort- und Phrasensuche, Textanalyse, Text Mining
erweiterter Standard I	Stufe 3a	Artikelseparierung automatisch mit manuellen Korrekturen der Artikelstruktur	Exakte Recherche auf Artikelebene, verbesserte Trefferrelevanz durch intellektuell geprüfte Strukturerkennung
	Stufe 3b	Artikelseparierung manuell mit Eingabe der Titelüberschriften der einzelnen Artikel	Exakte Recherche auf Artikelebene, genaue Suchergebnisse durch intellektuell korrigierte Überschriften
erweiterter Standard II	Stufe 4	Normdatenverknüpfung (Personen, Werke u.a.)	Auflösung von Pseudonymen, bio-bibliographische Erfassung wichtiger Beiträger; ermöglicht Verlinkung forschungsrelevanter Beiträge (etwa von Zeitungsrezensionen mit dem rezensierten Werk)
	Stufe 5	vertiefte sachliche Erschließung	fachliche und medientypologische Kontextualisierung und Verlinkung, Weiterverarbeitung
	Stufe 6	Bilderschließung, Bilderkennung Fachliche Bildindexierung	Illustrierte Zeitungen und Journale, Direktzugriff auf Bildurheber und Bildinhalte, Verlinkung, Weiterverarbeitung

Tabelle 5: Stufenmodell Erschließung

Die Auswahl der modularen Erschließungstiefen richtet sich nach den wissenschaftlichen Erkenntnisinteressen, die über Ziele und Methoden der Zeitungsdigitalisierung entscheiden. Das Stufenmodell (Tabelle 5) enthält sechs Stufen einer Grund- und Tiefenerschließung, denen unterschiedliche Anwendungsszenarien und Recherchebedarfe zugeordnet werden.
Erprobt wurden in den Pilotprojekten die Erschließungsstufen 1-4. Die Stufen 5 und 6 sind im Rahmen dieser Pilotphase nicht getestet worden, können aber im Rahmen einer Hauptphase herangezogen werden.

Grund- und Tiefenerschließung

Die Grunderschließung umfasst entsprechend dem Strukturdatenset des DFG-Viewers die Ausgabentrennung in Jahrgänge und Hefte, den Kalenderindex (Jahr, Monat, Tag) sowie das Paginieren, sofern die Vorlage es erfordert. Die Ausgabentrennung erfolgte in Bremen und Dresden manuell, in Halle und München halbautomatisch. Die Kalenderindexierung

erfolgte halb-automatisch in Dresden und Halle, manuell in München (u.a. wegen der oft wechselnden Erscheinungsverläufe langlebiger Zeitungsunternehmen sowie unregelmäßig erscheinender Beilagen) und in Bremen (hier wegen der Parallelität von gregorianischem und julianischem Kalender im 17. Jahrhundert, der Verwendung abweichender Monatsbezeichnungen sowie der gemischten Verwendung verschiedener Schrifttypen der Frakturfamilie innerhalb einer Ausgabe, die eine maschinelle Auswertung erschweren).

	SuUB Bremen	SLUB Dresden	ULB Sachsen-Anhalt	BSB München
Anzahl der Ausgaben	80.000	39.361	14.081	65.445
Anzahl der Seiten	375.000	464.162	145.454	601.051
Anzahl der Zeitungsunternehmen	500	7	1	2
Ausgabentrennung	x	x	x	x
Verfahren	manuell	manuell	halbautomatisch	halbautomatisch
Kalenderindexierung	x	x	x	x
Verfahren	manuell	halbautomatisch	halbautomatisch	manuell

Tabelle 6: Übersicht Grunderschließung, Stufe 1

Von den Pilotpartnern wurden unterschiedliche Tiefenerschließungen getestet. In Dresden und Halle wurde jeweils eine Zeitung auf Artikelebene erschlossen. München testete die Erschließung der Artikel in einem halbautomatischen Verfahren, welches mit einer OCR kombiniert wurde (s.u.). Aufgrund der hohen Anzahl von 80.000 manuell zu erschließenden Ausgaben war eine Tiefenerschließung in der SuUB Bremen nicht möglich. In Dresden wurden Namen von Autoren und Rezensenten mit der Normdatenbank (GND) verknüpft und dadurch Pseudonyme aufgelöst; das intellektuelle Verfahren ist zeitaufwendig und teuer. Es sollte bei entsprechendem wissenschaftlichem Bedarf künftig einer mit OCR kombinierten halb-automatischen Artikelerschließung nachfolgen und im Zuge der Eigennamenerkennung (Named Entity Recognition) in der Zukunft soweit als möglich automatisiert werden.[38]

	SLUB Dresden	ULB Sachsen-Anhalt	BSB München
Name der Zeitung	Dresdner Abend-Zeitung	Hallesches Tageblatt	Illustrierter Sonntag / Der Gerade Weg
Anzahl der Ausgaben	8.127	14.081	216
Anzahl der Artikel		191.186	12.882
Verfahren	manuell	manuell	halbautomatisch
Bearbeitungszeit für Tiefenerschließung (ggf. Artikelebene, Normdatenverknüpfung)	2.032 h 15min/Ausgabe	2.149 h 9,2min/Ausgabe	Keine Angabe möglich (Dienstleister)

[38] Eine GND-Ermittlung und -Verknüpfung kostete am Bespiel der „Dresdner Abend-Zeitung“ 10,36 EUR pro Verknüpfung. Es wurden beispielhaft 239 Verknüpfungen gesetzt.

Artikelebene	x	x	x
Artikelebene + OCR		x	x
Normdatenverknüpfung	x		

Tabelle 7: Übersicht Tiefenerschließung

Empfehlungen

Empfohlen wird eine halb-automatische Erschließung insbesondere bei standardisierten, gleichförmigen Titelverläufen. Eine manuelle kalendarische Erschließung auf Ausgabenebene empfiehlt sich bei oft wechselnden Erscheinungsverläufen langlebiger Zeitungsunternehmen, bei uneinheitlichem, oft wechselndem Layout, einer größeren Zahl an unregelmäßig erscheinenden Beilagen und bei uneinheitlicher, maschinell nicht auswertbarer Schrifttype.

Eine gesonderte Erschließung von regelmäßig erscheinenden Beilagen erfordert einen enormen Erschließungsaufwand (am Beispiel der „Augsburger Allgemeinen“ hätte dies den Personalaufwand verdoppelt, ohne dass ein nennenswerter Recherchevorteil entstehen würde) und ist nur zu empfehlen, wenn der besondere Quellenwert der Beilagen diesen Aufwand rechtfertigt.

OCR

In den Pilotprojekten wurden rund 448.000 Seiten vom Original und kleinere Mengen vom Film mit OCR bearbeitet. In Halle und München sind vorher umfangreiche OCR-Tests[39], in Dresden für den kleinen Anteil an Antiqua-Zeitungen weitere Tests durchgeführt worden. Im Folgenden werden kurz die bisherigen Benchmarks nach internationalem Forschungsstand (OCR-Methodik und -Qualitätsmessung), dann die Ergebnisse der Pilotprojekte vorgestellt. Berücksichtigt werden auch die Ergebnisse aus dem Europeana Newspapers-Projekt, bei dem rund 12 Mio. Seiten Volltexte erzeugt wurden.

Im Sinne einer Vergleichbarkeit richtet sich die Methodik zur Bestimmung der OCR-Qualität nach den in der wissenschaftlichen Community (International Association for Pattern Recognition, IAPR) etablierten Standards (gemäß Rice 1996[40]). Speziell zu den OCR-Qualitätsmessungen beim Medientyp Zeitung haben Holley 2009[41], Tanner 2009[42] und Tanner 2015[43] publiziert. Danach entsprechen 97% Zeichengenauigkeit in etwa einer Wortgenauigkeit von 80%, mit der wiederum ca. 98% der gesuchten Inhalte gefunden

[39] Vgl. die als Anlagen beigefügten Berichte, die auf der Grundlage der Studie „Volltext via OCR. Möglichkeiten und Grenzen“ von Maria Federbusch / Christian Polzin (Berlin 2013) entstanden; ferner: Sommer, Dorothea, Heiligenhaus, Kay, Pankratz, Manfred, Wippermann, Carola: Zeitungsdigitalisierung: eine neue Herausforderung für die ULB Sachsen-Anhalt: Werkstattbericht aus der Pilotphase des DFG-Projekts Digitalisierung historischer Zeitungen. *ABI Technik.* 34 (2014) 2, S. 75-85; Maria Wernersson, Evaluation von automatisch erzeugten OCR-Daten am Beispiel der Allgemeinen Zeitung, in: ABI Technik 35 (2015), S. 23–35.

[40] Stephen V. Rice (1996): Measuring the Accuracy of Page-Reading Systems, UNLV Dissertation, http://www.cs.olemiss.edu/~rice/rice-dissertation.pdf

[41] Holley, Rose (2009): How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs, D-Lib Magazine 15(3/4), March/April 2009, http://www.dlib.org/dlib/march09/holley/03holley.html

[42] Tanner, Simon, Trevor Muñoz, and Pich Hemy Ros (2009): Measuring Mass Text Digitization Quality and Usefulness: Lessons Learned from Assessing the OCR Accuracy of the British Library's 19th Century Online Newspaper Archive, D-Lib Magazine 15(7/8), July/August 2009, http://www.dlib.org/dlib/july09/munoz/07munoz.html

[43] Simon Tanner (2015): "OCR Accuracy Example", http://simon-tanner.blogspot.de/2015/06/text-capture-and-optical-character.html

werden können (Tanner 2009). Unterschieden wird dabei zwischen Zeichengenauigkeit (character accuracy) und Wortgenauigkeit (word accuracy) bzw. Genauigkeit signifikanter Wörter (significant word accuracy) unter Ausschluss von Stoppwörtern wie „und", „aber" etc.[44]

OCR-Ergebnisse Halle

In der ULB Sachsen-Anhalt kam zum Testzeitpunkt die Software ABBYY FineReader SDK in der Version 10 für Frakturschriften zum Einsatz. Für den OCR-Test wurden die Digitalisate von acht Bänden der Formate Oktav, Quart, Folio (in den Formaten 4° und 2°) der Jahrgänge 1800, 1801, 1832, 1833, 1856, 1868, 1872 und 1876 im Umfang von 9.544 Seiten zugrunde gelegt. Basierend auf einer zufällig ausgewählten Stichprobe von 40 Probewerten zu je 1.000 Zeichen wurde die Zeichengenauigkeit gemessen, indem die Erkennungsrate der richtig erkannten Zeichenanzahl in Prozent ausgezählt wurde, einmal für den OCR-erkannten Text basierend auf dem 300ppi-Digitalisat und zum Vergleich für den OCR-erkannten Text basierend auf dem 400ppi-Digitalisat der gleichen Textstelle.
Über ein statistisches Verfahren wurde ausgehend vom Mittelwert und der Standardabweichung der Stichprobe über die Studentsche t-Verteilung das Konfidenzintervall der tatsächlichen Erkennungsrate errechnet. Dabei ergaben sich bei einer Wahrscheinlichkeit von 99% (als Maß für die Güte der Schätzung) folgende Werte für die unteren Grenzen des Konfidenzintervalls: bei 400 ppi 98,53% und 98,36% bei 300ppi. Beide Werte liegen nah beieinander und über 97%, was in der zitierten Literatur als gut gilt (gut: 97–99,5%[45]). Das Ergebnis bezieht sich auf die Frakturschrift und nicht auf die Grundgesamtheit der Mischschriften (diese liegen vor allem in den Werbeanzeigen vor). Aktuell wird die OCR-Software ABBYY FineReader SDK in der Version 11 für Frakturschriften eingesetzt, deren Kosten pro Seite 0,14€ betragen.[46] Eine manuelle Nachkorrektur wurde bisher nicht vorgenommen, sie ist für besonders wichtige Seiten (z.B. die Jahrgangsregisterseiten bis Anfang der 1870er Jahre) empfehlenswert und kann jederzeit nachgeholt werden.

OCR-Ergebnisse München

Die BSB München hat in einem Vorprojekt, das ein eigenes Vergabeverfahren erforderte, mit dem Dienstleister Fraunhofer IAIS die OCR-Bearbeitung einer Testauswahl der „Allgemeinen Zeitung/Cotta'schen Zeitung" durchgeführt.[47] Die Ergebnisse wurden evaluiert und ausführlich dokumentiert.[48] Das auszuwertende Material bestand aus 11 Zeitungsbänden, überwiegend in Frakturschrift, mit insgesamt 11.354 Seiten (50 Stichproben mit jeweils 925–992 Zeichen, 1.000 mit Zeilenumbrüchen). Ziel dieser Evaluation war in erster Linie, die OCR in Hinsicht auf Zeichengenauigkeit, Lesefluss und Formaterkennung zu bewerten. Es wurde eine geschätzte Zeichengenauigkeit zwischen 94,70% und 97,65% (Konfidenzzahl: 95%) erreicht. Außer der Zeichengenauigkeit wurde

[44] https://sites.google.com/site/textdigitisation/qualitymeasures/computingerrorrates
[45] Die DFG-Praxisregeln sprechen von hier von groben Richtwerten. (Vgl. Praxisregeln DFG, S. 31 ff.).
[46] Eine Seite im Sinne der Lizenz ist eine DINA4-Seite, weshalb für eine Zeitungsseite mehrere (bis zu 4) Lizenzseiten bezahlt werden müssen, siehe hierzu auch: https://abbyyeu.com/rs/_media/rs2.0_faq_en_04_2008.pdf.
[47] Ursprünglich wurde von der BSB München beantragt, die Images- und Fraktur-OCR-Produktion als Gesamtpaket über ein einziges Vergabeverfahren an einen Dienstleister zu vergeben. Dies hätte allerdings die von der DFG empfohlenen Testreihen nicht ermöglicht. Die OCR-Produktion (mit Artikelseparierung) wurde deshalb nach Durchführung einer Testevaluierung in zwei separaten Vergabeverfahren beauftragt.
[48] Vgl. den als Anlage beigefügten OCR-Bericht München.

auch die Erkennung des Leseflusses und des Formats geprüft. Aufgrund dieses guten Ergebnisses wurde die automatische OCR-Bearbeitung des gesamten Erscheinungsverlaufes der Allgemeinen Zeitung (ca. 281.455 Images) vergeben und von Fraunhofer IAIS bearbeitet. Der Dienstleister setzte im Projekt für die reine OCR auf Textzeilenbasis ABBYY FineReader Engine 10.5 R3 ein.

Im Projekt erprobte die Bayerische Staatsbibliothek darüber hinaus bei der Wochenschrift „Der gerade Weg/Der illustrierte Sonntag" (3.145 S.) exemplarisch nach entsprechender Marktsichtung mit Fraunhofer IAIS sowie dessen Kooperationspartner ArchivInForm eine halb-automatische Artikelseparierung. Die automatische Analyse lieferte eine strukturierte OCR sowie eine Segmentierung in Einzelartikel. Diese Analyseergebnisse wurden in einem gesonderten Workflow halb-automatisch, teils manuell überarbeitet, um das Ziel einer Fehlerfreiheit der Artikelstruktur in Höhe von 97% zu erreichen.[49] Hierzu wurde vorab eine genaue Analyse der Wochenschrift „Der gerade Weg/Der illustrierte Sonntag" vorgenommen, um die Regeln, welche von Vorlage zu Vorlage differieren können, für die Nachkontrolle zu erstellen.

Als größeres Problem erwies sich, dass die Überschriften in sehr unterschiedlichen Schriftarten, Schriftstilen oder Schriftmischungen gehalten waren, was deren OCR-Genauigkeit und damit die spezifischen Vorteile einer Suche von bzw. in Artikeln beeinträchtigen kann. Werden beispielsweise wichtige Schlagworte in Überschriften nicht korrekt erkannt, können sie eine geringere Relevanz erzielen, da sich die Trefferzahl dann nur auf den laufenden Text bezieht. Des Weiteren kann es vorkommen, dass drucktypisch besonders gestaltete Überschriften als Bild erkannt werden und nicht als Text. Beide Testreihen verdeutlichen, dass die OCR für Fraktur die erwartete Genauigkeit geliefert hat, es jedoch gerade bei Überschriften (und auch Werbeblöcken) aufgrund der Mischschriftarten zu Fehlern kommt. Eine manuelle Nachkorrektur ist bei der enormen Datenmenge jedoch mit Aufwand und höheren Kosten verbunden.

Die SLUB Dresden hat zunächst zwei Zeitungen mit Antiqualettern im Volltext eingestellt. Die Tests bestätigen die positiven Ergebnisse des Europeana Newspapers-Projekt, die im Folgenden ausführlicher dargestellt werden.

	SLUB Dresden	ULB Sachsen-Anhalt	BSB München
Name der Zeitung	The Dresden Daily	Hallesches Tageblatt	Allgemeine Zeitung
	Jüdische Wochenschau		Der Gerade Weg
Anzahl der Seiten	6.886	145.454	295.051
Schrifttyp Antiqua	x		
Schrifttyp Fraktur		x	x
OCR Software	ABBYY FineReader 11	ABBYY FineReader 11	ABBYY FineReader 10.5 R3

Tabelle 8: Übersicht OCR

[49] Die Leistungen der manuellen Nachkontrolle umfassten: 1. Zusammenfügen und Löschen von Blöcken (merge, exclude), Zuweisung von Texttypen (Bild, Tabelle, Text), 2. Generierung, Zusammenführung und Trennung von Artikeln, Zuordnung zur Kategorie Redaktionell/ Nicht Redaktionell, 3. Label-Zuweisung (rooftitle, title, subtitle, textblock, caption), 4. Check / Fehler (Blockzählung für die Lesereihenfolge, Prüfung anhand Artikeltabelle: jedem Artikel muss ein Titel zugewiesen werden).

OCR- Ergebnisse des EU-Projekts

Parallel zu den Tests in den Pilotprojekten hat das Europeana Newspapers-Projekt in Massenverfahren rund 12 Mio. Volltextseiten in 14 Sprachen (ca. 60% nach Original- und 40% nach Filmvorlagen) generiert, darunter rund 3,2 Mio. Seiten aus 4 Berliner und 7 Hamburger Zeitungen.[50] Zusätzlich wurden 2 Mio. Seiten mit halbautomatischer Layout- und Artikelerkennung bearbeitet. Die zur signifikanten Reduktion der Datenmenge binarisierten Seiten wurden an der Universität Innsbruck mit der ABBYY FineReader Engine v11 bearbeitet, die Ergebnisse im ALTO-Format exportiert und im METS-Format strukturiert.
An einem repräsentativen Querschnitt von 600 Zeitungsseiten haben Pletschacher, Clausner und Antonacopoulos vom Pattern Recognition and Image Analysis Research Lab der University of Salford die OCR-Qualität getestet und im August 2015 Methodenmodelle zur Auswertung sowie quantitative und qualitative Ergebnisse vorgestellt.[51] Sie entwerfen einen komplexen Evaluationsworkflow, der Zeichen- und Wortgenauigkeitsprüfungen sowohl text- als auch layoutbasiert für unterschiedliche Anwendungsszenarien ermöglichen soll. Sie empfehlen bei Zeitungen, also bei großen Textmengen mit komplizierter Layoutstruktur, die Wortgenauigkeit ohne Berücksichtigung der Wortreihenfolge zu messen (Bag of Words). Durch Binarisierung zur Reduktion der Datenmenge werde nur 1% an Recherchequalität gemessen an der originalen Imagevorlage eingebüßt.
Sie kommen zu dem Ergebnis, dass die Wortgenauigkeit bei im Vorfeld normalisierten (von Sonderzeichen bereinigten) Texten etwa 3,2% höher liegt als bei originalen Textvorlagen. Im Ergebnis liegt die Wortgenauigkeit (Bag of Words) bei untersuchten Zeitungen in diesen sieben Sprachen jeweils über 80%: Niederländisch, Tschechisch, Englisch, Französisch, Deutsch, Ungarisch, Schwedisch. Die anderen Sprachen (Estnisch, Finnisch, Litauisch, Russisch, Serbisch-kyrillisch, Ukrainisch und Jiddisch) liegen zwischen 76,1 und 32,7% Wortgenauigkeit. Bei modernen deutschsprachigen Zeitungen werden 84%, bei historischen Sprachstufen („Old German“) 68,1% Wortgenauigkeit gemessen. Nach Font unterteilt wurde bei Berücksichtigung sämtlicher bearbeiteter Zeitungstitel in zahlreichen Sprachen bei Antiqua eine Wortgenauigkeit von 81,4%, bei Fraktur 67,3% und bei Mischschriften 64% erzielt. Die Autoren weisen darauf hin, dass die fortlaufende Frakturverbesserung inzwischen rund 70% Wortgenauigkeit erreiche und damit sinnvolle Volltextsuchen ermögliche: „From experience (discussions with library partners) it can also be said that success rates beyond 70% are usually good enough to provide an acceptable level of text search through a presentation system.”[52]
Die Ergebnisse der Pilotprojekte zeigen, dass bei Frakturschrift in Zeitungen, die zu den komplexesten und damit schwierigsten OCR-Vorlagen zählen, über 95% Zeichengenauigkeit erreichbar sind, die einer anzustrebenden Wortgenauigkeit von ca. 80% entsprechen.

Eine differenziertere fachwissenschaftliche Bewertung der Qualitätseinstufungen von Zeichen- und Wortgenauigkeit sollte mit dem von der DFG geförderten „Koordinierungsprojekt zur Weiterentwicklung von Verfahren der Optical Character

[50] Vgl. den Abschlussbericht, die detaillierte Evaluation der OCR/OLR nach Anwendungsszenarien sowie die Empfehlungen aus den einzelnen Arbeitspaketen: http://europeananewspapers.github.io/ http://www.europeana-newspapers.eu /public-materials/deliverables/ sowie hier http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D3.5_Performance_Evaluation_Report_1.0.pdf; http://www.europeana-newspapers.eu/wp-content/uploads/2012/04/D-2-2_Specification_of_requirements-2.pdf
[51] Vgl. http://primaresearch.org/publications/HIP2015_Pletschacher_OCRWorkflowEvaluation
[52] http://primaresearch.org/publications/HIP2015_Pletschacher_OCRWorkflowEvaluation, S. 44

Recognition (OCR)"[53] angestrebt werden, das die Verfahren der automatischen Texterkennung und Konzepte für optimale Workflows näher analysieren und verbessern will. Auch wenn der Focus dieses neuen Projekts auf methodischen, technischen und fachlichen OCR-Weiterentwicklungen für die Volltextgenerierung des Fraktur-Buchdrucks vom 16. bis 18. Jahrhundert liegen dürfte, sollte das Medium Zeitung mit seinen parallelen Schrift- und Layout-Entwicklungen rechtzeitig einbezogen werden.
Die Ergebnisse dieser Pilotphase sind ermutigend und zeigen, dass auch Zeitungen mit einer – gemessen an Drucken des 18. bis 20. Jahrhunderts – viel komplexeren Layoutstruktur sowohl mit Antiqua- als auch Fraktur-Lettern über OCR als Volltexte aufbereitet werden können und sollten.
Höhere Genauigkeiten (gem. Stufe 3) lassen sich durch Layouterkennung und (semi-) manuelle Nachkorrekturen erreichen. Softwarewerkzeuge und Richtlinien für die Evaluation von Text- und Layouterkennung wurden im Rahmen des Europeana Newspapers-Projekt für die freie Nutzung bereitgestellt.[54]

Empfehlungen

Für eine skalierbare Verarbeitung ist ein standardisierter Zeitungsworkflow notwendig. Für die spätere Konfiguration der OCR/OLR sind vorab Daten zu Sprache und Schrifttyp zu erheben. Es wird empfohlen, alle Zeitungen mit dem Grundstandard Stufe 1 (DFG-Viewer-Strukturdatenset) und Stufe 2 (OCR-Erkennung) zu erschließen. Hochgradig automatisierte Massenverarbeitungsverfahren[55] sind für Wortsuchen – unter Berücksichtigung zu differenzierender Anwendungsszenarien[56] – grundsätzlich geeignet.
Für eine wissenschaftlich verlässliche Texterkennung bleiben trotz der Fortschritte dennoch viele Wünsche offen. Erfolgversprechende Ansätze zur OCR-Optimierung sind die automatische OCR-Korrektur[57], aber auch die OCR-Korrektur durch Benutzer[58], wobei hier eine ganze Reihe auch technischer Herausforderungen zu meistern sind (Versionierung, Redaktionsgrad).
Grundsätzlich soll es möglich sein, OCR-Verfahren nach technischem Fortschritt zu wiederholen und/oder die Auswertung und Nutzung entstehender maschinenlesbarer Textkorpora projektspezifisch weiter zu entwickeln.[59]

---

[53] http://www.ocr-d.de/
[54] Vgl. http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D3.3_Evaluation_Tools_Final_Versions_1.0.pdf
[55] Vgl. zur Massenverarbeitung ab mehreren Millionen Seiten die Empfehlungen zur Organisation der Verarbeitungsschritte des Europeana Newspapers-Projekt: http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D-2.4_Recommendations_on_best_practices_for_refinement_1.0.pdf
[56] Vgl. http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/D3.5_Performance_Evaluation_Report_1.0.pdf
[57] Vgl. Martin Reynaert. 2008. Non-interactive OCR post-correction for giga-scale digitization projects. In: Proceedings of the 9th international conference on Computational linguistics and intelligent text processing (CICLing'08), Alexander Gelbukh (Ed.). Springer-Verlag, Berlin, Heidelberg, pp. 617-630. http://ilk.uvt.nl/downloads/pub/papers/CICLING08.TICCL.MRE.postpublication.pdf
[58] Vgl. Rose Holley. 2009. Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers, National Library of Australia, http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf
[59] Vgl. Günter Mühlberger, „Digitalisierung historischer Zeitungen aus dem Blickwinkel der automatisierten Text- und Strukturerkennung (OCR)“, Zeitschrift für Bibliothekswesen und Bibliographie : vereinigt mit Zentralblatt für Bibliothekswesen ; ZfBB ; Organ des wissenschaftlichen Bibliothekswesens 58, Nr. 1 (2011): 10–18.
Das Deutsche Textarchiv (DTA) hat mit insgesamt vier Zeitungen gezeigt, wie in Zusammenarbeit mit Linguisten Textcorpora zur Verfügung gestellt werden können. Auch an der BSB München liegen entsprechende Erfahrungen im Rahmen der Bereitstellung der Google-Digitalisate vor.

Stufe 3 mit Layouterkennung und Artikelseparierung zählt zum erweiterten Erschließungsstandard für wissenschaftliche Zwecke und ist insbesondere bei Zeitungen von überregionaler Bedeutung (Leitmedien, innovativen Zeitungen, fachspezifisch relevanten Zeitungen und Journalen) zu empfehlen. Die oft hochgradig komplexe Struktur von Zeitungen bringt die aktuellen Technologien für die Layouterkennung und Segmentierung noch an ihre Grenzen.[60] Der Bedarf an Forschung und Entwicklung erstreckt sich dabei auf die Erkennung von graphischen Elementen zur Trennung von Artikeln (Separatoren)[61], die Tabellenerkennung[62], die Strukturerkennung sowie die Bilderkennung und -Extraktion[63]. Die Verfügbarkeit valider Werkzeuge sollte zu einer deutlich besseren Erschließung und damit auch Präsentation führen.

Die Stufen 4 (Normdatenverknüpfung) und 5 (vertiefte sachliche Erschließung) sind für zahlreiche wissenschaftliche Fragestellungen (Auflösung von Pseudonymen, bio-bibliographische Erfassung, fachwissenschaftliche Kontextualisierungen) sinnvoll. Sie erfordern jedoch deutlich erhöhten Zeit- und Kostenaufwand. Eine Auswertung der Logfiles für die Suche im Zeitungsportal der Nationalbibliothek Wales zeigte im Jahr 2014, dass bis zu 90% der Suchanfragen im Zeitungskontext Personen oder Ortsnamen gelten.[64] Vor diesem Hintergrund ist eine Anreicherung von Zeitungsvolltexten mit einer Eigennamenerkennung (Named Entity Recognition, NER) sinnvoll. Hier liegen bereits erste Erfahrungen vor, die aber weiter entwickelt werden müssen.[65] Eine besondere Herausforderung stellt die Disambiguierung von Namen dar (Named Entity Disambiguation), auch dafür liegen bereits experimentelle Erfahrungen vor.[66] Ein weiteres wichtiges Arbeitsfeld werden Ortsnamen sein.

Insbesondere bei deutschsprachigem historischem Material findet sich eine große Vielfalt in der Rechtschreibung. Da sich Benutzer von digitalen Sammlungen typischerweise nur an der aktuellen Schreibweise orientieren, oder sich nicht aller der zahlreichen historischen Varianten bewusst sind, werden viele möglicherweise relevante Textstellen trotz korrekter OCR gar nicht erst gefunden. Die historischen Schreibvarianten folgen dabei oft einem bestimmten Muster wie y→i oder th→t. Beispiel: Theyl→ Theil→Teil. Die Computerlinguistik macht sich dies zunutze, um entsprechende historische Varianten von typischen OCR-Fehlern (rn→m) eindeutig unterscheiden zu können[67] und entsprechende Wörterbücher zu erarbeiten, die historische Schreibvarianten auf ihre modernen Lemmata abbilden. Entsprechende im Massenprozess nutzbare Werkzeuge vorausgesetzt ist der Vorteil

[60] Vgl. Apostolos Antonacopoulos, Christian Clausner, Christos Papadopoulos and Stefan Pletschacher. 2013. ICDAR2013 competition on historical newspaper layout analysis (HNLA13). http://www.primaresearch.org/www/assets/papers/ICDAR2013_Antonacopoulos_HNLA2013.pdf

[61] Ein vielversprechender Ansatz vgl. hierzu: David Hebert, Thomas Palfray, Stephane Nicolas, Pierrick Tranouez, and Thierry Paquet. 2014. Automatic article extraction in old newspapers digitized collections. In Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage (DATeCH '14). ACM, New York, NY, USA, pp. 3-8. DOI=http://dx.doi.org/10.1145/2595188.2595195.

[62] Stefan Klampfl, Jack Kris and Roman Kern. 2014. A Comparison of Two Unsupervised Table Recognition Methods from Digital Scientific Articles. In D-Lib Magazine 20, no. 11 (2014): p. 7. http://www.dlib.org/dlib/november14/klampfl/11klampfl.html; sowie: ICDAR2013 Table Recognition Competition, http://www.tamirhassan.com/competition.html.

[63] Zuletzt vielversprechend: Vgl. http://britishlibrary.typepad.co.uk/digital-scholarship/2013/10/peeking-behind-the-curtain-of-the-mechanical-curator.html

[64] Vgl. Paul Gooding. 2014. Exploring Usage of Digital Newspaper Archives through Web Log Analysis: A Case Study of Welsh Newspapers Online. In: Digital Humanities 2014, 2014-07-08, Lausanne. http://dharchive.org/paper/DH2014/Paper-310.xml

[65] Vgl. https://github.com/EuropeanaNewspapers/europeananp-ner

[66] Vgl. https://github.com/EuropeanaNewspapers/europeananp-dbpedia-disambiguation

[67] Vgl. Ulrich Reffle and Christoph Ringlstetter. 2013. Unsupervised profiling of OCRed historical documents. Pattern Recognition. 46, 5 (May 2013), pp. 1346-1357. DOI=http://dx.doi.org/10.1016/j.patcog.2012.10.002.

evident: Das Angebot einer Option „Suche nach historischen Varianten“ würde die Abfrage des Index um die im historischen Wörterbuch verzeichneten validen Varianten erweitern (query expansion) und so auch diese Fundstellen dem Benutzer als Treffer anbieten.[68] Diese Funktionalität wird so z.B. bereits im niederländischen Portal Delpher angeboten.[69]
Es gibt also im Bereich technikgetriebener Anreicherungsverfahren bereits eine Reihe vielversprechender Aktivitäten, deren Einbindung in eine Hauptphase mit einer deutlich steigenden Menge an Material eine Rolle spielen kann und gleichzeitig auch die Relevanz entsprechender Verfahren und deren Optimierung weiter erhöhen wird. In einer Hauptphase sollten sich deshalb Teilprojekte auch der Weiterentwicklung automatischer Erschließungs- und Anreicherungsverfahren widmen können, um die Relevanz der Ergebnisse aus der OCR für die (wissenschaftliche) Nutzung signifikant zu erhöhen. Die Projektpartner gehen davon aus, dass entsprechende Initiativen zudem im Bereich des DFG-Normalprogramms initiiert und bearbeitet werden können.
Um die Zeitungsdigitalisierung technisch und organisatorisch weiter zu entwickeln, erachten es die Pilotpartner als notwendig, ein „Kompetenznetzwerk Zeitungen“ im Zusammenhang mit der Deutschen Digitalen Bibliothek auszubauen, in dem strukturbildende Einrichtungen besondere Verantwortung übernehmen müssen. Dabei geht es um die dynamische Weiterentwicklung des Zusammenspiels von DDB und ZDB in Richtung einer virtuellen Zusammenführung, Koordination und Steuerung, und nicht zuletzt auch um den notwendigen Wissenstransfer, um die Vielzahl mittlerer und kleiner Einrichtungen über aktuelle Entwicklungen und Werkzeuge zu informieren.

URN-Granular: Persistente Adressierung von digitalisierten Zeitungen

Die DFG-Praxisregeln „Digitalisierung“ im Bereich der Wissenschaftlichen Literaturversorgungs- und Informationssysteme (LIS)[70] fordern die Sicherstellung einer persistenten Adressierbarkeit der im Netz bereitgestellten Ressourcen mit einer „größtmögliche[n] Granularität“.[71] Hierbei können verschiedene Persistenz-Verfahren genutzt werden (PURL, URN, DOI, Handle, etc.),[72] die Nutzung von Uniform Resource Names (URN) wird jedoch von den Praxisregeln „nachdrücklich empfohlen“[73].

Uniform Resource Names (URN)

Das im Jahr 2009 gemeinsam von der Deutschen Nationalbibliothek (DNB) und der ULB Sachsen-Anhalt erarbeitete Verfahren *URN granular* bedient sich einer Adressierungstechnik, die vor allem auf die granulare Adressierung monographischer Werke unter Maßgabe der oben zitierten Anforderungen der *DFG-Praxisregeln* abzielt.[74] Die

[68] Vgl. Annette Gotscharek, Andreas Neumann, Ulrich Reffle, Christoph Ringlstetter, and Klaus U. Schulz. 2009. Enabling information retrieval on historical document collections: the role of matching procedures and special lexica. In Proceedings of the Third Workshop on Analytics for Noisy Unstructured Text Data (AND '09). ACM, New York, NY, USA, pp. 69-76. DOI=http://dx.doi.org/10.1145/1568296.1568309.
[69] Vgl. http://www.digitisation.eu/blog/take-tools-within-succeed-project-implementation-inl-lexicon-service-delpher-nl/
[70] Aktuell ist die Fassung Februar 2013, im Internet zugänglich unter: http://www.dfg.de/formulare/12_151/12_151_de.pdf
[71] Ebd, S. 39.
[72] Ebd, S. 40.
[73] Ebd, S. 40.
[74] Vgl. Dorothea Sommer, Christa Schöning-Walter, Kay Heiligenhaus: URN Granular: Persistente Identifizierung und Adressierung von Einzelseiten digitalisierter Drucke; ein Projekt der Deutschen Nationalbibliothek und der Universitäts- und Landesbibliothek Sachsen-Anhalt. In: ABI Technik 28 (2008) 2, S. 106-114. http://dx.doi.org/10.1515/ABITECH.2008.28.2.106.

Nutzung dieses Verfahrens bei der Adressierung komplexerer Objekte ist aufgrund der starren Relation (Werk - Einzelseiten) jedoch nur eingeschränkt möglich.[75] Ziel des Teilprojektes URN granular 2 der ULB Sachsen-Anhalt war es folglich, das bestehende Adressierungsverfahren gemeinsam mit der DNB fortzuschreiben, um den komplexeren Herausforderungen bei der Bereitstellung und Adressierung von digitalisierten Zeitungen Rechnung zu tragen. Einen sinnvollen Ansatzpunkt hierzu sahen die Projektpartner in der Orientierung an einem Konzept, das in der generellen RFC-Spezifikation zu URIs als „fragment identifier component of a URI" bezeichnet wird.[76] Seit 2012 liegt hierzu ein RFC-Entwurf vor, der sich der Herausforderung der flexiblen Fragment-Adressierung widmet. Dieser hat allerdings noch keinen verabschiedeten Status.[77] Hieran anknüpfend hat die DNB im Projektverlauf eine eigenständige Implementierung entwickelt, die sich als flexibel genug erweist, die zuvor beschriebenen Limitierungen zu überwinden. Der von der DNB vorgelegte Entwurf[78] sieht zunächst die Erweiterung der URN-NBN-Syntax um einen reservierten URN-Bestandteil – /fragment/ – vor, dem beliebige weitere Parameter (in Form von Key-/Value-Pairs) folgen können.[79] Diese Parameter werden beim Resolving des übermittelten URN durch den DNB-Resolver nicht weiter ausgewertet, sondern an das lokale Repository weitergeleitet.

Digital Object Identifiers (DOI)

Auch das auf dem Handle-System aufbauende Verfahren der Digital Object Identifier (DOI) stellt eindeutige und dauerhafte Identifikatoren für digitale Ressourcen bereit und wird in der Praxis vor allem für die Referenzierung von Artikeln wissenschaftlicher Fachzeitschriften und neuerdings vermehrt für den persistenten Zugriff auf Forschungsprimärdaten verwendet.[80] Das DOI-System kennt mittlerweile auch granulare Adressierungstechniken für den Zugriff auf Teile von digitalen Ressourcen, wie sie für das Verfahren URN granular 2 beschrieben wurden.[81] Damit besteht grundsätzlich die Möglichkeit, in der digitalen Präsentation von Zeitungen DOIs zur Adressierung in analoger Form zu verwenden. Ausschlaggebend ist hier letztlich die grundlegende Persistenz-Strategie der digitalisierenden Einrichtung.

Umsetzung im Pilotprojekt

Aufgrund der langjährigen Vergabepraxis von URNs hat sich die ULB Sachsen-Anhalt entschieden, die bisherige Persistent-Identifier-Strategie aufrechtzuerhalten und das oben beschriebene Verfahren URN granular 2 für die persistente Adressierung von Teilobjekten digitalisierter Zeitungen (sowie aller weiteren Medientypen in Digitalisierungsprojekten an

[75] Vgl. Dorothea Sommer, Kay Heiligenhaus, Carola Wippermann, Manfred Pankratz: Zeitungsdigitalisierung: eine neue Herausforderung für die ULB Sachsen-Anhalt. Werkstattbericht aus der Pilotphase des DFG-Projekts „Digitalisierung historischer Zeitungen". In: ABI Technik 34 (2014) 2, S. 75–85. http://dx.doi.org/10.1515/abitech-2014-0013.
[76] Uniform Resource Identifier (URI): Generic Syntax. January 2005. https://tools.ietf.org/html/rfc3986#section-3.5.
[77] Uniform Resource Name (URN) Syntax draft-ietf-urnbis-rfc2141bis-urn-02. March 12, 2012. https://tools.ietf.org/html/draft-ietf-urnbis-rfc2141bis-urn-02.
[78] Vgl. Uta Ackermann, Kadir Karaca Koçer: *URN:NBN – Recent developments*. 7. May 2012. http://www.ratswd.de/ver/docs_PID_2012/Ackermann_Kocer_PID2012.pdf, Folie 4.
[79] Vgl. http://nbn-resolving.org/examples.
[80] Vgl. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn: nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung (Version 2.0), 2009: Kapitel 9.4.2. http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_335.pdf.
[81] Vgl. International DOI Foundation: *DOI® Handbook*. May 15, 2015, Kapitel 5.8. http://www.doi.org/doi_handbook/5_Applications.html#5.8.

der ULB Sachsen-Anhalt) zu nutzen. Nach diesem Verfahren wurde für die im Pilotprojekt digitalisierte Zeitung, das „Hallesche Tageblatt“,[82] ein URN

a.) auf Ebene der Zeitungsaufnahme in der ZDB vergeben und bei der DNB registriert sowie

b.) URNs auf der Ebene der jeweils digitalisierten Jahrgänge der Zeitung. Die persistente Adressierung

c.) jeder Einzelseite erfolgt mittels der lokalen Vergabe von Fragment-URNs, die nicht zentral registriert werden müssen.

Diese Granularität der URN-Vergabe folgt exakt dem im Verlauf des Pilotprojektes gemeinsam entwickelten vierstufigen Navigationsmodell für Zeitungen im Zusammenspiel von ZDB und DFG-Viewer: ZDB → Zeitung → Jahrgangskalender → Startseite der gewählten Ausgabe.[83]

Empfehlung

Für eine wissenschaftlich zuverlässige Zitierfähigkeit im Internet sind möglichst eine persistente seitengenaue Adressierbarkeit anzustreben und dafür erprobte Verfahren (z.B. URN granular 2, DOI) zu nutzen.

3.6 Qualitätssicherung

Die Qualitätssicherung erfolgte in den Projekten bei allen Schritten des Digitalisierungsworkflows.

Die Reproduktionen der Mikroformen der SuUB Bremen differieren in ihrem Erscheinungsbild von Ausgabe zu Ausgabe teils erheblich. Die Qualitätskontrolle sollte unter Berücksichtigung zeitökonomischer Faktoren sicherstellen, dass die Digitalisate die bestmögliche Entsprechung mit der Vorlage bieten. Neben den technischen Parametern (wie z.B. Auflösung und Farbtiefe), die automatisiert überprüft werden können, ist die vollständige Lesbarkeit von entscheidender Bedeutung.

Die ULB Sachsen-Anhalt führte laufend qualitätssichernde Maßnahmen während aller Schritte der Digitalisierung und Erschließung der Zeitung durch. Sie erfolgten durch eine Kombination von automatisierten Verfahren mit der Software Visual Library und intellektueller Kontrolle. Dies betraf zum einen notwendige Nachscan-Arbeiten, um bessere Ergebnisse der Bildschärfe für die OCR-Erkennung zu erzielen, zum anderen wurden während der Erschließungsarbeiten (Ausgabensegmentierung und Tiefenstrukturierung) die Seiten paginiert, um die Vollständigkeit und richtige Reihenfolge der Scans zu gewährleisten.

Die SLUB Dresden hat im Rahmen einer kostengünstigen Massendigitalisierung vom Film die Qualitätskontrolle auf das Trennen der Ausgaben nach dem Scanprozess und die

[82] Vgl. den Eintrag in der ZDB unter http://dispatch.opac.d-nb.de/DB=1.1/CMD?ACT=SRCHA&IKT=8506&TRM=2757736-3.

[83] Vgl. Ansicht auf Zeitungsebene: http://digitale.bibliothek.uni-halle.de/zd/id/9059307 und Ansicht auf Jahrgangsebene:http://digitale.bibliothek.uni-halle.de/zd/periodical/pageview/9339791 und Ansicht auf Seitenebene: http://digitale.bibliothek.uni-halle.de/urn/urn%3Anbn%3Ade%3Agbv%3A3%3A1-691276?&page=8965342

Seitenpaginierung konzentriert. Lückenergänzungen wurden nicht vorgenommen, Qualitätsschwankungen wurden in einer Excel-Tabelle und in einem Metadatenfeld notiert.

Die Qualitätskontrolle der Images im Teilprojekt der BSB München erfolgte kontinuierlich über den gesamten Produktionszeitraum jeweils zeitnah nach Erhalt der einzelnen Teillieferungen entsprechend der vorab via Pflichtenheft mit dem Dienstleister festgelegten Liefertermine. Die Kontrolle umfasste insbesondere die Überprüfung der Vollständigkeit, der Lesbarkeit und Ausrichtung sowie die Einhaltung der geforderten Scanparameter der Datenlieferungen. Dies geschah auch im Hinblick auf die weitere OCR-Produktion, deren Qualität maßgeblich von der Image-Produktion abhängt. Fehlerhaft gelieferte Images wurden beim Dienstleister zeitnah im Rahmen eines definierten Reklamationsworkflows beanstandet. Erst nach erneuter Prüfung und erfolgreicher Abnahme der korrigierten Digitalisate durch die BSB konnte der Dienstleister die Images in Rechnung stellen. Die Dokumentation der QS-Prüfschritte und die durchgeführten Reklamationen waren Bestandteil der Qualitätskontrolle. Die BSB empfiehlt, die Produktionsparameter und Vorgaben der jeweiligen projektspezifischen Leistungsbeschreibung durch eine umfassende Qualitätskontrolle der Images gegen zu prüfen und fehlerhafte Scans/Digitalisate im Rahmen eines Reklamationsworkflows korrigieren zu lassen. Eine seitengenaue Qualitätskontrolle ist mit Blick auf die Qualität zu generierender Struktur- und OCR-Volltextdaten einer stichprobenhaften Kontrolle vorzuziehen, jedoch nur bei überschaubaren Projektumfängen und ausreichenden Ressourcen durchgängig zu leisten.

Empfehlung

Eine Qualitätskontrolle während des gesamten Prozesses ist – abhängig von den zuvor möglichst präzise zu definierenden Qualitätsansprüchen an die jeweiligen Projektergebnisse – erforderlich. Es sollte nach Möglichkeit vor Beginn des Projektes eine genaue Analyse des Materials erfolgen, damit rein materialimmanente Fehler ausgeschlossen werden können. Außerdem sollte in dieser Planungsphase auch überlegt werden, ob und in welchem Umfang Lückenergänzungen und Nachscannen angestrebt werden bzw. notwendig sind, um die Kosten bereits im Vorfeld kalkulieren zu können.

3.7 Bildformate, Langzeitarchivierung

Die Digitalen Sammlungen der SuUB Bremen sind zum Zweck der Datensicherung an das Storage-Area-Network (SAN) der Staats- und Universitätsbibliothek Bremen angebunden. Die SuUB Bremen konzipiert die Langzeitarchivierung auf Basis des Archivierungssystems der Verbundzentrale des GBV in Göttingen. Bis zur Umsetzung dieses Archivierungssystems sind alle Daten der Digitalen Sammlungen der SuUB Bremen über klassische Datensicherungsmaßnahmen nach dem Stand der Technik gesichert. Aktuell ist gewährleistet, dass die binären Datenbestände wenigstens zehn Jahre vorgehalten werden können – ein Zeitraum, in dem ein Gesamtmodell zur Langzeitarchivierung konzeptioniert und in Betrieb genommen werden soll.
In der BSB München wurden die Daten (Scans und OCR) von rund 291.000 Seiten, die durch Dienstleister bearbeitet wurden, und die rund 306.000 Seiten, die durch Google digitalisiert wurden, im Rahmen des BSB-Digitalisierungsworkflows über die neue Zeitungs-ZEND-Instanz im Leibniz-Rechenzentrum langzeitarchiviert. Der hierfür benötigte Speicherplatzbedarf beläuft sich auf 19,21 TB. Wichtige Unterstützung leistet dabei das Rosetta Digital Preservation System der Firma Ex Libris, das gemeinsam mit der

Verbundzentrale des Bibliotheksverbundes Bayern seit 2012 produktiv betrieben wird und auch für die bayerischen Hochschulbibliotheken zur Verfügung steht. Die organisatorisch-technische Infrastruktur wurde bereits 2013 mit dem Data Seal of Approval für nachhaltige und vertrauenswürdige digitale Archive ausgezeichnet.
Die SLUB Dresden übernimmt die Langzeitarchivierung für die sächsischen Hochschulbibliotheken und kooperiert dabei mit dem Zentrum für Hochleistungsrechnen der TU Dresden. Im Einsatz ist ebenfalls das Rosetta Digital Preservation System. Für das Pilotprojekt sind 14 TB Speicherbedarf erforderlich. Gespeichert wird TIFF unkomprimiert, da es nahezu alle Bildverarbeitungsprogramme unterstützt und über mehrere Validierungs- und Reparaturtools verfügt.
Die ULB Sachsen-Anhalt hostet die TIFFs auf ihren eigenen Servern. Für das Projekt werden ca. 4,6 TB Speicherkapazität benötigt. In Abstimmung mit dem Rechenzentrum der Martin-Luther-Universität (ITZ) werden die TIFFs und Strukturdaten in ZIP-Containern, die über eine eindeutig zuordenbare persistente URN adressiert sind, im Back-Up-Archivsystem (IBM-Bandroboter 3584) gespeichert. Die eingesetzte Software ist IBM Tivoli Storage Manager (TSM) und wird auf einem gemeinsam mit der Firma semantics, dem ITZ und der ULB entwickelten Backupclient als Schnittstelle zwischen Visual Library und dem Backup-Archiv-System betrieben.
Einige europäische Partner, die auch beim Europeana Newspapers-Projekt mitwirkten, bevorzugen auf Grund erweiterter Darstellungsfunktionen (webbasierter Zoom, Rotation, Ausschnitte) sowie zur Einsparung von Speicherplatz das JPEG2000-Format, zumal es inzwischen auch eine Open-Source-Referenzimplementierung[84] des Formats sowie Validierungstools[85] und eine aktive Community gibt. Die Nationalbibliothek der Niederlande migrierte so z.B. ihre 9 Mio. Zeitungsseiten von TIFF nach JPEG2000, die British Library/National Library of Wales oder das neue Digitalisierungsprojekt in Dänemark verwenden standardmäßig JPEG2000 für Zeitungen. Da die Langzeitarchivierung im Rahmen dieses Pilotprojekts nicht getestet wurde, wird hier auf eine Empfehlung verzichtet..

3.8. Kostenkorridore der erprobten Verfahren [Stand: Mai 2017]

Die im Pilotprojekt erprobten unterschiedlichen Verfahren, Workflows, Mengen, Schwierigkeitsgrade und Erschließungstiefen ergeben ein differenziert zu betrachtendes Kostenspektrum, das sich nur bedingt verallgemeinern lässt. Im Folgenden werden relevante Kostenfaktoren in Abhängigkeit der entsprechenden Rahmenbedingungen dargestellt und mit konkreten Projektergebnissen unterlegt.

3.8.1 Kostenfaktoren der Zeitungsdigitalisierung

Die nachfolgend dargestellten Kostenfaktoren (Personal- und Sachkosten) korrespondieren mit dem entwickelten Digitalisierungsworkflow. Dabei ist zu beachten, dass diese Faktoren stets in Abhängigkeit von den konkreten Rahmenbedingungen und Projektzielen zu betrachten sind.

[84] Vgl. http://www.openjpeg.org/
[85] Vgl. http://openpreserve.github.io/jpylyzer/

Ermittelte Kostenfaktoren	Rahmenbedingungen
Aufgabenübergreifend	
• Projektleitung und –koordination	
Vorbereitung	
• Auswahl der zu digitalisierenden Inhalte und Abgleich mit der ZDB	Ggf. mit wissenschaftlicher Begleitung; Skaleneffekte des Mengengerüsts
• Beschaffung der Vorlagen	Nur Eigenbestand oder ggf. erhöhter Beschaffungsaufwand aufgrund von Lückenschluss mit Fremdbestand
• Prüfung der Vorlagenqualität und Entscheidung für eine Vorlagenart (Original vs. Mikrofilm)	Die gewählte Vorlagenart hat Einfluss auf den Kostenrahmen.
• Kollationierung bzw. Prüfung der Vollständigkeit und konservatorischen Eignung	
• Inhouse-Digitalisierung und/oder OCR-Bearbeitung/Tiefenerschließung: Prüfung der Eignung vorhandener Ausrüstung bzw. ggf. Beschaffung/Aufrüstung von Scannern und Software	
• Vergabe an Dienstleister (Digitalisierung und/oder OCR-Bearbeitung/Tiefenerschließung): Vorbereitung und Durchführung eines Vergabeverfahrens	
• Vorbereitung der Materialien: ggf. Lückenschluss, konservatorische Maßnahmen	Aufwand und Ausgestaltung abhängig von den Projektzielen
• Workflowplanung und Kostenkalkulation	
Digitalisierung	
• Art des Scannereinsatzes und Komplexität des Digitalisierungsvorgangs	Abhängig von der physischen Beschaffenheit der Vorlagen (gebunden/aufgeschnitten; Öffnungswinkel; konservatorische Merkmale; Zwischenblätter nach defekten oder fleckigen Seiten)
• Auflösung und Farbtiefe	Abweichung vom empfohlenen Standard ggf. in Abhängigkeit von der Vorlagenbeschaffenheit
• Qualitätskontrolle (bei Inhouse-Digitalisierung und in Zusammenarbeit mit dem Dienstleister)	
Erschließung	
• Bibliographische Erschließung	Ggf. jeweils Neuaufnahme/Korrektur für die aufeinander bezogenen Druck- und Reproduktionsformen erforderlich; RDA sieht für layoutgetreue Digitalisierung aktuell die identische Anlage von Druck- und Reproduktionskatalogisat vor. Hierdurch entstehen ggf.

	Zusatzaufwände für möglicherweise zahlreiche Titelsplits
• Strukturdatenerschließung	Aufwand und Grad der Automatisierbarkeit abhängig von der strukturellen Beschaffenheit der Vorlagen (z.B. Zahl unterschiedlicher Zeitungsausgaben, Beilagen, Grad der Einheitlichkeit der Ausgabenbezeichnungen, Erscheinungsfrequenz, wechselnde Kalendersysteme)
• OCR (Antiqua, Fraktur)	Die Verarbeitungskosten von Antiqua und Fraktur unterscheiden sich in der Regel, z.B. durch pauschale bzw. seitengenaue Errechnung der Lizenzkosten für die OCR-Software.
• Layouterkennung / Artikelseparierung	Abhängig von der strukturellen Beschaffenheit der Vorlage
• Qualitätskontrolle	
Bereitstellung	
• Einbinden in Präsentationsoberfläche / DFG-Viewer, ggf. mit Einrichtung entsprechender Schnittstellen	Abhängig von den jeweiligen Präsentationsvoraussetzungen
• Herstellung persistenter Adressierbarkeit	Granularität abhängig von den Projektzielen
• Qualitätskontrolle	
Archivierung	
• Datensicherung	Auflösung und Farbtiefe der Images haben Einfluss auf den benötigten Speicherplatz
• Qualitätskontrolle	

3.8.2 Ermittelte Kostenkorridore

Die nachfolgende Übersicht zeigt im Projekt konkret ermittelte Seitenpreise unter den jeweils genannten Rahmenbedingungen.

	SLUB Dresden	ULB Sachsen-Anhalt	BSB München	SuUB Bremen
Digitalisierungs-vorlage	*Mikrofilm*	*Original*	*Original*	*Sondermaterialien (Reproduktionen)*
Inhouse/Dienst-leister	*Inhouse*	*Inhouse*	*Dienstleister*	*Inhouse*
Seitenpreis (ohne OCR)	0,44 €	0,76 €	0,93 €	1,21 €
Seitenpreis mit OCR (Fraktur)	nicht durchgeführt	0,94 €	1,21 €	nicht durchgeführt
Seitenpreis mit OCR (Antiqua)	nicht repräsentativ	nicht durchgeführt	nicht durchgeführt	nicht durchgeführt
Seitenpreis mit OCR und Artikelseparierung	nicht durchgeführt	1,32 €	1,57 €	nicht durchgeführt

Die in der Aufstellung enthaltenen Seitenpreise, differenziert nach verschiedenen Erschließungstiefen, berechnen sich jeweils aus den Gesamtkosten des Projektes. Die Kosten enthalten damit auch insbesondere Ausgaben für Projektmanagement, Vorarbeiten des Scannens, Erfassen von über die Strukturdaten hinausgehenden Metadaten, sonstige Erschließungsaufwände, die Bereitstellung, aber auch Sachkosten wie Anschaffung eines Scanners, Softwarelizenzen oder Dienstleisterkosten. Nicht ausgewiesen sind die Scankosten im engen Sinn[86].

3.8.3 Spezifische Rahmenbedingungen der Pilotprojekte

Ein kurzer Abriss der jeweils spezifischen Rahmenbedingungen der Pilotprojekte, die in der Methodenwahl bewusst komplementär angelegt waren, illustriert die oben skizzierten Abhängigkeiten bei der Betrachtung der Kostenfaktoren.

Die SLUB Dresden erprobte ein Verfahren zur kostengünstigen Massendigitalisierung mit Mikrofilmen als Vorlage. Auch wenn bei diesem Teilprojekt die errechneten Kosten pro Seite am niedrigsten liegen, zeigt sich dennoch, dass der in anderen Kontexten oftmals als sehr gering angegebene Nettoseitenpreis des reinen Mikrofilmscans mit z.T. deutlich unter 0,10

[86] Also die Kosten für den eigentlichen Vorgang des Scannens selbst, d.h. Dienstleisterkosten bzw. anteilige Hardwarekosten sowie anteilige Personalkosten für Scannen, Strukturdatenerfassung und Qualitätskontrolle. Vgl. auch die Vorgaben auf S. 6, Fußnote 2, der DFG-Praxisregeln „Digitalisierung“ (Stand 12/16). Die Berücksichtigung der oben genannten Kostenfaktoren für die Erstellung einer belastbaren Kostenkalkulation ist aus Sicht des Konsortiums unabdingbar.

Cent nicht die wirklichen Kosten im Projekt widerspiegelt und weitere Faktoren einberechnet werden müssen. Im vorliegenden Projekt beinhaltete dies v.a. die Projektkoordination, die Anschaffung eines Scanners, die Ausgaben-/Artikelseparierung und die Erfassung / bibliographische Erschließung der Ausgaben mit Datum. Die Qualitätskontrolle war auf das Trennen der Ausgaben nach dem Scanprozess und die Seitenpaginierung konzentriert.

Die ULB Sachsen-Anhalt digitalisierte inhouse vom Original und erprobte die Artikelseparierung. Getestet wurden verschiedene Verfahren bei der Ausgabensegmentierung und Tiefenstrukturierung. Im Test überzeugte das selbstlernende halbautomatische Verfahren. Im Vergleich zu einer manuellen Strukturdatenerfassung, wie z.B. bei der BSB München oder SuUB Bremen, fließt dies kostenmindernd ein. Hauptfaktor dabei ist die Zeitersparnis. Qualitätssichernde Maßnahmen wurden gleichwohl während des gesamten Projektes durchgeführt. Hinzu kommt die Erprobung persistenter Adressierbarkeit auf feingranularer Ebene.

Die BSB München digitalisierte vom Original. Dabei wurden das Scannen, die OCR-Bearbeitung und die Artikelseparierung an einen Dienstleister vergeben. Folgende spezifischen Kostenfaktoren müssen bei einer Zusammenarbeit mit einem Dienstleister beachtet werden:

- Abwicklung des Ausschreibungsverfahrens: Erstellung der Leistungsbeschreibung (Pflichtenhefte) für die Ausschreibung der Scan- und OCR, Prüfung der Angebote
- Bestandsvorbereitung (Preprocessing): Erstellung von Transportprotokollen/Schadensprotokollen durch konservatorische Prüfung
- Dokumentation und Monitoring: Liefertermine, Reklamationsworkflow, Transportprotokolle
- Digitalisierung (Images und OCR): Kontrolle der vom Dienstleister nach festen Zeitplänen gelieferten Images (Überprüfung auf Vollständigkeit der Datenlieferung, der Lesbarkeit sowie der Einhaltung der geforderten Scanparameter; fehlerhaft gelieferte Daten sind zu reklamieren).

Nicht alle dieser Faktoren, wie z.B. Qualitätskontrolle, entfallen bei einer Inhousedigitalisierung. Vergabeverfahren, erhöhter Kommunikationsaufwand sowie Unklarheiten über den tatsächlichen Umfang des Reklamationsworkflows sind aber als Besonderheiten zu berücksichtigen. Die Kosten für einen Dienstleister sind allerdings auch immer von den konkreten Angeboten im Ausschreibungsverfahren und daher von der Marktsituation abhängig. Sie können jeweils deutlich divergieren. Nicht zwangsläufig ist eine Digitalisierung im Outsourcing günstiger als inhouse; es empfiehlt sich ggf. eine Marktsichtung.

Die SuUB Bremen digitalisierte mit den deutschsprachigen Zeitungen des 17. Jahrhunderts Sondermaterialien. Die Zeitungen wurden in Graustufe digitalisiert, bis zur Ausgabenebene strukturiert und manuell erschlossen. Die Ergebnisse dieses Projekts im Hinblick auf zu berücksichtigende Kostenfaktoren können nur bedingt auf andere Zeitungsdigitalisierungsprojekte übertragen werden. Als materialbedingten Besonderheiten sind hervorzuheben:

- Die physische Beschaffenheit der Digitalisierungsvorlagen: Reproduktionen von Mikroformen, Loseblattsammlungen

- Die Differenziertheit der Materialien: hohe Anzahl von Zeitungsausgaben (ca. 80.000) mit geringem Umfang, lückenhafte Überlieferung
- Die inhaltliche Komplexität der Materialien: insb. Berichtszeiträume und Erscheinungsdaten in zwei Kalendersystemen
- Die geringe Standardisierung der Materialien: wechselnde Titel, wechselndes Layout und Schriftarten innerhalb der Frakturfamilie oftmals innerhalb einer Ausgabe, teils handgeschriebene Zeitungen

Die physische Beschaffenheit der Digitalisierungsvorlagen führte dazu, dass die Vorlagen zwar größtenteils (ca. 60%) mit einem Durchzugsscanner gescannt werden konnten und somit der Aufwand für das Scannen selbst gering war. Doch war die Qualität der so entstehenden Digitalisate auch in technischer Sicht nicht hinreichend für eine automatische Volltexterkennung.

Die Differenziertheit, die inhaltliche Komplexität und die geringe Standardisierung der Materialien bedeuteten einen hohen Erschließungsaufwand bei gleichzeitiger Notwendigkeit einer manuellen Erschließung. Eine automatisierte Weiterverarbeitung war nicht möglich. Damit einhergehend waren hohe Aufwände in der wissenschaftlichen Projektbegleitung, die insb. beim Schließen von Bestandslücken und bei der notwendigen Zusammenarbeit mit der Fachwissenschaft zum Tragen kamen.

3.8.4 Fazit

Der Medientyp Zeitung, der im Rahmen einer möglichen Digitalisierungsinitiative über die Jahrhunderte betrachtet werden muss, zeichnet sich durch eine ausgeprägte Heterogenität in Bezug auf die physische und strukturelle Beschaffenheit der Vorlagen aus. Wie die Übersicht der Kostenfaktoren und ihre jeweils konkrete Ausprägung unter den Rahmenbedingungen der einzelnen Projekte zeigen, lassen sich zwar gewisse Kostenkorridore benennen, eine pauschalisierte Ableitung konkreter Kostensätze ist für die Zeitungsdigitalisierung hingegen nicht möglich.

Die oben angegebenen Seitenpreise können daher nur als erste Orientierung dienen. Für die konkrete Beantragung von Zeitungsdigitalisierungsprojekten wird empfohlen, nach der Bestandsprüfung das zu wählende Digitalisierungsverfahren abzuleiten (vgl. auch die Handreichungen zum Masterplan) und mittels Marktsichtungen (z.B. Scanner; Dienstleister), aktualisierten Informationen (z.B. OCR-Lizenzkosten) sowie Stichproben (v.a. Personalbedarf für Erfassung, Qualitätskontrolle etc.) den Kostenrahmen des Projektes anhand des entsprechenden Mengengerüstes zu kalkulieren. Festzuhalten ist, dass die unmittelbaren Kosten des Scannens nur einen Teilfaktor in der gesamten Kostenabbildung darstellen und sich so Unterschiede zwischen den konkreten Scanverfahren (Original / Mikrofilm; Outsourcing / Insourcing) relativieren.

3.9 Mengen- und Kostengerüst

Mit Stand Ende 2015 sind 21.583 Zeitungsunternehmungen des deutschen Sprachgebiets bis 1945 in der ZDB erfasst. Es gibt noch deutliche Unschärfen bei der Trennung zwischen Zeitungen und zeitungsähnlichen Journalen sowie noch nicht aufgenommene Bestände aus Archiven und Bibliotheken. Auch sind keine Seitenzahlen in der ZDB verzeichnet. Versucht man dennoch eine vage Hochrechnung der Seitenumfänge aufgrund der im österreichischen

ANNO-Portal enthaltenen 840 Zeitungen und zeitungsähnlichen Journale[87] mit rund 15 Mio. Seiten und der mit Google an der BSB München digitalisierten 1.050 Zeitungen mit rund 10 Mio. Seiten, so ergibt sich aus beiden Digitalisierungsgroßprojekten ein durchschnittlicher Seitenumfang pro Zeitung von 13.690 Seiten. Hochgerechnet auf 21.583 in der ZDB nachgewiesene Zeitungen ergäbe sich daraus rein rechnerisch ein geschätzter Gesamtumfang deutscher Zeitungen bis 1945 von 295 Mio. Seiten.

Bei den Drucken des 16., 17., und 18. Jahrhunderts wurde bzw. werden jeweils 50% der Titel mit DFG-Förderung digitalisiert, um eine kritische Masse für Forschung und Lehre zu erreichen und nachnutzbare Strukturen zu bilden. Wollte man einen ähnlichen Ansatz auf die Zeitungen übertragen, und reduziert aufgrund der Fülle des Materials insgesamt im Vergleich zu den VD-Projekten auf rund ein Drittel der deutschen Zeitungsmenge bis 1945, so wäre mit rund 7.000 Titeln und rund 98 Mio. Seiten eine kritische Masse deutscher Zeitungsvielfalt für Forschung und Lehre erreicht. Um zu einem solchen Umfang zu gelangen, ist ein Zusammenspiel vieler Partner notwendig (besitzende Einrichtungen international, national, regional und lokal + Google).

Generell ist festzustellen, dass die Anzahl der Digitalisate beständig wächst: Durch die Google-Projekte in München und Wien, durch anglo-amerikanische Projekte, durch EU-, DFG- und Regionalprojekte und nicht zuletzt durch das hier beschriebene Pilotprojekt mit dem Anspruch, zu einer nationalen Koordinierung und Strukturbildung beizutragen. Zu den weiteren DFG-geförderten Digitalisierungsprojekten zählen z.B. einige Satire-Zeitungen/Zeitschriften, eine Auswahl illustrierter Magazine der Moderne, die Exilzeitungen oder eine Auswahl von DDR-Zeitungen, die sämtlich in ein künftiges Zeitungsportal mit übergreifender Volltextsuche einzubinden sind.

Inzwischen gibt es schon über 4.000 Nachweise in der ZDB darüber, dass Zeitungsbestände digitalisiert wurden oder beabsichtigt ist, diese zu digitalisieren. Dabei handelt es sich allerdings auch um digitalisierte Teilbestände bis auf Einzelausgabenebene, so dass nicht darauf geschlossen werden kann, damit seien schon ca. 20% des Gesamtumfangs von rund 21.500 Zeitungen vollständig digitalisiert oder in Vorbereitung.

Geht man davon aus, dass die rund 4.000 digitalisierten bzw. sich in Vorbereitung befindlichen Zeitungen einen Umfang von 40 Mio. Seiten zählen, dann fehlten rund 58 Mio. Seiten, um die Digitalisierung eines Drittels der deutschen Zeitungen bis 1945 zu erreichen. Dies setzt voraus, dass die begonnenen bzw. angekündigten Digitalisierungsaktivitäten in einem nennenswerten Anteil weitergeführt bzw. abgeschlossen werden. Ausgehend von der genannten Zielzahl von 58 Mio. zu digitalisierenden Zeitungsseiten und der damit verbundenen Intensivierung der Zeitungsdigitalisierung sollten folgende weitere Überlegungen zu Kosten und Ausrichtung einer entsprechenden Förderinitiative zu Grunde gelegt werden.

Für Schwerpunkte einer DFG-Förderung haben im Rahmen des Projekts Wissenschaftlerinnen und Wissenschaftler Auswahlkriterien historischer Zeitungen benannt, um zu einem repräsentativen Sample digitalisierter deutscher Zeitungen bis 1945 zu gelangen, die als Quellen für Forschungsvorhaben dienen können. Dazu zählen fachlich definierte Sammlungen, etwa zur Presse des Kolonialismus oder der NS-Zeit, und aus

[87] Die ÖNB zählt in ANNO aktuell 480 Tageszeitungen mit 12,7 Mio. Seiten und 360 zeitungsähnliche Journale mit 3,1 Mio. Seiten.

presse- und mediengeschichtlicher Sicht typische Vertreter der Zeitungsvielfalt in angemessener chronologischer und geographischer Breite. Fachlich ist die Kerngruppe der zu digitalisierenden Zeitungen also umrissen.

Die Pilotprojekte haben Preise pro Seite zwischen 0,77 und 1,57 EUR für eine Farbdigitalisierung vom Original mit verschiedenen Erschließungstiefen/Weiterverarbeitungsschritten und mindestens 0,44 EUR (ohne OCR) vom Film ermittelt.[88] Setzt man einen Mittelwert von 1,08 EUR Seitenpreis für eine Digitalisierung vom Original und 0,64 EUR für die Digitalisierung vom Film, beide jeweils mit OCR, zugrunde, ließen sich bei einem Verhältnis 50:50 Scans in Graustufe vom Film und Farbscans von Originalen für rund 5,5 Mio. EUR Gesamtkosten pro Jahr innerhalb eines Jahres 6,45 Mio und innerhalb von 9 Jahren die angestrebte Gesamtzahl von 58 Mio. Zeitungsseiten digitalisieren. Nach einer Anlaufphase wird sich vermutlich im weiteren Verlauf der Durchsatz steigern lassen, wenn durch Massenverfahren im Bereich der Mikrofilmdigitalisierung und in der Nachverarbeitung des Materials die durchschnittlichen Seitenpreise gesenkt werden können. Für weitere Förderungen über das nun zur Förderung angepeilte Drittel des Zeitungsvolumens hinaus ist mit weiteren Skaleneffekten zu rechnen.

Unabhängig vom Beginn einer Hauptphase sollte innerhalb der ersten beiden Jahre die ZDB-Datenbasis so weit verbessert werden, dass präzisere Mengengerüste und verbesserte statistische Auswertungen zur Priorisierung von Beständen möglich sind.

Empfehlungen

Es wird empfohlen, angesichts der großen Menge, Vielfalt und Umfänge deutscher Zeitungen bis 1945 zunächst die Digitalisierung eines Drittels des geschätzten Gesamtumfangs anzustreben (das sind nach einer groben Hochrechnung unter Einrechnung bereits digitalisierter Zeitungsbestände rund 58 von insgesamt geschätzten 295 Mio. Seiten). Dazu sind neben Eigenleistungen der besitzenden Einrichtungen regionale und nationale Förderprogramme (und wo immer möglich PPP-Modelle) ergänzend zu einer DFG-Förderung notwendig.

Die Pilotbibliotheken empfehlen, mit einer Laufzeit von insgesamt 9 Jahren (aufgeteilt auf 3 x 3 Jahre) und einem Volumen von 5,5 Mio. EUR Gesamtkosten pro Jahr (aufzuteilen auf 1/3 Unterhaltsträger (= 1,85 Mio. EUR), 2/3 DFG (= 3,7 Mio. EUR) ein Förderprogramm „Digitalisierung historischer Zeitungen in Deutschland“ aufzulegen. Zu berücksichtigen ist dabei, dass das Digitalisierungsprogramm durch Fördermaßnahmen zur Verbesserung der Informationsinfrastrukturen (Zeitungsportal der DDB mit Einbindung aller bisher digitalisierten Zeitungen, Verbesserung der Datenbasis der ZDB) und durch Investitionen in die Weiterentwicklung technischer Erschließungsverfahren von Zeitungen (OCR, Named Entity Recognition, Bilderkennung etc.) begleitet werden sollte.

[88] Die Kosten für OCR aus der SLUB waren nicht repräsentativ, wir setzen aufgrund anderer Erfahrungen (SBB Berlin) und der sonst ermittelten Durchschnittswerte 20 Cent an.

4. Kriterien für eine Auswahl zu digitalisierender Zeitungen

4.1 inhaltlich

Die Wissenschaftlerinnen und Wissenschaftler des Workshops im Herbst 2014 in der SuUB Bremen[89] wünschten sich „so viel Zeitungsdigitalisierung wie möglich". Es sei „keine Zeit mehr zu verlieren", um eine breitere Quellenbasis für Textrecherchen und Forschungsfragen zu schaffen, um mehr Zeitungsanschauung für Forschung und Lehre zu ermöglichen, um Zeitungen als „gesellschaftlichem Phänomen sui generis" die notwendige Sichtbarkeit zu geben. Zur Gewinnung einer kritischen Masse relevanter Zeitungen seien die Anforderungen und Fragestellungen der verschiedenen Wissenschaftsdisziplinen zu berücksichtigen. Grundsätzlich sollten wie bei der digitalen Transformation der Drucke des 16. bis 18. Jahrhunderts der ganze geographische Raum und die ganze zeitliche Erstreckung angemessen repräsentiert sein. Bei Zeitungen als dem Leitmedium der Moderne seien neben dem gemeinfreien Zeitraum 1605 bis ca. 1920 auch Erscheinungszeiträume bis zur Gegenwart, also auch urheberrechtsgeschützte Zeitungsunternehmen, zu berücksichtigen (zur Lizenzierung von Zeitungen s. unter 4.2). Wissenschaftler historisch forschender Fachdisziplinen benötigten für Big Data-Analysen und zur Anwendung neuer Methoden der Digital Humanities vor allem Volltexte. Für Politikwissenschaftler oder Historiker seien aber auch schon über eine Kalendersuche zugängliche Images ein konstruktives Angebot, das die Arbeit wesentlich erleichtere. Ob eine Imagedigitalisierung als ausreichend bewertet wird oder eine digitale Volltexterschließung (OCR/OLR) notwendig wird, sei abhängig von den Anforderungen der jeweiligen wissenschaftlichen Forschungsfragestellung, wobei eine OCR-Erschließung für alle Disziplinen grundsätzlich wünschenswert sei. Zur Erforschung der in Zeitungen enthaltenen umfangreichen Bildanteile sei auf eine ausreichende Bildqualität zu achten.

Empfehlungen

Da aufgrund des Umfangs und der Kosten nicht alle Zeitungen in absehbarer Zeit digitalisiert werden können, wurde während des Workshops eine Digitalisierung nach folgenden systematischen Auswahlkriterien (vgl. auch Tabelle 11, S. 61f.) vorgeschlagen. Dieser Priorisierung schließen sich die Pilotpartner an:

1. Abdeckung des typologischen Spektrums (Zeitungen großer Zentren, genannt wurden für eine erste Phase konkret Hamburg, Köln, Berlin, Leipzig, Frankfurt am Main, München, Dresden, Breslau, Königsberg; Kreisblätter, Intelligenzblätter etc.)
2. Berücksichtigung von „Dauerbrennern" (Zeitungen mit langer Lebensdauer und von großer Reichweite)
3. Digitalisierung der Leitmedien (Zeitungen, denen historisch eine Leitfunktion zukam, bspw. durch Verbreitung, Leistung, Prominenz der Mitarbeiter, Reputation)
4. Berücksichtigung von „Innovatoren" (Zeitungen, die in der jeweiligen pressehistorischen Phase Neuerungen gebracht haben, z.B. Rheinischer Merkur 1814-1816, Kieler Blätter 1815-1819, Oppositionsblatt oder Weimarische Zeitung 1817, Rheinische Zeitung 1842/43, Parteizeitungen nach 1848)

[89] Teilnehmer waren: Prof. Fotis Jannidis (Linguist, Digital Humanities, Würzburg); Prof. Konrad Dussel (Medienhistoriker, Mannheim); Prof. Oliver Pfefferkorn (Germanist, Halle-Wittenberg); Prof. Patrick Rössler (Kommunikationswissenschaftler, Erfurt); Prof Jürgen Wilke (Medien- und Kommunikationshistoriker, Mainz); Dr. Bernd Florath (Behörde des Bundesbeauftragten für Stasi-Unterlagen, Berlin); Prof. Ulrich Johannes Schneider (Philosoph, Bibliothekar und Bibliothekshistoriker, Leipzig).

5. Thematische Kollektionen, etwa Kolonialzeitungen, NS-Zeitungen u.a.
6. Digitalisierung (presse)historischer Exponenten (Zeitungen, die für bestimmte Phasen exemplarische Bedeutung hatten, wie z.B. die erste Phase der deutschen Exilpresse im Vormärz mit Zentren in Straßburg, Paris und Zürich).
7. Die Abdeckung des politischen Spektrums, der Diversifizierung der Presse im 19. Jahrhundert entsprechend
8. Die Abdeckung des regionalen Spektrums, im 19. Jhdt. z.B. aus den Königreichen Preußen, Bayern, Hannover, Sachsen, Württemberg, den Großherzogtümern Baden und Hessen.

Es wird empfohlen, eine kritische Masse qualitativ wichtiger Zeitungen aus den vorgeschlagenen Segmenten zu digitalisieren und damit einen repräsentativen Querschnitt ins Netz stellen. Die am Workshop beteiligten Wissenschaftler empfahlen, zugunsten von Geschwindigkeit und Menge ggf. auch Abstriche bei der Qualität der Digitalisierung hinzunehmen. Notwendig sei die Digitalisierung möglichst vollständiger und langlebiger Zeitungsunternehmen. Um den Digitalisierungsprozess zu beschleunigen und kostengünstiger zu gestalten, sollte auf Vorarbeiten, insbesondere auf die Masterfilme von Archiven und Bibliotheken bzw. des Mikrofilmarchivs der deutschsprachigen Presse wann immer möglich zurückgegriffen werden.

4.2 rechtlich

Wie auch in der Roadmap for Improving Access to Newspapers[90] dargelegt, spielen vor allen technischen Erwägungen Fragen der rechtlichen Verfügbarkeit und die Nachnutzungsmöglichkeiten eine entscheidende Rolle. Die umfangreiche Verfügbarmachung von Volltexten unter freien Lizenzen (Public Domain Mark für Scans und Volltexte, CC0 für Metadaten), wie auch im Europeana Newspapers-Projekt erfolgt, ermöglicht neue Forschungsvorhaben, wie sie in den Bremer Zeitungsworkshops von Wissenschaftlern gefordert wurden und im Rahmen des EU-Projekts in Interviews mit Wissenschaftlern dokumentiert sind. [91]

Angesichts der 70jährigen Schutzfrist gem. §64 UrhG sind mehr als zwei Drittel der Druckwerke des 20. Jahrhunderts noch nicht gemeinfrei. Dieser Urheberrechtsschutz gilt auch für Lichtbildwerke. Die weitreichenden urheberrechtlichen Einschränkungen erschweren es, Zeitungen des 20. Jahrhunderts für Zwecke von Forschung und Lehre digital frei zugänglich zu machen. Werke von 1945 verstorbenen Autoren werden am 1.1.2016 gemeinfrei. Bis 1945 verfasste Beiträge namentlich genannter Autoren, die nach 1945 lebten, sind also noch urheberrechtlich geschützt.[92] Die Gesetzesnovelle zur Verlängerung der Bildschutzrechte im Jahre 1985 verlängerte den Urheberschutz für Lichtbildwerke auf 70 Jahre, wenn ihre 25jährige Schutzfrist bis 1985 noch nicht abgelaufen war. Bilder in Zeitungen vor 1945 sind demnach nicht mehr urheberrechtlich geschützt.

[90] http://www.europeana-newspapers.eu/wp-content/uploads/2015/05/Roadmap_for_Improving_Access_to_Newspapers_final.pdf
[91] Vgl. http://www.europeana-newspapers.eu/category/interviews-with-researchers/
[92] Zur Problematik des Urheberrechts siehe: Klimpel, Paul: Urheberrecht, Praxis und Fiktion. Rechteklärung beim kulturellen Erbe im Zeitalter der Digitalisierung. In: Klimpel, Paul/Ellen, Euler (Hrsg.): Der Vergangenheit eine Zukunft. Kulturelles Erbe in der Digitalen Welt, Berlin 2015, S. 168-188. Derzeit hat die Deutsche Digitale Bibliothek einen Think Tank zu Rechtsfragen im Digitalen Zeitalter gegründet. Vgl. https://www.deutsche-digitale-bibliothek.de/content/ueber-uns/aktuelles/recht-und-kulturelles-erbe-im-digitalen-zeitalter-deutsche-digitale-bibliothek-veroeffentlicht-dritte-folge-ihrer-thementrailerreihe (abgerufen am 10.12.2015).

Eine digitale Bereitstellung von Zeitungen nach 1945 ist rechtlich kompliziert und kann faktisch nur durch die Verlage selbst erfolgen. Die Staatsbibliothek zu Berlin konnte mit den Rechtsnachfolgern von drei DDR-Zeitungen eine Vereinbarung schließen, die es Nutzern ermöglicht, diese Zeitungen nach ihrer Anmeldung und Authentifizierung über das Zeitungsportal der Staatsbibliothek zu Berlin frei einzusehen.[93]
Aus der Zeit des Nationalsozialismus 1933-1945 sind in Deutschland bislang nur wenige zeitungsähnliche Serien durch Bibliotheken digitalisiert und online bereitgestellt worden. Die Nationalbibliothek der Niederlande in Den Haag ist mit ihrer Open Access-Strategie in die Offensive gegangen und hat im Zeitungsportal Delpher rund 300.000 Seiten aus 1.000 Zeitungen der Kriegszeit 1940-45 zugänglich gemacht, Zeitungen also, die von der Zeit der deutschen Besatzung geprägt sind. In einer öffentlichen Diskussion hat die Nationalbibliothek der Niederlande die Veröffentlichung als Teil einer umfangreichen wissenschaftlichen Aufklärung über die Zeit des Nationalsozialismus gerechtfertigt. Nachfahren der in den Zeitungen genannten Personen hätten durchaus auch Verständnis für diese Form öffentlicher Aufarbeitung gezeigt.[94]

Die SLUB Dresden hat die Dresdner NS-Parteizeitung „Der Freiheitskampf" (1930-1945) digitalisiert, die wohl letzte, noch bis zum 8. Mai 1945 gedruckte NS-Zeitung in Deutschland. Diese Zeitung wird vom Hannah Arendt-Institut für Totalitarismusforschung (HAIT) Dresden intensiv erforscht, das zwischen 2009 und 2014 rund 70.000 EUR in die Erschließung[95] von 4 der 16 Jahrgänge investiert hat. Angesichts des unmittelbaren Forschungsbedarfs lag es nahe, diese Zeitung in eine Pilotphase zur Digitalisierung und Erschließung wissenschaftsrelevanter Zeitungen einzubeziehen.

Die juristische Fakultät der Technischen Universität Dresden hat mit zwei Stellungnahmen das Vorhaben beratend begleitet. In einer strafrechtlichen Einschätzung kommt sie zu dem Schluss, dass eine mögliche Strafbarkeit wegen des Verbreitens von Propagandamitteln und Kennzeichen verfassungswidriger Organisationen, wegen Volksverhetzung, wegen des öffentlichen Zugänglichmachens von Gewaltdarstellungen und wegen Bekenntnisbeschimpfung nicht vorliegt, wenn mit der Veröffentlichung eine angemessene Distanzierung und Kontextualisierung erfolgt.

Die Zugänglichmachung von NS-Quellen unterstützt Forschung und Lehre und ist von hoher wissenschaftlicher Bedeutung. Eine zweite ausführliche Stellungnahme zu den urheberrechtlichen Rahmenbedingungen der Veröffentlichung des „Freiheitskampfs" kommt zu dem Schluss, dass knapp 25% der Beiträge namentlich gezeichnet sind und deshalb Zustimmungen der Journalisten bzw. ihrer Erben notwendig sind, wenn die Beiträger nach 1945 gelebt haben. Die Ermittlung der Autoren bzw. ihrer Erben ist ohne unverhältnismäßigen Aufwand jedoch nicht zu leisten, und sie ist auch nicht zielführend, da zu erwarten ist, dass die Autoren bzw. Nachfahren eine Veröffentlichung eines Beitrags aus der NS-Zeit auf Anfrage eher ablehnen als dieser zustimmen dürften.[96] Auch stellen

[93] Vgl. http://zefys.staatsbibliothek-berlin.de/ddr-presse/ (abgerufen am 10.12.2015).
[94] Vgl. https://www.kb.nl/nieuws/2010/achtergronden-van-de-foute-kranten-in-de-kb-krantenwebsite.
[95] Das HAIT hat die Jahrgänge in einer Datenbank bis auf Artikelebene nach einem Themenkatalog verschlagwortet. Auch eine Personendatei wurde angelegt. Die Datenbank ist zurzeit über einen Einzelleseplatz zugänglich.
[96] Zum Vergleich: Die zwei Digitalisierungsprojekte der Deutschen Nationalbibliothek "Exilpresse digital" und "Jüdische Periodika aus NS-Deutschland" wurden aufgrund der Rechtslage auf Einzelplatzlösungen im Lesesaal beschränkt. Von 239.270 digitalisierten Artikeln im Projekt "Exilpresse digital" waren bei 168.347 Artikeln keine

Vergütungen von Beiträgen keine nennenswerten Anreize dar. Bei einer Veröffentlichung ohne Zustimmung würde Autoren und Erben kein nennenswerter wirtschaftlicher Schaden zugefügt, etwaige Schadensersatzansprüche dürften entsprechend gering ausfallen. Würde ein Verfasser oder Erbe jedoch die Verletzung seiner Urheberrechte geltend machen, dürften seine Beiträge zukünftig nicht mehr öffentlich zugänglich gemacht werden. Dies würde dem wichtigsten Ziel widersprechen, die Quelle insgesamt und uneingeschränkt für Lehre und Forschung frei nutzbar zu machen. Deshalb und wegen des aktuell aufgeheizten aggressiven Rechtspopulismus in Deutschland und Europa will die SLUB eine Freischaltung vorerst zurückstellen, um diese dann rechtssicher umsetzen zu können. Konkret sollten die geplanten Veränderungen bei den Regelungen zu vergriffenen Werken und zur Verbesserung der Wissenschaftsschranke forciert werden. Es kann politisch nicht richtig sein, dass urheberrechtliche Regelungen, die ihren intendierten wirtschaftlichen Zweck verfehlen, die Freiheit der Wissenschaft bei der Erforschung der Quellen des Nationalsozialismus drastisch einschränken.[97] Auch ist durch wissenschaftliche und publizistische Arbeit anzustreben, dass die politischen Diskussionen um die 2016 erfolgte wissenschaftliche Veröffentlichung von Adolf Hitlers „Mein Kampf" und die parallel laufenden Diskussionen über das Verbotsverfahren der NPD zu einem politischen und wissenschaftlichen Verständnis führen, dass Veröffentlichungen von Quellen des Nationalsozialismus auch via Internet notwendig sind und rechtssicher vorgenommen werden können.[98] Die SLUB und das Hannah Arendt-Institut richten bis zur weiteren juristischen und politischen Klärung jeweils einen öffentlich zugänglichen Einzelleseplatz ein.

Die SLUB empfiehlt, im Rahmen einer Hauptphase zur Zeitungsdigitalisierung eine Kollektion „Zeitungen der NS-Zeit" zu digitalisieren und diese dann zusammen mit dem „Freiheitskampf" – wissenschaftlich kontextualisiert und publizistisch begleitet – rechtssicher zu veröffentlichen.[99]

Autoren verzeichnet, 12.881 wurden unter Pseudonym und 58.042 unter Realnamen veröffentlicht. Für ca. 13.000 Autoren müsste daher der Rechtestatus geklärt werden, was nicht leistbar war. Vgl. Asmus, Sylvia/Zechmann, Dorothea: Exilpress digital und Jüdische Periodika aus NS-Deutschland. Zwei Digitalisierungsprojekte der Deutschen Nationalbibliothek. In: Klimpel, Paul/Ellen, Euler (Hrsg.): Der Vergangenheit eine Zukunft. Kulturelles Erbe in der Digitalen Welt, Berlin 2015, S. 226-234, hier S. 230. . [Anmerkung Mai 2017: Diese Beschränkung wurde inzwischen wieder aufgehoben.]

[97] Zur Relevanz der Quellenuntersuchungen vgl. u.a. Peter Longerich: NS-Propaganda in Vergangenheit und Gegenwart. Bedeutung der nationalsozialistischen Tagespresse für Zeitgenossen und Nachgeborene. In: Kuchler, Christian (Hg.): NS-Propaganda im 21. Jahrhundert zwischen Verbot und öffentlicher Auseinandersetzung. Köln 2014, S. 15-26; Karl Christian Führer: Die deutsche Tagespresse im Zweiten Weltkrieg. Fakten und Fragen zu einem unerforschten Abschnitt der NS-Mediengeschichte. In: ZfG 60 (2012), S. 417-440; Christian A. Braun: Nationalsozialistischer Sprachstil. Theoretischer Zugang und praktische Analysen auf der Grundlage einer pragmatisch-textlinguistisch orientierten Stilistik. Heidelberg 2007.

[98] Zur Diskussion um Hitlers "Mein Kampf" vgl. Hitlers "Mein Kampf", Aus Politik und Zeitgeschichte (APuZ), 65.Jg. (2015), Heft 43-45.

[99] In Baden-Württemberg wird mit dem Ansatz der Public History die NS-Zeit der Ministerien erforscht. NS-Quellen werden kommentiert und wissenschaftlich begleitet öffentlich einem breiten Publikum zugänglich gemacht. Das Portal geht sogar noch einen Schritt weiter: "Der Projektteilbereich ‚Public History' hat sich zum Ziel gesetzt, über den im historischen Arbeitsprozess üblichen Austausch zwischen Experten aus den Bereichen Wissenschaft, Archivwesen und Verwaltung hinaus gerade auch die bis dato nur in sehr wenigen Fällen ausgeschöpften Interaktionsprozesse mit Bürgerinnen und Bürgern zu fördern." Arendes, Cord: Moderne Wissenschaftskommunikation als Informations- und Interaktionsprozess: Start der App „NS-Ministerien in BW", am 1. September 2015, (URL: http://ns-ministerien-bw.de/2015/09/moderne-wissenschaftskommunikation-als-informations-und-interaktionsprozess-start-der-app-ns-ministerien-in-bw/).

Empfehlungen

Grundsätzlich sollen Zeitungsdigitalisate einschließlich Volltexten unter freien Lizenzen für die uneingeschränkte Nachnutzung verfügbar sein.

Um Zeitungen der Zeit 1933 bis 1945 einschließlich der Exilpresse frei ins Netz stellen zu können, sollen die politischen Anstrengungen auf allen Ebenen verstärkt werden, bei den angestrebten Lösungen zu den vergriffenen Werken Zeitungen ausdrücklich einzubeziehen.

Der Wunsch der Wissenschaftler, auch urheberrechtsbewehrte Zeitungen bereitzustellen, wird von den Arbeitsgruppen zu regionalen und nationalen Lizenzkonsortien aufgegriffen. Empfohlen wird bei nationalen Lizenzierungen von Zeitungen die Einbindung der Recherchefunktion in ein künftiges Zeitungsportal.

4.3 konservatorisch

Digitalisierung vom Original oder vom Film?

Die internationalen Zeitungsportale in den USA, in England und in Australien setzen vielfach auf Mikrofilmen auf, die seit den 60er Jahren in großem Umfang hergestellt wurden. Diese Mikrofilmaktionen hatten den Publizisten Nicholson Baker zu seinem Bestseller „Double Fold. Libraries and the Assault on Paper" veranlasst (2002, dt.: Der Eckenknick, oder wie die Bibliotheken sich an den Büchern versündigen, 2005). Darin beschreibt er, wie zahlreiche englische und amerikanische Bibliotheken, darunter die British Library und die Library of Congress, Zeitungen verfilmten, um anschließend die Originale aus Platz- und Kostengründen kommerziell (z.B. als Geburtstagszeitungen) verwerten zu lassen oder zu makulieren. Zeitungen würden, so Baker, von zahlreichen Bibliothekaren als nicht erhaltbar eingestuft und deshalb vernachlässigt. Dabei würde die Einzigartigkeit dieser Quelle übersehen, darunter auch die der umfangreichen (farbigen) Bildbeilagen, die oft nur mangelhaft verfilmt worden seien. Baker klagte insbesondere an, dass die Verlage schlechte Mikrofilme teuer verkauft hätten – und für den Schutz der Originale nichts getan worden sei.

Systematische Makulierungsaktionen wie die von Baker beklagten sind in den zuständigen Staats- und Landesbibliotheken in Deutschland nicht bekannt geworden. Allerdings haben viele Einrichtungen enorme Platzprobleme. Gebundene Zeitungen sind schwer, haben Überformate und müssen aus konservatorischen Gründen liegend aufbewahrt werden, was viele Einrichtungen aus Raumnot nicht leisten können. In Archiven sind häufig auch ungebundene Zeitungen vorzufinden. Die Deutsche Nationalbibliothek, die über umfangreiche Bestände seit ihrer Gründung 1913 verfügt, hatte sich nach Einführung der Mikroverfilmung entschlossen, Zeitungen nur als Mikroformen zu sammeln. Inzwischen sammelt sie Tageszeitungen in elektronischer Form.

Die Sammlung von Pflichtexemplaren ist in Deutschland gesetzlich geregelt, nicht aber speziell der Umgang mit Zeitungen. Eine Koordinierung der Bestandserhaltung von Zeitungen fehlt bislang, entsprechend unklar ist der Umgang mit Mehrfachexemplaren. Eine aktuelle Erhebung über „Die Erhaltung des schriftlichen Kulturguts in Archiven und Bibliotheken in Deutschland. Bundesweite Handlungsempfehlungen für die Beauftragte der Bundesregierung für Kultur und Medien und die Kultusministerkonferenz"[100] ist im Oktober 2015 erschienen und hat die Überlieferung von bis zum Jahr 1945 erschienenen Zeitungen

[100] Vorgelegt von der Koordinierungsstelle für die Erhaltung des schriftlichen Kulturguts (KEK) an der Staatsbibliothek zu Berlin - Preußischer Kulturbesitz. Berlin 2015, 103 S.

in die Umfragen einbezogen.[101] Danach gibt es in den abgefragten Bibliotheken mehr als 66.000 Zeitungstitel in mehr als 800.000 Bänden, in den Archiven der Kommunen, Länder und des Bundes mehr als 75.000 lfm Zeitungen. Von den Zeitungen in den Bibliotheken der Länder gelten 47,2% als stabil und benutzbar, 42,8% als gebräunt und brüchig, so dass eine Sekundärform dringend notwendig ist. 10% der Zeitungen sind bereits so geschädigt, dass von ihnen eine Sekundärform nicht mehr hergestellt werden kann.
In der Staatsbibliothek zu Berlin und in der Deutschen Nationalbibliothek gelten 50 bzw. 70% als verbräunt und brüchig, so dass Sekundärformen dringend notwendig sind; 5 bzw. 15% gelten als extrem brüchig, so dass die Originale nicht mehr nutzbar sind.
Vor diesem Hintergrund ist es sinnvoll und notwendig, schützenswerte wertvolle, insbesondere unikale Zeitungen zu digitalisieren. Wenn es von einer Zeitung mehrere Exemplare gibt und ein Mikrorollfilm sowie ein gutes Digitalisat vorliegen, muss es künftig auch möglich sein, nicht mehr benötigte Mehrfachexemplare aus Kosten- und Platzgründen zu makulieren.
Im Rahmen nationaler Verfilmungsprogramme von Bund und Ländern sind, auch mit Förderung der Volkswagen Stiftung und der Deutschen Forschungsgemeinschaft, in Deutschland zahlreiche Zeitungen verfilmt worden. Die Sicherungsverfilmungen sollten die fragilen Originale schützen und zugleich deren Inhalte nutzbar machen. Es wurden Masterfilme erstellt, von denen Nutzungskopien (Silberduplikatfilme und davon Diazokopien) angefertigt wurden. Bereits 1965 wurde in Hamburg das Mikrofilmarchiv der deutschsprachigen Presse e.V. (MFA) gegründet, das heute an das Institut für Zeitungsforschung in Dortmund angebunden ist. Die Masterfilme dieses Archivs werden also geschützt und stehen für die Erstellung von Nutzungsfilmen und für Zwecke der Digitalisierung zur Verfügung. Eine hauseigene Datenbank erfasst die verfilmten Bestände.[102] Das Archiv bewahrt aus aktuell 132 Mitgliedsarchiven und –bibliotheken rund 12.000 ganz oder teilweise verfilmte Zeitungstitel auf (42.000 Filmrollen). Davon wurden 50% zwischen 1960 und 1990, weitere 50% zwischen 1990 und 2015 hergestellt (darunter viele Nachkriegszeitungen). Es wird eingeschätzt, dass 50% der Filme von historischen Zeitungen, insbesondere die, die von der DFG und der Volkswagen Stiftung gefördert wurden, in guter Qualität vorliegen. Wie viele und welche digitalisiert werden sollten, ließe sich zuverlässig erst nach einer Integration der MFA-Daten in die ZDB ermitteln, in der dann alle Daten über originale, verfilmte und digitalisierte bzw. für eine Digitalisierung vorgemerkte Bestände zusammengeführt und gut überschaubar sind.
Während die frühen Filme teilweise schlecht lesbar sind, erfolgen die jüngeren Verfilmungen und Duplizierungen nach DIN 19057 und weiteren Normen durch erfahrene und zertifizierte Fachbetriebe in sehr guter Qualität. Häufig führt der schlechte Erhaltungszustand der Originale zu Wiedergabeproblemen, die für Verfilmung und Digitalisierung gleichermaßen zutreffen.
Im Rahmen von Bestandserhaltungsprogrammen einzelner Bundesländer wurden in den letzten Jahrzehnten regionale Filmarchive aufgebaut, die mit dem MFA kooperieren.
Wenn eine normgerechte Verfilmung eines Zeitungstitels vorliegt, ist zu empfehlen, zur Schonung des Originals und zur Dämpfung der Kosten vorhandene Mikrorollfilme als Digitalisierungsvorlagen zu wählen. Diese Empfehlung bedeutet gleichzeitig, die auf diesem Weg erzielbare Qualität als ‚kleinsten gemeinsamen Nenner' zu akzeptieren. Darüber hinaus

[101] Die Umfrage fand in den Bibliotheken staatlicher Trägerschaft statt, die im Handbuch der historischen Buchbestände aufgeführt sind.
[102] Vgl. http://www.mfa-dortmund.de/bestand.php

gehende Forderungen sind dann unrealistisch, wenn für einen Teil der verfilmten Zeitungstitel aus konservatorischen Gründen keine erneute Bearbeitung mehr möglich ist – auch keine Digitalisierung vom Original. In solchen Fällen sind vorhandene Mikrofilme die einzige noch nutzbare Quelle.
Die Wissenschaftler haben beim Bremer Workshop im Jahr 2014 vorgeschlagen, den Zeitungsbestand des Mikrofilmarchivs der deutschsprachigen Presse e.V. soweit möglich und sinnvoll als Nukleus zu digitalisieren. Dabei sei zu beachten, dass das reiche Bildmaterial, das ab dem frühen 19. Jahrhundert viele Zeitungen und Beilagen füllt, durch Mikrofilme oft nur unzureichend reproduziert worden sind.
Es liegt aufgrund des bisher Gesagten nahe, die in dieser Datenbank enthaltenen Informationen mit der ZDB zusammenzuführen, um so in einem Nachweissystem eine konsistente Sicht auf den Gesamtbestand zu haben. Die ZDB ist hierbei als das führende Instrument anzusehen; dies wird auch von MFA bestätigt. Deshalb sollte die Datentransformation in die ZDB so schnell wie möglich erfolgen und gefördert werden.

Empfehlungen

1. Der Mikrofilm ist als Archivmedium gut geeignet, für die Nutzung jedoch so schnell wie möglich durch digitale Angebote zu ersetzen.
2. Aus konservatorischen Gründen und Gründen der Langzeitsicherung kann es sinnvoll sein, Film und Digitalisat in einem Bearbeitungszusammenhang herzustellen.
3. Zur Schonung fragiler Originale und aus Kostengründen sollte bei der Digitalisierung auf Mikrorollfilme zurückgegriffen werden, wenn diese den qualitativen Standards bestmöglicher Wiedergabe und Lesbarkeit entsprechen (s.a. die entsprechende Handreichung).
4. Zur Digitalisierung von reich illustrierten Zeitungen und Beilagen ab dem frühen 19. Jahrhundert reicht die Qualität der Mikrofilme oftmals nicht aus, hier sollte nach Möglichkeit auf die Originale zurückgegriffen werden.
5. Die Titeldaten zu den Filmen des Mikrofilmarchivs (MFA) sollten so schnell wie möglich in die ZDB integriert werden.
6. Auf der Grundlage der in der ZDB zusammengeführten Daten sollte die Auswahl zu digitalisierender Filmbestände in Zusammenarbeit mit dem MFA erfolgen.

5. Die Rolle der ZDB für Zeitungsdigitalisierungen [Stand: Mai 2017]

5.1 Der ZDB Katalog

Die Betaversion des ZDB-Katalogs (http://beta.zdb-katalog.de/index.xhtml) wird nach aktueller Einschätzung im Juli 2017 durch eine Produktivversion abgelöst.

Mit der im Rahmen der Pilotprojekte erfolgten Erstellung des neuen ZDB-Katalogs sind SBB und DNB den Wünschen nach optimierten Recherchemöglichkeiten und verbesserter Darstellung von Ergebnissen und Kontexten nachgekommen. In methodischer Hinsicht wurde der Katalog einerseits an moderne Nutzerwartungen angepasst. Stichworte sind vor allem: „Google-Suchfenster“ und die Einschränkung von Treffermengen über Facetten.

Andererseits wurde verstärkt auf Visualisierungen gesetzt, um die Komplexität der Daten möglichst intuitiv erfassbar zu machen. Die Visualisierungen wurden von den existierenden Daten abgeleitet: die Titelhistorie – ein lang beklagtes Desiderat – und Titelrelationen werten z.B. die mit einem Datensatz gekoppelten „Vorgänger-“ und „Nachfolgertitel“ sowie deren Beilagen und Parallelausgaben aus und visualisieren diese. Der Bestandsvergleich visualisiert die Bestandsangaben der ZDB-Teilnehmer und macht sie für den Nutzer leicht verstehbar.

Weil SBB und DNB an möglichst frühen Rückmeldungen der Pilotbibliotheken und weiterer Tester interessiert waren, wurde der neue ZDB-Katalog so früh wie möglich als „Beta-Version“ veröffentlicht und fortlaufend weiter optimiert. „Beta“ bezieht sich ausschließlich darauf, dass einige wenige Funktionen des alten OPACs noch nicht vollständig im neuen Katalog implementiert sind. Gleichwohl ist festzuhalten, dass die Beta-Version bereits zum jetzigen Zeitpunkt eine gute Grundlage zur Unterstützung von Digitalisierungsprojekten bietet und auf den gesamten ZDB-Datenbestand zugreift; die Suchergebnisse sind tagesaktuell. Der Großteil der im Abschlussbericht zugesagten geplanten Optimierungen wurde inzwischen umgesetzt (Stand Mai 2017):

- einfache Vorabeinschränkung nach Zeitungen durch Checkbox in der erweiterten Suche,
- Facette und Karte „Verbreitungsorte Zeitungen“,
- eindeutige URLs für Bestandsvergleich, Titelhistorie und Titelrelationen, so dass diese Sichten in der Detailansicht als Lesezeichen gespeichert und von Katalogen/Portalen aus referenziert werden können.
- Möglichkeit zur unmittelbaren Einsicht von detaillierten Bestandsangaben im Bestandsvergleich.

Vor Produktivname wird noch folgendes Feature implementiert:

- URL-Verweis in Kurzanzeige zur Unterstützung eines schnellen Direktzugriffs auf frei verfügbare Volltexte.

5.2 ZDB und die Verbesserung der Datengrundlage

Die neuen Recherche- und Präsentationsmöglichkeiten des ZDB-Katalogs basieren zum größeren Teil auf den existierenden, qualitativ hochwertigen ZDB-Daten. Im Einzelfall genügen die existierenden Daten den neuen Zwecken aber nicht vollständig. Dies gilt für die folgenden Bereiche:

1. Ortspräsentationen auf geographischen Karten erfordern vormals nicht existente Geo-Koordinaten.
   ZDB-Maßnahme: Für die Bestandskarte und die Standortkarten wurden von der SBB Koordinaten zu sämtlichen in der ZDB nachgewiesenen Bibliotheksstandorten in den Daten der ISIL-Agentur ergänzt.
2. Verbreitungsorte werden bislang nicht vollständig angegeben.
   ZDB-Maßnahme: Für die Verbreitungsortkarte hat die DNB die 200 am häufigsten verlinkten Orte Deutschlands mit Geo-Koordinaten angereichert; ca. 7.000 Orte verfügten bereits zu Projektbeginn über Koordinaten. Die ZDB-Teams werden sich unabhängig von anderen Aktivitäten um die Ergänzung der noch fehlenden Koordinaten für Verbreitungsorte bemühen. Aufgrund der zu beobachteten Long-Tail-

Distribution ist bereits jetzt ein guter Abdeckungsgrad erreicht, so dass die Verbreitungsortkarte mit aussagekräftigen Ergebnissen genutzt werden kann.

3. Zur Visualisierung von Bestandsangaben und zur Suche nach in Bibliotheken nachgewiesenen Jahrgängen werden Jahresangaben in normierter Form benötigt, damit erkennbar ist, für welche Jahre der Bestand Lücken aufweist. Diese Bestandsübersicht ist insbesondere für Digitalisierungsprojekte bedeutsam.
   ZDB-Maßnahme: Entsprechende Datenfelder zur Erfassung normierter Jahresangaben stehen zur Verfügung. In den existierenden Daten sind nicht in jedem Fall normierte Jahresangaben enthalten, eine automatisierte Umsetzung nicht-normierter in normierte Jahresangaben ist wegen der Vielfalt der eingetragen Werte nicht durchgängig möglich. Die ZDB prüft gegenwärtig, inwieweit die vorliegenden Altdaten zu Zeitungen durch software-unterstützte Redaktionsarbeit aufgefangen werden kann.
4. Zur Visualisierung von zeitlichen Verläufen von Zeitungen in der Zeitungshistorie und der Suche nach Erscheinungsjahren von Zeitungen müssen normierte Jahresangaben in der ZDB hinterlegt werden.
   ZDB-Maßnahme: Die entsprechenden Datenfelder stehen zur Verfügung.
5. Für die korrekte Anzeige von Titelrelationen müssen Titelverknüpfungen bidirektional angelegt werden, d.h. z.B. dass Vorgänger als auch Nachfolger eines Zeitungstitels genannt sein müssen. Einige Titelrelationen weisen diese Informationen zurzeit nicht vollständig nach.
   ZDB-Maßnahme: Die ZDB prüft im Laufe des Jahres 2017 Möglichkeiten der nachträglichen semi-automatischen Datenanreicherung.

5.3 Rolle der ZDB in einer Hauptphase

Das Ausmaß, in dem die ZDB die ihr zugedachte Rolle als Steuerungsinstrument ausfüllen kann, hängt ganz wesentlich von der einheitlichen Dateneingabe der digitalisierenden Bibliotheken ab. Dabei ist die Eintragung geplanter Digitalisierungsprojekte nur ein Aspekt des Themas. Das „standardisierte“ Einpflegen gleichförmiger Daten über Zeit durch verschiedenste Bearbeiter erfordert ein hohes Maß an Konsistenz, die aus Sicht der ZDB nur bei Vorlage exakter Erfassungs- und Katalogisierungsanweisungen gewährleistet werden kann. Die Berliner Spezialisten haben deshalb geeignete Anweisungen erarbeitet, die unter http://www.zeitschriftendatenbank.de/zeitungsdigitalisierung öffentlich zugänglich sind. Dort behandelte Aspekte beinhalten u.a.:

- Ankündigung geplanter Digitalisierungsvorhaben,
- Erstellen von O-Aufnahmen (i.e. Titelaufnahmen für den digitalisierten Zeitungstitel), in seltenen Fällen ggf. Erzeugung neuer Titelaufnahmen,
- Eingabe von Bestandsangaben,
- Löschung eingetragener Digitalisierungsvorhaben nach Realisierung.

Die Anweisungen richten sich gleichermaßen an Einrichtungen, die unter DFG-Bedingungen digitalisieren als auch an solche, die dies außerhalb einer DFG-Förderung tun.

6. Ein nationales Zeitungsportal auch für Deutschland?

Prioritär wünschen Wissenschaftlerinnen und Wissenschaftler und die besitzenden Einrichtungen eine Verbesserung der Transparenz und leichten Zugänglichkeit der verstreuten und heterogenen Zeitungsangebote. Im Zuge dieses Pilotprojekts wurden einige dazu notwendige Werkzeuge weiter entwickelt. Das größte Desiderat besteht jedoch weiterhin: Es fehlt ein nationales Zeitungsfenster, das den Zeitungsportalen anderer Länder vergleichbar komfortable Überblicke und Sucheinstiege bietet. Nach Prüfung verschiedener Optionen und ausführlichen Diskussionen zwischen allen zu beteiligenden Institutionen als auch mit den Wissenschaftlerinnen und Wissenschaftlern ist ein Zeitungsfenster innerhalb der Deutschen Digitalen Bibliothek (DDB) nachdrücklich zu empfehlen. Es könnte ähnlich wie das Archivportal-D innerhalb der DDB errichtet werden und müsste diese Anforderungen erfüllen:

- eine übergreifende Volltextsuche in den digitalisierten Zeitungsbeständen,
- zusätzlich Einstiegspunkte über Zeitungstitel, Kalender, Erscheinungsorte, Verbreitungsgebiete,
- Integration des Zeitungsviewers in die Portalumgebung mit stufenloser Zoom-Funktion, mit Highlighting von Suchtreffern und der Möglichkeit, Textabschnitte im Volltext per Copy&Paste direkt zu kopieren und weiter zu verwenden,
- Nachnutzung der konsolidierten Daten- und Präsentationsstrukturen der ZDB,

- eine konsistente Möglichkeit, Zeitungen bzw. Einzelausgaben persistent zu referenzieren und sie somit zitierfähig zu machen.

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 und Verarbeitung verteilter digitaler Bestände und Sammlungen einerseits und einer vertrauenswürdigen und stabilen Betriebssituation beim technischen Betreiber der DDB (FIZ Karlsruhe) andererseits,
- die langfristige Perspektive der DDB als gesamtstaatliches Vorhaben, dessen dauerhafte Weiterführung seitens des Bundes und der Länder als Unterhaltsträger derzeit von niemandem in Frage gestellt wird,
- die erklärte Bereitschaft der Verantwortlichen innerhalb der DDB, eine um zeitungsspezifische Komponenten erweiterte technische und organisatorische Infrastruktur dauerhaft zu betreiben,
- die Verpflichtung oder Empfehlung an alle mit öffentlichen Mitteln geförderte Einrichtungen, digitalisierte Sammlungen an die DDB zu liefern, sowie die bereits integrierten Zeitungsbestände.

[Aktualisierung Mai 2017: In zeitlicher Unabhängigkeit von der angestrebten DFG-Förderlinie zur Digitalisierung historischer Zeitungen soll bei der DFG die Förderung für ein Projekt zur Realisierung eines nationalen Zeitungsportals beantragt werden - und zwar voraussichtlich im Rahmen der Förderlinie „e-Research-Technologien". Dieses Zeitungsportal soll auf der Basis der technologischen und organisatorischen Infrastruktur der Deutschen Digitalen

Bibliothek (DDB) realisiert werden und den zentralen Zugang zu den in Deutschland digitalisierten Zeitungen einschließlich adäquater Recherche- und Navigationsmechanismen ermöglichen. Dabei sollen bestehende Technologien genutzt werden - insbesondere den DFG-Viewer - und einschlägige Standards und Best Practices berücksichtigt werden (DFG-Praxisregeln, METS/ALTO, OAI-PMH). Außerdem ist eine Verzahnung mit der ZDB vorgesehen - insbesondere für die Übernahme von Titel- und Bestandsdaten in das Zeitungsportal.

Zur Vorbereitung des Antrages hat der an der DNB ansässige DDB-Geschäftsbereich „Technik, Entwicklung, Service" weitere Planungen vorgenommen. Unter Federführung der DDB/DNB wird gemeinsam mit weiteren Projektpartnern der konzeptionelle Rahmen abgesteckt. Es wird angestrebt, noch im Frühsommer einen Antrag einzureichen.]

Empfehlung

Der DFG wird empfohlen, zeitnah ein nationales Zeitungsportal innerhalb der DDB zu fördern, das an den Anforderungen der Wissenschaft orientiert ist. Dieses sollte von der technischen und organisatorischen Basis des Gesamtsystems einschließlich der Betriebsinfrastruktur profitieren und die bislang etablierten Prozesse und Werkzeuge für Datenlieferung, Aufbereitung, Transformation, Speicherung und Bereitstellung nutzen. Die Anforderungen könnten in Projekten mit zwei- oder dreijähriger Laufzeit umgesetzt werden; eine erste grobe Abschätzung ergibt eine Kostenspanne von ca. 750.000 EUR bis 950.000 EUR. Eine genauere Kostenschätzung setzt weitere Analyse- und Planungsschritte der DDB voraus, die – gemeinsam mit den potentiellen Projektpartnern – im Vorfeld eines möglichen konkreten DFG-Antrags erbracht werden müssten.

7. Zusammenfassung der wichtigsten Empfehlungen

Im Ergebnis der Pilotphase des Zeitungsdigitalisierungsprojekts geben die Partnereinrichtungen einvernehmlich zusammenfassend folgende Empfehlungen:

7.1 Die Deutsche Digitale Bibliothek (DDB) sollte so bald als möglich ein nationales Zeitungsportal mit dem Zugang zu allen digitalisierten Zeitungen in Deutschland mit den in diesem Masterplan beschriebenen Features entwickeln. Eine Förderung ist dringlich, um die zahlreichen überregionalen, regionalen und lokalen Aktivitäten wissenschaftskonform nutzbar zu machen (zentrale Sucheinstiege: Kalender- und Volltextsuche über alle Zeitungen).

7.2 Als Datenbasis und Steuerungsinstrument für die Zeitungsdigitalisierung und Zeitungspräsentation dient die Zeitschriftendatenbank (ZDB). Die Datenbasis sollte systematisch weiter ausgebaut und verbessert werden, indem

- die umfangreichen Daten des Mikrofilmarchivs der deutschsprachigen Presse e.V. integriert werden,
- die Datenqualität durch Komplettierung noch unzureichend erfasster Verbreitungsorte von Zeitungen für wissenschaftliche Auswertungen und Visualisierungen verbessert wird,
- Anreizförderungen ermöglicht werden, um fehlende Nachweise wichtiger Bestände insbesondere aus Archiven und kleineren Bibliotheken in koordinierter Form nachzutragen.

7.3 Es wird empfohlen, nach den in der Pilotphase entwickelten wissenschaftlichen Auswahlkriterien, den erprobten Erschließungsstandards und der Verbesserung der zugehörigen Informationsinfrastrukturen die Zahl wissenschaftsrelevanter digitalisierter Zeitungen signifikant zu erhöhen. Wünschenswert ist ein Förderprogramm „Digitalisierung historischer Zeitungen in Deutschland“ mit einer Laufzeit von insgesamt 9 Jahren (aufgeteilt auf 3 x 3 Jahre) und einem Volumen von 5,5 Mio. EUR Gesamtkosten pro Jahr (aufzuteilen auf 1/3 Unterhaltsträger (= 1,85 Mio. EUR), 2/3 DFG (= 3,7 Mio. EUR). Zu berücksichtigen ist dabei, dass das Digitalisierungsprogramm durch Fördermaßnahmen zur Verbesserung der Informationsinfrastrukturen (Zeitungsportal der DDB mit Einbindung aller bisher digitalisierten Zeitungen, Verbesserung der Datenbasis der ZDB) und durch Investitionen in die Weiterentwicklung technischer Erschließungsverfahren von Zeitungen (OCR, Named Entity Recognition, Bilderkennung etc.) begleitet werden sollte.

7.4 Nach den Kriterien des Förderstufenmodells (Tabelle 11, S. 61) sollte die Digitalisierung vom Original, aber auch vom Mikrofilm („Qualität und Quantität“) gefördert werden. Die Digitalisierung von mikroverfilmten Zeitungen erlaubt ein auf Masse und schnelle Bereitstellung ausgerichtetes Verfahren (Schwerpunkt Masse). Die Digitalisierung vom Original eröffnet alle Möglichkeiten der forschungsgeleiteten, qualitätsvollen Volltext- und Bilderschließung (Schwerpunkt Klasse).

Es wird empfohlen, qualitativ hochwertige Filme von Zeitungen aus dem urheberrechtsfreien Zeitraum aus den Beständen des Mikrofilmarchivs der deutschsprachigen Presse e.V. zu digitalisieren.

Zeitungsexponenten (z.B. Leitmedien, Innovatoren, Zeitungen mit langen Laufzeiten und großen Reichweiten, mit spezifischen, wissenschaftsrelevanten Inhalten) sollten vom Original digitalisiert und nach dem Stufenmodell bedarfsgerecht erschlossen werden.[103]

[103] Bei den Bremer Wissenschaftlerworkshops 2009 und 2014 wurden als Beispiele die unten genannten sechs langlebigen Zeitungen mit großer Reichweite und einem geschätzten Umfang von insgesamt 3,2 Mio. Seiten genannt, deren Digitalisierung vom Original mit Volltexterschließung 3,5 Mio. EUR, mit zusätzlicher Layout- und Artikelerschließung rd. 4,2 Mio. EUR kosten würde und von denen bislang nur einzelne Jahrgänge digital zur Verfügung stehen:

- Königlich privilegirte Berlinische Zeitung von Staats- und gelehrten Sachen (1785-1911; mit Vorläufern bis ins 17. Jh.) als bedeutender Vorgänger der bereits digital vorliegenden Vossischen Zeitung
- Münchener Neueste Nachrichten (1848-1945; digitalisiert bis 1875) als Vorgänger der Süddeutschen Zeitung
- Kölnische Zeitung (1802-1945; mit Vorläufern bis ins 18. Jh.)
- Frankfurter Handelszeitung (1856-1943)
- Leipziger Post- und Ordinar-Zeitung (1673-1945)
- Erlanger Real-Zeitung (1792-1829; mit Vorläufern im 18. Jh.)

Erschließungstiefe Stufe 1-6	Grundstandard		erweiterter Standard I		erweiterter Standard II		
Auswahlkriterien	Stufe 1 Images mit Struktur-daten	Stufe 2 OCR-Volltexte	Stufe 3a Artikelseparierung (automatisch, halbautomatisch), manuelle Korrekturen bei Titelüberschriften	Stufe 3b Artikelseparier-ung manuelle Eintragung der Kapitelüber-schriften	Stufe 4 Normdaten-verknüpfung Named entity recognition	Stufe 5 vertiefte sachliche Erschließ-ung	Stufe 6 Bilder-schließung, Bild-erkennung
Bereitstellung eines repräsentativen typologischen Spektrums							
„Dauerbrenner“ (Langlebige Zeitungen von großer Reichweite)							
Leitmedien mit hoher Auflage							
„Innovatoren“ (Zeitungen mit pressehistorischen Neuerungen)							
Thematische Kollektionen							
Exponenten von exemplarischer Bedeutung							
Politischer Querschnitt, Diversifizierung der Presse, 19. Jhdt.							
Regionaler und lokaler Querschnitt, bes. 19. Jhdt.							

Förderstufenmodell
Auswahlkriterien und Erschließungstiefen

Erschließungsstufen 1 und 2 sollten Grundstandard sein

Erschließungsstufen 3 bis 6 je nach Vorlage und wissenschaftlichem Bedarf

Tabelle 11: Förderstufenmodell Auswahlkriterien und Erschließungstiefen

7.5 Alle DFG-geförderten Projekte sollten als Mindeststandard mit Strukturdaten und OCR erschlossen werden (Grundstandard Stufe 1 und 2). Der erweiterte Erschließungsstandard I und II (Erschließungsstufen 4-6) sollte in einer Hauptphase bei medien- und fachspezifischem Bedarf (z.B. Exponenten, Rezensionszeitungen, illustrierte Zeitungen und Journale, satirische Blätter u.a.) genutzt werden.

7.6 In einer Hauptphase sollten sich Teilprojekte auch der Weiterentwicklung automatischer Erschließungsverfahren widmen können (soweit diese nicht im Normalprogramm förderbar sind). Zeitungen bieten unabsehbar viele, noch zu entdeckende Daten und Informationen

und allen historisch interessierten Wissenschaften neue Wege zu Quellen, Deutungen, veröffentlichter Meinung. Valide (technische) Ergebnisse in Form geeigneter Verfahren und Tools sollten sukzessive in die Praxisstandards der DFG für (Zeitungs-)Digitalisierungen aufgenommen werden.

7.7 Für eine enge Kooperation bei der Zeitungsdigitalisierung zwischen DDB, ZDB und regionalen Zentren, insbesondere auch mit dem Ziel einer Einbindung von Beständen kleinerer Einrichtungen, ist eine Steuerung und Koordinierung notwendig, die in bestehende Strukturen effektiv eingebunden werden sollte.

Anhang

Öffentlichkeitsarbeit, Vorträge und Publikationen

*2*015

- Albers, Christoph/Förster, Caroline /Hubrich, Jessica /Meyer, Sebastian /Müller, Uwe/ Seiderer, Birgit/Sommer, Dorothea: Zeitungsdigitalisierung in Deutschland. Zum Stand des DFG-Pilotprojektes und zur Entwicklung eines Masterplanes. Workshop mit Impulsreferaten auf dem Bibliothekartag in Nürnberg (26.05.2015).
- Förster, Caroline: Das DFG-Pilotprojekt „Digitalisierung historischer Zeitungen“. Vortrag bei der Dresdner Summerschool (12.10.2015).
- Förster, Caroline: Digitizing Historical Newspapers in Germany, Workshop Digitization am IO Regensburg (27.04.2015).
- Förster, Caroline/Hermes, Maria: Digitalisierung historischer Zeitungen - das DFG-Projekt (Vortrag Hermes) Vortrag in Frankfurt am Main (24.09.2015).
- Hubrich, Jessica: Visualisierung von Titelzusammenhängen: Netzwerkgraph und Titelhistorie im neuen ZDB-Katalog. Vortrag auf dem österreichischen Bibliothekartag in Wien (17.09.2015).
- Hubrich, Jessica/Stei, Andrea: Vortrag auf dem Bibliothekartag in Nürnberg im Rahmen des ZDB-Anwendertreffens (28.05.2015).
- Hubrich, Jessica: Zeitungssuche in der Zeitschriftendatenbank. Vortrag auf dem Bibliothekartag in Nürnberg (26.05.2015).
- Lieder, Hans-Jörg: Vortrag zum OPAC auf der Fachtagung des Mikrofilmarchivs der deutschsprachigen Presse (28.04.2015).
- Stei, Andrea: Der neue ZDB-Opac. Präsentation beim 18. EZB-Anwendertreffen am 13.10.2015 in Regensburg (13.10.2015).

2014

- Bürger, Thomas: Digitalisierung historischer Zeitungen in Deutschland, Vortrag bei den Europeana Newspaper Information Days, Berlin, 27.2.2014.
- Gehrhardt, Timon/Hubrich,Jessica: Zeitungssuche in der Zeitschriftendatenbank. Vortrag in Bremen auf dem Workshop „Zeitungsdigitalisierung in Deutschland – Anforderungen der Wissenschaftlerinnen und Wissenschaftler an ein zukünftiges Digitalisierungsprogramm der Deutschen Forschungsgemeinschaft“ (06.10.2014).
- Gehrhardt Timon/Czwinkalik, Karin: Neugestaltung des ZDB-OPAC: Aktueller Stand und Ausblick. Vortrag auf dem Bibliothekartag in Bremen im Rahmen des ZDB-Anwendertreffens (03.06.2014).
- Hermes, Maria: Digitalisierung der vollständigen deutschsprachigen Zeitungen des 17. Jahrhunderts an der SuUB Bremen – Ein Projekt mit Komplexität, Vortrag bei den Europeana Newspaper Information Days, Berlin, 27.2.2014.
- Hubrich, Jessica: Informationsveranstaltung für eine Delegation chinesischer Bibliothekare (12 Personen) in der DNB (03.12.2014).
- Hubrich, Jessica: Informationsveranstaltung in der DNB zum Thema ZDB-Zeitungsportal für Referendare (12.09. 2014).
- Lieder, Hans-Jörg/Hubrich, Jessica: Zeitungssuche interaktiv: das ZDB-Zeitungsportal. Vortrag auf dem Bibliothekartag in Bremen im Rahmen der Sitzung „Herausforderungen von Zeitungen als besonderes Sammelobjekt" (05.06.2014).

- Müller, Maria Elisabeth: Die ältesten Zeitungen digital – eine Hausforderung der digitalen Transformation eines Jahrhunderts. Vortrag in der AG Regionalbibliotheken, Darmstadt, 30.09.2014.
- Müller, Maria Elisabeth: Digitalisierung der vollständigen deutschsprachigen Zeitungen des 17. Jahrhunderts an der SuUB Bremen – Ein Projekt mit Komplexität, Vortrag beim Workshop zum Thema Historische Zeitungen aus dem 18. und 19. Jahrhundert im Institut für deutsche Sprache, Mannheim, 21.11.2014.
- Müller, Maria Elisabeth/Maria Hermes: Digitalisierung der deutschsprachigen Zeitungen des 17. Jahrhunderts: Das Kompliziertеste zuerst. Vortrag im Rahmen des 103. Bibliothekartags in Bremen, 04.06.2014.

Publikationen

- Bürger, Thomas: Bibliothek als Forschungsinfrastruktur. Aktuelle Herausforderungen und Chancen. Hrsg. von Thomas Bürger und Uwe Rosemann. Zeitschrift für Bibliothekswesen und Bibliographie 61 (2014), H. 4-5, S. 189-289.
- Gehrhardt, Timon/Hubrich, Jessica: Neue Zugänge zum Medium Zeitung: Projekt „Relaunch des ZDB-Katalogs". In: Dialog mit Bibliotheken 2 (2014), S. 47–51. URL: http://d-nb.info/1068614862/34.
- Hermes, Maria/Manfred Nölte: Neue Einblicke in den früheren Alltag – Digitalisierung historischer Zeitungen der SuUB. In: Impulse aus der Forschung, Universität Bremen, 2014, Heft 01, S.16–19.
- Hermes, Maria/Müller, Maria Elisabeth: Digitalisierung der vollständigen deutschsprachigen Zeitungen des 17. Jahrhunderts an der SuUB Bremen – Ein Werkstattbericht. In: o-bib das offene Bibliotheksportal, 2014/1, S. 265 – 279,https://www.o-bib.de/article/view/2014H1S265-279/1168
- Lieder, Hans-Jörg/Hubrich; Jessica: Zeitungssuche interaktiv: der neue ZDB-Webkatalog. Dezember 2014 in o-bib 2014/1, S. 305 –311. URL: https://www.o-bib.de/article/view/2014H1S305–311.
- Matzerath, Josef: Zeitungen als historische Quellen. Anregungen und Wünsche eines Historikers. In: BIS – Das Magazin der Bibliotheken in Sachsen, 5 (2012), Nr. 3, S. 189-191 http://www.qucosa.de/recherche/frontdoor/?tx_slubopus4frontend[id]=9665
- Müller, Maria Elisabeth: Digitalisierung der vollständigen deutschsprachigen Zeitungen des 17. Jahrhunderts an der SuUB Bremen – Ein Projekt mit Komplexität. In: Bibliotheksdienst, 2014, Bd. 48, Heft 12, S. 1000–1013.
- Sommer, Dorothea/Heiligenhaus, Kay/Pankratz, Manfred /Wippermann, Carola: Zeitungsdigitalisierung: eine neue Herausforderung für die ULB Sachsen-Anhalt. Werkstattbericht aus der Pilotphase des DFG-Projekts Digitalisierung historischer Zeitungen. ABI Technik. 34 (2014) 2. S. 75–85.

OCR-Bericht Halle [wird noch ergänzt]

OCR-Bericht München

[siehe gesondertes PDF-File]

Goobi-Bericht Berlin

Abschlussbericht Goobi-Production (Kurzfassung)

1. Allgemeine Angaben:

DFG-Geschäftszeichen:	SCHN 743/14-1
Antragsteller:	Barbara Schneider-Kempf, Generaldirektorin Staatsbibliothek zu Berlin – Preußischer Kulturbesitz Potsdamer Str. 33, 10785 Berlin Tel.: 030 – 266-432323, Fax: 030-266-331301 E-Mail: barbara.schneider-kempf@sbb.spk-berlin.de
Thema des Projektes:	Funktionsanpassung Goobi.Production
Projektzeitraum (geplant)	01.07.2013 – 30.06.2014 (12 Monate Projektlaufzeit)
Projektzeitraum (realisiert)	01.07.2013 – 06.02.2015
Berichtszeitraum:	01.07.2013 – 06.02.2015
Internetadresse des Vorhabens:	keine Webpräsenz, da nur Software-Entwicklung
Erstbewilligung der DFG für das Rahmenprojekt vom:	19.12.2012 Hiz
Bewilligte Mittel für das Teilprojekt Sachmittel Programmpauschale	57.353 € 11.500 €

2. Arbeits- und Ergebnisbericht

2.1 Ausgangslage und Zielsetzung des Projektes

Ziel des Projektes war die Weiterentwicklung des Goobi-Softwaremoduls *Goobi.Production* für die DFG-Viewer-kompatible Digitalisierung von Zeitungen. Für diese Materialart gibt es im Bezug auf die Digitalisierung spezifische Anforderungen, die eine Reihe von Erweiterungen und Anpassungen der vorhandenen Workflow-Software erfordern. Diese Erweiterungen wurden in die sogenannte *Community-Edition* integriert, so dass sie für alle Goobi-Anwender nachnutzbar sind. Neben der Staatsbibliothek zu Berlin als Antragsteller wird die Workflow-Software Goobi derzeit u.a. auch in der SLUB Dresden, der SUB Göttingen und in der SuUB Hamburg eingesetzt. Das Projekt wurde im Rahmen des Gemeinschaftsantrags der fünf Pilotbibliotheken (Berlin, Bremen, Dresden, Halle, München) durchgeführt und mit diesen abgestimmt.

2.2 Ausschreibung und Vergabe des Projektauftrages

Die öffentliche Ausschreibung zur „Erweiterung der Open-Source Software *Goobi - Digital Library Modules* wurde am 28.08.2013 im Deutschen Ausschreibungsblatt, mit der Ausschreibungs-Nr. 836661 veröffentlicht. Die reguläre Frist für die Abgabe von Angeboten bis zum 14.10.2013 wurde aus besonderem Grund noch einmal bis zum 28.10.2013 verlängert. Es gingen Angebote von zwei voneinander unabhängigen Bietern ein. Die Auswertung und Bewertung der Angebote wurde am 05.11.2013 abgeschlossen, dem erfolglosen Bieter wurde am 02.12.2013 abgesagt und nach Ablauf einer Einspruchsfrist von 14 Kalendertagen erfolgte der Zuschlag für die ausgewählte Firma Zeutschel GmbH mit Sitz

in Tübingen-Hirschau am 17.12.2013. Der Vertrag mit dem Dienstleister wurde am 21.01.2014 geschlossen. Am selben Tag fand in der Staatsbibliothek zu Berlin ein *Kick-off-Meeting* statt, an dem die Projektpartner sowie ein Vertreter des Goobi-Releasemanagements für die Community-Edition (verantw. SLUB Dresden) teilnahmen. Für das Projekt waren auf Seiten der Firma Zeutschel GmbH als Auftragnehmer Herr Frank-Ulrich Weber (Projektleiter) und Herr Matthias Ronge (Software-Entwicklung zuständig, auf Seiten der Staatsbibliothek zu Berlin als Auftraggeber waren es Herr Christoph Albers (Zeitungsabteilung) und Herr Gerrit Gragert (Abteilung Informations- und Datenmanagement).

2.3 Projektplanung, Meilensteine und abgerufene Mittel:

Arbeitsschritt	Von – Bis (geplant)	Von – Bis (real)	Mittelabruf (real)
Analysephase (Vertragsabschluss / Kick-off-Meeting / Erstellung und Abnahme Pflichtenheft)	21.01.14 - 11.02.14	21.01.14 - 11.02.14	13.427,66 € (25%)
Realisierungsphase (Fertigstellung aller Komponenten)	12.02.14 – 06.05.14	12.02.14 – 03.07.14	32.226,39 € (60%)
Abnahmephase (Test aller Funktionen und Fehlerkorrektur)	06.05.14 – 30.05.14	03.07.14 – 05.02.15	
Abnahme der vertraglich vereinbarten Dienstleistung	31.05.14	06.02.15	8.056,60 € (15%)
Merge Release-Management mit *Goobi-Community Edition*	01.06.14 – 30.06.14	06.02.15 - 26.03.15	
Summe			**53.710,65 € (100%)**

2.4 Technischer Ansatz, Methoden und Verfahren

Die Erweiterungen in dem Software-Modul Goobi.Production wurden streng nach dem Open-Source-Konzept auf der Plattform Launchpad <https://launchpad.net/goobi-production> bzw. Github <https://github.com/goobi/goobi-production> betrieben, was eine Einbindung der Fachcommunity, insbesondere der Goobi-Anwender und der Pilotprojekte zur Digitalisierung von Zeitungen während der Realisierungsphase ermöglichte. In der Realisierungsphase fand nach jedem Arbeitsschritt eine Prüfung und Abstimmung mit dem Goobi-Releasemangement sowie ein sich hieran anschliessender Funktionstest mit Abnahme durch die Staatsbibliothek zu Berlin als Auftraggeber statt.

2.5 Projektergebnisse

Alle Erweiterungen und Anpassungen wurden fester Bestandteil der *Community-Edition* von Goobi, so dass diese seit der Freigabe durch das Goobi-Releasemanagement (verantw. SLUB Dresden) am 26.03.2015 für alle Goobi-Anwender zur Verfügung stehen. Dies sind im Detail:

2.5.1 Erweiterte Katalogsuche (nun auch über Titel und ZDB-ID)

Der Bildschirmdialog „Einen neuen Vorgang anlegen“ wurde so erweitert, dass die Katalogsuche nun auch in den Feldern „Titel“ und „ZDB-ID“ des konfigurierten Kataloges (z.B. auch Zeitschriftendatenbank) möglich ist. Werden mehrere Treffer gefunden, wird die Zahl der Treffer angezeigt, und es erscheint eine selektierbare Trefferliste mit bibliographischen Angaben.

2.5.2 Importieren einer Ovu-PPN-Liste (Massenimport-Plugin)

Diese Funktion ist Bestandteil der allgemeinen Goobi-Version und wurde für die Materialart „Zeitungen“ entsprechend erweitert.

2.5.3 Erscheinungsverlauf-Editor für die Erfassung von Zeitungen

Zur Erfassung von Erscheinungsverläufen von Zeitungen wurde ein graphischer, kalendarischer Erscheinungsverlauf-Editor implementiert. Für die Digitalisierung von Hauptausgaben und Beilagen von Zeitungen können ein oder mehrere Erscheinungsverläufe (mit Ausnahmen) erfasst werden. Im kalendarischen Editor kann ein zuvor erzeugter und als XML-Datei gespeicherter Erscheinungsverlauf eingelesen werden, so dass bereits erfasste Daten nicht noch einmal neu eingegeben werden müssen.

2.5.4 Granularität / Aggregationsstufe für Zahl der Vorgänge festlegen

Auf der Basis der geschätzten Gesamtseitenzahl des erfassten Erscheinungsverlaufs lassen sich verschiedene Aggregationsstufen (Ausgaben, Tage, Wochen, Monate, Quartale und Jahrgänge) für die jeweiligen Vorgänge festlegen. Nicht alle Stufen sind auf Grund der zu erwartenden Seitenanzahl pro Vorgang praktikabel, sinnvoll oder machbar.

2.5.5. Normdatenschnittstelle zur GND für Personen und Körperschaften

Goobi Production wurde um ein Eingabefeld für einen URI zu den Normdaten erweitert. Eine GND-Nummer wird direkt in der Eingabemaske aufgelöst, sodass der Name in den Datensatz eingetragen wird. Im Metadaten-Editor kann zu einer Person der URI eines Normdatensatzes erfasst werden. Vor- und Nachname lassen sich aus der GND einlesen. Der URI wird mit den Metadaten gespeichert und auch exportiert.

2.5.6. Erweiterung des Moduls für Batches

Die Verwaltung der Batches wurde erneuert. Um die Logistik des Aushebens, des Transports und der Zuführung zur Scan-Station zu unterstützen, können mehrere Vorgänge zu einem Logistik-Batch zusammengefasst werden, ohne dass diese Vorgänge in einen sachlichen Zusammenhang gesetzt werden. Da für die Digitalisierung einer Zeitung mit vielen Ausgaben, die Aufteilung auf mehrere (oder sogar sehr viele) Vorgänge notwendig ist, können diese Vorgänge wieder zu logischen Batches zusammengefasst werden. In der gleichen Weise können auch mehrere Vorgänge von Fortlaufenden Sammelwerken zu logisch zusammengehörigen Batches zusammengefasst werden.

2.6 Gewonnene Erkenntnisse (Lessons learned)

- Der Einsatz der Kommunikations-Plattform Github <https://github.com/goobi/goobi-production> hat sich für die Software-Entwicklung nach dem Open-Source-Konzept bewährt.
- Die im Projektplan festgelegten Jour-Fix-Termine (teilweise via Skype), sowie die jederzeit enge Abstimmung per E-mail und Telefon zwischen allen am Projekt beteiligten Partnern ist essentiell für den Erfolg des Projekts.

- Verzögerungen im Projektablauf entstanden aus vielerlei Gründen, die wichtigsten waren:
- Die Open-Source Software war an bestimmten Stellen des Quellcodes nicht oder nicht ausreichend gut dokumentiert (z.B. bestimmte Funktionalitäten, die von anderer Seite programmiert worden waren, konnten vom beauftragten Dienstleister nicht nachvollzogen werden);
- Teile der sogenannten UGH-Bibliothek standen nicht in der aktuellen Fassung zum Linken und Compilieren des Source-Codes zur Verfügung (Problem das im wirtschaftlichen Wettbewerb stehende Anbieter am selben Open-Source-Produkt arbeiten).
- Das in der Leistungsbeschreibung beschriebene Verfahren bzw. der gewählte technische Lösungsansatz liess sich aus programmtechnischer Sicht nicht umsetzen, so dass eine neue, alternative Lösung mit derselben Funktionalität gefunden werden musste.
- Unterschiedliche Hard- und Software-Umgebungen beim Dienstleister als Auftragnehmer und bei der Staatsbibliothek zu Berlin als Anwender der Software erschwerten die Fehlersuche;
- Nicht im Leistungsumfang enthaltene Veränderung/Anpassung von Funktionen der allgemeinen Goobi-Version (z.B. Massen-Plugin), die im Bezug auf die beauftragten neuen Funktionalitäten für Zeitungen erforderlich wurden.

3. Fazit und Ausblick

Im Rahmen der programmtechnischen Erweiterung der Workflow-Software Goobi für die Materialart Zeitungen wurde – soweit möglich – eine zeitgemäße Benutzerführung (GUI) implementiert. Gleichwohl war es im Rahmen dieses (Erweiterungs-)Auftrages nicht möglich, die Benutzeroberfläche und Bedienerführung des Hauptprogramms allgemein entsprechend den heutigem Standard anzupassen. Dies sollte im Rahmen zukünftiger (Weiter-)Entwicklungen der Software dringend erfolgen.

Berlin, den 24.07.2015

Christoph Albers
Staatsbibliothek zu Berlin

Textfassung des Web-Werkzeugs „Online-Wegweiser“

Wegweiser für die Digitalisierung historischer Zeitungen

Sie interessieren sich für die Digitalisierung von Zeitungen? Dieser Wegweiser möchte Sie bei den wichtigsten Schritten Ihres Digitalisierungsprojekts begleiten und, wesentlich orientiert an den praktischen Arbeiten, konkrete Hilfestellungen geben. Die Grundlage der präsentierten Informationen und Empfehlungen stellt dabei der vom DFG-Pilotprojekt "Digitalisierung historischer Zeitungen" erarbeitete Masterplan dar.

Vorbereitung

Digitalisierungsstandards

Was bedeutet eigentlich "Digitalisierung" im konkreten Einzelfall? Bei der Digitalisierung von Zeitungen sind ganz unterschiedliche Bearbeitungsniveaus denkbar, die jeweils für sich einen Endpunkt darstellen oder als Schritt auf dem weiteren Bearbeitungsweg angesehen werden können. Abhängig von Ihren Zielen im Digitalisierungsprojekt, können die folgenden Bearbeitungsstufen (vgl. Masterplan, S. 28 ff.) aufeinander aufbauend sämtlich oder teilweise durchgeführt werden:

Grundstandard

- Stufe 1: Erzeugung von digitalen Abbildungen der Zeitungsseiten einschließlich Strukturdaten zur Ermöglichung einer einfachen Navigation seitens der Nutzer
- Stufe 2: Erzeugung von OCR-Volltexten

Erweiterter Standard 1

- Stufe 3a: (halb)automatische Artikelseparierung, manuelle Korrekturen von Überschriften
- Stufe 3b: (halb)automatische Artikelseparierung, manuelle Erfassung von Überschriften

Erweiterter Standard 2

- Stufe 4: Normdatenverknüpfung, Named Entity Recognition (NER)
- Stufe 5: vertiefte sachliche Erschließung je nach Anforderung
- Stufe 6: Bilderkennung und -erschließung

Das minimale Ziel sollte das Erreichen des beschriebenen Grundstandards der Stufe 1 sein, über die übrigen Stufen ist im Einzelfall nach Vorlage und Bedarf zu entscheiden.

Original oder Mikrofilm

Die Digitalisierung vom gedruckten Original ermöglicht eine Reproduktion in bester Faksimilequalität und vermittelt einen optimalen Gesamteindruck des Originals. Für eine spätere OCR sind die besten Voraussetzungen gegeben. Insbesondere unikale und seltene Zeitungen sowie Zeitungen von besonderem kulturhistorischem Wert, z.B. solche mit wichtigen Bildanteilen, sollten daher vom Original in Farbe oder in Graustufe gescannt werden. Allerdings ist dies auch die zeitaufwändigste und damit kostenträchtigste Vorgehensweise, unabhängig davon, ob inhouse oder mit einem Dienstleister gearbeitet wird.

Die Digitalisierung vom Mikrofilm, sofern vorhanden, schont die fragilen Originale. Gleichzeitig können die mit einer Verfilmung bereits durchgeführten Arbeiten (Verzeichnung, ggf. Lückenergänzung, vgl. Masterplan, S. 27 f.) teilweise nachgenutzt werden. Eine Massendigitalisierung kann relativ schnell und preiswert erfolgen. Bei Verwendung hochwertiger Mikrofilme jüngeren Datums sind bei einer OCR-Bearbeitung vergleichbare Ergebnisse wie bei der Digitalisierung vom Original zu erwarten.

In jedem Fall muss vorab geprüft werden, ob die gewählte Vorlage – Papieroriginal oder Mikrofilm – für eine Digitalisierung geeignet ist. Bei der Digitalisierung vom Mikrofilm darf der Nachbearbeitungsauswand nicht vergessen oder unterschätzt werden. Abhängig von der Qualität der Vorlage/der Mikroverfilmung fallen mehr oder weniger zeitaufwändige Korrekturen der automatischen Seitenerkennung an.

Weitere Informationen und Bewertungskriterien zur dieser Frage können Sie dem Masterplan (ab S. 52) sowie den gesonderten Entscheidungshilfen zur Digitalisierung von Original oder Mikrofilm entnehmen.

Inhouse vs. Outsourcing

Bei der Wahl zwischen einer Digitalisierung mit eigenem Personal und Equipment ("Inhouse") und einer Vergabe der Digitalisierung an einen Dienstleister ("Outsourcing") handelt es sich oftmals um eine Grundsatzentscheidung, bei der jenseits des einzelnen Projekts z.B. auch die strategische und personelle Ausrichtung einer Einrichtung eine wichtige Rolle spielen kann. Zahlreiche Faktoren sind von Bedeutung und werden je nach Ausgangslage und Zielvorstellung ganz anders zu gewichten sein.

Wird im eigenen Haus digitalisiert, sind nicht nur die unmittelbar anfallenden Ausgaben zu berücksichtigen, sondern auch die Folgekosten. Wartung, Reparaturen, Austausch von Geräten, Software-Updates und ähnliches erzeugen nicht nur häufig übersehene Kosten, sondern verursachen auch Stillstände in der Produktion. Auf der anderen Seite darf nicht vergessen werden, dass auch die Vergabe an einen Dienstleister erhebliche Ressourcen vor Ort bindet. So entstehen Aufwände durch Ausschreibungsverfahren, Definition der

Anforderungen, Materialauswahl, -aushebung, -übergabe, Qualitätskontrolle der Lieferungen etc.

Bei der Abwägung Ihrer Entscheidung kann diese Checkliste nützlich sein.

Kostenkalkulation

Die unterschiedlichen Verfahren, Workflows, Mengen, Schwierigkeitsgrade und Erschließungstiefen bei der Zeitungsdigitalisierung ergeben ein differenziert zu betrachtendes Kostenspektrum, das sich nur bedingt verallgemeinern lässt. Die ermittelten Kostenfaktoren (vgl. Masterplan, S. 41 ff.) in Abhängigkeit der entsprechenden Rahmenbedingungen sind dabei:

- **Aufgabenübergreifend**
  - Projektleitung und –koordination
- **Vorbereitung**
  - Auswahl der zu digitalisierenden Inhalte und Abgleich mit der ZDB
  - Beschaffung der Vorlagen
  - Prüfung der Vorlagenqualität und Entscheidung für eine Vorlagenart (Original vs. Mikrofilm)
  - Kollationierung bzw. Prüfung der Vollständigkeit und konservatorischen Eignung
  - Inhouse-Digitalisierung und/oder OCR-Bearbeitung/Tiefenerschließung: Prüfung der Eignung vorhandener Ausrüstung bzw. ggf. Beschaffung/Aufrüstung von Scannern und Software
  - Vergabe an Dienstleister (Digitalisierung und/oder OCR-Bearbeitung bzw. Tiefenerschließung): Vorbereitung und Durchführung eines Vergabeverfahrens
  - Vorbereitung der Materialien: ggf. Lückenschluss, konservatorische Maßnahmen
  - Workflowplanung und Kostenkalkulation
- **Digitalisierung**
  - Art des Scannereinsatzes und Komplexität des Digitalisierungsvorgangs
  - Qualitätskontrolle (bei Inhouse-Digitalisierung und in Zusammenarbeit mit dem Dienstleister)
- **Erschließung**
  - Bibliographische Erschließung
  - Strukturdatenerschließung
  - OCR (Antiqua, Fraktur)
  - Layouterkennung / Artikelseparierung
- **Bereitstellung**
  - Einbindung in Präsentationsoberfläche / DFG-Viewer, ggf. mit Einrichtung entsprechender Schnittstellen
- **Archivierung**
  - Herstellung persistenter Adressierbarkeit
  - Datensicherung / Langzeitarchivierung

Der Medientyp Zeitung zeichnet sich durch eine ausgeprägte Heterogenität in Bezug auf die physische und strukturelle Beschaffenheit der Vorlagen aus. Für die konkrete Planung von Zeitungsdigitalisierungsprojekten wird empfohlen, nach der Bestandsprüfung das zu wählende Digitalisierungsverfahren abzuleiten und mittels Marktsichtungen (z.B. Scanner; Dienstleister), aktualisierten Informationen (z.B. OCR-Lizenzkosten) sowie Stichproben (v.a. zur Ermittlung des Personalbedarfs für Erfassung, Qualitätskontrolle etc.) den Kostenrahmen des Projektes anhand des entsprechenden Mengengerüstes zu kalkulieren.

Detailliertere Informationen zu den im Rahmen des Pilotprojekts ermittelten Kostenkorridoren lassen sich im Masterplan (S. 43) finden.

Vollständigkeitsprüfung

Vollständigkeit sollte ein wichtiges Ziel jedes Digitalisierungsprojekts sein. Damit kann einerseits die Vollständigkeit aller Ausgaben eines Zeitungstitels gemeint sein. Andererseits wird, insbesondere bei größeren, langlebigen Zeitungsunternehmungen, lediglich die Vollständigkeit eines festzusetzenden Erscheinungszeitraums anzustreben sein. Im Idealfall enthalten die erzeugten Digitalisate des gewählten Titels oder Zeitraums jede bekannte gedruckte Seite, deren Existenz bspw. durch durchgehende Seiten- oder Ausgabenzählung bekannt ist oder angenommen werden darf (vgl. Masterplan, S. 27 f.).

Vor Beginn der Digitalisierung sollte deshalb die Vollständigkeit der zu digitalisierenden Zeitungen geprüft werden. Werden Lücken im eigenen Bestand identifiziert, können fehlende Bestandsteile, sofern in anderen Einrichtungen vorhanden, über entsprechende Recherchen in der ZDB gefunden werden. Eine eigens für einen einrichtungsübergreifenden Bestandsvergleich eingerichtete Visualisierung bietet einen schnellen und präzisen Überblick. Im Fall der Digitalisierung vom Mikrofilm prüfen Sie zusätzlich die im Mikrofilmarchiv der deutschsprachigen Presse e.V. und im EROMM-Register nachgewiesenen Bestände.

ZDB

Die Zeitschriftendatenbank (ZDB) ist das wichtigste Erfassungs- und Nachweisinstrument für Zeitschriften, Zeitungen und andere Periodika in Deutschland und Österreich und wurde im Rahmen des Projekts zum wissenschaftsfreundlichen Nachweis- und Steuerungsinstrument weiterentwickelt (vgl. Masterplan, S. 11 ff.). Es eröffnet den Zugang zu den Beständen von mehr als 4.000 Kultur- und Wissenschaftseinrichtungen.

Bitte beachten Sie, dass die ZDB keine Präsentationsumgebung für Bilddateien oder Volltexte ist.

Im Kontext von Zeitungsdigitalisierungsprojekten in Deutschland erfüllt die ZDB mehrere wichtige Funktionen:

- als Rechercheinstrument zur Klärung der Bestandslage und zur Identifizierung möglicher Partnereinrichtungen im Fall von existierenden Bestandslücken in der digitalisierenden Einrichtung
- als Ort der Bekanntmachung geplanter Digitalisierungen zur Vermeidung von Doppelarbeit
- als Ort der Katalogisierung der digitalisierten Zeitungen
- als Instrument der verstärkten Sichtbarmachung und Benutzerzuführung zu den digitalen Sammlungen der digitalisierenden Einrichtungen
- als übergreifendes Rechercheinstrument mit zahlreichen modernen Such-, Stöber- und Visualisierungsfunktionen für Endnutzer
- als Datenprovider bei späteren Portalisierungen digitaler Zeitungsbestände etwa in der Deutschen Digitalen Bibliothek oder in der Europeana.

Die jeweiligen Funktionen der ZDB werden an den entsprechenden Textstellen dieses Wegweisers sowie umfänglich im Masterplan ab S. 55 erläutert.

Digitalisierung

Bilddateien

- TIFF
  Bildmaster von Graustufen oder Farbbildern sollten nach dem derzeitigen Kenntnisstand im Format "**TIFF uncompressed**" langzeitgesichert werden. Das Format TIFF gibt es schon seit den 1980er Jahren. Es hat sich als einer der wichtigsten de-facto-Standards etabliert und es ist damit zu rechnen, dass es auch in Zukunft von allen Standardprogrammen unterstützt wird. So haben auch die Projektpartner durchgängig TIFF in ihren Zeitungsdigitalisierungsprojekten genutzt (vgl. Masterplan, S. 40 f.). Neben TIFF kann auch TIFF-LZW oder JPEG2000 in seiner **verlustfreien Form** als Format für den Bildmaster verwendet werden.
- JPEG2000
  Im internationalen Umfeld hat sich JPEG2000 in den letzten Jahren zu einer ernst zu nehmenden Alternative für TIFF entwickelt (vgl. Masterplan, S. 41). Vorteile sind vor allem der geringere Speicherplatzbedarf durch verlustfreie/-behaftete Kompression und der insbesondere für großformatige Zeitungen praktische Zoom in gängigen Web-Browsern. Inzwischen existiert mit OpenJPEG auch eine quelloffene Referenzimplementierung des "Part 1" der JPEG2000-Spezifikation. Für die Speicherung von Mastern im JPEG2000-Format ist daher darauf zu achten, dass nur die lizenzfreien Bereiche von JPEG2000 Verwendung finden.

Für die Digitalisierung von Zeitungen wird eine Auflösung von 300ppi und eine Farbtiefe von mindestens 8-Bit (256 Graustufen) im Format TIFF empfohlen.

Struktur- und Metadaten

- METS/MODS
  Die Bereitstellung der Metadaten zur weiteren Nutzung gemäß den materialspezifischen Standards ist verpflichtend: Die DFG-Praxisregeln "Digitalisierung" sehen dabei für gedruckte Textwerke die Nutzung von METS/MODS vor.
- ENMAP
  ENMAP ist ein METS/ALTO Profil für Zeitungen das vom Europeana Newspapers Projekt entwickelt wurde und das insbesondere nützliche Hinweise für eine Feinstrukturierung der formalen und inhaltlichen Zeitungsbestandsteile enthält. Bitte beachten Sie jedoch, dass aufwendige Feinstrukturierungen möglicherweise ausschließlich in lokalen Umgebungen Mehrwerte erbringen und in überregionalen Nachweisinstrumenten (z.B. DDB, Europeana) nicht nachgenutzt werden können.

Für Meta- und Strukturdaten von digitalisierten Zeitungen wird die Verwendung von METS/MODS empfohlen.

Volltexte

- ALTO
  Für Volltexte hat sich ALTO als de-facto-Standard etabliert. ALTO wird von der Library of Congress gepflegt und ist ein speziell für die Anforderungen von OCR und OLR entwickeltes XML-Schema in dem u.a. pixelgenaue Koordinaten für die erkannten Zeichen und Layoutelemente sowie Konfidenzwerte der Erkennungsqualität abgelegt werden können.
- TEI
  Neben ALTO hat sich - vor allem im Bereich der Wissenschaft - TEI als Standard für die Kodierung auszuzeichnender Volltexte durchgesetzt, das auch die Erfassung von Koordinaten und typographischen Merkmalen erlaubt.
- hOCR
  Schließlich sei auch noch hOCR genannt, welches u.a. bei der Digitalisierung im Rahmen von Google Books Projekten sowie in diversen Open Source OCR Softwareprogrammen Verwendung findet.

Für Zeitungsvolltexte wird die Verwendung von ALTO empfohlen.

Qualitätskontrolle: Digitalisierung

Eine Qualitätskontrolle der Digitalisierung sollte mindestens stichprobenhaft durchgeführt werden (siehe dazu auch den Masterplan, S. 39 ff.). Die zu prüfenden Aspekte sind:

- **Scanparameter** - sind Kontrast/Farbe optimal auch über mehrere Zeitungen hinweg?
- **Technische Parameter** - wurden die Digitalisierungsparameter eingehalten?

- **Lesbarkeit** - sind die Texte gut lesbar?
- **Ausrichtung** - sind alle Seiten gerade ausgerichtet?
- **Vollständigkeit** - sind Bilddateien für alle Zeitungsseiten der Vorlageform vorhanden?
- **Reihenfolge** - liegen alle Bilddateien in der korrekten Reihenfolge vor?

Speziell bei der Digitalisierung vom Mikrofilm ist darüber hinaus auf Folgendes zu achten:

- **Rahmensetzung** - sind alle Seiten erkannt und die Rahmen korrekt gesetzt worden?
- **Dubletten** - sind bei der Digitalisierung überflüssige, d.h. dublette Seiten erfasst worden?

  (Anmerkung: dies kommt vor allem bei Mikrofilmen vor, auf denen einzelne schwer lesbare Seiten häufig mit verschiedenen technischen Parametern verfilmt wurden, um die Lesbarkeit der Texte zu gewährleisten)

Erschließung

Grundstandard (Stufe 1)

Wie bereits an anderer Stelle erwähnt wurde, kann die Erschließungstiefe von digitalisierten Zeitungen mit unterschiedlichen legitimen Begründungen von Projekt zu Projekt differieren.

Zunächst aber zu den Gemeinsamkeiten. Aus der Publikationslogik von Zeitungen – Titel, Jahrgang, Ausgabe, Seite – ergibt sich der Bedarf einer spezifischen Präsentation in online-Umgebungen. Üblicherweise wird mindestens eine Kalenderfunktion (Jahr, Monat, Tag) angeboten, mit der Ausgaben einzelner Titel tagesgenau angesteuert werden können. Um dies zu ermöglichen, sollten die entsprechenden Termini des Strukturdatensets des DFG-Viewers verwendet werden.

Allen Digitalisierungsprojekten gemein ist somit Stufe 1 des **Grundstandards** der Erschließung (vgl. Masterplan, S. 16 ff.): Erzeugung von digitalen Abbildungen der Zeitungsseiten einschließlich Strukturdaten zur Ermöglichung einer einfachen Navigation seitens der Nutzer.

Sind weitergehende Erschließungsarbeiten sinnvoll und erwünscht, so finden diese auf dem Weg der **Tiefenerschließung** statt.

Texterkennung (OCR)

Texterkennung - oder auch OCR - bezeichnet die automatisierte Erkennung des Textes innerhalb von Bildern. Texterkennung ist notwendig, da Scanner oder Digitalkameras als

Ergebnis ausschließlich Rastergrafiken liefern können, d.h. in Zeilen und Spalten angeordnete Punkte unterschiedlicher Färbung (Pixel). Texterkennung bezeichnet dabei die Aufgabe, die so dargestellten Buchstaben als solche zu erkennen, d.h. zu identifizieren und ihnen den Wert zuzuordnen, der ihnen nach üblicher Textcodierung zukommt (Unicode).

Insbesondere im Bereich älterer historischer Zeitungen ist damit zu rechnen, dass die Ergebnisse einer OCR-Bearbeitung nicht vollständig korrekt sein werden. Zu den typischen Gründen dafür gehören mindere Papier- und/oder Druckqualitäten oder qualitativ schlechte Mikrofilme, im Extremfall Textverluste bereits in der Vorlage, während der Digitalisierung erzeugte Phänomene (z.B. Seitenwölbungen), schwer lesbare bzw. erkennbare Textbereiche (z.B. Impressum, Tabellen, Diagramme etc.) sowie uneinheitliche Schrifttypen gerade im Bereich der Frakturschrift. Welche Ergebnisse letztlich als qualitativ ausreichend akzeptiert werden kann, hängt von einigen oder all diesen Faktoren ab.

Weitere Informationen zu OCR und zu erwartenden Qualitätsstufen entnehmen Sie bitte dem Masterplan (S. 31 ff.).

Layoutanalyse (OLR)

Layoutanalyse oder auch OLR bezeichnet die automatisierte Erkennung der Struktur eines Dokuments bzw. einer Seite. Die Layoutanalyse kann als integraler Bestandteil der OCR durchgeführt werden oder als separater bzw. nachträglicher Bearbeitungsschritt.

Insbesondere bei Zeitungen spielt die Qualität der OLR eine entscheidende Rolle, um z.B. Spalten und Artikel zu erkennen und so die Lesbarkeit einzelner logischer Elemente auf einer Seite zu bewahren. Gerade für historische Zeitungen besteht hier jedoch noch Entwicklungsbedarf, z.B. hinsichtlich der Erkennung und Klassifikation von unterschiedlichen Regionen (Artikel, Werbung, Metainformationen wie Preis und Herausgeber, grafische Elemente und Abbildungen, Tabellen, usw.).

Tiefenerschließung

Neben der unbedingt durchzuführenden Erzeugung von Struktur- und Metadaten, die den erwähnen Erschließungs-Grundstandard bilden, und die Navigation innerhalb der Zeitungssammlung ermöglichen, können die Digitalisate tiefergehender und granularer erschlossen werden (vgl. Masterplan, S. 29 f.).

Jenseits von OCR und OLR sind weitere Arbeitsschritte denkbar: Automatische Auswertung typographischer Merkmale (z.B. zur Erfassung von Schlagzeilen), Erfassung oder Korrekturen von Überschriften, Bilderkennung und -erschließung, semantische Anreicherungen usw.

Die hier unter Vorbereitung/Digitalisierung beschriebenen Bearbeitungsstufen stellen eine Empfehlung dar und sind nicht im Sinne von Ausschließlichkeit zu verstehen. Allerdings ist zu beachten, dass eine granulare Erschließung auch eine granulare Such- und Präsentationsumgebung benötigt, um die mit viel Aufwand erzeugten Datenfacetten oder -anreicherungen auch angemessen recherchierbar und darstellbar zu machen.

ZDB-Erschließung

Die ZDB ist ein wichtiges Katalogisierungs- und Nachweisinstrument im Kontext der Zeitungsdigitalisierung in Deutschland. Bevor Sie die Arbeit beginnen, sollte Ihr Vorhaben in der ZDB angekündigt bzw. dokumentiert werden, damit Doppelarbeit vermieden werden kann. Nach Abschluss der Digitalisierung sollten die Nachweise der Digitalisate ebenfalls in der ZDB erfasst und nachgewiesen werden.

Von den bibliographischen Titeldatensätzen der ZDB aus können Benutzer direkt in die lokalen oder überregionalen Rechercheumgebungen geleitet werden. Somit unterstützt die ZDB die Sichtbarkeit Ihrer digitalen Bestände in ganz erheblicher Weise.

Für digitalisierende Einrichtungen stellt die ZDB detaillierte Erfassungsanweisungen zur Verfügung, die sie hier einsehen können.

Qualitätskontrolle: Erschließung

Zur Qualitätskontrolle der Erschließung bieten sich die folgenden Verfahren an:

- **Grundstandard** - eine Überprüfung des Grundstandards der Erschließung kann bspw. vor der Veröffentlichung im Web erfolgen, indem stichprobenhaft einzelne Ausgaben via Kalender aufgerufen und die Seiten durchgeblättert werden.
- **OCR** - Eine Kontrolle der Qualität von OCR lässt sich bspw. durch eine Evaluation mittels sog. "Ground Truth" (manuell erzeugte, zu 100% korrekte Transkriptionen des Text und Layouts eines Dokuments/einer Seite) ermitteln. Die Erstellung geeigneten Ground Truth-Materials ist jedoch mit erheblichem Aufwand verbunden. Weitere Informationen und Werkzeuge für die OCR Evaluation mit Ground Truth sind im Masterplan (S. 31 ff.) sowie hier zu finden (Englisch).
- **OLR** - Analog zur Qualitätskontrolle der OCR lassen sich auch die Ergebnisse von OLR mit Ground Truth evaluieren. Hierbei ist insbesondere auf eine an den Benutzeranforderungen orientierte Evaluation zu achten.

Bereitstellung

Präsentation

Digitalisierte Zeitungen sind ein ganz besonderes Material, das spezifische Anforderungen an die digitalen Recherche- und Präsentationsumgebungen stellt. Selbst wenn nach dem Grundstandard Stufe 1 (Erzeugung von digitalen Abbildungen einschließlich Strukturdaten) erschlossen wird, wird mindestens eine Kalenderfunktion zur Navigation benötigt. Eine weitergehende Erschließung erfordert weitere Funktionen - die OCR-Erschließung erfordert einen Ort zur Darstellung der Texte, Textauszeichnungen müssen nachvollziehbar dargestellt werden können, semantische Anreicherungen müssen gefunden und verstanden werden können, und nicht zuletzt erfordern die originalen Zeitungsformate und die Größe moderner Bildschirme eine Zoomfunktion. Es ist zu bedenken, dass solche spezifischen Umgebungen nicht ohne erhebliche Aufwände geschaffen werden können.

So naheliegend es ist, bei der Verwertung der erzeugten Daten zunächst Ihre eigenen Online-Umgebungen und -Services im Auge zu halten, so sollte doch auch daran gedacht werden, dass mit Ihren Zeitungsdaten in der Regel umso sinnvoller gearbeitet werden kann, je größer die durchsuchbaren Datenmengen sind. Mit anderen Worten: digitale Daten aus Zeitungssammlungen anderer Einrichtungen erhöhen den Nutzen Ihrer eigenen Sammlung. Ein gemeinsamer, überregionaler Nachweis liegt also durchaus im Interesse der einzelnen digitalisierenden Einrichtung.

Für einen solchen überregionalen Nachweis sind Schnittstellen von großer Bedeutung, mittels derer die entsprechenden Portale, z.B. die DDB, Ihre Daten automatisiert einsammeln können (vgl. Masterplan, S. 57 f.). Zunehmend stellen die Kultur- und Wissenschaftseinrichtungen ihre Daten über geeignete Schnittstellen aber auch Endnutzern zur Verfügung. Insbesondere im Bereich der Digital Humanities favorisieren viele Endnutzer den Download der angebotenen Daten, um diese in eigenen digitalen Umgebungen optimal analysieren zu können.

DFG-Viewer

Der DFG-Viewer ist ein Browser-Webdienst zur Anzeige von Digitalisaten aus dezentralen Bibliotheksrepositorien. Im Rahmen des DFG-Pilotprojekts zur Digitalisierung historischer Zeitungen wurde der DFG-Viewer um Funktionen für die speziellen Erfordernisse des Gattungstyps Zeitungen erweitert (vgl. Masterplan, S. 16 ff.). Diese umfassen:

- generische Umsetzung der dreistufigen Kalendernavigation (Titel, Jahrgang, Ausgabe)
- stufenloser Zoom für alle Zeitungsformate über OpenLayers
- freie Bildpositionierung (Panning) über OpenLayers
- verteilte Volltextsuche (SRU-Schnittstelle, ALTO-Format)
- Überarbeitung der Formatdokumentationen für METS und MODS
- Erstellung von Beispielen für den Demonstrator

Der Quellcode des DFG-Viewers wurde auf der Entwicklungsplattform GitHub unter der Open-Source-Lizenz GPL3 veröffentlicht und kann frei nachgenutzt werden. Der von der SLUB Dresden betriebene DFG-Viewer ist ein freier Webdienst, der ohne lokale Installation verwendet werden kann.

Schnittstellen

Die Bereitstellung der Digitalisate sollte über die reine Präsentation im Webportal hinaus auch über geeignete Schnittstellen, zudem möglichst ohne rechtliche Beschränkungen hinsichtlich der Nutzbarkeit, erfolgen. Schnittstellen, über die Daten bereitgestellt werden, dienen mindestens zwei Zwecken:

- Bereitstellung der Daten für Endnutzer. Insbesondere im Bereich der Digital Humanities favorisieren viele Endnutzer den Download der angebotenen Daten, um diese in eigenen digitalen Umgebungen optimal analysieren zu können.
- Überregionale Nachweisportale, z.B. die DDB, können Ihre Daten über Schnittstellen automatisiert einsammeln.

Gängige Schnittstellenprotokolle sind:

- OAI-PMH für die Bereitstellung von Metadaten. Die Bereitstellung der deskriptiven Metadaten über eine OAI-PMH-Schnittstelle ist verpflichtend, wahlweise im eigenen System oder über ein geeignetes Zielportal.
- SRU für die Bereitstellung von durchsuchbaren Volltexten. Liegen die Volltexte im Format ALTO vor, so unterstützt der DFG-Viewer die Volltextsuche via SRU.
- IIIF ist ein relativ neuer Standard für die Bereitstellung von Images, Volltexten und Annotationen.

Persistenz

Digitalisate müssen, um von anderen Objekten oder Datenbanken aus erreichbar zu sein, eindeutig angesprochen werden können. Dazu ist über die übliche Zitierform hinaus, die durch ein Angebot in der Navigationssoftware als klassische Form weitergenutzt werden kann und sollte, die Festlegung und die online zugängliche Dokumentation von Adressierungstechniken erforderlich. Sichergestellt werden müssen die Erreichbarkeit und Zitierbarkeit einer Ressource als Ganzes und die Erreichbarkeit und Zitierbarkeit von einzelnen physischen Seiten dieses Werkes.

Die DFG-Praxisregeln "Digitalisierung" fordern die Sicherstellung einer persistenten Adressierbarkeit der online bereitgestellten Ressourcen mit einer "größtmögliche[n] Granularität". Hierbei können verschiedene Persistenz-Verfahren genutzt werden (PURL,

URN, DOI, Handle etc.). Die Nutzung von URNs wird jedoch von den Praxisregeln "nachdrücklich empfohlen".

Das 2009 von der Deutschen Nationalbibliothek und der ULB Halle erarbeitete Verfahren **URN granular** bedient sich einer Adressierungstechnik, die auf die granulare Adressierung monographischer Werke unter Maßgabe der oben zitierten Anforderungen der DFG-Praxisregeln abzielt. Die Nutzung dieses Verfahrens bei der Adressierung komplexerer Objekte ist jedoch nur eingeschränkt möglich. Basierend auf einem Konzept, das in der generellen RFC-Spezifikation zu URIs als "fragment identifier component of a URI" bezeichnet wird, wurde daher **URN granular 2** entwickelt, um den komplexeren Herausforderungen bei der Bereitstellung und Adressierung von digitalisierten Zeitungen Rechnung zu tragen.

Eine andere Methode sind **DOI**s, die eindeutige und dauerhafte Identifikatoren für digitale Ressourcen bereitstellen und in der Praxis vor allem für die Referenzierung von Artikeln wissenschaftlicher Fachzeitschriften verwendet werden. Das DOI-System kennt granulare Adressierungstechniken für den Zugriff auf Teile von digitalen Ressourcen, wie sie für das Verfahren URN granular 2 beschrieben wurden. Weitere Details hierzu finden Sie auch im Masterplan auf S. 37 ff.

Archivierung

LZA

Die Langzeitarchivierung der in einem Projekt entstehenden Digitalisate ist eine der grundlegenden Bedingungen, die für eine Förderfähigkeit gestellt werden. Damit ist nicht nur die Sicherung der entstehenden Dateien vor Kompromittierung oder Verlust gemeint, sondern auch ihre stabile Adressierung und Zitierbarkeit im Netz, zuverlässige Zugangssysteme und komfortable Nachnutzungsmöglichkeiten. Hinzu kommt die langfristige Sicherung der Nutzbarkeit der Dateien in zukünftigen Systemumgebungen, die andere Dateiformate als die heute verbreiteten voraussetzen: Es müssen rechtzeitig Migrationen in andere Formate vorgenommen werden, die über die erforderlichen Features für die Nutzung der Dateien verfügen. Auch andere Formen der langfristigen Absicherung der Nutzbarkeit der Daten sind denkbar.

In der Regel wird eine digitalisierende Einrichtung diese Aufgabe nicht selbst übernehmen, sondern an einen geeigneten Dienstleister übergeben. Dabei ist es wichtig, in der vertraglichen Vereinbarung mit diesem die Anforderungen an die Qualität der Langzeitarchivierung zu vereinbaren und Sicherheit darüber zu erlangen, welche wechselseitigen Leistungen erwartet werden. Es empfiehlt sich, eine solche Vereinbarung bereits im Vorfeld eines Digitalisierungsprojekts abzuschließen, um sich daraus ergebende Vorbedingungen insbesondere für die Aufbereitung der im Projekt entstehenden Daten bereits frühzeitig festlegen zu können.

In Deutschland hat sich zur Beratung und wechselseitigen Unterstützung das nestor-Netzwerk etabliert, in dem auch alle wesentlichen Anbieter von Dienstleistungen auf dem Gebiet der Langzeitverfügbarkeit vertreten sind. Ein wichtiger Indikator für die Vertrauenswürdigkeit eines Archivservices ist daher auch die Zertifizierung des jeweiligen Archivs, die mit dem "nestor-Siegel" dokumentiert wird.

Entscheidungshilfe:
Mikrofilm oder Papieroriginal als Vorlage für die Digitalisierung

Vor jeder Digitalisierung steht zunächst die Identifikation und Sichtung der für ein Projekt angedachten Bestandsgruppe. Die Vorlagenform ist entsprechend der Digitalisierungsziele (bspw. Massendigitalisierung vs. Herstellung hochwertiger Reproduktionen von Abbildungen) zu bestimmen. Dabei ist zu ermitteln, ob bereits ein Mikrofilm vorhanden ist und ggf. zu überprüfen, inwieweit der Mikrofilm auch für eine Digitalisierung geeignet ist.

Von der Printausgabe zu scannen ermöglicht eine Reproduktion in bester Faksimilequalität mit einem optimalen Gesamteindruck des Originals und bietet zudem sehr gute Voraussetzungen für eine spätere OCR. Insbesondere unikale und seltene Zeitungen sowie Zeitungen von besonderem kulturhistorischem Wert, z.B. mit wichtigen Bildanteilen, sollten vom Original in Farbe oder in Graustufe gescannt werden. Es ist eine relativ zeitaufwändige Variante, sowohl Inhouse als auch durch einen Dienstleister. Ob ein vorliegendes Original für eine Digitalisierung geeignet ist, ist vorab zu prüfen (s.u. Checkliste „Original“).

Von Mikrofilmen zu scannen ermöglicht es, bereits sicherungsverfilmte fragile Originale zu schonen und die mit einer Verfilmung durchgeführten Vorleistungen (Verzeichnung, ggf. Lückenergänzung) zu nutzen. Von Filmen kann schneller und preisgünstiger als von Originalen eine Massendigitalisierung erfolgen. Dies gilt allerdings nur für qualitativ stark homogene bzw. standardisierte Mikrofilme. Die Qualität der Filme ist daher vor einer Digitalisierung zu prüfen (s.u. Checkliste „Mikrofilm“).“

In manchen Fällen empfiehlt sich die Digitalisierung einer Testcharge, um daran stichprobenhaft die erzielbare Qualität und zu erwartenden Aufwände für u.a. Nachbearbeitung zu ermitteln. Der Umfang einer Stichprobe ist immer in Abhängigkeit von der Anzahl der für die Digitalisierung vorgesehenen Zeitungen zu bestimmen, dabei sollte jedoch zumindest von jedem Titel auch eine Ausgabe in der Stichprobe berücksichtigt werden. Auch im Falle von grundlegenderen Änderungen im Erscheinungsbild einer Zeitung sollte jeweils eine entsprechende Ausgabe in der Stichprobe enthalten sein. Bei der Digitalisierung von Zeitungen mit langen Erscheinungsverläufen empfiehlt sich die Abdeckung unterschiedlicher Zeiträume für die Zusammenstellung einer Testcharge.

(1) Checkliste Mikrofilm

Hinweis: Alle Fragen sollen mit Ja oder Nein beantwortet werden. Je mehr positive Antworten gegeben werden, desto geeigneter sind die Mikrofilme für eine Digitalisierung. Eine gute und homogene Qualität der ausgewählten Mikrofilme ist Voraussetzung für ein besonders wirtschaftliches Digitalisierungsverfahren. Die Fragen differenzieren „Muss"- und „Soll"-Anforderungen. Erfüllen die ausgewählten Mikrofilme die „Muss"-Anforderungen nicht, ist mit Mehrkosten zu rechnen. Sind die ausgewählten Mikrofilme zudem von schlechter Aufnahmequalität, ist eine Digitalisierung vom Original zu erwägen (s.u. Checkliste „Original"). Die „Soll"-Anforderungen sind als optional anzusehen. Mikrofilme, die das jeweilige Kriterium nicht erfüllen, sind nicht prinzipiell ungeeignet für eine Digitalisierung, doch sollte in jedem Einzelfall kritisch geprüft werden, ob Qualitätsabstriche oder erhöhter Arbeitsaufwand in Kauf genommen werden sollen.[104]

1. Gibt es Masterfilme, die für die Digitalisierung genutzt werden können? Zu prüfen sind mindestens die folgenden Nachweissysteme:
a) lokaler Bestand einer Bibliothek
b) ZDB
c) MFA
d) EROMM (Muss)
Ja ☐
Nein ☐

2. Handelt es sich bei den Mikroformen um einen vollständigen Bestand von weitestgehend homogener Qualität? (Muss)
Ja ☐
Nein ☐
(Anmerkung: nur stichprobenhafte Prüfung sinnvoll; zu ermitteln über Abgleich des Erscheinungsverlaufs eines Titels in der ZDB mit den Mikrofilm-Beständen zu diesem Titel; ggf. Identifizierung von Bestandslücken und Ermittlung entsprechender Bestände in anderen Einrichtungen mittels ZDB, MFA und EROMM)

3. Sind auf dem Film Metadaten vorhanden, die eine eindeutige Zuordnung der Abbildungen zum Zeitungstitel ermöglichen? Das gilt auch für Nachhol-, Berichtigungs- und Wiederholungsaufnahmen. (Muss)
Ja ☐
Nein ☐
(Anmerkung: Entsprechende Angaben sind üblicherweise auf dem Vorspann des Mikrofilms zu finden; Achtung: Metadaten müssen nicht immer korrekt sein. In jedem Fall empfiehlt sich eine

[104] Ein ähnliches Bewertungsraster findet sich auch bei: https://www.archivschule.de/uploads/Forschung/Digitalisierung/Handreichungen/Checkliste_fuer_Mikroformen.pdf

abschließende Vollständigkeitskontrolle der Metadaten, um das Weiterführen unbemerkter Bestandslücken zu vermeiden.)

4. Sind auf dem Film Testtafeln mit Graufeldern (s/w-Film) bzw. Colorcharts (Farbfilm) vorhanden? (Muss)
Ja ☐
Nein ☐

5. Sind die Aufnahmen der Zeitungsseite auf dem Film gut lesbar und ist die Schärfe ausreichend? (Muss)
Schärfe und Lesbarkeit des Films müssen gut sein. Beides kann mit der ISO-Testtafel nach DIN 19051-1 stichprobenhaft überprüft werden. Mindestens das Testzeichen 84 muss lesbar sein. Testtafeln befinden sich i.d.R. auf den Vorspännen der Mikrofilme.
Ja ☐
Nein ☐

6. Durch die Weiterentwicklung der Verfilmungstechnik und -standards sind Mikroformen jüngeren Entstehungsdatum häufig von besserer Qualität. Sie sollten in der Regel nicht älter als 25 Jahre sein.
Wurden die Mikroformen nach 1990 erstellt? (Soll)
Ja ☐
Nein ☐
(Anmerkung: Das Herstellungsdatum ist üblicherweise auf dem Vorspann des Mikrofilms zu finden)

7. Hat der Film ausreichende unbelichtete Vor- und Abspänne, so dass die eingesetzten Scangeräte Filmanfang und –ende vollständig erfassen können? (Soll)
Ja ☐
Nein ☐
(Anmerkungen: ggf. müssen Vor- und Abspann nachträglich angebracht werden)

8. Sind die Zeitungsseiten auf dem Film vollständig abgebildet (in der überwiegenden Zahl der Fälle keine abgeschnittenen Ränder, Textverlust durch zu engen Falz o.ä.)? (Muss)
Ja ☐
Nein ☐
(Anmerkungen: Vollständigkeit in der Darstellung kann sich innerhalb eines Titels, abhängig vor allem von der gebundenen Vorlage der Verfilmung, ganz unterschiedlich darstellen; vorab ist lediglich eine stichprobenhafte Prüfung möglich)

9. Eine Mikrofilm-Abbildung besteht typischerweise aus 2 Zeitungsseiten. Können die einzelnen Seiten automatisiert getrennt werden? (Muss)
Ja ☐
Nein ☐
(Anmerkung: Üblicherweise werden Mikrorollfilmscanner mit entsprechender Software geliefert, die eine automatisierte Identifikation und Trennung von Einzelseiten ermöglicht; spezifische Aufnahmebedingungen, z.B. schief abgefilmte Seiten oder Seiten mit Textverlust, erfordern gelegentlich das manuelle Nachjustieren der automatisierten Rahmensetzungen oder machen es in

seltenen Fällen unmöglich, Rahmen überhaupt sinnvoll zu setzen. Hier wären stichprobenhaft einzelne Filmrollen zu prüfen.)

10. Ist die Aufnahme der abgebildeten Zeitungsseiten frei von störenden Effekten (z.B. Verzerrungen/gewellte Zeilen, Verschmutzungen oder Beschädigungen des Films)? (Soll)
Ja ☐
Nein ☐
(Anmerkung: aufgrund der Testdigitalisierung einer Stichprobe kann entschieden werden, ab welcher Häufung von störenden Effekten Nachteile bei einer späteren OCR-Bearbeitung zu erwarten sind.)

11. Ermöglicht der Seitenzustand des Mikrofilms hinsichtlich seiner optischen Beschaffenheit Scans, die keine relevanten OCR-Nachteile erwarten lassen?
Ja ☐ (Muss)
Nein ☐
(Anmerkung: Zu prüfen ist, ob die Mikrofilmseiten in den Textbereichen in einem signifikanten Mengenbereich Textstellen aufweisen, die sich z.B. durch geringen Kontrast (insb. bei Graustufenfilmen) negativ auf eine spätere OCR-Bearbeitung auswirken könnten.)

(2) Checkliste Original

Hinweis: Alle Fragen sind *mit Ja oder Nein* zu *beantworte*n*. Je mehr positive Antworten gegeben werden, desto geeigneter sind die Originale für eine Digitalisierung. Erfüllen die ausgewählten Originale die „Muss"-Anforderungen nicht, ist mit Mehrkosten zu rechnen. Originale, die das jeweilige Kriterium nicht erfüllen, sind nicht prinzipiell ungeeignet für eine Digitalisierung, doch sollte in jedem Einzelfall kritisch abgewogen werden, ob mindere Qualität oder erhöhter Arbeitsaufwand in Kauf genommen werden sollen. Insbesondere wertvolle und/oder unikale Bestände rechtfertigen u.U. auch bei Vorliegen größerer Schäden eine Digitalisierung.*

1. Ist das Original in physischer Hinsicht für die Digitalisierung geeignet?
Ja ☐ (Muss)
Nein ☐
(Anmerkung: Zu prüfen ist durch Bandautopsie des Originals, ob der konservatorische Zustand (Bindung, Einzelseiten hinsichtlich Schäden [Gebrauchsschäden, Papierzersetzung/-beschädigung, Wasserschäden, Schimmel, Verschmutzungsgrad usw.]) die Digitalisierung ohne erheblichen Zusatzaufwand erlaubt) oder ob ggf. auch durch eine Digitalisierung entstehende Schäden vertretbar sind (z.B. wegen dubletter Bestände).

2. Ermöglicht der Seitenzustand des Originals hinsichtlich seiner optischen Beschaffenheit Scans, die keine relevanten OCR-Nachteile erwarten lassen?
Ja ☐ (Muss)
Nein ☐
(Anmerkung: Zu prüfen ist, ob die Originalseiten in den Textbereichen in einem signifikanten Mengenbereich Textstellen aufweisen, die sich z.B. durch stark welliges Papier, Falten, Verschmutzungen, Schäden im Papier, stark durchscheinende Seiten, negativ auf eine spätere OCR-Bearbeitung auswirken.)

3. Liegen die Originale vollständig vor oder sind vollständig zu beschaffen?
Ja ☐ (Soll)
Nein ☐
(Anmerkung: zu ermitteln über einen einrichtungsübergreifenden Bestandsvergleich in der ZDB)

4. Lässt das gebundene Original einen Öffnungswinkel von mindestens 90 Grad zu?
Ja ☐ (Soll)
Nein ☐
(Anmerkung: Die Digitalisierung mit einem 180-Grad-Öffnungswinkel ist die kostengünstigere Variante)

5. Wenn die Digitalisierung im eigenen Haus erfolgen soll: Lässt das Format der Zeitung eine Digitalisierung mit der vorhandenen Scanhardware zu?
Ja ☐ (Muss)
Nein ☐

Checkliste:
Digitalisierung in-house oder mit Dienstleister („Outsourcing“)

Bei der Wahl zwischen einer Digitalisierung mit eigenem Personal und Equipment („In-house“) und einer Vergabe der Digitalisierung an einen Dienstleister („Outsourcing“) handelt es sich oftmals um eine Grundsatzentscheidung, bei der jenseits des einzelnen Projekts z.B. auch die strategische und personelle Ausrichtung einer Einrichtung eine wichtige Rolle spielen kann. Zahlreiche Faktoren sind von Bedeutung und werden je nach Ausgangslage und Zielvorstellung ganz anders zu gewichten sein.

Wird im eigenen Haus digitalisiert, sind nicht nur die unmittelbar anfallenden Ausgaben zu berücksichtigen, sondern auch die Folgekosten. Wartung, Reparaturen, Austausch von Geräten, Software-Updates und ähnliches erzeugen nicht nur häufig übersehene Kosten, sondern verursachen auch Stillstände in der Produktion. Auf der anderen Seite darf nicht vergessen werden, dass auch die Vergabe an einen Dienstleister erhebliche Ressourcen vor Ort bindet. So entstehen Aufwände durch Ausschreibungsverfahren, Definition der Anforderungen, Materialauswahl, -aushebung, -übergabe, Qualitätskontrolle der Lieferungen etc.

Bei der Abwägung Ihrer Entscheidung können die folgenden Überlegungen nützlich sein.

Gründe für eine In-house Digitalisierung

- eine ausreichende Infrastruktur – Personal, Scangeräte für die entsprechenden Formate, Rechenleistung, Speicherplatz – steht bereits zur Verfügung und kann für die gesamte Projektdauer genutzt werden
- eine ausreichende Infrastruktur steht nicht zur Verfügung, die Einrichtung ist aber langfristig am Aufbau einer solchen interessiert
- eine strukturelle Finanzierung für Aufbau und nachhaltigen Betrieb von technischer und personeller Infrastruktur steht zur Verfügung
- die eigenen Zeitungsbestände und die eigenen Digitalisierungsvorhaben sind von erheblicher Größe, so dass sich die Kosten teurer Geräte (z.B. Mikrofilmscanner) amortisieren können
- eigene Kapazitäten für Softwareentwicklung oder -anpassung sind vorhanden, die Digitalisierung kann direkt mit den sich aus dem dauerhaften Betrieb ergebenden technischen Anforderungen abgestimmt werden

- die zu digitalisierenden Zeitungsbestände sind so fragil, dass ein Transport nicht in Frage kommt; notwendige Restaurierungsarbeiten können vor Ort durchgeführt werden (aber: manche Dienstleister digitalisieren auch vor Ort)

Gründe für ein Outsourcing

- eine ausreichende Infrastruktur steht nicht zur Verfügung
- Ihr Digitalisierungsvorhaben ist eine einmalige Unternehmung
- für eine Digitalisierung stehen ausschließlich einmalige Finanzmittel zur Verfügung
- die eigenen Zeitungsbestände sind so klein, dass sich die durch den Aufbau der notwendigen Infrastruktur entstehenden Kosten nicht amortisieren können

ZDB-Erfassungsanweisung

Nutzung der ZDB für Zeitungsdigitalisierungsprojekte

Zur ZDB

Die ZDB ist das zentrale Katalogisierungs- und Nachweissystem für periodisch erscheinende Publikationen, die in ca. 4.000 deutschen und österreichischen Kultur- und Wissenschaftseinrichtungen verfügbar sind. Dieser hohe Abdeckungsgrad macht die ZDB zum wichtigen Hilfsmittel bei der Vorbereitung von Zeitungsdigitalisierungsprojekten in Deutschland und zum Ort, an dem Ihre Projektergebnisse bekannt gemacht werden.

In diesem Dokument sollen digitalisierenden Einrichtungen erste Hinweise zur Nutzung der ZDB für die Katalogisierung der eigenen Digitalisierungsaktivitäten erhalten. Weiterführende Links finden Sie am Ende dieses Dokuments. Grundsätzlich ist die Teilnahme an der ZDB kostenfrei und steht jeder Einrichtung offen. Bitte beachten Sie, dass die ZDB ein System ist, in dem viele Menschen mit unterschiedlichen Berechtigungen arbeiten und nicht alle BearbeiterInnen alles machen können.

ZDB und Zeitungsdigitalisierungen

In der ZDB sind aktuell (Stand Mai 2017) ca. 61.500 Zeitungstitel verzeichnet. Hiervon sind ca. 22.000 als „Deutsche Historische Zeitungen“ zu betrachten, d.h. diese Titel sind zuerst zwischen 1600 und 1944 im Deutschen Reich oder in Deutschland erschienen oder sie sind, unabhängig vom Druckort, in deutscher Sprache verfasst.

Bei Ihrer Digitalisierungsarbeit ist die ZDB an verschiedenen Stellen Ihres Workflows zu berücksichtigen:

1. Nachdem Ihr Digitalisierungsprojekt geplant, aber bevor es begonnen wurde, machen Sie Ihr Vorhaben bitte bekannt, indem Sie es in der ZDB ankündigen. Damit vermeiden Sie mögliche Doppelarbeiten.
2. Nach erfolgter Digitalisierung weisen Sie bitte in der ZDB Ihre Ergebnisse nach.
3. Ebenfalls nach erfolgter Digitalisierung löschen Sie die unter 1. erwähnte Digitalisierungsabsicht bitte wieder aus der ZDB.

Die ZDB unterscheidet **Titeldaten**, also bibliographische Daten, die einen Zeitungstitel beschreiben, und **Exemplardaten**, also Daten, die das Bestandssegment beschreiben, das in einer bestimmten Einrichtung zu einem bestimmten Titel vorliegt. In der Regel wird für den Zeitungstitel, den Sie digitalisieren möchten, in der ZDB bereits ein entsprechender Titeldatensatz der gedruckten Zeitung vorhanden sein. Einem solchen Titeldatensatz für die gedruckte Zeitung, der sog. **A-Aufnahme** (A wie analog) muss für den Nachweis der neu erstellten digitalen Ausgaben eine sog. **O-Aufnahme** (O wie online) zur Seite gestellt werden.

Für die Katalogisierung in der ZDB stehen zwei unterschiedliche Verfahren bereit, die beide eine Anmeldung bei dem ZDB-Benutzerservice erfordern.

1. **WinIBW**

Die in zahlreichen Bibliotheken verwendete Redaktionsschnittstelle WinIBW ist ein komplexes Expertensystem, das zur Bedienung spezifische Kenntnisse voraussetzt. Bitte erkundigen Sie sich ggf. im Kreise der Kollegen und Kolleginnen, ob solche Kenntnisse in Ihrer Einrichtung schon existieren.
Die Erzeugung einer O-Aufnahme des von Ihnen gewählten Zeitungstitels erfolgt automatisiert mittels eines in der WinIBW hinterlegten Scripts. Die so erstellte O-Aufnahme enthält alle relevanten Informationen der A-Aufnahme.

Einzelschritte in der WinIBW

- Ankündigung der Digitalisierungsabsicht:
  - Erfassung im PICA-Feld 4260 der A-Aufnahme
- Erfassung der von Ihnen digitalisierten Bestandssegmente
  - Erzeugung einer O-Aufnahme
  - Erfassung einer URL, die zu Ihren Digitalisaten führt, im PICA-Feld 4085 der Titeldaten (O-Aufnahme)
  - Erfassung des Bestandsverlaufs der Digitalisate im PICA-Feld 4085 der Titeldaten (O-Aufnahme)
  - Erfassung des Bestandsverlaufs der Digitalisate im PICA-Feld 8032 der Exemplardaten
    (Anm.: der Bestandsverlauf der Digitalisate ist sowohl in den Titeldaten als auch in den Exemplardaten zu erfassen; dies ist notwendig, damit andere Bibliotheken direkt aus den Titeldaten die entsprechenden Informationen entnehmen können – wichtig z.B. für Digitalisierungsprojekte, die von mehreren Einrichtungen durchgeführt werden.)
  - Ggf. Erfassung ergänzender Informationen zum Titel
- Löschung der Digitalisierungsabsicht aus dem PICA-Feld 4260 der Titeldaten (A-Aufnahme)

Detaillierte Erfassungsanweisungen für die Arbeit mit der WinIBW liegen in verschiedenen Zuschnitten vor:

- Dokumentation des vollständigen ZDB-Formats: http://www.zeitschriftendatenbank.de/erschliessung/zdbformat/
- Dokumentation der Besonderheiten der Zeitungskatalogisierung: http://www.zeitschriftendatenbank.de/fileadmin/user_upload/ZDB/dokumente/rda/modul5/Modul_5B_15_Zeitungen_pica_20170328.pdf
- Dokumentation der Besonderheiten von Original und Reproduktion (hier Digitalisat):
- http://www.zeitschriftendatenbank.de/fileadmin/user_upload/ZDB/pdf/zdbformat/4256.pdf

2. WebCat

Für eine vereinfachte Katalogisierung stellt die ZDB eine webbasierte Redaktionsschnittstelle – WebCat – zur Verfügung, die die wichtigsten Katalogisierungsfacetten unterstützt. Sollten die Funktionalitäten des WebCat im Einzelfall nicht genügen, so stehen kompetente MitarbeiterInnen der Zeitungs-Redaktionsteams in Berlin zu Ihrer Unterstützung bereit. Die Erzeugung einer O-Aufnahme des von Ihnen gewählten Zeitungstitels erfolgt automatisiert mittels eines Auswahlmenüs. Die so erstellte O-Aufnahme enthält alle relevanten Informationen der A-Aufnahme.

Einzelschritte im WebCat

- Ankündigung der Digitalisierungsabsicht:
    - Mit WebCat nicht möglich; bitte informieren Sie die Zeitungs-Redaktion der ZDB, die Ihre Digitalisierungsabsicht für Sie in das PICA-Feld 4260 der Titeldaten (A-Aufnahme) eintragen werden
    - Erfassung einer URL, die zu Ihren Digitalisaten führt, in den Titeldaten (O-Aufnahme)
    - Erfassung des Bestandsverlaufs der Digitalisate in den Titeldaten (O-Aufnahme)
    - Erfassung des Bestandsverlaufs in den Exemplardaten
      (Anm.: der Bestandsverlauf der Digitalisate ist sowohl in den Titeldaten als auch in den Exemplardaten zu erfassen; dies ist notwendig, damit andere Bibliotheken direkt aus den Titeldaten die entsprechenden Informationen entnehmen können – wichtig z.B. für Digitalisierungsprojekte, die von mehreren Einrichtungen durchgeführt werden.)
    - Ggf. Erfassung ergänzender Informationen zum Titel
- Löschung der Digitalisierungsabsicht aus dem PICA-Feld 4260 der A-Aufnahme durch die Zeitungs-Zentralredaktion

Dokumentation der WebCat-Funktionen:
http://www.zeitschriftendatenbank.de/erschliessung/webcat/).

Anmeldungen von Zugängen zu WinIBW und WebCat:

Tel.: +49 30 266 434444
Email: zdb-hotline@sbb.spk-berlin.de

Für Fragen zu Zeitungen, Datenformaten, Katalogisierungsdetails etc. wenden Sie sich bitte an:

Tel.: +49 30 266 434255
Email: carmen.thomas@sbb.spk-berlin.de