

Die Berliner Volkszeitung digital erforschen: Digitales Kuratieren, Metadaten, Text Mining: Praktiken und Potentiale historischer Presseforschung in digitalen Kontexten

Koenen, Erik; Krause, Falko; Sax, Simon

Erstveröffentlichung / Primary Publication

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Koenen, E., Krause, F., & Sax, S. (2022). Die Berliner Volkszeitung digital erforschen: Digitales Kuratieren, Metadaten, Text Mining: Praktiken und Potentiale historischer Presseforschung in digitalen Kontexten. In C. Schwarzenegger, E. Koenen, C. Pentzold, T. Birkner, & C. Katzenbach (Hrsg.), *Digitale Kommunikation und Kommunikationsgeschichte: Perspektiven, Potentiale, Problemfelder* (S. 187-223). Berlin <https://doi.org/10.48541/dcr.v10.7>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:
<https://creativecommons.org/licenses/by/4.0>

Empfohlene Zitierung: Koenen, E., Krause, F., & Sax S. (2022). Die Berliner Volkszeitung digital erforschen. Digitales Kuratieren, Metadaten, Text Mining: Praktiken und Potentiale historischer Presseforschung in digitalen Kontexten. In C. Schwarzenegger, E. Koenen, C. Pentzold, T. Birkner & C. Katzenbach (Hrsg.), *Digitale Kommunikation und Kommunikationsgeschichte: Perspektiven, Potentiale, Problemfelder* (S. 187–223). <https://doi.org/10.48541/dcr.v10.7>

Zusammenfassung: Mit digitalisierten Zeitungen und digitalen Zeitungsportalen hat sich im Laufe der letzten zwei Jahrzehnte für die historische Presseforschung eine grundlegend neue Forschungs- und Quellensituation herausgebildet. Trotzdem fallen digitale Methoden und Werkzeuge in der pressehistorischen Methodenpraxis bislang kaum ins Gewicht und ebenso wird die Quellenpraxis mit digitalisierten Zeitungen selten thematisiert. In diesem Sinne stellt der Beitrag die Fragen, welche spezifischen epistemischen und quellenkritischen Praktiken für den Umgang mit Zeitungen vonnöten sind, wenn sie als digitale Objekte bzw. Ressourcen vorliegen, und wie pressehistorische Forschungsprozesse in digitalen Kontexten modelliert und operationalisiert werden können. Als Beispiel dient hierfür die digitalisierte Ausgabe der *Berliner Volkszeitung* (1890–1930), die das Zeitungsinformationssystem ZEFYS der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz zum europäischen Zeitungsportal *Europeana Newspapers* metadaten- und volltexterschlossen beigesteuert hat. Neben einer medien- und zeithistorischen Einordnung werden entlang der praktischen digitalen Forschung mit dem Zeitungskorpus drei Forschungsdimensionen systematisch diskutiert und mit Beispielen illustriert: (1.) Digitales Kuratieren, (2.) Metadaten und (3.) Text Mining.

Lizenz: Creative Commons Attribution 4.0 (CC-BY 4.0)

Erik Koenen, Falko Krause & Simon Sax

Die *Berliner Volkszeitung* digital erforschen

Digitales Kuratieren, Metadaten, Text Mining: Praktiken und Potentiale historischer Presseforschung in digitalen Kontexten

1 Einleitung

Angesichts der grundlegend neuen Forschungs- und Quellensituation, die sich mit digitalisierten Zeitungen und digitalen Zeitungsportalen herausgebildet hat, steht die historische Presseforschung vor der Aufgabe, sich intensiv mit den Möglichkeiten von Forschung in digitalen Kontexten auseinanderzusetzen und die Potentiale digitaler Methoden und Werkzeuge für die Forschungspraxis auszuloten (vgl. Koenen, 2018). Damit sind konkret zwei Fragen aufgerufen: *Erstens* ist zu diskutieren, welche Forschungsfragen und -interessen sich mit Zeitungen verfolgen lassen, wenn diese digitalisiert als Images, Metadaten und Volltexte vorliegen. *Zweitens* geht es um die Konzeption, Modellierung und Operationalisierung konkreter Forschungsprozesse für den praktischen Umgang mit digitalisierten Zeitungsressourcen.

Digitale Methoden und Werkzeuge fallen in der pressehistorischen Methodenpraxis bislang kaum ins Gewicht und sind wenig elaboriert. Mit vereinzelten Methodenversuchen befindet sich die historische Presseforschung in Punkto digitaler Methodenpraxis noch immer in einem Pionierstadium des Ausprobierens und

Testens. Gründe hierfür sind: (1.) die neuartigen methodischen Herausforderungen digitaler Methoden verbunden mit (2.) einer verbreiteten Skepsis gegenüber deren Leistungen im Vergleich zur herkömmlichen lesend-verstehenden Hermeneutik sowie (3.) nicht zuletzt die Verfügbarkeit entsprechender Quellenressourcen. „Gemessen an der Gesamtzahl der seit 1605 erschienenen Zeitungen“, so Astrid Blome (2018, S. 1), „liegt der digital verfügbare Anteil deutschsprachiger Titel bei wenigen Prozent.“ In der Tat täuschen die Millionen digitalisierter Zeitungssseiten sehr, mit denen Zeitungsportale wie zum Beispiel *digiPress* oder *ZEFYS* werben. Insgesamt sind von der vier Jahrhunderte umfassenden reichen deutschsprachigen Zeitungstradition bloß einige wenige Zeitungen digitalisiert. Hinzu kommt, dass Zeitungen kaum über ihren gesamten Erscheinungsverlauf hinweg digitalisiert werden. Und schließlich werden digitalisierte Zeitungen selten als digitale Ressourcen für Forschungszwecke frei zugänglich gemacht bzw. zur Verfügung gestellt.

Im Mittelpunkt des Beitrags steht die digitalisierte Ausgabe der *Berliner Volkszeitung* – eine von vier Zeitungen, die das Zeitungsinformationssystem *ZEFYS* der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz zum europäischen Zeitungsportal *Europeana Newspapers* metadaten- und volltexterschlossen beigesteuert hat. Angesichts der neun Jahrzehnte umspannenden Erscheinungsgeschichte der bereits 1853 aus der kurz zuvor verbotenen *Urwähler-Zeitung* hervorgegangenen und 1944 mit der *Berliner Morgenpost* fusionierten *Berliner Volkszeitung* liegt mit der für den Zeitraum 1890 bis 1930 digitalisierten Ausgabe allerdings nur ein kleiner Ausschnitt als digitale Ressource vor, die im Rahmen des Kultur-Hackatons *Coding da Vinci Berlin* 2017 unter CCO-Lizenz bereitgestellt wurde. Die (undokumentierten, für die Quellenkritik digitalisierter Zeitungen jedoch nicht unwesentlichen) Gründe für die Digitalisierung gerade dieses Zeitraums (bei der zudem Bestandslücken nicht virtuell ergänzt und geschlossen wurden) liegen wohl vor allem in „praktischen Erwägungen“ (Blome, 2018, S. 11), die dem oftmals nur als Projekt zu realisierenden aufwendigen und kostenintensiven Prozess der Zeitungsdigitalisierung immer wieder enge Grenzen setzen: zu geringe Drittmittel, zu kurze Projektlaufzeiten, zu wenig Personal. „Entsprechend heterogen sind die realisierten Projekte“, so Blome (2018, S. 11), und „umfassen in der Mehrzahl regional, thematisch oder forschungspraktisch selektierte Quellenkorpora sowie Rara.“

Im Prozess von Retrodigitalisierung und Retroedition werden Zeitungen generell als Quelle neu formatiert und „remediated‘ and not just reproduced“ (Nicholson, 2013, S. 64), sie werden zu Einsen und Nullen verdatet, werden zu Dateien,

Dokumenten und digitalen Korpora, zu „in Form gebrachter Information“ (Fickers, 2016) als Grundlage für eine nun auch mögliche maschinelle Lesbarkeit und Verarbeitung. Um mit diesen neuartigen Datenquellen umzugehen, bildet auch in der digitalen Methodendiskussion der historischen Presseforschung vor allem ein Ansatz den zentralen Bezugspunkt: Franco Morretis Konzept des „Distant Reading“ (Moretti, 2000, 2016; vgl. Koenen, 2018, S. 448–550). Im Gegensatz zum hermeneutisch-verstehenden textnahen „Close Reading“ werden beim „Distant Reading“ die Lese-prozesse an den Computer delegiert, was Forschenden einerseits erlaubt, immens große Mengen Text sehr schnell ‚zu lesen‘, und andererseits eine neue „spezifische Form von Erkenntnis“ (Moretti, 2009, S. 7) ermöglicht: „Weniger einzelne Elemente bedeuten eine bessere Übersicht über ihr Abhängigkeiten untereinander. Umrisse, Beziehungen und Strukturen werden so deutlich, Formen, letztlich Modelle.“

Morettis Forderung, „we know how to read texts, now let’s learn how not to read them“ (Moretti, 2000, S. 57), war und ist für viele in den Geistes- und historischen Wissenschaften eine gehörige Provokation. In den vergangenen Jahren hat sich die anfangs sehr aufgeregte und scharfe Diskussion um das „Für und Wider“ und die erkenntnispraktischen Leistungen und Vorteile von „Close“ vs. „Distant Reading“ beruhigt und in Richtung einer beide Seiten vermittelnden Kombination verschoben, wie sie u.a. unter Stichworten wie „Blended Reading“, „Scalable Reading“ oder „zooming in und zooming out“ ausprobiert und erörtert wird (vgl. Stulpe & Lemke, 2016; Weitin, 2017; Schruhl, 2018; Schwandt, 2018). Dahinter steht nicht zuletzt die Einsicht, dass es sich bei „Close“ und „Distant Reading“ ja eher um methodische Perspektiven und noch nicht „um konkrete Methoden oder klar umrissene Analyseprogramme“ handelt (Schruhl, 2018, S. 10). Sie verweisen zunächst einmal nur auf methodische Optionen und klären mitnichten die ganz forschungspraktische Frage, „wie Objekte konstituiert, befragt und behandelt werden“, müssen also erst einmal erkenntnistheoretisch, methodisch und technologisch als Praktiken der „Ausweitung von Aufmerksamkeit“ und spezifische „Objektumgangsweisen“ operationalisiert werden (Schruhl, 2018, S. 4, 6). Hybride Ansätze wie „Blended Reading“, „Scalable Reading“ oder „Zooming“ stehen hierbei für ein „integriertes Verständnis aller Akte des Lesens und der Analyse“ und resultieren aus der evidenten Erfahrung, dass sich Forschung in digitalen Kontexten eigentlich immer „innerhalb einer weiten *scale* medialer Formen und analytischer Aufbereitungen“ abspielt und konstituiert (Weitin, Gilli & Kunkel, 2016, S. 115). Was eine solche hybride Methodenpraxis für die historische Presseforschung bedeutet

und welche spezifischen „Objektumgangsweisen“ (Schruhl, 2018, S. 4) Zeitungen ebenso erfordern wie ermöglichen, wenn sie als digitale Objekte vorliegen, ist am konkreten Beispiel der *Berliner Volkszeitung* Thema dieses Beitrags. Auch digitale Quellen brauchen historische Einordnung und so wird im folgenden Abschnitt die *Berliner Volkszeitung* erst einmal ganz klassisch analog medien- und zeithistorisch kontextuiert. Danach werden entlang von (1.) digitalem Kuratieren, (2.) Metadaten und (3.) Text Mining unterschiedliche Praktiken im Umgang mit der *Berliner Volkszeitung* als digitaler Ressource vorgestellt. Zuletzt werden im Resümee Probleme und Potentiale dieser Methodenpraxis reflektiert.

2 Die Berliner Volkszeitung analog: medien- und zeithistorische Einordnung

Dass digitales Forschen im Kern „hybrides“ Forschen ist, das „in vielfacher Hinsicht von Praktiken der *analogen* Zeit geprägt“ wird (Föhr, 2018, S. 92), d.h. auf diesen aufbaut und eng mit diesen verschränkt ist, offenbart sich unmittelbar beim Begleitwissen, das Zeitungsportale zu den von ihnen vorgehaltenen digitalisierten Zeitungen anbieten und welches notwendig ist, um mit diesen komplexen digitalen Quellen überhaupt zu arbeiten und sie daten- und quellenkritisch zu reflektieren. Ebenso wie die „Auswahlkriterien“ für die Digitalisierung von Zeitungen und der Prozess der Zeitungsdigitalisierung selbst (Metadaten schemata, verwendete Software, Vorlagen etc.) selten dargelegt bzw. dokumentiert werden, fehlt in den Portalen in der Regel eine quellenkritische Einordnung der digital präsentierten Zeitungstitel in die „historische Medienlandschaft“ (Blome, 2018, S. 1, 11) – das gilt genauso für die Digitalausgabe der *Berliner Volkszeitung*, zu der man leider weder bei *Europeana Newspapers* noch *ZEFYS* Näheres erfährt.

Ohne dieses Vorwissen bleibt digitale Forschung jedoch nur ein Fischen im trüben Datenmeer. Recherchiert man deshalb ganz konventionell *analog* nach solchem Kontextwissen, so wird zugleich deutlich, dass die mangelhafte fachwissenschaftliche Unterfütterung der Zeitungsportale mit einem gravierenden Defizit an pressehistorischer Grundlagenforschung korrespondiert. Hochgradig selektiv, punktuell, unkoordiniert sind sicherlich berechnete Kritikpunkte an der bisherigen deutschen Zeitungsdigitalisierungsstrategie, wenn man erwartet, damit die von der Entfesselung der Massenpresse hervorgebrachte

ungeheure Menge an Zeitungstiteln in den Griff zu bekommen und als kulturelles Erbe zu sichern – umgekehrt hat sich die historische Presseforschung hierzulande bislang selbst recht wenig um eine verlässliche und systematische Dokumentation der Presselandschaft des 19. und 20. Jahrhunderts gekümmert. So ist die Fortführung der von Holger Böning (vgl. Böning & Moepps, 1996) initiierten, wegweisenden biobibliographischen Handbücher zur Geschichte der deutschsprachigen periodischen Presse für die Zeit nach 1815 ein gewichtiges Desiderat. Damit fehlt der pressehistorischen Forschung nicht bloß ein zentrales Informations- und Nachweisinstrument, sondern ebenso ein valides Steuerungsinstrument für eine Zeitungsdigitalisierungsstrategie nach wissenschaftlichen Entscheidungskriterien, die sich langfristig am Ziel der „digitalen Verfügbarkeit eines umfangreichen, repräsentativen historischen Pressekorpus“ (Hagenah, 2010, S. 183) orientiert. Stattdessen gehen derzeit schon der Rekonstruktion einfachster Basisdaten und der oftmals wechselhaften Erscheinungsgeschichte (mit Angaben und Hinweisen zu Herausgebern und Redakteuren, zu Auflagen, Erscheinungsweise, Format, Umfang, Preis, vorliegender Forschung usw. usf.) für den Großteil der Zeitungen immer noch sehr mühselige Recherchen diversester und oft weit verstreuter Daten, Literatur und Quellen und nicht zu vergessen intensives Zeitungslernen voraus – dies trifft gleichermaßen für die *Berliner Volkszeitung* zu.

2.1 *Eine unter Tausenden: Die Berliner Volkszeitung in der historischen Medienlandschaft*

Als eine unter tausenden von Zeitungen des Deutschen Kaiserreichs und der Weimarer Republik, die die in der zweiten Hälfte des 19. Jahrhunderts einsetzende Entfesselung der Massenkommunikation hervorgebracht hat (vgl. Wilke, 2000, S. 155, 252–291), muss die *Berliner Volkszeitung* zuallererst in diese historische Medienlandschaft eingeordnet werden. Um einen ersten Eindruck von den Dimensionen der deutschen Zeitungslandschaft in diesem Zeitraum zu bekommen, für den die *Berliner Volkszeitung* digitalisiert vorliegt, kann man sich zunächst die Anzahl der Zeitungstitel zu einigen Stichjahren anschauen. Demnach erschienen 1891 bereits 3.005 Zeitungen, 1914 gab es 4.221 Zeitungen, 1928 waren es 3.356 und 1932 schließlich 3.723 Zeitungen (siehe Abb. 1;

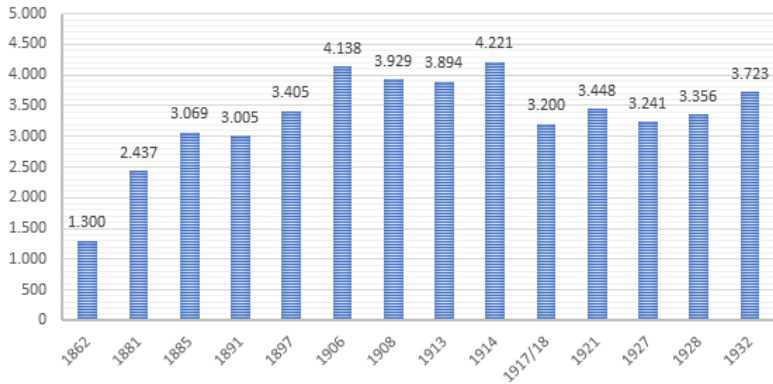
Wilke, 2000, S. 260, 345). Der Zeitungskundler Robert Brunhuber gibt aus zeitgenössischer Sicht mit dem Faktor Reichweite noch einen wichtigen Hinweis zur Einschätzung dieser Zahlen:

Wie die Ziffer auch genau lauten mag – von Belang ist ihre exakte Höhe ja nicht –, sicher ist, daß ungefähr 4.000 Zeitungen gegenwärtig in Deutschland erscheinen, von denen man allerdings gut 90 Prozent als Zwerg- und Mittelblätter bezeichnen kann, während von den verbleibenden 400 größeren Zeitungsunternehmen etwa 300 lokaler Natur sind und nur etwa 100 Zeitungen Anspruch auf allgemeine politische, wirtschaftliche oder kulturelle Bedeutung machen können. (Brunhuber, 1908, S. 10)

Als engerer medienhistorischer Kontext ist deshalb der Berliner Zeitungsmarkt relevant, der nicht allein die unmittelbare lokale Konkurrenz für die *Berliner Volkszeitung* darstellte. Zur Kaiserzeit war Berlin vielmehr das „Epizentrum der größten Zeitungsvielfalt Deutschlands“ (Schilling, 2011, S. 33). Um 1900 versammelte Berlin einen Großteil jener seinerzeitigen „Avantgarde der journalistischen Medien“ (Birkner, 2012, S. 267), die laut Brunhuber (1908, S. 10) überhaupt „Anspruch auf allgemeine politische, wirtschaftliche oder kulturelle Bedeutung“ erheben konnten: „Die journalistischen Institutionen, die hier investierten und am Puls der Zeit agierten, waren am besten auf die nun folgenden Entwicklungen vorbereitet.“ (Birkner, 2012, S. 216) Zeitungsverlage wie *Mosse*, der 1904 die *Berliner Volkszeitung* übernahm, *Scherl* oder *Ullstein* waren die Innovatoren moderner Tageszeitungsproduktion und entwickelten die „Blaupause der modernen Presse“ (Stöber, 2008, S. 410): formal und inhaltlich in Aufmachung und Gestaltung, technisch, organisatorisch, wirtschaftlich, im Tempo und nicht zuletzt im Ressourceneinsatz, um die „einmal entfesselte Nachrichtenflut“ (Wilke, 2012, S. 70) zu bändigen und zu bewältigen. Zugleich bedeutete die „außerordentliche mediale Dichte“ des Berliner Zeitungsmarkts eine „politische und wirtschaftliche Konkurrenzlage“, die sich „auf dem relativ engen städtischen Raum als außerordentlich scharf erwies“ (Sösemann, 1992, S. 146, 147). Für das Jahr 1928 gibt eine von dem Zeitungskundler Günter Ost (vgl. Ost, 1929) unternommene statistische Untersuchung interessante Einblicke in diese besondere Konkurrenzsituation. Ost verzeichnete insgesamt 86 parallel in Berlin erscheinende Zeitungen. Pro Tag wurden dort nach Ost rund 3,75 Millionen Zeitungsexemplare gedruckt: „eine erstaunlich hohe Zahl, an jedem Tag für jeden Leserkundigen eine Zeitungsnummer; auch dann noch, wenn man bedenkt,

daß ein großer Teil der Auflage für die Provinz, für das Reich bestimmt ist“ (Ost, 1929, Sp. 532).

Abbildung 1: Entwicklung der Zeitungslandschaft in Deutschland 1862–1932



Quelle: Wilke, 2000, S. 260, 345

Diese Zeitungen mit überregionaler Reichweite waren wiederum vor allem politische Zeitungen. Bereits 1896 erschienen in Berlin 36 politische Zeitungen „aller möglichen Schattierungen“, die die „Vormachtstellung“ der Pressemetropole auch auf dem deutschen politischen Meinungsmarkt behaupteten (Wilke, 2000, S. 269). Trotz der in der Weimarer Republik grundlegend veränderten politischen Kultur ermittelte Ost noch für 1928 37 politische Zeitungen. Von den 24 vor 1918 gegründeten Zeitungen waren dabei nur zehn Zeitungen sogar schon vor 1874 gegründet worden (Ost, 1929, Sp. 531) – und selbst wenn Ost leider keine konkreten Zeitungstitel nennt, gehörte die *Berliner Volkszeitung* mit Sicherheit zu diesem kleinen Kreis besonders einfluss- und traditionsreicher politischer Zeitungen.

2.2 Ein „Organ für Jedermann aus dem Volke“: Kurze Charakteristik und Geschichte der Berliner Volkszeitung

In der Tat war die 1853 von Aaron Bernstein und Franz Duncker als Nachfolgerin der kurz zuvor verbotenen *Urwähler-Zeitung* ins Leben gerufene *Berliner Volkszeitung* jahrzehntelang auf dem politischen Meinungsmarkt eine feste Größe und speichert so in ihren Seiten und Spalten „ein bemerkenswertes Kapitel“ (Mendelssohn, 1959, S. 116) deutscher Zeit- und Zeitungsgeschichte. Entsprechend heißt es schon 1916 in einer Firmen-Festschrift für die Annoncen-Expedition Rudolf Mosses, der die Zeitung seit 1904 verlegte:

Die Geschichte der Berliner Volkszeitung darstellen, hieße die Historie Preußen-Deutschlands, die innenpolitische und die auswärtige, vom Jahre 1848 bis auf unsere Tage erörtern. Hieße zeigen, wie Zeit und Zeitung in wechselndem Wandel politischer und sozialer Entwicklung sich bedingen und beeinflussen, hieße erweisen, welch außerordentliches politisches Erziehungs- und Machtmittel eine Zeitung in der Hand politisch charaktervoller und mit gefesteter Weltanschauung ausgerüsteter Verleger ist, denen die Herausgabe einer Zeitung nicht nur wirtschaftliche Betätigung, sondern zugleich sittliche Mission ist. (Annoncen-Expedition Rudolf Mosse, 1916, S. 127)

Ihren guten Ruf verdankte das Blatt allerdings (auch wenn Lokalgeschehen und -politik immer einen großen Raum einnahmen) neben den Berliner Lesern und Leserinnen einer ebenso treuen wie zahlreichen Leserschaft in ganz Deutschland. Während sogar „viele Berliner gar nicht wussten, dass es sie gab“, so Peter de Mendelssohn (1959, S. 115), hatte die *Berliner Volkszeitung* „fern vom großen weltstädtischen Verkehr eine sehr ansehnliche Verbreitung, um die sie manche namhaftere Zeitung beneidet hätte.“ Seit den 1850er Jahren war sie als „bedeutendes linksliberales Blatt“ (Koszyk, 1966, S. 113) bekannt, das sich beständig durch seine freisinnig-demokratische Grundhaltung und „couragierte Sprache“ (Mendelssohn, 1959, S. 115) auszeichnete, wie sich Franz Mehring, zeitweise einer der leitenden Redakteure, erinnert:

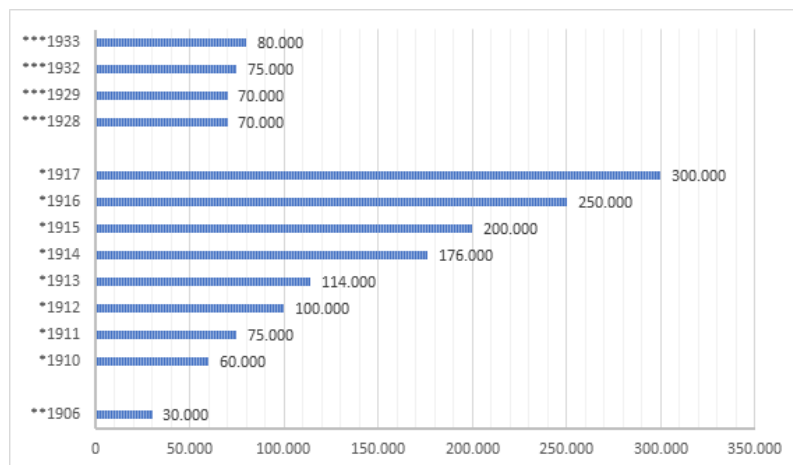
Entschieden demokratisch, aber keineswegs republikanisch, war sie am wenigsten Parteiorgan im Sinne des politischen Tageskampfes, sondern, wie es schon an der Spitze des Blattes hieß, ‚Organ für Jedermann aus dem Volke‘. Das kam auch schon an der äußerlichen Ausgestaltung zum Ausdruck. Der täglich an der Spitze des Blattes stehende ‚Leitartikel‘ war eine Neuerung, die bahnbrechend für die junge politi-

sche Tagesschriftstellerei in Deutschland wurde. Bis in sein hohes Alter hinein hat Bernstein mit äußerst seltenen, nur durch Krankheit veranlassten Unterbrechungen seinen Leitartikel tagaus, tagein geschrieben. (Mendelssohn, 1959, S. 116–117)

1885 kam die *Berliner Volkszeitung* in den Besitz des Verlegers Emil Cohn (Schwager von Rudolf Mosse), 1904 wurde schließlich Rudolf Mosse selbst Verleger der Zeitung. Journalistisch oder politisch standen für die *Berliner Volkszeitung* mit diesem Übergang ins Zeitungshaus Mosse keine großen Veränderungen auf der Tagesordnung – eher im Gegenteil. Sie wurde, „ohne ihren Charakter aufzugeben, zu einer Art kleinen Bruders des großmächtigen *Berliner Tageblatts*“, blieb, „trotz ihres unscheinbaren Auftretens, bei aller Volkstümlichkeit eine Zeitung von bemerkenswertem Niveau“ (Mendelssohn, 1959, S. 117) und profitierte zudem nachhaltig vom verlegerischen Geschick Mosses und der Kapitalkraft und den Ressourcen seines Unternehmens (vgl. Koszyk, 1966, S. 279–283; Scharf 1975). Nicht zuletzt die „technischen Vorzüge eines großen, glänzend durchorganisierten Verlages ermöglichten ihr den Aufstieg, den sie genommen hat“ (BVZ, Nr. 424, 9.9.1920). Wie das ursprüngliche Gründerduo der Zeitung Bernstein und Duncker fühlte sich auch Mosse den „Ideen von 1848 verpflichtet“ (Scharf, 1975, S. 204) und kümmerte sich seit 1906 sogar höchstpersönlich um die *Berliner Volkszeitung*. Mit der in der Folge eindrucksvoll steigenden Auflage, die sich binnen zehn Jahren nahezu „verzehnfacht“ hatte (siehe Abb. 2), machte er sich dabei einmal mehr einen Namen als erfolgreicher Zeitungsverleger und zeigte, „dass konkurrierende Blätter durch systematische Pflege des Marktes auf der Grundlage technischer und redaktioneller Verbesserungen zu gleichmäßigem Erfolg entwickelt werden können“ (Koszyk, 1966, S. 282).

Gerade weil *Berliner Tageblatt* und *Berliner Volkszeitung* für eine liberale Politik standen, spielte die Art und Weise der Aufbereitung und Darstellung der journalistischen Inhalte im Zusammenspiel mit dem Gesicht der Zeitung eine entscheidende Rolle, um trotzdem unterschiedliche Leserkreise anzusprechen und beide auf dem Markt zu behaupten. Und während so das *Berliner Tageblatt* in Aufmachung und Schriftbild geradezu zeitlos auf der Titelseite mit viel Text, arrangiert und verteilt auf drei Spalten verlässlich und seriös „nüchterne Klarheit“ (Schilling, 2011, S. 216) ausstrahlte und an diesem „journalistischen Kompositionsprinzip“ (vgl. Vicari, 2014) über die Jahrzehnte weitgehend festhielt, probierte man bei der *Berliner Volkszeitung* immer wieder Neues aus, um „einfallsreich und vielfältig [...] der Monotonie des Alltags eine kurzweilige Darstellung vom Tage entgegensetzen“

Abbildung 2: Auflagenentwicklung der Berliner Volkszeitung 1906–1933



Quellen: *Berliner Volkszeitung, 1890–1930; **Scharf, 1975, S. 211; ***OEAW, 1996

(Schilling, 2011, S. 251). Mit klar aufgeräumten, leicht lesbaren, übersichtlichen Zeitungsseiten konnten sich die Leser der *Berliner Volkszeitung* „also allein rein äußerlich betrachtet auf eine sehr abwechslungsreiche, geradezu unkonventionelle Lektüre einstellen“ (Schilling, 2011, S. 236). Ganz praktisch hieß dies: Anders als das *Berliner Tageblatt*, das noch in den 1920er Jahren auf der „Titelseite so gut wie nie eine graphische Abbildung enthielt“ (Schilling, 2011, S. 217), war die *Berliner Volkszeitung* schon sehr früh eine Bildzeitung. Das erste Bildelement auf einer Titelseite war eine Übersichtskarte des Nord-Ostsee-Kanals in der *Berliner Volkszeitung* vom 12. Juni 1895 (BVZ, Nr. 270, 12.6.1895). Und ebenso früh nutzte man Schlagzeilen – die ersten finden sich im Zusammenhang der Erdbebenkatastrophe in Italien Ende Dezember 1908 (vgl. BVZ, Nr. 609, 30.12.1908; BVZ, Nr. 611, 31.12.1908; BVZ, Nr. 612, 31.12.1908).

Die Überschrift „Neue Form – aber: Wir bleiben die Alten!“ (BVZ, Nr. 559, 27.11.1930), unter der die Macher der *Berliner Volkszeitung* Ende November 1930 wieder einmal ein neues Gewand und Gesicht der Zeitung ankündigten, bringt es sehr gut auf den Punkt, dass die im Laufe der Zeitungsgeschichte erfolgten zahlreichen Relaunches in Layout und Schriftbild keineswegs wahllos passierten

oder einen Wandel in politischer Haltung, Gesinnung und Tendenz signalisierten – ganz im Gegenteil:

Im Laufe der Jahrzehnte hat die *Berliner Volkszeitung* verschiedentlich Aussehen und Ausdrucksform geändert, nie aber ist der Boden des Volksblattes geistig verlassen worden. [...] Zeitung ist immer Spiegel der Zeit, darum muss auch die *Berliner Volkszeitung* in Tagen, die, wie die unsrigen, voll von Wandlung und Entwicklung sind, [...], Ausdruck des Zeitgeistes sein. (BVZ, Nr. 559, 27.11.1930)

Entsprechend blieb die Zeitung kontinuierlich ihren „liberaldemokratischen“ (Mendelssohn, 1959, S. 307) Positionen treu und trat in der Weimarer Republik offen als eine jener Zeitungen hervor, „die für die Erhaltung und Stärkung der parlamentarischen Demokratie, der Republik, eintraten“ (Mendelssohn, 1959, S. 308), und zeigte sich sogar als „glühende Verfechterin republikanischer Ziele“ (Schilling, 2011, S. 248). In den Zeitungsspalten fand diese „demokratisch-republikanische“ (OEAW, 1996) Gesinnung u.a. Niederschlag in einer exklusiven Sparte für den republikanischen Schutzbund „Reichsbanner Schwarz-Rot-Gold“ (Schilling, 2011, S. 249; vgl. weiter: Elsbach, 2019), die den Verteidigern der Weimarer Republik regelmäßig einen prominenten Platz einräumte.

Wenig verwunderlich ist somit, dass sich unmittelbar nach der vollzogenen Machtergreifung im Frühjahr 1933 das NS-Regime des renommierten liberalen *Mosse-Verlags* und seiner Zeitungen „bemächtigt“ (Scharf, 1975, S. 213) hat. Wie andere liberale Zeitungen wurde auch die *Berliner Volkszeitung* jedoch nicht gleich eingestellt, sondern vorerst ‚nur‘ gleichgeschaltet und inhaltlich und redaktionell in Fesseln gelegt (vgl. Hale, 1965; Koszyk, 1972, S. 346–443). Ihr traditionsreicher Name, an dem sich in der verödenen NS-Zeitungslandschaft immer noch viele Leser festhielten und so der *Berliner Volkszeitung* weiter eine gewisse Reichweite sicherten, „stellte noch ein gewisses Kapital dar“, das man nicht „verschleudern“ wollte (Mendelssohn, 1959, S. 398). Nichtsdestotrotz ist über das letzte Jahrzehnt der Zeitung kaum etwas bekannt und erforscht. In der Wikipedia (2020) findet man noch den Hinweis, dass sie laut der Monatsberichte des Geschäftsführers des *Deutschen Verlags* Max Wießner an den NS-Zentralverlag *Franz Eher Nachfolger* 1937 zu Berlins meistgelesener Boulevard- und Straßenverkaufszeitung umfunktioniert wurde (vgl. Oels, 2014, 2015). Schließlich wurde sie 1944 mit der *Berliner Morgenpost* zusammengelegt, was dem endgültigen Aus für die *Berliner Volkszeitung* gleichkam.

3 Die Berliner Volkszeitung digital erforschen

Digitalisierte Zeitungen sind hochkomplexe digitale Objekte bzw. Quellen, die auf ein weites Spektrum digitaler Formatierung und Datenniveaus verweisen. Zeitungen werden im Prozess der Zeitungsdigitalisierung nicht einfach digitalisiert, sondern sie werden als digitale Objekte konstruiert und als solche wiederum für unterschiedlichste Zwecke bereitgestellt und modelliert. Entsprechend finden sich die Daten der *Berliner Volkszeitung* als Metadaten in Bibliothekskatalogen und bibliographischen Indizes wie der ZDB¹ oder als Images und Volltexte in den Zeitungsportalen *Europeana Newspapers* und *ZEFYS*. Im Rahmen unserer Forschungen haben wir direkt mit dem unaufbereiteten Datensatz der *Berliner Volkszeitung* gearbeitet, weil die genannten Portale eben doch vor allem als digitale Lesesäle funktionieren und weniger als virtuelle Forschungsumgebungen konzipiert sind (vgl. Birkner, Koenen & Schwarzenegger, 2018). Hierfür waren spezifische Praktiken für den Umgang mit der *Berliner Volkszeitung* als digitaler Ressource zu entwickeln, die wir im Weiteren entlang unseres Forschungsprozesses näher vorstellen möchten.

3.1 Digitales Kuratieren

Andreas Fickers hat darauf hingewiesen, dass „der Prozess der Digitalisierung die klassische ‚Kontrollzone‘ der Archive verändert und ein neues Handlungsfeld erzeugt, in dem Datenproduzenten, Datenhüter und Datennutzer auf je eigene Weise in die Konstruktion des epistemischen Objekts ‚Quelle‘ eingebunden sind“ (Fickers, 2016; vgl. weiter: Lagoze, 2014). Die Ausformung von solchen neuen Kontrollzonen für die epistemische Aneignung von und den praktischen Umgang mit digitalen Objekten als Quellen lässt sich bspw. auf dem deutschen Kultur-Hackathon *Coding da Vinci* beobachten, zu dem mittlerweile mehrmals im Jahr Datengeber*innen, Designer*innen, Entwickler*innen und Fachwissenschaftler*innen zusammenkommen, um „gemeinsam aus offenen Daten mit eigenen Ideen neue digitale Anwendungen wie Apps, Dienste und

1 Vgl. für die *Berliner Volkszeitung* die Angaben unter <https://zdb-katalog.de/title.xhtml?idn=014054299>, 31. Januar 2021.

Visualisierungen zu skizzieren und umzusetzen“ (Fischer, 2019, S. 419). Sehr deutlich offenbart sich in der Praxis eines solchen Hackathons, dass es bei der epistemischen Konstruktion von Daten, digitalen Objekten und Ressourcen als Quellen „neben dem Coden, dem Programmieren“ immer um viele weitere fachwissenschaftlich-informativ-hybride „Praktiken wie das Evaluieren des Daten-Outputs oder dessen statistische Weiterverarbeitung, [...] das Kopieren und Parsen von Dateien, um die Konfiguration der Datenhaltungsinfrastruktur und dergleichen mehr“ geht (Trilcke & Fischer, 2018, S. 4).

Unsere digitalen Forschungen mit der Digitalausgabe der *Berliner Volkszeitung*, die auf dem *Coding da Vinci Berlin 2017* angelaufen sind, in dessen Rahmen diese Ressource bereitgestellt wurde, haben konkret gezeigt, dass solche Projekte nicht in einer klaren Arbeitsteilung zwischen Fachwissenschaftler*innen, die allein analysieren und interpretieren, und Informatiker*innen, die nur coden und programmieren, aufgehen. Das beginnt schon beim Kennenlernen und Verstehen der digitalen Objekte. Trotzdem die *Berliner Volkszeitung* eine der wenigen Zeitungen des ZEFYS-Portals ist, deren komplette Rohdaten frei für Forschungszwecke zur Verfügung stehen, stößt man auch hier wieder auf das Problem vieler Zeitungsportale, dass man als Anwender*in und Nutzer*in über diese Daten, ihre Generierung, ihre Formatierung, ihre Qualität und Tiefe, so gut wie nichts erfährt. Häufig entspringen ja „Digitalisierungsaktivitäten [...] nicht einem spezifischen Forschungsinteresse, sondern [...] der jeweiligen digitalisierenden Institution“ (Schöch, 2017, S. 227) – und so haben die damit verbundenen komplexen „data assemblages“ (vgl. Kitchin, 2014, S. 24–26) nicht nur einen wesentlichen Einfluss auf die Datengenese, sondern sind ebenso ein wichtiger Hintergrund für den weiteren Umgang mit den so produzierten Daten. Wie bei vielen Digitalisierungsprojekten, so war aber leider auch im Falle der *Berliner Volkszeitung* mit der Datenaufbereitung, -bereitstellung und -haltung für die Datenproduzent*innen die Arbeit getan und damit kam uns als Datennutzern gleichsam gezwungenermaßen als erste Aufgabe zu, mit viel Unterstützung vom verantwortlichen Datengeber und -hüter Clemens Neudecker (vgl. Neudecker, 2017) erst einmal Kenntnisse über die Datenkonstruktion und somit nach und nach Kontrolle über diese Daten zu gewinnen.² Entscheidend war zunächst die Anlage, Formatierung, Organisation und

2 Vgl. die knappe Datenbeschreibung sowie den Zugang zum Zeitungsdatenset unter <https://codingdavinci.de/daten/digitalisierte-historische-zeitungen>, 31. Januar 2021.

Struktur der Zeitungsdaten zu verstehen, die in der Auszeichnungssprache XML nach dem kombinierten Metadaten-Schema ALTO (Analyzed Layout and Text Object) und METS (Metadata Encoding and Transmission Standard) jede einzelne

Abbildung 3: Die Seite Eins der Berliner Volkszeitung vom 16. Januar 1919 im Europeana Newspapers Interface und modelliert als digitales Objekt



Zeitungssseite in einer Datei als digitales Objekte beschreiben und modellieren (siehe Abb. 3). Dies war die Voraussetzung, um mittels eines eigens für dieses Projekt von Falko Krause entwickelten XML-Parsers die Metadaten aus weit über hundertvierzigtausend Dateien abzufragen, auszulesen, anzureichern und in einem eigenen Metadaten set zu erfassen.³ Idee und Funktionsweise des Parsers orientieren sich an dem Verfahren, das Jean-Philippe Moreux entworfen hat, um für die französischen Zeitungen in *Europeana Newspapers* spezielle Metadaten sets zu generieren (vgl. Moreux, 2015).

Praktiken der Aufbereitung und Anreicherung digitaler Objekte werden im Forschungsdatenmanagement unter dem Konzept des „Digitalen Kuratierens (*digital curation*)“ zusammengefasst (vgl. Mauer & Recker, 2019, S. 117–118). Kuratierung wird dabei „als Oberbegriff für alle Aktivitäten im Lebenszyklus von digitalen Objekten verstanden, die nicht nur darauf abzielen, diese zu erhalten und nutzbar zu machen, sondern insbesondere auch, die digitalen Objekte mit einem Mehrwert (*added value*) zu versehen“ (Mauer & Recker, 2019, S. 118). Der entscheidende Unterschied zu einem allein datentechnischen Preprocessing liegt also darin, dass es beim digitalen Kuratieren nicht bloß um die „Arbeit an den Daten selbst“ (Auslesen, Bereinigen, Normalisieren) geht, sondern ebenso und im Wechselspiel mit Praktiken des Preprocessings um die „inhaltliche Einordnung“ bzw. vielmehr inhaltliche Validierung der Daten (Mauer & Recker, 2019, S. 118). Digitales Kuratieren wird somit zu einem wesentlichen Element digitaler Quellenkritik: Es hilft erst einmal, „Ungenauigkeiten und Unschärfen“ (Jannidis, 2017, S. 107) in den verdateten Quellen zu erkennen und ggf. zu bereinigen; im Weiteren ist es wichtig, um den Gehalt, die Qualität und Tiefe der Daten zu beurteilen und auf dieser Grundlage dann ganz konkret die Grenzen, Potentiale und Praxis digitaler Forschung *mit* diesen Datenquellen zu diskutieren.

Mit der Datenanalyse und -visualisierungs-Software Tableau 2020.4, in die wir unser Metadaten set der Berliner Volkszeitung eingespeist haben, konnten wir unser mehrdimensionales Datenkorpus erkunden und prüfen. Im Folgenden wollen wir das quellenkritische Potential digitalen Kuratierens anhand der praktischen Arbeit mit unseren Daten beispielhaft illustrieren:

3 Die Scripte finden sich unter <https://github.com/shoutrlabs/berliner-schlagzeilen/tree/master/metadataStatistics>, 31. Januar 2021.

- *Konsistenz des Datenkorpus.* Visualisiert man die Anzahl der Datensätze (= Ausgaben) im Zeitverlauf, so lassen sich sehr rasch Auffälligkeiten, Fehler oder Lücken im Datenkorpus identifizieren. Wenn man weiß, dass es ab 1894 von der *Berliner Volkszeitung* in der Regel Montags und Sonntags eine Ausgabe und an allen anderen Wochentagen zwei Ausgaben pro Tag gab, dann erfordern die Ausreißer mit drei Ausgaben pro Tag zumindest eine Nachprüfung, ob es sich dabei um Fehler oder ggf. um Sondernummern handelt (siehe Abb. 4). In unserem Falle hat die manuelle Überprüfung aller betreffenden Datensätze ergeben, dass es sich bei den Ausreißern durchweg um technisch produzierte Fehler in den Metadaten handelt, die wir dann bereinigt haben. Weiter lassen sich durch solche Visualisierungen der Anzahl der Datensätze im Zeitverlauf leicht Lücken im Korpus entdecken. In der Gesamtübersicht findet man fehlende Jahrgänge und Monate und mit einer Kalenderansicht, lassen sich sogar ausgaben- und taggenau Lücken identifizieren (siehe Abb. 5).

Abbildung 4: Datenrohkorpus der Digitalausgabe der Berliner Volkszeitung: Ausgaben pro Tag

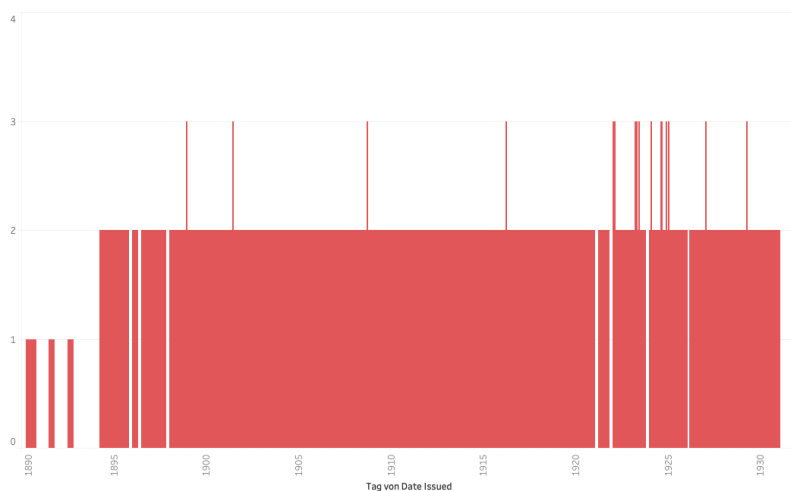
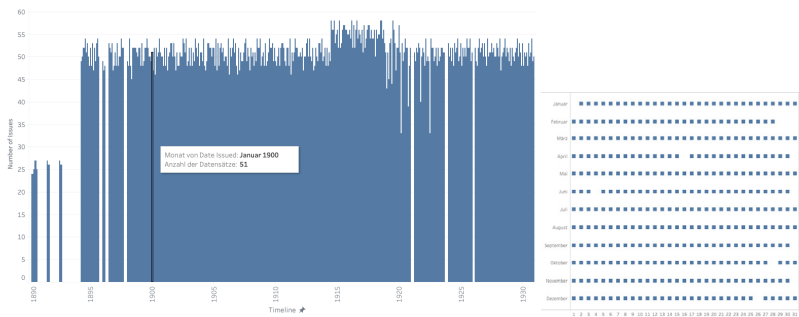
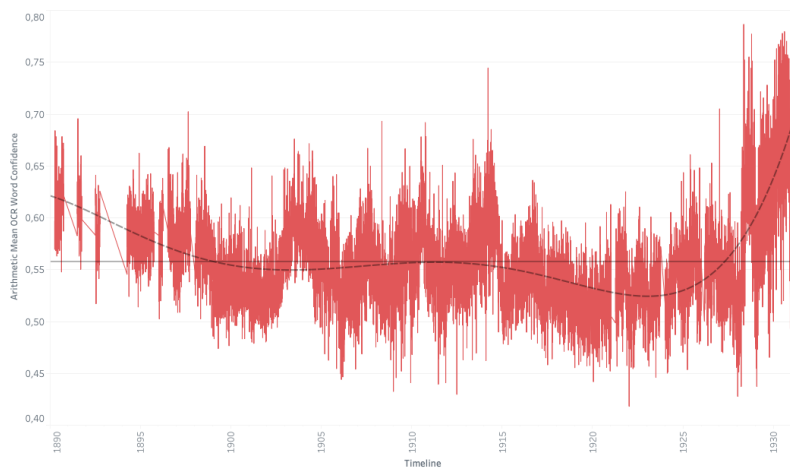


Abbildung 5: Kuratierter Datenkorpus der Digitalausgabe der Berliner Volkszeitung: Ausgaben pro Monat – Ausgaben pro Monat und Tag für 1900



- *Qualität der Texterkennung.* Was die Qualität der Texterkennung betrifft, so haben digitalisierte Zeitungen gemeinhin einen eher schlechten Ruf. Die Gründe dafür sind vielfältig und reichen von der Beschaffenheit und technischen Einrichtung der Digitalisierungsvorlagen, der verwendeten Software und ihrer objektspezifischen Konfiguration bis zur „komplexen Materialität“ (Müller & Hermes, 2014, S. 274), die Zeitungen eigen ist (vgl. Holley, 2009; Strange, Wodak & Wood, 2011; weiter speziell für deutschsprachige Zeitungen u.a.: Mühlberger, 2011; Kugler, 2018). Orientiert man sich an den Metadaten für die Word Confidence (WC), mit der die für die Texterkennung eingesetzte Software sich selbst überprüft hat, so muss unser Volltext entsprechend der DFG-Praxisregeln, nach denen „erst Texte ab einer Genauigkeit von 99,5 Prozent als wissenschaftlich zuverlässig gelten können“ und alles darunter eher „fragwürdig“ ist (DFG, 2013, S. 13), als sehr unzuverlässig eingestuft werden. Mit vielen Schwankungen liegt das Worterkennungsvertrauen, das auf einer Skala von 0 (sehr schlecht) bis 1 (sehr gut) gemessen wird, für die Seite Eins im Mittel bei lediglich 0,5573 (siehe Abb. 6).

Abbildung 6: OCR-Qualität: Durchschnittliche Word Confidence pro Tag für Seite Eins



3.2 Metadaten

Metadaten helfen nicht bloß beim digitalen Kuratieren eines Zeitungskorpus, sondern sie eröffnen der historischen Presseforschung sogar eine ganz neue Forschungsdimension. Die meisten pressehistorischen Forscher*innen sehen die Vorteile digitalisierter Zeitungen und Zeitungsportale vor allem in der Möglichkeit, große Mengen an Zeitungen und Zeitungstext „broadly, deeply and quickly“ (Blome 2018, S. 12) zu durchsuchen. Dagegen werden die im Digitalisierungsprozess anfallenden, über den eigentlichen Inhalt hinausgehenden Metadaten bzw. „information in the form of numbers“ (Liddle, 2012, S. 230) weitgehend ignoriert und sind somit hinsichtlich ihres Erkenntnispotentials noch recht wenig ergründet. Übersehen wird dabei insbesondere, „that the advent of electronic newspaper databases [...] has created and continues to create volumes of metadata that must necessarily reflect important qualities of the textual data“ (Liddle, 2012, S. 236–237). Metadaten modellieren das äußerst sperrige und zudem massenhaft vorliegende Materialobjekt Zeitung als genuines „physical, typographical, and information artifact“ (Liddle, 2012, S. 231) und versetzen uns so in die Lage, die charakteristische „semantisch-materiell-visuelle Mehrdimensionalität“ (Podewski,

2018, S. 9) von Zeitungen zu erforschen. Das können dann u.a. „Aufbereitungsrou-
tinen für die wechselnden Inhalte“ (Bucher, 2016, S. 285) in Aufmachung und Rub-
rizierung oder verschiedene Elemente periodischer und visueller „mediumspezifi-
scher Ordnung“ (Podewski, 2018, S. 6) wie Erscheinungsrhythmus, Format, Layout
und Umfang sein – also Aspekte, die Kevin G. Barnhurst und John Nerone in der
Frage nach der Geschichte der „Form“ von Zeitungen zusammengefasst haben:

Form includes the things that are traditionally labelled layout and design and typo-
graphy; but it also includes habits of illustration, genres of reportage, and schemes
of departmentalization. *Form* is everything a newspaper does to present the look of
the News. [...] Put another way, the form includes the way the medium imagines its-
elf to be and to act. In its physical arrangement, structure, and format, a newspaper
reiterates an ideal for itself. (Barnhurst & Nerone, 2001, S. 3)

Mit Metadaten lässt sich demnach endlich auf einer ebenso breiten wie „ge-
nauen *rechnerischen* Grundlage“ (Weber, 2001, S. 322) das einlösen, was schon Max
Weber in seinem Programm zu einer Soziologie des Zeitungswesens angeregt hat:

zu messen, mit der Schere und mit dem Zirkel, wie sich denn der Inhalt der
Zeitungen in quantitativer Hinsicht verschoben hat [...], [...] im Inseratenteil, im
Feuilleton, zwischen Feuilleton und Leitartikel, zwischen Leitartikel und Nachricht,
zwischen dem, was überhaupt an Nachrichten gebracht wird und was heute nicht
mehr gebracht wird (Weber, 1988, S. 441).

So kann bereits ein recht einfaches, bislang aber sehr schwer im großen
Maßstab ermittelbares Metadatum wie die Wortanzahl der Artikel, das von ei-
nigen Zeitungsportalen automatisiert generiert wird, Einblicke in die formale
Entwicklung journalistischer Genres geben. Dallas Liddle ist in einem kleinen
Experiment mit solchen Metadaten der Entwicklung des Aufmachers in der
Times nachgegangen und konnte sichtbar machen, dass sich ab den 1850er
Jahren die Form des Leitartikels (gemessen am Umfang) mit durchschnittlich
1.500 Wörtern stabilisierte (vgl. Liddle, 2012, S. 234–236). Ferner sind in diesem
Zusammenhang die Experimente von Jean-Philippe Moreux mit dem franzö-
sischen Zeitungskorpus in *Europeana Newspapers* interessant, die weitere Mög-
lichkeiten presserhistorischer Forschung mit Metadaten demonstrieren (vgl.
Moreux, 2016). Ein großer Vorzug der von Moreux verwendeten Metadaten ist
es, dass sie auch den Aufbau, die Aufmachung und das Layout der einzelnen

Zeitungsausgaben und -seiten betreffen. Dank Artikelsegmentierung und automatisierter Klassifikation von Inhalten verfügt Moreux über ein komplexes Set an Metadaten, um dem vielschichtigen Formwandel von Zeitungen nachzuspüren. So gibt die langfristig-vergleichende Entwicklung der durchschnittlichen Anzahl der Artikel pro Ausgabe (siehe Abb. 7) nicht bloß Hinweise auf die Informationsdichte und den unterschiedlichen Stellenwert textlicher Informationen in den Aufbereitungsroutinen der einzelnen Zeitungen, sondern stößt einen ebenso auf das Wechselspiel von Informationsgehalt und Zeitgeschehen. Deutlich erkennbar ist bspw. nach 1914 „the impact of the Great War on the activity of the press“ (vgl. Moreux, 2016, S. 5), woraus sich wiederum Anschlussfragen nach den Ursachen ergeben: Was für eine Rolle spielte Papierknappheit für die Reduktion von Inhalten und Umfang und welche die Zensur?

Wenn man noch tiefer in diese Metadaten geht, lässt sich sogar der innere Formwandel einzelner Zeitungen anhand automatisiert klassifizierter Inhaltssegmente im langfristigen Trend verfolgen (siehe Abb. 8): Wann und in welchem Umfang haben sich neue Formen herausgebildet und sind Formelemente wie Anzeigen, Illustrationen oder Tabellen neben dem reinen Text genuiner inhaltlicher Teil von Zeitungen geworden?

Abbildung 7: Entwicklung der durchschnittlichen Anzahl der Artikel pro Ausgabe für die französischen Zeitungen in Europeana Newspapers (Quelle: Moreux, 2016, S. 3)

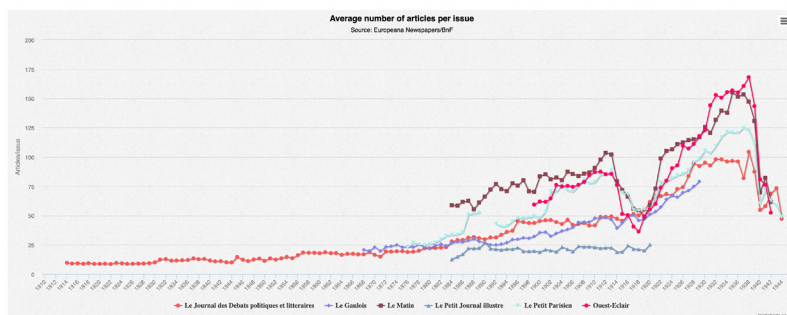
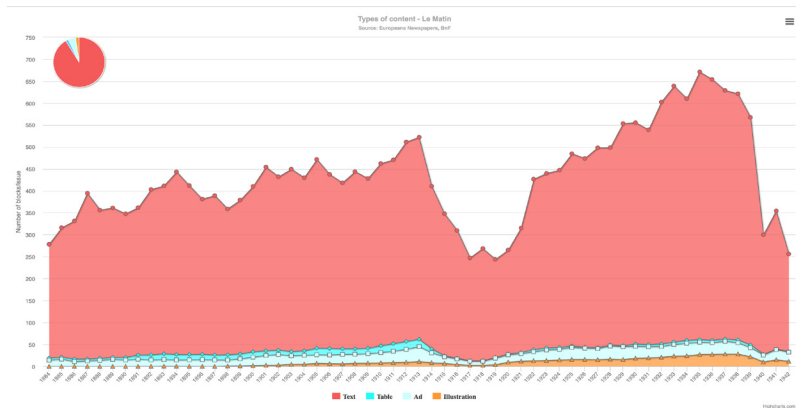


Abbildung 8: Entwicklung unterschiedlicher Inhaltssegmente (Illustrationen, Tabellen, Text, Werbung) im *Le Matin* (Quelle: Moreux, 2016, S. 5)



Trotz der innovativen Forschungsperspektiven, die Metadaten prinzipiell für die Form- und Publikationsgeschichte von Zeitungen ermöglichen, sind solche Forschungen momentan noch sehr anspruchsvoll und schwierig zu operationalisieren. Neben den mangelnden Erfahrungen und Vorarbeiten der pressehistorischen Forschung in dieser Richtung (vgl. Wilke 2000, S. 272–273; Wilke 2008) bilden die Metadaten selbst dabei das größte forschungspraktische Hindernis. Weder für die Aufnahme und Formatierung noch für den Umfang dieser Metadaten gibt es bislang einheitliche, portalübergreifende Standards. So halten die Autorinnen des *Atlas of Digitised Newspapers and Metadata* M.H. Beals und Emily Bell fest:

Each database instantiation that we have surveyed uses different language to describe physical objects (for example a newspaper issue, edition or volume), layout terminology (article or advertisement, title or heading), and more abstract concepts (genre, document type). Moreover, each database organizes these terms into different hierarchies of classification. Layers of nested items, containers, and technical metadata unique to different standards (and often unique to specific repositories) raise challenging questions about what data matters and what data can be dismissed as too technical to be of interest to the digital humanities researcher. (Beals & Bell, 2020, S. 2)

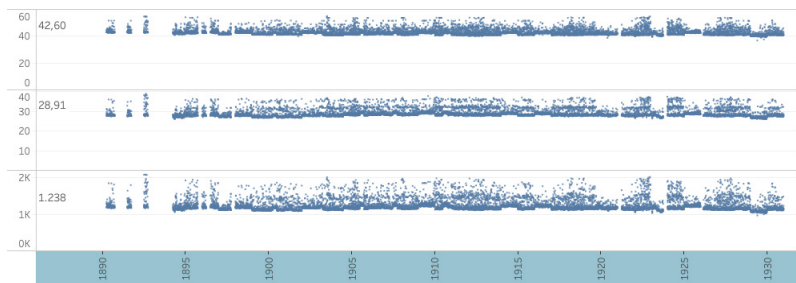
Noch sind also viel Experimentiergeist und Kreativität gefragt und „researchers must solve several technical and philosophical challenges“ (Beals & Bell, 2020, S. 2), um solche Metadaten auszulesen und weiter für speziell pressehistorische Fragestellungen aufzubereiten und zu verarbeiten.

Ein Beispiel aus unseren Analysen des Metadatensets der *Berliner Volkszeitung* soll die Praxis, Probleme und Potentiale des Metadata Reading für die pressehistorische Forschung veranschaulichen:

- *Format*. Eine Formatgeschichte von Zeitungen gibt es bislang nicht. Erst in der zweiten Hälfte des 19. Jahrhundert hat sich die Bandbreite an Zeitungsformaten herausgebildet, wie wir sie heute kennen (vgl. Groth 1928–1930, Bd. 1, S. 327–331). Ausschlaggebend waren hier u.a. technische Fortschritte in Druck sowie Papierherstellung und damit verbunden der Rückgang der Preise für Zeitungspapier, was gemeinsam mit wachsender Stoffmenge vor allem einen Trend zur Vergrößerung des Formats auslöste. Auf einem Zeitungsmarkt mit großer Konkurrenz war das Format aber auch ein wichtiges Moment der Lesersprache und -bindung, weil es Individualität und Unterscheidbarkeit von Zeitungen gewissermaßen materialisierte. Daten zum Format von Zeitungen gibt es nun jedoch nur in Ausnahmefällen und eine retrospektive Rekonstruktion ist sehr schwierig. Das beginnt schon beim Eintragen und Verzeichnen von Zeitungen in Bibliothekskatalogen und bibliographischen Nachweisinstrumenten, wo das Format in der Regel nicht erfasst wurde und wird. Selbst bei den unterschiedlichen bibliothekarischen Konservierungsstrategien von Zeitungen spielt das Format keine Rolle: Beim Binden werden Zeitungen beschnitten und ein weiteres Mal geht das Format bei der Mikroverfilmung verloren. Zeitungsbände und -mikrofilme sind beliebte Ausgangsformate für die Retrodigitalisierung von Zeitungen, in deren Prozess das Zeitungsformat wiederum nun als Metadatum des digitalen Objekts nicht registriert wird. So gibt es auch von der *Berliner Volkszeitung* (trotz „layoutgetreuer Digitalisierung“, so der Katalogeintrag im ZEFYS), keine entsprechenden Informationen. Abhilfe bot im ersten Schritt nur der antiquarische Kauf eines Exemplars der Morgenausgabe der *Berliner Volkszeitung* vom 6. Januar 1917 und das nun quasi empirische Wissen, dass dieses Exemplar im Format (Breite/Höhe) 32 x 47 cm erschienen ist, was dem damals noch nicht standardisierten Berliner

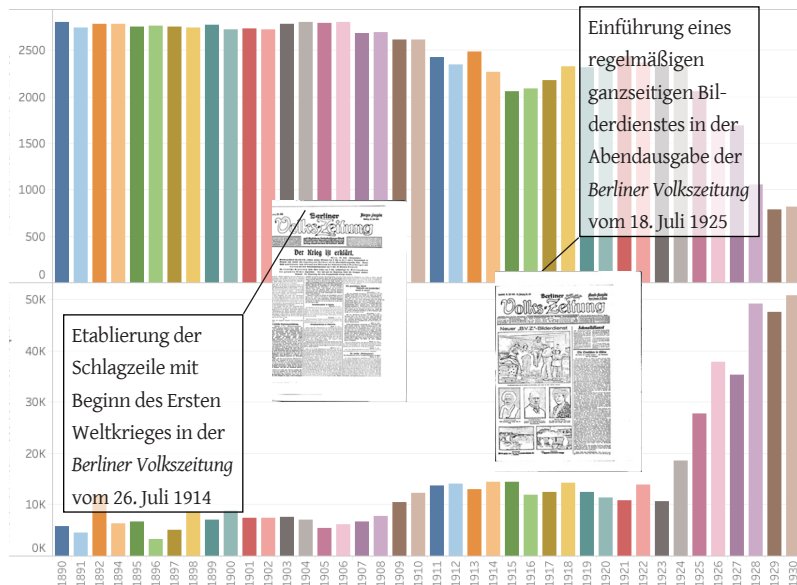
Format entsprechen dürfte. Weiter finden sich versteckt in den (undokumentierten) Metadaten verschiedene Maße über die Größe des digitalen Objekts in Pixel, mit denen sich dann wenigstens behelfsweise die Druckfläche (im Durchschnitt 1.238 cm²) und der Druckspiegel (28,91 x 42,60 cm) berechnen lassen (siehe Abb. 9) – Maße, die mit dem gekauften Exemplar übereinstimmen.

Abbildung 9: Entwicklung von Druckfläche und Druckspiegel (Breite, Höhe) für Seite Eins



- Was sagen uns nun diese Metadaten und sind sie mehr als nur eine Spielerei? Sieht man einmal von den vielen Streuungen ab, die auf eine Vielzahl druckextrinsischer (u.a. enge Bindung, Flecken oder Risse im Druckbild) und -intrinsischer (unsauberer Druck, verrutschter Druckspiegel usw.) Faktoren zurückgehen (vgl. Federbusch & Polzin 2013, S. 26–30; & Kugler, 2018, S. 36–40), die die maschinelle Erfassung der Größe permanent verzerren, so lässt die Visualisierung immerhin keine Sprünge erkennen – ein Indiz, dass Druckfläche und Format der *Berliner Volkszeitung* zumindest für den digitalisierten Zeitraum gleich geblieben sind. Somit hat man mit der Druckfläche zugleich eine Referenzgröße, um der Frage nachzugehen, wie dieser Raum für Bild und Text genutzt wurde, und kann so bspw. markante Wechsel in Aufmachung, innerer Form und Layout identifizieren (siehe Abb. 10).

Abbildung 10: Entwicklung der Anzahl der Wörter (oben) und der Fläche für Fotos, Grafiken und Illustrationen (unten) für Seite Eins



3.3 Text Mining

Wie Metadaten erschließen die Volltexte eines Zeitungskorpus der historischen Presseforschung über die „practical revolution“ (Nicholson, 2013, S. 61) einer einfacheren und effektiveren inhaltlichen Recherche von Zeitungen hinaus originär neue Forschungsperspektiven. Schon die mittlerweile in vielen Zeitungsportalen als Standard implementierte Durchsuchbarkeit digitalisierter Zeitungen nach Stichworten macht es möglich, „to trace the development and movement of ideas and discursive formations in ways that were once impossible“ (Nicholson, 2013, S. 63). Insbesondere durch diese Möglichkeit der Volltextdurchsuchbarkeit sind Zeitungen mehr und mehr für unterschiedlichste Fragestellungen und Forschungen von unmittelbarem Interesse geworden, die sich

dabei zunutze machen, dass in Zeitungen über lange Zeiträume die ganze Breite und Vielfalt gesellschaftlicher und soziokultureller Entwicklungen und Diskurse Ausgabe für Ausgabe, Jahrgang für Jahrgang abgelagert und überliefert werden. Durch die Volltexterschließung werden Zeitungen, so Bob Nicholson (2013, S. 70–71), quasi zu einem digitalen „living dictionary, one which allowed me to access the minutiae of everyday cultural discourse and track the complex ways in which words and phrases circulated, how their meanings evolved and how their usage changed on an almost daily basis.“

Komplexere Forschungen in dieser Richtung nutzen freilich digitale Methoden und Techniken des Text Mining (vgl. Ignatow & Mihalcea, 2018; Tonkin & Tourte, 2016), um in den Zeitungskorpora langfristige Trends sozialen Wandels und Strukturen öffentlicher Diskurse zu untersuchen (vgl. u.a. Dzogang et al., 2016; Lansdall-Welfare et al., 2017). Prinzipiell verarbeiten die verschiedenen Verfahren des Text Mining Texte nicht bloß maschinell, sondern stellen verschiedene Techniken der Textextraktion bereit und erzeugen unterschiedliche Textsichten. Grob gesagt geben sie (1.) korpusstatistische Zusammenfassungen; helfen (2.) spezifische Informationen zu extrahieren, etwa geographische Angaben, Konzepte oder vorkommende Personen; sind nützlich, um (3.) Verwendungsmuster von einzelnen Worten oder Kombinationsmuster mehrerer Worte zu erkennen oder (4.) Themen und (5.) Stimmungen in Texten zu identifizieren.

Um jedem die Möglichkeit zu geben, selbst einen ersten Eindruck von gut vierzig Jahrgängen Volltext der *Berliner Volkszeitung* zu gewinnen, hat Falko Krause ein Tool für eine browserbasierte „Automatic Digital Discourse“-Analyse entwickelt, das für jeden Monat eine Wordcloud basierend auf den fünfzig häufigsten Wörtern auf Seite Eins der *Berliner Volkszeitung* errechnet und visualisiert.⁴ Wordclouds als „Bag of Words“-Ansatz sind sicherlich die einfachste und simpelste Form der Aufbereitung und Visualisierung von Textdaten via Text Mining, wobei ihr größter Nachteil ist, dass die einzelnen Wörter aus dem Kontext von Satz und Text gerissen werden (Graham, Milligan & Weingart, 2016, S. 74–77):

However, historians can find utility in word clouds. If we are concerned with change over time, we can trace how words evolved in historical documents. While this is fraught with issues – words change meaning over time, different terms are

4 Online unter <https://bs.rockdapus.org/word-cloud.html>, 31. Januar 2021.

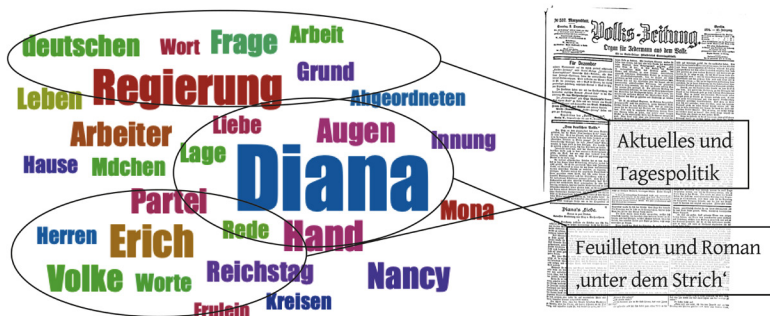
used to describe similar concepts [...] – we can arguably still learn something from this.“ (Graham, Milligan & Weingart, 2016, S. 76)

Wordclouds sind also vor allem eine einsteigerfreundliche Visualisierungsform, deren wichtigste heuristische Erkenntnisleistung darin liegt, ganz intuitiv und naiv das Entdecken von großen Textmengen zu ermöglichen und eingefahrene Perspektiven auf Texte herauszufordern und zu verschieben. Sie geben einen Gesamtüberblick, markieren durch den Fokus und die Reduktion auf häufig vorkommende Worte und die Visualisierung des Wortvorkommens Schlag- und Schlüsselwörter in Diskursen oder machen auf zeitgenössische öffentlichkeitswirksame Ereignisse und Geschehnisse aufmerksam. Neben der gleichzeitig explorativen wie komplexitätsreduzierenden Leistung wohnt diesen Visualisierungen ein starkes erkenntnisproduktives Irritationspotential inne, weil sie gewohnte Leseerwartungen enttäuschen bzw. stören, den Blick auf Texte verändern und mit neuen Textsichten überraschen. Sie sind so „Provokationen im eigentlichen Wortsinn: Sie rufen einen Eindruck hervor, der zum Interpretieren anregt“ (Schwandt, 2018, S. 124). Wir wollen diese Erkenntnis- und Interpretationspraxis im Folgenden mit Beispielen weiter verdeutlichen:

- *Liebe und Politik*. Unser erstes Beispiel illustriert recht anschaulich das Irritationspotential, das Wordclouds auszeichnet und verweist auf die Notwendigkeit, solche Visualisierungen stets im Kontext des konkreten Publikationszusammenhangs zu interpretieren, d.h. „Distant Reading“ und „Close Reading“ miteinander zu kombinieren und zu vermitteln. Zumindest waren wir selbst beim Durchsehen der Wortwolken sehr überrascht, als im Dezember 1894 auf einmal das Wort „Liebe“ auftauchte (siehe Abb. 11) – Grund genug für uns, den Erkenntnisfokus von Ferne auf Nähe zu wechseln und im ursprünglichen Veröffentlichungszusammenhang eine Erklärung zu suchen. Dabei wurde sehr deutlich, dass Computer und digitale Textmethoden Forschende eben keineswegs von der Interpretation entlasten. Vielmehr fordern sie permanent zum Interpretieren heraus, um den mithilfe von Computer und Software entdeckten Textmustern überhaupt einen Sinn zu verleihen. Entsprechend beschreiben Pim Huijnen und Melvin Wevers den Interpretationsprozess in digitalen Kontexten treffend als „a constant to and fro between distant reading and close reading – the interpretation of data visualizations and the zooming in on potentially interesting texts“ (Huijnen & Wevers, 2015). In der Praxis eines solchen

hin- und hergehenden „zooming in und zooming out“ (Schwandt, 2018), das sich auf unterschiedliche „mediale Formen und analytische Aufbereitungen“ (Weitin, Gilli & Kunkel, 2016, S. 115) bezieht und in unserem Beispiel zwischen dem Imagedigitalisat und der Wordcloud als Textextrakt vermittelt, entfaltet sich eine vermehrt experimentell-kreative Interpretations- und Textpraxis, die digital delegiertes Lesen und konventionelles hermeneutisch-verstehendes Lesen miteinander trianguliert.

Abbildung 11: Wortwolke Seite Eins Dezember 1894



- In dieser Weise erklärt erst die Interpretation im Kontext des ursprünglichen Veröffentlichungszusammenhang diese Wortwolke und offenbart, dass sich in dieser verschiedene Textdimensionen vermischen, womit vor dem Hintergrund der eigentümlichen „semantisch-materiell-visuellen Mehrdimensionalität“ (Podewski, 2018, S. 9) von Zeitungen immer zu rechnen ist. Seit den 1890er Jahren teilten sich so Aktuelles, Feuilleton, Kultur und Politik die erste Seite der *Berliner Volkszeitung*. Dazu gehörte in der Morgenausgabe lange Zeit regelmäßig ein Roman, der in diesem Falle ganz einfach „Dianas Liebe“ hieß. Erst mit der *Berliner Volkszeitung* vom 16. September 1907 (BVZ, Nr. 434, 16.9.1907) nutzte die Redaktion die Titelseite fortan exklusiv für tagesaktuellen Journalismus.

Abbildung 12: Wortwolke Seite Eins April 1912



Abbildung 13: Wortwolke Seite Eins August 1914



- *Medienereignisse.* Gerade durch ihre enorme komplexitätsreduzierende Leistung sind diese Wortwolken zudem intuitive Einstiegspunkte, um die vielschichtige Medienberichterstattung der *Berliner Volkszeitung* zu entdecken. Insbesondere nach der Reservierung des redaktionellen Raumes der Titelseite für informationsorientierten Journalismus und Tagespolitik, so zeigen unsere Beispiele zum Ausbruch des Ersten Weltkriegs und dem Untergang der Titanic (siehe Abb. 12 und 13), werden die Wortwolken zu einem robusten Indikator von Medienereignissen und deren Thematisierung.

4 Reflexion

Die Folgen der neuen digitalen Forschungs- und Quellensituation erschöpfen sich für die historische Presseforschung nicht in einem einfachen Transfer von analogen in nun digitale Kontexte, sondern machen vielmehr eine konsequente Reflexion der bisherigen Methodenpraxis erforderlich. Bislang eher unerhebliche informatorisch-technische Aspekte wie Datenformate, Recherche- und Suchinterfaces oder Programmierschnittstellen werden für die pressehistorische Forschung nicht allein relevant, weil sie für den Zugang zu Zeitungen als digitalen Ressourcen entscheidend sind. Informatorische Herausforderungen erstrecken sich über den gesamten Forschungsprozess und bilden zum Beispiel beim digitalen Kuratieren oder den Umgang mit Meta- und Textdaten genuin neue Epistemologien digitaler pressehistorischer Praktiken heraus, die konkrete Erkenntnisinteressen und die spezifischen Eigenschaften bzw. die Konstitution der jeweils zugrundeliegenden Daten miteinander vermitteln. Bemerkenswert ist, dass sich alle hier beschriebenen „Objektumgangsweisen“ (Schruhl, 2018, S. 4), die wir für den praktischen Umgang mit der *Berliner Volkszeitung* als digitaler Ressource entwickelt und genutzt haben, im Zuge eines sehr quellennahen Ausprobierens, Vorantastens und Testens herausgebildet haben. Ein gründliches Kennenlernen digitaler Ressourcen und ihrer zunächst experimentell-praktischen Handhabung als digitale Quellen ist unseres Erachtens die Voraussetzung, um digitale Forschungen produktiv auf das Niveau des „Distant Reading“ zu skalieren. Das liegt nun nicht nur unbedingt daran, dass zur Aneignung, Anwendung und dem Einsatz digitaler Methoden kaum Erfahrungswerte vorliegen. Als eine entscheidende Problemstelle digitaler Forschung haben sich in unserem Projekt die Grundlagen herausgestellt, auf denen solche Forschung fußt. Das betrifft die mehrfach kritisierte mangelhafte Dokumentation von Datengenesen und -konstruktion sowie des Prozesses der Zeitungsdigitalisierung und meint zudem das disparate und mangelhafte Kontext- und Vorwissen, das seitens der pressehistorischen Forschung selbst hierfür bereitsteht. Der Erfolg digitaler Methoden zum Beispiel in der Literaturwissenschaft als einem Kernfeld der Digital Humanities beruht nicht zuletzt auf einer traditionsreichen Interpretations-, Text- und Quellenpraxis, die als Vorwissen steuernd in den digitalen Erkenntnisprozess einfließt. „So irreführend diese Vorstellung ist“, betont der Literaturwissenschaftler Thomas Weitin

(2017, S. 2) in diesem Zusammenhang, „so deutlich zeigen gerade diejenigen Arbeiten, die bestimmten quantitativen Verfahren zur literaturwissenschaftlichen Korpusanalyse zum Durchbruch verholfen haben, dass die AutorInnen ihr Korpus jeweils sehr gut kannten.“ In diesem Sinne muss mit dem Einsatz digitaler Methoden für die pressehistorische Forschung eine engagierte methodologische Debatte über die Eigenart, Kritik und Nutzung von Zeitungen als Quellen überhaupt einhergehen, die deren spezifische „semantisch-materiell-visuelle Mehrdimensionalität“ (Podewski, 2018, S. 9) herausarbeitet – eine Debatte, die bis heute noch stark deskriptiv und fokussiert auf „singularisierende Quellenarbeit“ (Wilke, 1987, S. 49) und die Inhalte der Zeitung (vgl. zuletzt Opgenoorth, 1997, S. 106–109; Ackermann, 1992; Kraus 2002) geführt wird und somit wenig hilfreich für eine ebenso forschungspraktische wie profunde Quellenkritik ist, die sich gleichermaßen für Form, Inhalt, Materialität der Massenquelle Zeitung wie die hinter der Gestalt und den Inhalten einer Zeitung liegenden redaktionellen und verlegerischen Strukturen interessiert (vgl. Murphy, 2014). Unter dieser Maßgabe können digitale Methoden und Zeitungen als digitale Objekte und Ressourcen für eine innovative Methoden- und Quellenpraxis historischer Presseforschung ein enormes Potential entfalten, weil sie im Spannungsfeld zwischen Gesamtkorpus, unterschiedlichen Quellendimensionen und exemplarischen Text originelle und ebenso kombinierbare wie skalierbare Forschungsperspektiven für Zeitungen als komplexe Forschungsobjekte markieren und möglich machen.

Dr. Erik Koenen ist Universitätslektor (in Vertretung) am Zentrum für Medien-, Kommunikations- und Informationsforschung – ZeMKI der Universität Bremen, ekoenen@uni-bremen.de

Simon Sax ist Doktorand und Wissenschaftlicher Mitarbeiter am Zentrum für Medien-, Kommunikations- und Informationsforschung – ZeMKI der Universität Bremen, sax@uni-bremen.de

Falko Krause ist Informatiker und lebt in Berlin.

Quellenverzeichnis

- Ackermann, V. (1992). Presseartikel. In B.-A. Rusinek, V. Ackermann & J. Engelbrecht (Hrsg.), *Einführung in die Interpretation historischer Quellen. Schwerpunkt: Neuzeit* (S. 233–252). Paderborn: Ferdinand Schöningh.

- Annoncen-Expedition Rudolf Mosse (1916). *Fest-Schrift zur Feier des fünfzigjährigen Bestehens der Annoncen-Expedition Rudolf Mosse*. Berlin: Rudolf Mosse.
- Barnhurst, K.G., & Nerone, J. (2001). *The Form of News. A History*. New York, London: The Guilford Press.
- Beals, M.H., & Bell, E. (2020). *The Atlas of Digitised Newspapers and Metadata. Reports from Oceanic Exchanges*. Loughborough University. <https://doi.org/10.6084/m9.figshare.11560059>
- Birkner, T. (2012). *Das Selbstgespräch der Zeit. Die Geschichte des Journalismus in Deutschland 1605-1914*. Köln: Herbert von Halem.
- Birkner, T., Koenen, E. & Schwarzenegger C. (2018). A Century of Journalism History as Challenge: Digital Archives, Sources and Methods. *Digital Journalism*, 6(9), 1121–1135.
- Blome, A. (2018). Zeitungen. In L. Busse, W. Enderle, R. Hohls, T. Meyer, J. Prellwitz & A. Schuhmann (Hrsg.), *Clio Guide – Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften* (B.6, S. 1–36). Berlin: Clio-online.
- Böning, H., & Moepps, E. (1996). *Hamburg. Kommentierte Bibliographie der Zeitungen, Zeitschriften, Intelligenzblätter, Kalender und Almanache sowie biographische Hinweise zu Herausgebern, Verlegern und Druckern periodischer Schriften*. Stuttgart, Bad Cannstatt: Frommann-Holzboog.
- Brunhuber, R. (1908). *Das deutsche Zeitungswesen*. Leipzig: G.J. Göschen.
- Bucher, H.-J. (2016). Ein „Pictorial Turn“ im 19. Jahrhundert. Überlegungen zu einer multimodalen Mediengeschichte am Beispiel der Illustrierten Zeitungen. In S. Geise, T. Birkner, K. Arnold, M. Löblich & K. Lobinger (Hrsg.), *Historische Perspektiven auf den Iconic Turn. Die Entwicklung der öffentlichen visuellen Kommunikation* (S. 280–317). Köln: Herbert von Halem.
- DFG – Deutsche Forschungsgemeinschaft (2013). DFG-Praxisregeln „Digitalisierung“, https://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung_2013.pdf, 31. Januar 2021.
- Dzongang, F., Lansdall-Welfare, T., & Cristianini, N. (2016). Discovering periodic patterns in historical news. *PLoS One*, 11(11), <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0165736>, 31. Januar 2021.
- Elsbach, S. (2019). *Das Reichsbanner Schwarz-Rot-Gold. Republikerschutz und politische Gewalt in der Weimarer Republik*. Stuttgart: Franz Steiner.

- Europeana Collections – Newspapers (2020). *Berliner Volkszeitung / Volkszeitung – digital 1890–1930*, <https://www.europeana.eu/de/collections/topic/18-newspapers>, 31. Januar 2021.
- Federbusch, M., & Polzin, C. (2013). *Volltext via OCR. Möglichkeiten und Grenzen Testszenarien zu den Funeralschriften der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz*. Berlin: Staatsbibliothek zu Berlin – Preußischer Kulturbesitz
- Fickers, A. (2016). Digitale Metaquellen und doppelte Reflexivität, <https://www.hsozkult.de/debate/id/diskussionen-2954>, 31. Januar 2021.
- Fischer, B. (2019). Coding da Vinci oder für eine Renaissance der Allmendekultur mit digitalen Mitteln. In L. Pöllmann & C. Herrmann (Hrsg.), *Der digitale Kulturbetrieb. Strategien, Handlungsfelder und Best Practices des digitalen Kulturmanagements* (S. 415–430). Wiesbaden: Springer Gabler.
- Föhr, P. (2018). *Historische Quellenkritik im Digitalen Zeitalter*. Dissertation vorgelegt der Philosophisch-Historischen Fakultät der Universität Basel, https://edoc.unibas.ch/64111/1/Föhr_Pascal-Historische_Quellenkritik_im_Digitalen_Zeitalter-2018.pdf, 31. Januar 2021.
- Graham, S., Milligan, I., & Weingart, S. (2016). *Exploring big historical data. The historian's macroscope*. London: Imperial College Press.
- Groth, O. (1928–1930). *Die Zeitung. Ein System der Zeitungskunde (Journalistik)*. 4 Bde. Mannheim, Leipzig, Berlin: J. Bensheimer.
- Hagenah, U. (2010). Retrodigitalisierung von Zeitungen durch Regionalbibliotheken. Gedanken zu einer Momentaufnahme vom März 2010. *Zeitschrift für Bibliothekswesen und Bibliographie*, 57(3/4), 183–189.
- Hale, O.J. (1965). *Presse in der Zwangsjacke 1933–1945*. Düsseldorf: Droste Verlag.
- Holley, R. (2009). How good can it get? Analysing and improving OCR accuracy in large scale historic newspaper digitisation programs. *D-lib Magazin*, 15(3/4), <http://www.dlib.org/dlib/march09/holley/03holley.html>, 31. Januar 2021.
- Huijnen, P., & Wevers, M. (2015). Digital Deconstruction. The digital turn and the use of news media as sources for historical research. *Officina della Storia*, 13, <https://www.officinadellastoria.eu/it/2015/09/14/digital-deconstruction-the-digital-turn-and-the-use-of-news-media-as-sources-for-historical-research/>, 31. Januar 2021.
- Ignatow, G., & Mihalcea, Rada F. (2018). *An introduction to text mining. Research design, data collection, and analysis*. Thousand Oaks: Sage.

- Jannidis, F. (2017). Grundlagen der Datenmodellierung. In F. Jannidis, H. Kohle & M. Rehbein (Hrsg.), *Digital Humanities. Eine Einführung* (S. 99–108). Stuttgart: J.B. Metzler Verlag.
- Kitchin, R. (2014). *The data revolution. Big data, open data, data infrastructures and their consequences*. Los Angeles, London, New Delhi, Singapore, Washington: Sage.
- Koenen, E. (2018). Digitale Perspektiven in der Kommunikations- und Mediengeschichte. Erkenntnispotentiale und Forschungsszenarien für die historische Presseforschung. *Publizistik*, 63(4), 535–556.
- Koszyk, K. (1966). *Deutsche Presse im 19. Jahrhundert*. Berlin: Colloquium Verlag.
- Koszyk, K. (1972). *Deutsche Presse 1914–1945*. Berlin: Colloquium Verlag.
- Kraus, H.-C. (2002). Zeitungen, Zeitschriften, Flugblätter, Pamphlete. In M. Maurer (Hrsg.), *Aufriss der Historischen Wissenschaften. Bd. 4: Quellen* (S. 373–401). Stuttgart: Philipp Reclam jun.
- Kugler, A. (2018). Automatisierte Volltexterschließung von Retrodigitalisaten am Beispiel historischer Zeitungen. *Perspektive Bibliothek*, 7(1), 33–54, <https://doi.org/10.11588/pb.2018.1.48394>
- Lagoze, C. (2014). Big data, data integrity, and the fracturing of the control zone. *Big Data & Society*, 1(2), 1–11.
- Lansdall-Welfare, T., Sudhahar, S., Thompson, J., Lewis, J., & Cristianini, N. (2017). Analysis of 150 years of British periodicals. *Proceedings of the National Academy of Sciences*, 114(4), <https://www.pnas.org/content/114/4/E457>, 31. Januar 2021.
- Liddle, D. (2012). Reflections on 20,000 victorian newspapers. „distant reading“ the times using the times digital archive. *Journal of Victorian Culture*, 17(2), 230–237.
- Mauer, R., & Recker, J. (2019). Data Sharing. Von der Sicherung zur langfristigen Nutzung der Forschungsdaten. In U. Jensen, S. Netscher & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten. Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 115–133). Opladen, Berlin, Toronto: Verlag Barbara Budrich.
- Mendelssohn, P. de (1959). *Zeitungsstadt Berlin. Menschen und Mächte in der Geschichte der deutschen Presse*. Berlin: Ullstein Verlag.
- Moretti, F. (2000). Conjectures in world literatures. *New Left Review*, 1, 54–68.
- Moretti, F. (2009). *Kurven, Karten, Stammbäume. Abstrakte Modelle für die Literaturgeschichte*. Frankfurt/Main: Suhrkamp.
- Moretti, F. (2016). *Distant Reading*. Konstanz: University Press.

- Moreux, J.-P. (2015). *Data mining historical newspaper metadata. Old news teaches history*, http://altomotor.github.io/EN-data_mining/, 31. Januar 2021.
- Moreux, J.-P. (2016). *Innovative approaches of historical newspapers. Data Mining, data visualization, semantic enrichment. Paper, IFLA News Media Section, Lexington 2016*, http://www.euklides.fr/blog/altomotor/EN-DM/article_en2.pdf, 31. Januar 2021.
- Mühlberger, G. (2011). Digitalisierung historischer Zeitungen aus dem Blickwinkel der automatisierten Text- und Strukturerkennung (OCR). *Zeitschrift für Bibliothekswesen und Bibliographie*, 58(1), 10–18.
- Müller, M.E., & Hermes, M. (2014). Digitalisierung der vollständigen deutschsprachigen Zeitungen des 17. Jahrhunderts in der SuUB Bremen. Ein Werkstattbericht. *o-bib*, 1(1), 265–279, <https://doi.org/10.5282/o-bib/2014H1S265-279>
- Murphy, J.S. (2014). Introduction to the Special Issue: Visualizing Periodical Networks. *The Journal of Modern Periodical Studies*, 5(1), III–XV.
- Neudecker, C. (2017). *Digitalisierte historische Zeitungen der Staatsbibliothek zu Berlin. Präsentation, Coding da Vinci, Berlin 2017*, https://codingdavinci.de/sites/default/files/2020-08/1515_BO2_SBB_Zeitungen.pdf, 31. Januar 2021.
- Nicholson, B. (2013). The digital turn. Exploring the methodological possibilities of digital newspaper archives. *Media History*, 19(1), 59–73.
- OEAW – Österreichische Akademie der Wissenschaften (1996). *Hypress – Datenbank Berliner Tageszeitungen 1918–1933*, www.oeaw.ac.at/cmc/hypress/bz.htm, 31. Januar 2021.
- Oels, D. (2014) Die Monatsberichte Max Wießners an den Zentralpartei-Verlag der NSDAP Franz Eher Nachf. über wichtige Geschäftsvorfälle im Deutschen Verlag 1940–1945. Teil 1: 1940–1941. In: *Archiv für Geschichte des Buchwesens*, 69, 153–234.
- Oels, D. (2015) Die Monatsberichte Max Wießners an den Zentralpartei-Verlag der NSDAP Franz Eher Nachf. über wichtige Geschäftsvorfälle im Deutschen Verlag 1940–1945. Teil 2: 1942–1945. Mit einem kommentierten Register von Carolin Antes. In: *Archiv für Geschichte des Buchwesens*, 70, 107–239.
- Opgenoorth, E. (1997). *Einführung in das Studium der neueren Geschichte*. Paderborn: Ferdinand Schöningh.
- Ost, G. (1929). Die Berliner Zeitungen am Jahresschluß 1928. *Zeitungs-Verlag*, 30(10), 531–532.

- Podewski, M. (2018). „Kleine Archive“ in den Digital Humanities. Überlegungen zum Forschungsobjekt „Zeitschrift“. *Zeitschrift für digitale Geisteswissenschaften*, Sonderbd. 3, Beitrag 11, 1–19, https://doi.org/10.17175/sb003_010
- Scharf, W. (1975). Rudolf Mosse (1843–1920). In H.-D. Fischer (Hrsg.), *Deutsche Presseverleger des 18. bis 20. Jahrhunderts* (S. 204–213). Pullach b. München: Verlag Dokumentation.
- Schilling, K. (2011). *Das zerstörte Erbe. Berliner Zeitungen der Weimarer Republik im Porträt*. Norderstedt: Books on Demand.
- Schöch, C. (2017). Aufbau von Datensammlungen. In F. Jannidis, H. Kohle & M. Rehbein (Hrsg.), *Digital Humanities. Eine Einführung* (S. 223–233). Stuttgart: J.B. Metzler Verlag.
- Schruhl, F. (2018). Objektumgangsnormen in der Literaturwissenschaft. *Zeitschrift für digitale Geisteswissenschaften*, Sonderbd. 3, Beitrag 12, 1–16, https://doi.org/10.17175/sb003_012
- Schwandt, S. (2018). Digitale Methoden für die Historische Semantik. Auf den Spuren von Begriffen in digitalen Korpora. *Geschichte und Gesellschaft*, 44(1), 107–134.
- Sösemann, B. (1992). Momentaufnahmen – Epochensignaturen. Hauptstädtische Kultur und Zeitungen in den zwanziger Jahren. In K. Siebenhaar (Hrsg.), *Das poetische Berlin. Metropolenkultur zwischen Gründerzeit und Nationalsozialismus* (S. 139–161). Wiesbaden: Deutscher Universitäts-Verlag.
- Staatsbibliothek zu Berlin – Preußischer Kulturbesitz Zeitungsinformationssystem Zefys (2020). *Berliner Volkszeitung / Volkszeitung – digital 1890–1930*, <http://zefys.staatsbibliothek-berlin.de/index.php?id=start>, 31. Januar 2021.
- Stöber, R. (2008). Modernisierungszwang und Massenpresse. Ein Beitrag zur Cultural-Lag-Theorie (nebst Hinweisen zur Frühgeschichte der Pressefotografie). In M. Welke, J. Wilke (Hrsg.), *400 Jahre Zeitung. Die Entwicklung der Tagespresse im internationalen Kontext* (S. 409–430). Bremen: Edition Lumière.
- Strange, C., Wodak, J., & Wood, I. (2014). Mining for the meanings of a murder. The impact of OCR quality on the use of digitized historical newspapers. *Digital Humanities Quarterly*, 8, <http://www.digitalhumanities.org/dhq/vol/8/1/000168/000168.html>, 31. Januar 2021.

- Stulpe, A., & Lemke, M. (2016). Blended Reading. Theoretische und praktische Dimensionen der Analyse von Text und sozialer Wirklichkeit im Zeitalter der Digitalisierung. In M. & G. Wiedemann (Hrsg.), *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse* (S. 17–61). Wiesbaden: Springer VS.
- Tonkin, E. L., & Tourte, G. J. L. (2016). *Working with text. Tools, techniques and approaches for text mining*. Cambridge: Chandos Publishing, Elsevier.
- Trilcke, P., & Fischer, F. (2018). Literaturwissenschaft als Hackathon. Zur Praxeologie der Digital Literary Studies und ihren epistemischen Dingen. In: *Zeitschrift für digitale Geisteswissenschaften*, Sonderbd. 3, Beitrag 4, 1–16, https://doi.org/10.17175/sb003_009
- Vicari, J. (2014). *Blätter machen. Bausteine zu einer Theorie journalistischer Komposition*. Köln: Herbert von Halem.
- Weber, M. (1988). Soziologie des Zeitungswesens (1910). In M. Weber (Hrsg.), *Gesammelte Aufsätze zur Soziologie und Sozialpolitik* (S. 434–441). Tübingen: J. C. B. Mohr (Paul Siebeck).
- Weber, M. (2001). Vorbericht über eine vorgeschlagene Erhebung über die Soziologie des Zeitungswesens (1908). In H. Pöttker (Hrsg.), *Öffentlichkeit als gesellschaftlicher Auftrag. Klassiker der Sozialwissenschaften über Journalismus und Medien* (S. 316–326). Konstanz: UVK.
- Weitin, T. (2017). Scalable Reading. *Zeitschrift für Literaturwissenschaft und Linguistik*, 47(1), 1–6.
- Weitin, T., Gilli, T., & Kunkel, N. (2016). Auslegen und Ausrechnen. Zum Verhältnis hermeneutischer und quantitativer Verfahren in den Literaturwissenschaften. *Zeitschrift für Literaturwissenschaft und Linguistik*, 46(1), 103–115.
- Wikipedia (2020). *Berliner Volkszeitung*, https://de.wikipedia.org/wiki/Berliner_Volks-Zeitung, 31. Januar 2021.
- Wilke, J. (1987). Quantitative Methoden in der Kommunikationsgeschichte. In M. Bobrowsky & W. R. Langenbucher (Hrsg.), *Wege zur Kommunikationsgeschichte* (S. 49–57). München: Verlag Ölschläger.
- Wilke, J. (2000). *Grundzüge der Medien- und Kommunikationsgeschichte. Von den Anfängen bis ins 20. Jahrhundert*. Köln; Weimar; Wien: Böhlau Verlag.
- Wilke, J. (2008). Inhalt und Form der Zeitung im Wandel. In M. Welke & J. Wilke (Hrsg.), *400 Jahre Zeitung. Die Entwicklung der Tagespresse im internationalen Kontext* (S. 355–378). Bremen: Edition Lumière.

Wilke, J. (2012). Nachrichtenflut. In K. Beyrer, H. Gold (Hrsg.), *Das Zeitsparbuch* (S. 67–71). Mainz: Hermann Schmidt Verlag.