

Marcus Schröter

## Historische Volltextdatenbanken

### 1. Historische Volltextdatenbanken und digitale Geschichtswissenschaften

#### 1.1 Einführung

Text, Mensch, Welt – Perspektiven geschichtswissenschaftlichen Arbeitens

Historikerinnen und Historiker arbeiten mit vielfältigsten Texten, in denen sie die Geschichte des Menschen und der Welt lesen und ihre Einsichten der Forschungscommunity und der Öffentlichkeit vorzugsweise schreibend kommunizieren. Zu diesen Texten gehört das breite Portfolio an Forschungsliteratur genauso wie die Vielfalt historischer Quellen in Gestalt von Texten – seien es handschriftliche oder gedruckte textuelle Überlieferungen in Urkunden, historiografischen und literarischen Texten oder in nicht-textuellen kulturellen Artefakten. Sowohl Forschungsliteratur als auch Quellen liegen zunehmend in Volltextdatenbanken vor, die das digitale geschichtswissenschaftliche Arbeiten in Studium, Forschung und Lehre grundlegend prägen.<sup>1</sup>

Volltextdatenbanken sind daher einerseits Archive handschriftlicher und gedruckter Textquellen, andererseits Repositorien und Räume geschichtswissenschaftlicher Diskurse in selbstständigen<sup>2</sup>

---

<sup>1</sup> Im Vergleich zu den Voraufgaben aus den Jahren 2016 und 2018 (<https://guides.clio-online.de/guides/sammlungen/historische-volltextdatenbanken/2016>, <https://guides.clio-online.de/guides/sammlungen/historische-volltextdatenbanken/2018>) musste der Autor diesen Artikel grundlegend überarbeiten, da neue Technologien die Möglichkeiten und dadurch die Relevanz von Volltextdatenbanken für geschichtswissenschaftliches Arbeiten signifikant verändert haben. Sämtliche Online-Ressourcen wurden am 16.06.2024 geprüft.

<sup>2</sup> Beispielsweise Monografien, Zeitschriften oder Sammelbände.

und unselbstständigen<sup>3</sup> Forschungsbeiträgen, die zunehmend vernetzt und interaktiv ausgestaltet werden. Dieser Guide legt im Folgenden den Fokus auf Volltextdatenbanken historischer Quellen. Historische Volltextdatenbanken können entweder auf genuin digitalen oder retrodigitalisierten einzelnen Texten oder Sammlungen von Büchern und Objekten in Bibliotheken, Archiven und Museen basieren. Die Zusammenführung von Einzeltexten oder Textcorpora orientiert sich häufig an chronologischen, regionalen oder thematischen Kriterien, die auch für die Unterscheidung der Guides dieses Handbuchs bestimmend sind. Als Ergebnis dieser Aufbereitung von Texten präsentiert sich die historische Volltextdatenbank als eine in der Regel Einzeltexte oder Einzelbestände übergreifende digitale Textsammlung. Folgerichtig ist der Guide „Historische Volltextdatenbanken“ in den Teil „Sammlungen“ eingebettet.

Eng verbunden mit Volltextdatenbanken sind die Digital Humanities: Diese entwickeln komplexe Werkzeuge und Methoden, mit denen Volltexte generiert, analysiert, visualisiert und bearbeitet

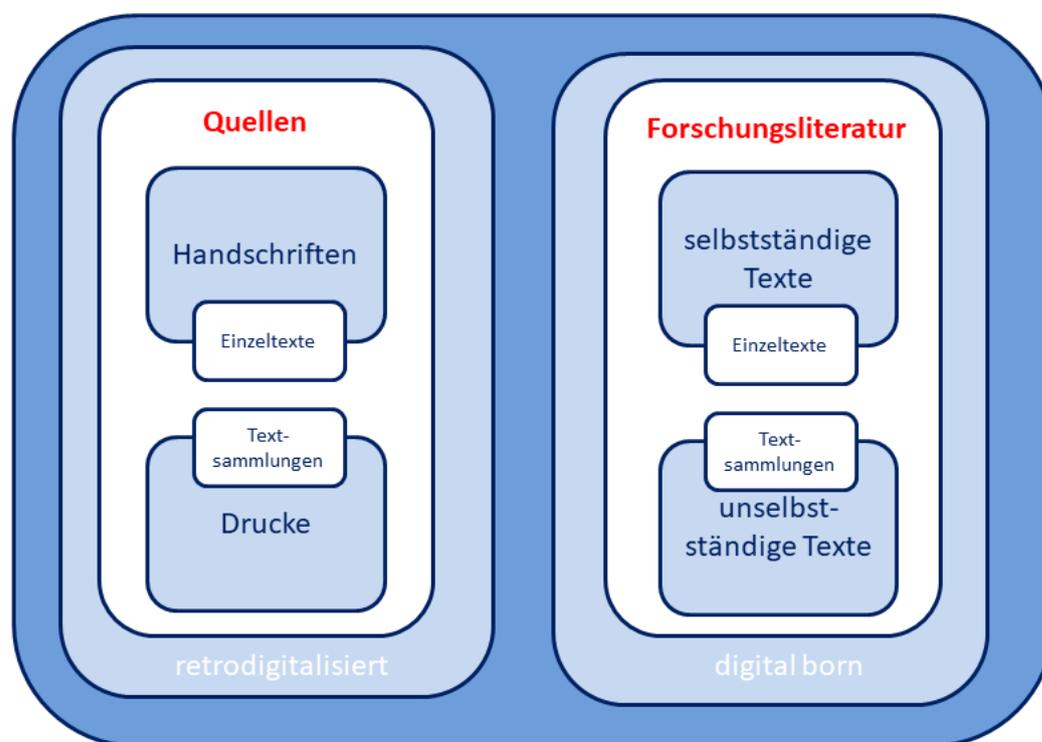


Abbildung 1. Typen von Volltexten

<sup>3</sup> Beispielsweise Aufsätze in Zeitschriften, Sammelbänden, Fachportalen, Blogs.

werden können, um zu historischen Erkenntnissen zu gelangen, die im Zeitalter analoger Texte und Textüberlieferung so nicht möglich gewesen wären. Zugleich erfordert der Umgang mit digitalen Quellen eine spezifisch digitale Quellenkritik. Welche Rolle historische Volltextdatenbanken in diesen komplexen Prozessen spielen, möchte dieser Guide zeigen.

### Manuskriptkultur, «Gutenberg-Galaxis» und Gedächtnisinstitutionen

Die Geschichte der Geisteswissenschaften ist nicht nur eng mit der Geschichte der Institutionen des kulturellen Gedächtnisses, den Archiven, Bibliotheken und Museen, sondern auch mit der Geschichte der schriftbasierten Kulturtechniken Lesen und Schreiben verbunden.<sup>4</sup> Nach Justus Lipsius (1547–1606)<sup>5</sup> wurde die Bibliothek gemeinsam mit der Schrift erfunden.<sup>6</sup> Daher soll das Thema Historische Volltextdatenbanken in den engen Kontext der Schrift-, Buch- und Bibliotheksgeschichte gestellt werden. Die kulturgeschichtliche Leistung der Gedächtnisinstitutionen ist die planvolle Sammlung und Aufbewahrung handschriftlicher und gedruckter Texte oder Artefakte menschlicher Kultur. Bibliotheken stellen zugleich Strukturen bereit, in denen Menschen, die mit Texten arbeiten, sich bewegen. Die Geschichte von Texten und Büchern ist zugleich die Geschichte von Leserinnen und Lesern, aus denen im Raum realer und virtueller Bibliotheken Autorinnen und Autoren werden.

Eine erste Herausforderung für Archive, Bibliotheken und Museen war das Ende der Manuskriptkultur im 15. Jahrhundert – bis dahin wurden Texte einzeln handschriftlich in den Skriptorien der Klöster kopiert und in den daran angeschlossenen Bibliotheken aufbewahrt.

---

<sup>4</sup> Lauer, Gerhard, Was ist Buchwissenschaft, wenn sie eine Disziplin ist?, in: Archiv für Geschichte des Buchwesens 77 (2022), S. 173–178, hier S. 173.

<sup>5</sup> Lipsius, Justus, Syntagma de bibliothecis, Antwerpen 1602, S. 9.

<sup>6</sup> Stäcker, Konversion des kulturellen Erbes für die Forschung, S. 223.

Mit dem Beginn des nach dem Mainzer Buchdrucker Johannes Gutenberg als Gutenberg-Galaxis<sup>7</sup> bezeichneten Buchzeitalters begann das exponentielle Wachstum gedruckter Texte und damit die wachsende Komplexität der Gedächtnisinstitutionen: Bibliotheken wuchsen nicht nur physisch, auch das Sammeln von Büchern und das Organisieren des in diesen dokumentierten Wissens stellte sie vor immer neue Herausforderungen. Da keine einzelne Institution das gesamte, in Büchern fixierte Weltwissen sammeln konnte, sondern dieses über zahllose Archive, Bibliotheken und Museen hinweg verstreut war, lag die Herausforderung für Forschende darin, die für sie relevanten Bücher überhaupt erst zu finden. Praktisch bedeutete dies, sich auf Reisen zu begeben und die Texte zu finden, zu lesen und zu exzerpieren. Forschende wurden vielfach selbst zu Autorinnen und Autoren, deren neue Erkenntnisse in eigenen gedruckten Texten wiederum neue Lektüren und Texte anstießen.

Zu den vor-digitalen, analogen Grundlagen historischer Volltextdatenbanken gehören nicht nur handschriftliche und gedruckte Schriftquellen, sondern auch ihre Editionen in zum Teil umfassenden Corpora. Sie sind Ergebnisse forschenden Reisens zu Bibliotheken. Viele dieser insbesondere seit dem 19. Jahrhundert entstandenen Monumente der Wissenschaftsgeschichte wie die *Monumenta Germaniae Historica* (MGH) oder die *Regesta Imperii* (RI) sind heute in Volltextdatenbanken transformiert und ermöglichen Methoden eines *distant reading*.

Vielfältige Innovationen bei der Reproduktion von Texten im 20. Jahrhundert waren ein erster Schritt zur Überwindung der Gutenberg-Galaxis. Sie veränderten nicht nur die Sammeltätigkeit von Bibliotheken, sondern auch das Arbeiten mit ihnen. Die analogen Vorläufer historischer Volltextdatenbanken entstanden aus der Reproduktion handschriftlicher und gedruckter Einzeltexte oder Textsammlungen in umfangreichen, kumulierten Corpora in Mikroform-Archiven (Mikrofilm, Mikrofiche). Bereits hier wurde angestrebt, chronologisch, regional oder thematisch verwandte Texte zusammenzuführen. An ihrer Erstellung waren Bibliotheken in

---

<sup>7</sup> McLuhan, Marshall, Die Gutenberg-Galaxis. Das Ende des Buchzeitalters, München 1968.

Kooperation mit Verlagen nicht nur aktiv beteiligt, sie erweiterten durch die Erwerbung derartiger Mikroform-Kopien der in anderen Bibliotheken verfügbaren Bestände ihre eigenen Sammlungen signifikant. Dennoch blieb dieser Schatz an Texten und an Wissen häufig verborgen – lag (und liegt) doch eine Barriere für die Benutzung in der Notwendigkeit, wenig komfortable Lesegeräte verwenden zu müssen. Volltextrecherchen, wie sie Volltextdatenbanken ermöglichen, waren im Falle von Mikroform-Archiven nicht möglich.

#### Von der Digitalisierung zur digitalen Transformation

Überwand die „Gutenberg-Galaxis“ im 15. Jahrhundert die Manuskriptkultur der Vormoderne, so wird seit Beginn des 21. Jahrhunderts die ein halbes Jahrtausend währende Epoche der „schwarzen Kunst“ durch das Zeitalter der Digitalisierung und der elektronischen Medien abgelöst. Historische Volltextdatenbanken entstanden – und entstehen – im konkreten Kontext der Entwicklung von der Digitalisierung zur digitalen Transformation während der noch andauernden Revolution unserer Schriftkultur. Meilensteine der Digitalisierung in deutschen wissenschaftlichen Bibliotheken war die Gründung der durch die *Deutsche Forschungsgemeinschaft (DFG)*<sup>8</sup> geförderten Digitalisierungszentren an der Bayerischen Staatsbibliothek München und an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen im Jahr 1997. Beide widmeten sich von Beginn an der Retrodigitalisierung ihrer historischen Bestände – den Quellen geschichtswissenschaftlichen Arbeitens. Diese sind heute Grundlage historischer Volltextdatenbanken für Quellen.

Da die Qualität eines Digitalisats entscheidend für dessen wissenschaftliche Weiterverarbeitung ist, wurden von Anfang an verbindliche Standards nicht nur für technische Parameter und Prä-

---

<sup>8</sup> <https://www.dfg.de/index.jsp>

sensation der Digitalisate, sondern auch für ihre formale und inhaltliche Erschließung in Praxisregeln für die Digitalisierung<sup>9</sup> festgelegt. Das *Göttinger Digitalisierungszentrum (GDZ)*<sup>10</sup> widmete sich unter anderem im Projekt *Gutenberg Digital*<sup>11</sup> der Digitalisierung des Göttinger Exemplars der berühmten Mainzer Bibel<sup>12</sup>, mit der die Gutenberg-Galaxis ihren Ausgang nahm: Sie war eines der ersten Bücher, die in das digitale Zeitalter transformiert wurden. Das *Münchener Digitalisierungszentrum (MDZ)*<sup>13</sup> legte einen Fokus unter anderem auf die Digitalisierung seiner weltweit einmaligen Inkunabelsammlung. Beide Institutionen trugen in vielfältigen Kooperationen dazu bei, dass die Digitalisierung historischer Bestände schrittweise in den Routinebetrieb auch kleinerer Bibliotheken mit historischen Beständen überführt werden konnte.

Seit Beginn der systematischen Digitalisierung historischer Bibliotheksbestände mit dem Ziel der Unterstützung von Studium, Lehre und Forschung der Zukunft wurden schrittweise analoge Einzelobjekte oder ganze Sammlungen in digitale Formate transformiert. Die so entstandenen digitalen Forschungsressourcen ergänzen – wie bereits zuvor die Mikroform-Archive – die in der individuellen Bibliothek physisch vorhandenen Texte und Sammlungen.

Digitalisate analoger Texte sind jedoch weit mehr als bloße Reproduktionen: Sie können im Volltext durchsucht, automatisch transkribiert, übersetzt, vorgelesen und digital weiterbearbeitet werden. Insofern prägt die Form der Vorlagen die Praktiken der Leserinnen und Leser sowie die Arbeitsweisen und -methoden der Historikerinnen und Historiker – und umgekehrt.

---

<sup>9</sup> Vgl. hierzu: [https://www.dfg.de/formulare/12\\_151/](https://www.dfg.de/formulare/12_151/);  
<https://zenodo.org/record/7435724>.

<sup>10</sup> <https://gdz.sub.uni-goettingen.de/>

<sup>11</sup> <http://www.gutenbergdigital.de/>

<sup>12</sup> Das Exemplar der Niedersächsischen Staats- und Universitätsbibliothek Göttingen gehört seit 2001 zum Weltdokumentenerbe „Memory of the World“ der UNESCO.

<sup>13</sup> <https://www.digitale-sammlungen.de/de/>

---

Die Frage, „welchen Stellenwert das Analoge im Digitalen hat, verändert doch die digitale Transformation auch ihren analogen Ausgangspunkt“<sup>14</sup>, ist zu diskutieren. Den Weg von der Digitalisierung zur digitalen Transformation beschreiten die einzelnen Wissenschaftsdisziplinen in unterschiedlichen Geschwindigkeiten und mit eigenen Schwerpunkten. Für die Geschichtswissenschaften ist die Entwicklung der Digital Humanities mit ihren spezifischen Konzepten und Methoden zentral – exakt hier spielen die historischen Volltextdatenbanken eine zentrale Rolle.

### Von der digitalen Transformation zu den Digital Humanities

Das Bewusstsein, dass das kulturelle Erbe das „*unschätzbare Gewebe*“<sup>15</sup> sei, das Europa zusammenhält, ließ die Europäische Kommission erstmals das Jahr 2018 als „*Europäisches Jahr des Kulturerbes*“<sup>16</sup> erklären. Hierbei wurde insbesondere die Digitalisierung für die Kulturvermittlung hervorgehoben, durch die die *Europeana*<sup>17</sup> als Schaufenster des europäischen Kulturerbes erst möglich wurde.

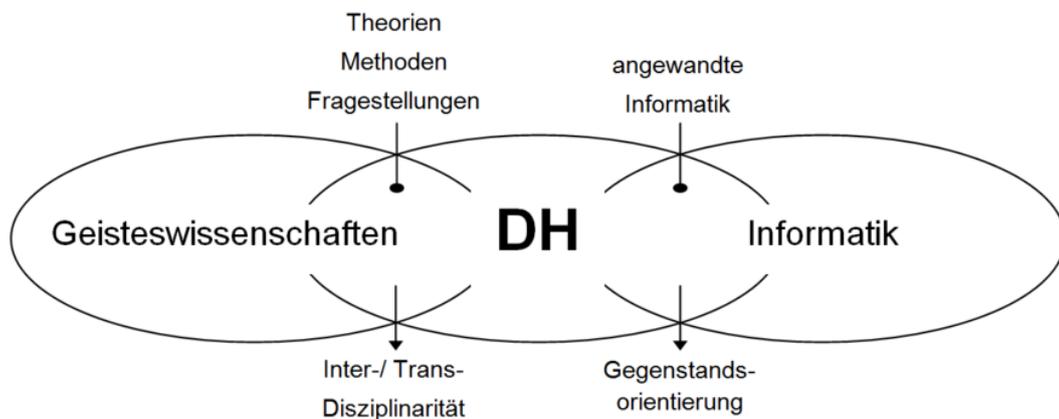


Abbildung 2. Digital Humanities als Brückendisziplin zwischen Geisteswissenschaften und Informatik (Grafik von Patrick Sahle, <https://www.geschichte.uni-wuppertal.de/de/lehrgebiete/digital-humanities/>, 16.07.2024)

<sup>14</sup> Steinhauer, Eric W., Die Bibliothek 2040 – eine Einrichtung der digitalen Transformation mit vielen Büchern?, in: Bibliothek: Forschung und Praxis 47 (2023) 1, S. 31, <https://doi.org/10.1515/bfp-2023-0014>.

<sup>15</sup> [https://dewiki.de/Lexikon/Europäisches\\_Kulturerbejahr](https://dewiki.de/Lexikon/Europäisches_Kulturerbejahr)

<sup>16</sup> [https://ec.europa.eu/commission/presscorner/detail/de/MEMO\\_17\\_5066](https://ec.europa.eu/commission/presscorner/detail/de/MEMO_17_5066)

<sup>17</sup> <https://www.europeana.eu/de>

Dieser entscheidende Schritt von der Digitalisierung des schriftlichen Kulturerbes hin zur digitalen Transformation bedeutet für die Geistes- und Kulturwissenschaften einen Paradigmenwechsel, durch den die traditionellen hermeneutischen Methoden durch digitale Methoden und Werkzeuge erweitert werden.<sup>18</sup>

In den Geistes- und Kulturwissenschaften hat diese Entwicklung zur Etablierung der Digital Humanities (DH) geführt, die sich mit der Anwendung und Reflexion computergestützter Verfahren und Instrumenten zur Bearbeitung digitaler Ressourcen befassen. Typische Arbeitsfelder sind qualitative Verfahren wie die Beschreibung, Transkription, Edition und Annotation digitalisierter Quellen beispielsweise durch digitale Manuskriptanalyse<sup>19</sup> einerseits oder quantitative Methoden wie die Analyse digitalisierter Texte durch Text Mining, Topic Modeling,<sup>20</sup> Textvisualisierung,<sup>21</sup> Stilometrie<sup>22</sup> oder Netzwerkanalyse andererseits.<sup>23</sup> Die dabei entstehenden technischen und inhaltlichen Metadaten sind als Forschungsdaten Rohstoff für künftige Analysen und Bearbeitungen.

---

<sup>18</sup> Stäcker, Konversion des kulturellen Erbes für die Forschung.

<sup>19</sup> Horstmann, Jan, Digitale Manuskriptanalyse, in: forTEXT. Literatur digital erforschen, 2018, <https://fortext.net/routinen/methoden/digitale-manuskriptanalyse>. Ist eine einzelne Handschrift gemäß den technischen Standards digitalisiert, lässt sie sich mit Verfahren der digitalen Manuskriptanalyse durch strukturierte und normierte Annotation oder automatische Transkription bearbeiten; ist das digitalisierte Manuskript nicht für eine automatische Texterkennung und Transkription geeignet, können die Bilddateien durch Metadaten detailliert erschlossen werden.

<sup>20</sup> Horstmann, Jan, Topic Modeling, in: forTEXT. Literatur digital erforschen, 2018, <https://fortext.net/routinen/methoden/topic-modeling>.

<sup>21</sup> Horstmann, Jan; Stange, Jan-Erik, Textvisualisierung, in: forTEXT. Literatur digital erforschen, 2018, <https://fortext.net/routinen/methoden/textvisualisierung>.

<sup>22</sup> Horstmann, Jan, Stilometrie, in: forTEXT. Literatur digital erforschen, <https://fortext.net/routinen/methoden/stilometrie>.

<sup>23</sup> Schumacher, Mareike, Netzwerkanalyse, in: forTEXT. Literatur digital erforschen, 2018, <https://fortext.net/routinen/methoden/netzwerkanalyse>.

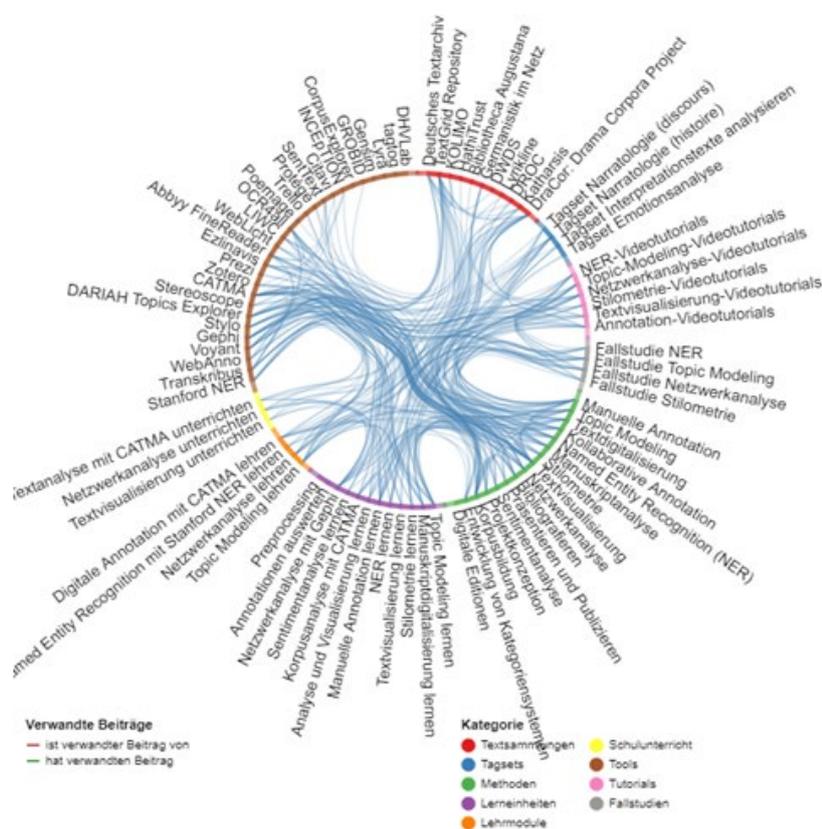


Abbildung 3. Arbeiten mit digitalen Texten in den Digital Humanities (Grafik der Inhalte von forTEXT, <https://fortext.net/>, 16.07.2024)

Perspektivisch lassen sich die Methoden und Tools der DH nicht nur in einen erweiterten Kanon der traditionellen historischen Grundwissenschaften<sup>24</sup> einbinden, sondern in eine digitale Quellenkritik und Hermeneutik, die Ergebnis der weiteren digitalen Transformation der Geschichtswissenschaften sein wird.

Die DH entwickeln sich gegenwärtig einerseits als eigenständige, die Geistes- und Kulturwissenschaften umfassende interdisziplinäre akademische Disziplin, andererseits finden ihre Konzepte

<sup>24</sup> Bösch, Frank; Schlotheuber, Eva, Quellenkritik im digitalen Zeitalter: Die Historischen Grundwissenschaften als zentrale Kompetenz der Geschichtswissenschaft und benachbarter Fächer, in: Historische Grundwissenschaften und die digitale Herausforderung, Hrsg. für H-Soz-Kult von Rüdiger Hohls, Claudia Prinz und Eva Schlotheuber, Historisches Forum 18 (2016), S. 16–20, <https://edoc.hu-berlin.de/handle/18452/19491>; Digital Humanities in den Geschichtswissenschaften, hrsg. von Christina Antenhofer, Christoph Kühberger und Arno Strohmeyer, Wien 2024.

Eingang in die einzelnen akademischen Fächer – wie beispielsweise in die Geschichtswissenschaften.<sup>25</sup> Hier werden sie den traditionellen Kanon der Grundwissenschaften substanziell ergänzen.<sup>26</sup>

#### Texterkennung für Volltextdatenbanken und digitale Editionen

Die Transformation des handschriftlichen und gedruckten schriftlichen Kulturerbes nach dem Ende der Gutenberg-Galaxis in eine maschinenlesbare Form wurde am Beginn des Computerzeitalters durch die Eingabe intellektuell erstellter Abschriften in elektronische Dokumente realisiert. Grundlage waren entweder die Originalquellen oder ihre Digitalisate. Während im ersten Fall ein Bearbeiter das Original aufsuchte, um dieses abzuschreiben, ermöglichte im zweiten Fall der ubiquitäre Zugriff auf Digitalisate eine Transkription auch durch beliebig viele Transkribierende in einem kollaborativen Dokument – beispielsweise in einem Transkriptionstheon.<sup>27</sup> Mit dieser Übertragung eines Textes aus einer analogen oder digitalen Quelle in ein elektronisches Textverarbeitungsprogramm war der erste Schritt hin zu einer beliebig weiten Verbreitung digital codierter Texte gelungen – ein in der Geschichte der Schriftkultur revolutionärer Vorgang. In Volltextdatenbanken können diese transformierten Texte systematisch analysiert und weiterverarbeitet werden.

Dieses Verfahren individuell oder kollaborativ intellektuell erstellter elektronischer Transkriptionen von Texten aus Originalquellen wird künftig vollständig automatisiert. Waren die schrittweise digitalisierten Textquellen in Form von Bilddateien zwar

---

<sup>25</sup> Kobel, Esther; Volp, Ulrich, Distant reading – Perspektiven einer digitalen Zeit. Eine Einführung, in: *Journal of Ethics in Antiquity and Christianity* 4 (2022), S. 5–10.

<sup>26</sup> Diese Gedanken wurden unter anderem formuliert in der Diskussion anlässlich des Papiers von Bösch; Schlotheuber, *Quellenkritik*, 2016; Johrendt, Jochen, Digitalisierung als Chance, in: Bösch/Schlotheuber, *Quellenkritik*, 2016, S. 41–42, Rehbein, Malte, Digitalisierung braucht Historiker/innen, die sie beherrschen, nicht beherrscht, in: Bösch/Schlotheuber, *Quellenkritik*, 2016, S. 45–52.

<sup>27</sup> Ein Beispiel für ein digitales Crowd-Sourcing-Projekt ist „Faithful Transcriptions“ (<https://lab.sbb.berlin/events/faithful-transcriptions/>).

---

weltweit zugänglich, so mussten sie doch erst durch Menschen gelesen und elektronisch transkribiert werden, bevor sie in Form von Volltexten weiterbearbeitet werden konnten. Doch wie lassen sich die Bilddateien softwaregestützt in prozessierbare Volltexte verwandeln? Unproblematisch ist die *Optical Character Recognition* (OCR)<sup>28</sup> einer Antiqua-Standard-Schrift, die moderne Drucke auszeichnet. Der überwiegende Anteil digitalisierter Schriftquellen und Textobjekte liegt jedoch handschriftlich oder in gebrochenen Druckschriften vor, die größte Varianz besitzen und nur mit komplexerer Transkriptionssoftware und Modellen der OCR oder *Handwritten Text Recognition* (HTR) zu bearbeiten sind.

Ziel der Texterkennung durch HTR/OCR ist es, die innerhalb der durch Scanner oder Digitalkameras produzierten Bilddigitalisate erfassten Buchstaben, Wörter und Sätze einer handschriftlichen oder gedruckten Buchseite zu erkennen, sie als bedeutungstragende Zeichen zu interpretieren und in maschinenlesbare Texte umzuwandeln. Diese Umwandlung von Bildern in Texte findet in komplexen Schritten statt: Nach einer Vorverarbeitung des Bilddigitalisats (Preprocessing mit Bildoptimierung, Binarisierung) für die Texterkennung findet zunächst eine Optical Layout Recognition (OLR) mit Identifizierung der Layoutelemente wie Textblöcke, Absätze, Überschriften, Grundlinien und Zeilen statt, anschließend werden die Einzelzeichen von einer Texterkennungs-Engine automatisch voneinander separiert, klassifiziert und ein Abgleich der identifizierten Worte mit einem Wörterbuch durchgeführt. Voraussetzung einer solchen automatischen Transkription ist jedoch die intellektuelle Vorbereitung und Kontrolle dieses Prozesses, indem zunächst eine Anzahl von Seiten der vorstrukturierten Bilddateien transkribiert wird. Auf Grundlage dieser so genannten Grund-

---

<sup>28</sup> In diesem Guide wird der etablierte Begriff Optical Character Recognition (OCR) verwendet, auch wenn inzwischen „Text durch neuronale Netze verarbeitet [wird], die nicht Zeichen für Zeichen lesen, sondern ganze Zeilen statt einzelner Zeichen erkennen“ (<https://ocr-bw.bib.uni-mannheim.de/was-ist-ocr/>). Grundsätzlich beschreibt OCR den technischen Prozess, in dem in einem digitalen Bild Schriftzeichen (characters) erkannt werden. Die Qualität des Image-Digitalisats ist für die OCR-Verarbeitung von entscheidender Bedeutung.

Truth-Daten<sup>29</sup> werden Modelle trainiert, mit denen automatische Transkriptionen möglich werden. Abschließend wird das HTR/OCR-Ergebnis korrigiert, bevor eine Langzeitarchivierung der Transkriptionen in geeigneten Repositorien stattfindet.

Aktuelle Verfahren automatischer Texterkennung zielen auf die Entwicklung standardisierter Modelle, um möglichst vielfältige Schreibstile und Drucktypen zu erkennen. Eine Volltexterkennung gedruckter Texte mit der freien OCR-Software *Tesseract*<sup>30</sup> oder dem kommerziellen *Abbyy FineReader*<sup>31</sup> erzielt bereits hervorragende Ergebnisse.<sup>32</sup> Auf das Projekt *OCR-D*<sup>33</sup> wird in Kapitel 1.3 eingegangen. Die Texterkennung für Handschriften basiert auf kommerzieller HTR-Software wie *Transkribus*<sup>34</sup> oder dem Open Source-Programm *eScriptorium*<sup>35</sup>.<sup>36</sup> Die Zukunftsvision, das gesamte in Handschriften und Drucken überlieferte schriftliche Kulturerbe, sofern es digitalisiert ist, automatisch lesbar zu machen, gewinnt bereits

---

<sup>29</sup> Ground Truth (GT)-Daten sind der automatischen Zeichenerkennung zugrunde liegenden Referenzdaten und meinen die fehlerfreien, manuell korrigierten Text- und Strukturdaten, auf deren Grundlage OCR/HTR-Modelle trainiert werden können, vgl. <https://ocr-d.de/de/gt-guidelines/trans/index.html>.

<sup>30</sup> Hinter dem selbstständig lernenden Open-Source-System Tesseract-OCR (<https://github.com/tesseract-ocr>) steht inzwischen eine breite Community.

<sup>31</sup> Schumacher Mareike, *Abbyy FineReader*, in: *forTEXT. Literatur digital erforschen*, 2019; <https://fortext.net/tools/tools/abbyy-finereader>.

<sup>32</sup> Mit dem von Stefan Weil entwickelten Tesseract-OCR-Modell *frak2021* steht ein leistungsfähiges Tool für die OCR-Erkennung historischer Zeitungen zur Verfügung ([https://ub-backup.bib.uni-mannheim.de/~stweil/tesstrain/frak2021/tessdata\\_best/](https://ub-backup.bib.uni-mannheim.de/~stweil/tesstrain/frak2021/tessdata_best/)).

<sup>33</sup> <https://ocr-d.de/>

<sup>34</sup> <https://readcoop.eu/de/transkribus/>

<sup>35</sup> <https://gitlab.com/scripta/escriptorium>

<sup>36</sup> Nach Vorarbeiten im Projekt *tranScriptorium* (2013–2015) wurde *Transkribus* im EU-geförderten Projekt *READ* (2016–2019) weiterentwickelt und anschließend als Genossenschaft kommerziell selbstständig. Beispielsweise bietet *Transkribus* inzwischen für Kurrentschriften generische Modelle, die gute automatische Transkriptionen ohne Nachbearbeitung ermöglichen. Es ist möglich, Ground-Truth-Daten im PAGE-XML-Format aus *Transkribus* zu exportieren und nach *eScriptorium* zu importieren.

jetzt eine immer konkretere Gestalt, da Entwicklungen der Künstlichen Intelligenz (KI) aus unterschiedlichen Fachperspektiven vorangetrieben werden.

Um Konzepte und Instrumente der DH vollumfänglich auf das schriftliche Kulturerbe anzuwenden, müssen ihre Forschungsgegenstände, insbesondere die in historischen Quellen überlieferten Texte, vollständig digital verfügbar sein. Damit ist nicht nur die Herstellung digitaler Faksimiles gemeint, sondern die Transkription und Auszeichnung der Quellentexte in digitalen maschinenlesbaren Formaten einer Volltexttransformation.

In der digitalen Transformation textbasierter historischer Kulturwissenschaften müssen Texte nicht nur *transitiv*<sup>37</sup> sein, indem sie von ihren materiellen Trägern abgelöst und beliebig kopiert werden können, Texte müssen auch *prozessierbar*<sup>38</sup> sein, indem sie indexiert, durchsucht und mit Methoden der DH, beispielsweise in digitalen Editionen, bearbeitet und erforscht werden können. Dies ist erst dann möglich, wenn die auf Grundlage der Digitalisate intellektuell oder automatisch generierten Transkriptionen eine TEI/XML-Auszeichnung erhalten.<sup>39</sup>

## 1.2 Nationale bibliothekarische Infrastrukturen und Nachweissysteme

System der überregionalen Versorgung für Forschungsressourcen und Dokumentation des deutschen schriftlichen Kulturerbes

Die digitale Transformation der Geistes- und Kulturwissenschaften benötigt leistungsfähige Informationsinfrastrukturen, zu denen auch Gedächtnisinstitutionen wie Bibliotheken gehören. Die Entwicklung historischer Volltextdatenbanken für Quellen und Forschungsliteratur ist häufig mit diesen Infrastrukturen verbunden. Gegenstand des folgenden Kapitels sind zentrale Institutionen, die

---

<sup>37</sup> Stäcker, Thomas, Konversion des kulturellen Erbes für die Forschung, S. 1ff.

<sup>38</sup> Ebda.

<sup>39</sup> Einen hervorragenden Überblick über Konzepte und Methoden digitaler Textanalyse im Kontext der Digital Humanities bietet *forTEXT*: <https://fortext.net/>.

gemeinsam mit maßgeblichen Autorinnen und Autoren sowie mit Verlagen relevante Projekte mit Bezug auf historische Volltextdatenbanken für Quellen und Forschungsliteratur tragen.<sup>40</sup>

Nach Bernhard Fabian<sup>41</sup> haben Bibliotheken für die Geisteswissenschaften eine vergleichbare Funktion, wie sie Labore für die Naturwissenschaften besitzen. Da es sich bei den von Bibliotheken bereitgestellten geisteswissenschaftlichen Ressourcen vornehmlich um Texte handelt, liegt eine Herausforderung historischen Arbeitens darin, mit textbasierten Forschungsressourcen und Nachweis-systemen realer und virtueller Bibliotheken umzugehen, Texte zu finden, zu lesen, zu analysieren und zu interpretieren. Historische Volltextdatenbanken sind Teil dieser komplexen Textwelten.

Die bibliothekarische Infrastruktur in Deutschland gründet auf zwei Konzepten: Verteilte deutsche Nationalbibliothek und Verteilte nationale Forschungsbibliothek. Während die Verteilte deutsche Nationalbibliothek das gesamte deutsche schriftliche Kulturerbe als Quellen sammelt, katalogisiert, digitalisiert und bewahrt, verfolgt die Verteilte nationale Forschungsbibliothek dasselbe Ziel für die relevante Forschungsliteratur aus allen wissenschaftlichen Disziplinen – auch für die Geschichtswissenschaften. In beiden Fällen entstehen Volltextdatenbanken für Studium, Forschung und Lehre.

---

<sup>40</sup> Klaus Gantert, Clio-Guide: Bibliotheken, in: Clio Guide – Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften, hrsg. von Silvia Daniel, Wilfried Enderle, Rüdiger Hohls, Thomas Meyer, Jens Prellwitz, Claudia Prinz, Annette Schuhmann, Silke Schwandt, 3. erw. und aktualisierte Aufl., Berlin 2023, <https://doi.org/10.60693/10nt-2y42>.

<sup>41</sup> Fabian, Bernhard, Buch, Bibliothek und geisteswissenschaftliche Forschung. Zu Problemen der Literaturversorgung und Literaturproduktion in der Bundesrepublik Deutschland, Göttingen 1983.

---

Verteilte deutsche Nationalbibliothek und retrospektive deutsche Nationalbibliografie: Handschriften, Inkunabeln, Alte Drucke des 16. bis 18. Jahrhunderts

Sammlungen in Bibliotheken als Spiegel deutscher Geschichte

Die Verteilte deutsche Nationalbibliothek lässt sich unmittelbar aus der deutschen Geschichte herleiten. Im Gegensatz zu anderen Nationen wurde in Deutschland eine Nationalbibliothek erst 1912 gegründet, um als zentrale Archivbibliothek das schriftliche Kulturerbe in deutscher Sprache aus Deutschland und dem Ausland nicht nur zu sammeln, sondern als nationalbibliografisches Zentrum auch zu verzeichnen. Hierin liegt die Aufgabe der heutigen *Deutschen Nationalbibliothek (DNB)*<sup>42</sup>. Für das vorangegangene Jahrtausend mit seiner reichen Schriftkultur in Gestalt von Manuskripten und Drucken indessen ist der Nachweis der Überlieferung höchst disparat und durch die Geschichte der deutschen Territorien geprägt, aus der einzelne Bibliotheken mit besonderer Sammlungsgeschichte hervorgingen. Diesen wurde später für die Jahrhunderte vor 1912 die retrospektive Sammlung und Verzeichnung des deutschen schriftlichen Kulturerbes anvertraut (s.u.).

Handschriften

Die zentrale Quellengattung vormoderner historischer Epochen gründet in der reichen Klosterkultur, in der Texte erdacht, abgeschrieben und aufbewahrt wurden. Das Fenster in diese Welt der Manuskripte ist das *Handschriftenportal (HSP)*<sup>43</sup>, das 2023 *Manuscripta Mediaevalia* ablöste. Es weist Handschriften und Handschriftenfragmente aus Mittelalter und Neuzeit in deutschen Sammlungen zentral nach (vgl. Kap. 2.6). Im Hinblick auf das mittelalterliche Handschriftenerbe geht man von etwa 60.000 überlieferten Dokumenten aus. Als Forschungsinfrastruktur für das in deutschen

---

<sup>42</sup> [https://www.dnb.de/DE/Home/home\\_node.html](https://www.dnb.de/DE/Home/home_node.html)

<sup>43</sup> <https://handschriftenportal.de/>

Sammlungen aufbewahrte und im HSP präsentierte Handschriftenerbe haben sich sechs *Handschriftenzentren*<sup>44</sup> an den national bedeutenden Altbestandsbibliotheken etabliert: Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Universitätsbibliothek Frankfurt am Main, Universitätsbibliothek Leipzig, Bayerische Staatsbibliothek München, Württembergische Landesbibliothek Stuttgart, Herzog August Bibliothek Wolfenbüttel.

### Inkunabeln

Sind Handschriften als historische Quellen für einen Zeitraum von etwa eineinhalb Jahrtausenden von der Spätantike bis in die jüngste Zeitgeschichte relevant, so prägen Inkunabeln lediglich die kurze Epoche zwischen der Mitte des 15. und dem Beginn des 16. Jahrhunderts.<sup>45</sup> Mediengeschichtlich stehen sie zwischen Handschriften und Alten Drucken: Zwar überwand die Inkunabeln mit ihrer technischen Innovation des Druckes mit beweglichen Lettern und der dadurch möglichen hundertfachen Vervielfältigung von Texten das handgeschriebene mittelalterliche Buch – im Hinblick auf Gestaltung und Layout blieb dieses jedoch noch mächtiges Vorbild, an dem sich die Inkunabeldrucker häufig bis ins Detail orientierten.

Schätzungen zufolge sind weltweit etwa 550.000 Exemplare von rund 30.000 unterschiedlichen Werken erhalten, darunter etwa 125.000 Exemplare in Deutschland.<sup>46</sup> Über die national und international relevanten Nachweissysteme und Volltextdatenbanken für Inkunabeln bietet Kapitel 2.6 einen detaillierten Überblick. Die

---

<sup>44</sup> <https://www.handschriftenzentren.de/>

<sup>45</sup> Während die Drucke der Jahre bis 1500 als Inkunabeln (lat. *Windeln, Wiege*) bezeichnet wurden, spricht man von den Drucken der Jahre 1501–1520 von *Postinkunabeln* (Schmitz, Wolfgang, Grundriss der Inkunabelkunde. Das gedruckte Buch im Zeitalter des Medienwechsels, Stuttgart 2023).

<sup>46</sup> Eisermann, Falk, The Gutenberg Galaxy's Dark Matter: Lost Incunabula, and Ways to Retrieve Them, in: Flavia Bruni und Andrew Pettegree (Hrg.), Lost Books. Reconstructing the Print World of Pre-Industrial Europe. Leiden/Boston: Brill 2016, S. 31, Anm. 2.; <https://www.blb-karlsruhe.de/sammlungen/inkunabeln/>.

historische Epoche der Frühen Neuzeit umfasst die mediengeschichtlichen Epochen der Inkunabeln und der Alten Drucke zwischen dem Ende der Handschriftenzeit Mitte des 15. Jahrhunderts und dem Beginn der industriellen Buchproduktion im Zeitalter der Industrialisierung im 19. Jahrhundert.

#### Alte Drucke des 16. bis 18. Jahrhunderts

Die Epoche der sogenannten Alten Drucke begann in dem Moment, als sich das gedruckte Buch nach dem Übergangsmedium der Inkunabel von seinem kulturellen Vorbild, der Handschrift, vollends zu emanzipieren begann. Die systematische Dokumentation und Sammlung der deutsch(sprachig)en Schriftkultur beruht auf unterschiedlichen Säulen: Auf der seit 1969 als Druckausgabe erstellten retrospektiven Nationalbibliografie für Drucke der Jahre 1501–1600<sup>47</sup> bauten seit 1996 die retrospektive Nationalbibliografie für Drucke der Jahre 1601–1700 (VD17)<sup>48</sup> sowie seit 2009 die retrospektive Nationalbibliografie für Drucke der Jahre 1701–1800 (VD18)<sup>49</sup> auf (vgl. Kap. 2).

Dieses System ist Grundlage der Verteilten deutschen Nationalbibliothek. Sie wird koordiniert durch die *Arbeitsgemeinschaft Sammlung Deutscher Drucke (AG SDD)*<sup>50</sup>, dem 1989 gegründeten Zusammenschluss der sechs bedeutendsten deutschen Altbestandsbibliotheken, die gemeinsam das gedruckte deutsche Kulturerbe sammeln, erschließen, digitalisieren und bewahren:

- 1450–1600: Bayerische Staatsbibliothek, München
- 1601–1700: Herzog August Bibliothek, Wolfenbüttel
- 1701–1800: Niedersächsische Staats- und Universitätsbibliothek, Göttingen
- 1801–1870: Universitätsbibliothek Johann Christian Senckenberg, Frankfurt am Main

---

<sup>47</sup> <https://bvbat01.bib-bvb.de/TP61/start.do?View=vd16&SearchType=2>

<sup>48</sup> <http://www.vd17.de/>

<sup>49</sup> <https://vd18.gbv.de/viewer/index/>

<sup>50</sup> [https://www.ag-sdd.de/Webs/agsdd/DE/Home/home\\_node.html](https://www.ag-sdd.de/Webs/agsdd/DE/Home/home_node.html)

- 1871–1912: Staatsbibliothek zu Berlin – Preußischer Kulturbesitz
- 1913ff.: Deutsche Nationalbibliothek, Leipzig und Frankfurt am Main

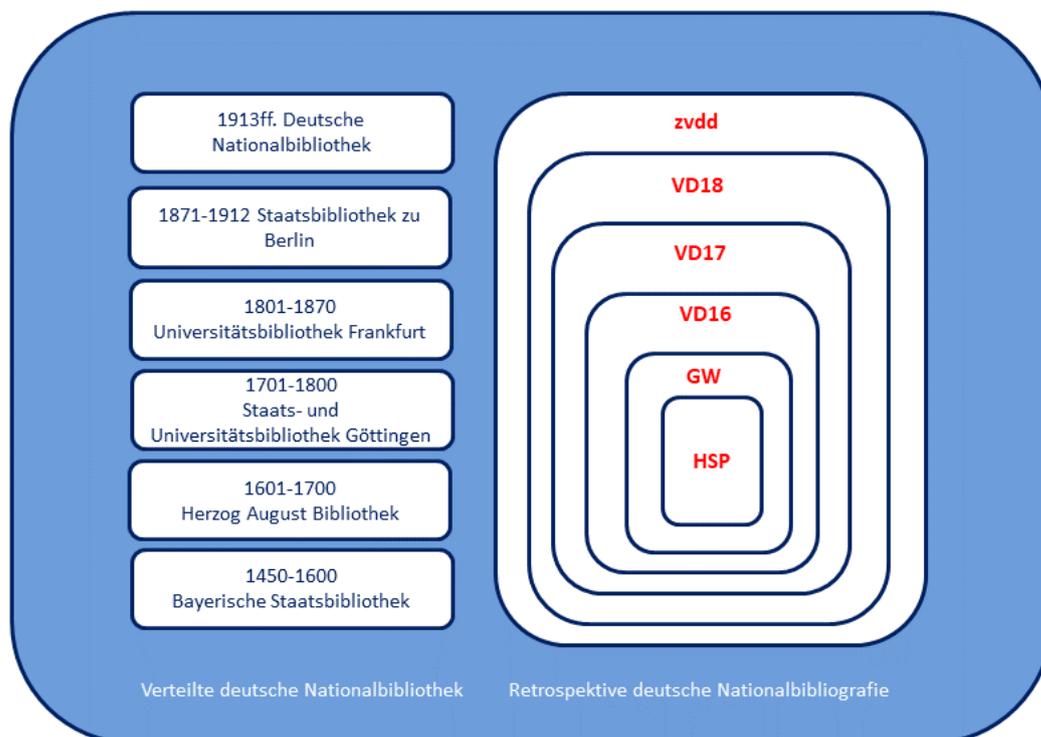


Abbildung 4. Verteilte deutsche Nationalbibliothek und retrospektive deutsche Nationalbibliografie

### Verteilte nationale Forschungsbibliothek – Sondersammelgebiete (SSGs) und Fachinformationsdienste (FIDs)

Die Idee einer Verteilten nationalen Forschungsbibliothek<sup>51</sup> entstand nach dem Zweiten Weltkrieg, als Studium, Forschung und Lehre an den Universitäten vor dem Hintergrund begrenzter Ressourcen effizient wieder aufgenommen werden sollten. Bereits damals war es keiner einzelnen deutschen Bibliothek möglich, die national und international publizierte Forschungsliteratur auch nur annähernd umfassend zu erwerben. Daher wurde 1949 das kooperativ organisierte Modell unterschiedlicher *Sondersammelgebiete*

<sup>51</sup> Ein Überblick über die im Folgenden besprochenen Sondersammelgebiete (SSGs) und Fachinformationsdienste (FIDs) findet sich auf der Seite Webis – Sammelschwerpunkte an deutschen Bibliotheken ([https://wikis.sub.uni-hamburg.de/webis/index.php/Verteilte\\_nationale\\_Forschungsbibliothek](https://wikis.sub.uni-hamburg.de/webis/index.php/Verteilte_nationale_Forschungsbibliothek)).

(SSGs)<sup>52</sup> etabliert, durch das führende wissenschaftliche Bibliotheken mit Unterstützung der DFG sicherstellten, dass die relevanteste internationale Forschungsliteratur in Deutschland in mindestens einem gedruckten Exemplar zur Verfügung stand. Dieses konnte im Rahmen der überregionalen Literaturversorgung per Fernleihe deutschlandweit ausgeliehen werden. Die SSGs waren nicht nur printbasiert, sondern verfolgten auch eine forschungsunabhängige Erwerbungspolitik, um auch diejenigen Publikationen zu akquirieren, die erst künftiger Forschung dienen könnten.

Angesichts der digitalen Transformation in Wissenschaft und Bibliothek musste sich das etablierte kooperative System der überregionalen Literaturversorgung grundlegend wandeln. Die SSGs wurden seit 2015 zu *Fachinformationsdiensten (FIDs)*<sup>53</sup> weiterentwickelt, die zwar weiterhin die lokalen Bibliotheken an Universitäten und Forschungseinrichtungen durch zentrale Erwerbung gedruckter Literatur ergänzen, die aber insbesondere auf die ortsunabhängige Versorgung der Wissenschaft mit digitalen Forschungsressourcen und Dienstleistungen auch für aktuelle Forschungstrends zielen. Die geschichtswissenschaftlichen SSGs stellen inzwischen ein reiches Portfolio von Volltextdatenbanken für Forschungsliteratur – und für Quellen – zur Verfügung.

---

<sup>52</sup> [https://wikis.sub.uni-hamburg.de/webis/index.php/Kategorie:Ehemalige\\_SSGs](https://wikis.sub.uni-hamburg.de/webis/index.php/Kategorie:Ehemalige_SSGs)

<sup>53</sup> [https://wikis.sub.uni-hamburg.de/webis/index.php/Fachinformationsdienste\\_für\\_die\\_Wissenschaft](https://wikis.sub.uni-hamburg.de/webis/index.php/Fachinformationsdienste_für_die_Wissenschaft)

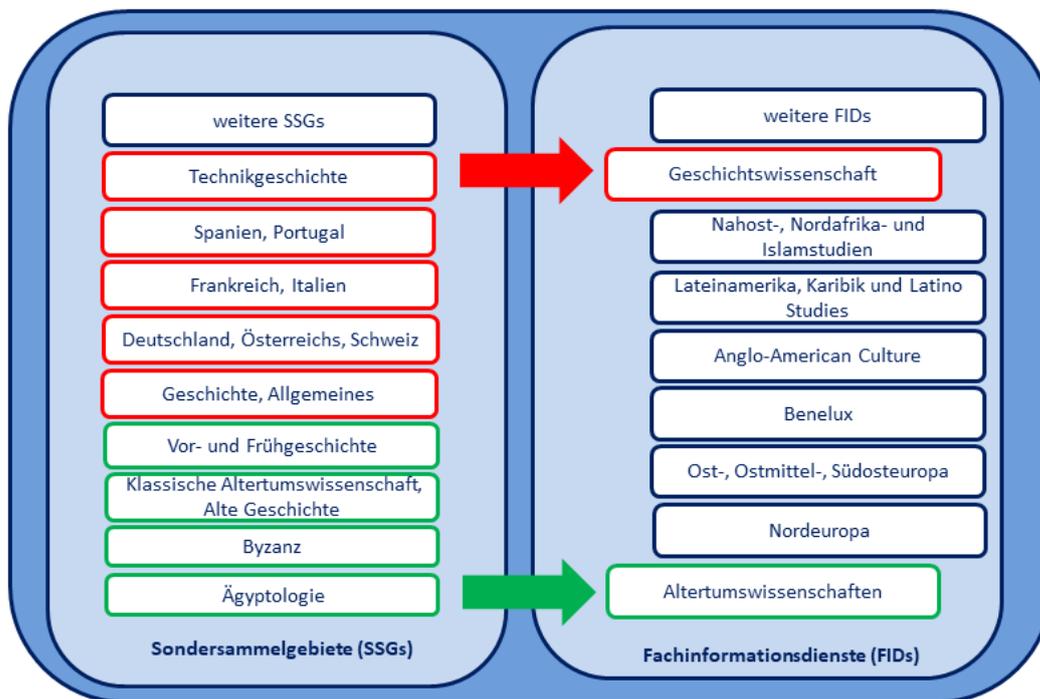


Abbildung 5. Verteilte nationale Forschungsbibliothek – Beispiele für die Umwandlung geschichtswissenschaftlicher SSGs zu FIDs

### 1.3 Digitalisierung und Volltexttransformation von Handschriften, Drucken und Forschungsliteratur – Institutionen, Projekte, Forschungsförderung

Im Folgenden wird die Entwicklung der Volltexterkennung (Kap. 1.1) in Bezug auf die Konzepte der Verteilten deutschen Nationalbibliothek und der Verteilten nationalen Forschungsbibliothek (Kap. 1.2) für historische Volltextdatenbanken für Quellen und Forschungsliteratur beschrieben. Dabei werden zunächst zentrale Institutionen vorgestellt, die relevante Projekte zu Digitalisierung, Volltexterkennung, Volltextdatenbanken und Digitalen Editionen tragen.

Die Philosophische Fakultät der Universität zu Köln gehört zu den führenden Standorten im Bereich digitale Geschichtswissenschaften und DH in Deutschland: Hier wurde in der Tradition der historischen Fachinformatik 1997 die Professur für *Historisch-Kulturwissenschaftliche Informationsverarbeitung* (HKI) eingerichtet, die von Manfred Thaller geprägt wurde, der auch den Prozess der Digitalisierung des schriftlichen Kulturerbes an Bibliotheken beglei-

tete. Das *Institut für Digital Humanities (IDH)*<sup>54</sup> widmet sich der historisch-kulturwissenschaftlichen und sprachlichen Informationsverarbeitung und arbeitet mit dem *Data Center für Digital Humanities (DCH)*<sup>55</sup> zusammen, das Geisteswissenschaftlerinnen und Geisteswissenschaftler bei der Sicherung, Verfügbarkeit und Präsentation von Forschungsdaten und -ergebnissen berät. Es ist darüber hinaus an geisteswissenschaftlichen Konsortien der *Nationalen Forschungsdateninfrastruktur (NFDI)*<sup>56</sup> beteiligt. Beide Institutionen sind Teil des 2009 gegründeten *Cologne Center for eHumanities (CCeH)*<sup>57</sup>, das als Kompetenzzentrum für DH-Projekte über Köln hinauswirkt.

Mit der Kölner Schule Thallers verbunden ist Patrick Sahle, der mit zahlreichen Publikationen und Projekten der DH<sup>58</sup> hervorgetreten ist und diesen Forschungsbereich an der *Bergischen Universität Wuppertal*<sup>59</sup> vertritt.

Wie Köln besitzt das 1998 gegründete *Kompetenzzentrum – Trier Center for Digital Humanities*<sup>60</sup> eine reiche Tradition. Mit diesem kooperiert das Fach *Computerlinguistik und Digital Humanities*<sup>61</sup> im Fachbereich II Sprach-, Literatur- und Medienwissenschaften der Universität.

Neben Köln, Wuppertal und Trier besitzt die Westfälische Wilhelms-Universität Münster einen *etablierten Schwerpunkt im Bereich der DH*<sup>62</sup>. Während das *Center for Digital Humanities (CDH)*<sup>63</sup> einen Interessensverbund von digital Forschenden der Fachbereiche

---

<sup>54</sup> <https://dh.phil-fak.uni-koeln.de/>

<sup>55</sup> <https://dch.phil-fak.uni-koeln.de/>

<sup>56</sup> <https://www.nfdi.de/>

<sup>57</sup> <https://cceh.uni-koeln.de/>

<sup>58</sup> Sahle, Patrick, *Digitale Editionsformen: zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bände, Norderstedt 2013.

<sup>59</sup> <https://www.geschichte.uni-wuppertal.de/de/lehrgebiete/digital-humanities/>

<sup>60</sup> <https://tcdh.uni-trier.de/de>

<sup>61</sup> <https://www.uni-trier.de/universitaet/fachbereiche-faecher/fachbereich-ii/faecher/computerlinguistik-und-digital-humanities/home>

<sup>62</sup> <https://www.uni-muenster.de/DH/>

<sup>63</sup> <https://www.uni-muenster.de/DH/cdh/index.html>

---

Geisteswissenschaften und Informatik bildet, bietet das *Service Center for Digital Humanities (SCDH)*<sup>64</sup> konkrete Unterstützung bei Planung und Durchführung einschlägiger Projekte. Die institutionelle Einbindung des SCDH in die *Universitäts- und Landesbibliothek Münster*<sup>65</sup> zeigt die enge Verbindung zwischen Bibliothek und Fachwissenschaften auf dem Feld der DH.

Aus einer ähnlichen Verbindung ist das Würzburger *Zentrum für Philologie und Digitalität „Kallimachos“ (ZPD)*<sup>66</sup> hervorgegangen. Als zentrale wissenschaftliche Einrichtung der Universität verbindet es Geisteswissenschaften, Informatik und DH in hervorragender Weise, wie die hier entstandenen digitalen Editionen und Volltextdatenbanken zeigen.<sup>67</sup>

Ein Beispiel für ein außeruniversitäres Forschungsinstitut, das in DH-basierten Projekten über lange Erfahrungen verfügt, ist das *Leibniz-Institut für Bildungsmedien|Georg-Eckert-Institut (GEI)*<sup>68</sup>. Seine bis in das 17. Jahrhundert zurückgehende Sammlung historischer Schulbücher wird seit 2009 digitalisiert über die digitale Schulbuchbibliothek *GEI-Digital*<sup>69</sup> publiziert.

Auf der Digitalisierung der Verzeichnisse deutscher Drucke des 16., 17. und 18. Jahrhunderts innerhalb der Verteilten deutschen Nationalbibliothek (Kap.1.2) setzt seit 2014 das Projekt *OCR-D*<sup>70</sup> auf.<sup>71</sup> Da große Teile der Nationalbibliografien VD16, VD17 und VD18 inzwischen mit Volldigitalisaten angereichert sind, ist durch

---

<sup>64</sup> <https://www.uni-muenster.de/DH/scdh/index.html>

<sup>65</sup> <https://www.ulb.uni-muenster.de/>

<sup>66</sup> <https://www.uni-wuerzburg.de/zpd/startseite/>

<sup>67</sup> Narragonien digital. Digitale Edition von Ausgaben und Bearbeitungen des „Narrenschiffs“ vor 1500 (<https://www.narragonien-digital.de/exist/index.html>); Camerarius digital. Lexikon und Volltexterschließung der Werke von Joachim Camerarius d.Ä. (<http://www.camerarius.de/>).

<sup>68</sup> <https://www.gei.de/>

<sup>69</sup> <https://gei-digital.gei.de/viewer/index/>

<sup>70</sup> <https://ocr-d.de/>

<sup>71</sup> Zum technischen Hintergrund vgl. Engl, Elisabeth, OCR-D kompakt. Ergebnisse und Stand der Forschung in der Förderinitiative, in: *Bibliothek – Forschung und Praxis* 44 (2020) 2, S. 218–230, <https://doi.org/10.1515/bfp-2020-0024>.

die Entwicklung im Bereich OCR eine Volltexttransformation des gesamten gedruckten deutschen schriftlichen Kulturerbes möglich geworden, indem aus den als Bilddateien gespeicherten Digitalisaten durchsuchbare Textdateien erzeugt werden: Die wissenschaftliche Nutzbarkeit digitalisierter Drucke insbesondere im Kontext der DH setzt zwingend maschinenlesbare Volltexte voraus. Künftig sollen innerhalb der digitalen Verteilten deutschen Nationalbibliothek nicht nur umfassende Volltextsuchen, sondern auch differenzierte Analysen und Bearbeitungen der Textquellen mit Werkzeugen der DH im *distant reading* möglich sein.

Mit der Open Source *OCR-D-Software*, deren Prototyp 2020 fertiggestellt wurde, können Modelle für eine automatische Transkription von Texten entwickelt werden, die mit der Vielfalt der historischen Layoutvarianten, der Drucktypen, der Orthographie und der Sprache umgehen können.

Um Technologien der automatischen Texterkennung möglichst niederschwellig einsetzen zu können, arbeitet das *Zentrum für Philologie und Digitalität „Kallimachos“* (ZPD) gegenwärtig daran, in den Projekten *OCR4all*<sup>72</sup> und *OCR4all-libraries – Volltexterkennung historischer Sammlungen*<sup>73</sup> unterschiedliche freie OCR-Tools in einem standardisierten Workflow zusammenzuführen.

Sämtliche Projekte, die von universitären und außeruniversitären Instituten sowie von Bibliotheken konzipiert und realisiert werden, bedürfen einer zuverlässigen Finanzierung: Hierbei spielt der DFG-Förderbereich *Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS)*<sup>74</sup> eine herausragende Rolle. Die DFG-geförderten Projekte und Forschungsinitiativen zu Digitalisierung und Volltexterkennung als Grundlage für die Entwicklung von Volltextdatenbanken dokumentiert die Datenbank *GEPRIS*<sup>75</sup>.

Neben dem DFG-geförderten nationalen Projekt *OCR-D* wird die Forschung zum Thema Volltexterkennung auch auf Landesebene

---

<sup>72</sup> <https://www.ocr4all.org/>

<sup>73</sup> <https://gepris.dfg.de/gepris/projekt/460665940>

<sup>74</sup> <https://www.dfg.de/foerderung/programme/infrastruktur/lis/>

<sup>75</sup> <https://gepris.dfg.de/gepris>

öffentlich unterstützt: Das an den Universitätsbibliotheken Mannheim und Tübingen angesiedelte Projekt *OCR-BW*<sup>76</sup> beispielsweise hat zum Ziel, Bibliotheken, Universitäten und Forschende in Baden-Württemberg bei der Implementierung und Anwendung von automatischer Texterkennungs- und Transkriptionssoftware zu unterstützen.<sup>77</sup> Neben der Software OCR-D erforscht OCR-BW die Software-Lösungen und OCR-Engines *Tesseract*<sup>78</sup>, *OCRmyPDF*<sup>79</sup>, *Ocropus*<sup>80</sup>, *Kraken*<sup>81</sup>, *Calamari*<sup>82</sup>, *eScriptorium*<sup>83</sup> und *Transkribus*<sup>84</sup>.

#### 1.4 Publikationen, Verbände, Kommunikation

Zum Abschluss dieses Kapitels informieren wenige Hinweise darüber, wie man sich über aktuelle Themen, Projekte sowie wichtige Autorinnen und Autoren innerhalb des Fachdiskurses zu Volltextdatenbanken im Kontext digitaler Geschichtswissenschaften informieren kann. Relevant ist insbesondere die spezifische inhaltliche Perspektive: Handelt es sich um eine eher geschichtswissenschaftliche, informationswissenschaftliche oder bibliothekarische Fragestellung?

Aus geschichtswissenschaftlicher Sicht bietet der FID *historicum.net*<sup>85</sup> einen der wichtigsten Rechercheeinstiege auch für Methoden und Werkzeuge der digitalen Geschichtswissenschaften. Der FID *Buch-, Bibliotheks- und Informationswissenschaft (FID BBI)*<sup>86</sup>

---

<sup>76</sup> <https://ocr-bw.bib.uni-mannheim.de/>

<sup>77</sup> Huff, Dorothee; Stöbener, Kristina, Projekt OCR-BW: Automatische Texterkennung von Handschriften, in: O-Bib. Das Offene Bibliotheksjournal 9 (2022) 4, S. 1–19, <https://doi.org/10.5282/o-bib/5885>.

<sup>78</sup> <https://github.com/tesseract-ocr>

<sup>79</sup> <https://ocrmypdf.readthedocs.io/en/latest/index.html>

<sup>80</sup> <https://github.com/ocropus/ocropy>

<sup>81</sup> <https://github.com/mittagessen/kraken>

<sup>82</sup> <https://github.com/Calamari-OCR/calamari>

<sup>83</sup> <https://gitlab.com/scripta/escriptorium>

<sup>84</sup> <https://readcoop.eu/de/transkribus/>

<sup>85</sup> <https://www.historicum.net/home/>

<sup>86</sup> <https://katalog.fid-bbi.de/>

informiert über Spezialliteratur und Forschungsressourcen aus drei Kerndisziplinen, deren Anwendung auch in den digitalen Geschichtswissenschaften relevant sein können.

Während diese beiden Portale für die Recherche nach selbstständiger (Monografien) und unselbstständiger (Aufsätze) Literatur geeignet sind, bietet die *Zeitschriftendatenbank (ZDB)*<sup>87</sup> einen vollständigen Überblick über die periodisch erscheinende Literatur der betreffenden Disziplinen.

Ein Beispiel ist die *Zeitschrift für digitale Geisteswissenschaften (ZfdG)*<sup>88</sup>. Sie bietet einen Einblick in den interdisziplinären Fachdiskurs der mit digitalen Ressourcen, Methoden und Konzepten arbeitenden Geisteswissenschaften, zu denen die Geschichtswissenschaften gehören.

Ein anderes Beispiel ist *RIDE – A review journal for digital editions and resources*<sup>89</sup> mit einem Überblick über Konzepte und Technologien aktueller digitaler Editionen. In seiner Schriftenreihe (*SIDE*<sup>90</sup>) behandelt das *Institut für Dokumentologie und Editorik (IDE)*<sup>91</sup> die Anwendung innovativer Informationstechnologien für die Arbeit mit historischen Dokumenten und Texten. In Kooperation mit dem IDE und der Bergischen Universität Wuppertal wird von Patrick Sahle der *Catalog of Digital Scholarly Editions*<sup>92</sup> herausgegeben, der eine gezielte Recherche nach digitalen Editionen ermöglicht.

Aktuelle Informationen über die Entwicklung historischer Volltextdatenbanken bietet auch der *Blog*<sup>93</sup> des Verbandes *DHd – Digital Humanities im deutschsprachigen Raum*. Er entstand auf Initiative

---

<sup>87</sup> <https://zdb-katalog.de/index.xhtml>

<sup>88</sup> <https://www.zfdg.de/>

<sup>89</sup> <https://ride.i-d-e.de/>

<sup>90</sup> <https://www.i-d-e.de/publikationen/schriften/>

<sup>91</sup> <https://www.i-d-e.de/>

<sup>92</sup> [https://www.digitale-edition.de/exist/apps/editions-browser/\\$app/index.html](https://www.digitale-edition.de/exist/apps/editions-browser/$app/index.html)

<sup>93</sup> <https://dhd-blog.org/>

---

der Forschungsverbände *TextGrid*<sup>94</sup> und *DARIAH-DE*<sup>95</sup> sowie des Max-Planck-Instituts für Wissenschaftsgeschichte Berlin.

Neben dem in Fachjournalen und Wissenschaftsblogs schriftlich geführten Fachdiskurs ist der direkte persönliche Austausch in Fachverbänden entscheidend. Für den Fachdiskurs in Deutschland am wichtigsten ist der 2012 an der Universität Hamburg gegründete Verband *DHd – Digital Humanities im deutschsprachigen Raum*<sup>96</sup>. Dieser richtet jährliche Tagungen aus, ist Herausgeber der genannten *Zeitschrift für digitale Geisteswissenschaften (ZfdG)* und bearbeitet in zahlreichen Arbeitsgruppen zentrale Themen der DH – die *AG OCR*<sup>97</sup> beispielsweise Volltextdatenbanken. Innerhalb des *Verbands der Historiker und Historikerinnen Deutschlands (VHD)*<sup>98</sup> befasst sich die *Arbeitsgemeinschaft Digitale Geschichtswissenschaft*<sup>99</sup> mit den Themen DH, Forschungsdaten, digitale Methoden oder digitale Quellenkritik.

### 1.5 Von der Theorie in die Praxis – Nutzererwartungen an historische Volltextdatenbanken

Die vorausgegangene Skizze des State of the Art der Digitalisierung, digitalen Transformation und Methodenentwicklung in Bibliotheken und digitalen Geschichtswissenschaften zeigt, in welche Nutzungsszenarien historische Volltextdatenbanken eingebettet sind. *Historische Volltextdatenbanken für Quellen und Forschungsliteratur* werden als integrale Bestandteile offener, vernetzter Text- und Datenwelten verstanden, die über folgende Qualitäten verfügen sollten:

- Digitalisierung nach DFG-Standards.

---

<sup>94</sup> <https://textgrid.de/>

<sup>95</sup> <https://de.dariah.eu/>

<sup>96</sup> <https://dig-hum.de/>

<sup>97</sup> <https://dig-hum.de/ag-ocr>

<sup>98</sup> <https://www.historikerverband.de/>

<sup>99</sup> <https://www.historikerverband.de//arbeitsgruppen/ag-digitale-gw.html>

- Beschreibung der Digitalisate durch Metadaten im Format METS/MODS für den Datenaustausch über APIs.
- Downloadmöglichkeit der Digitalisate mitsamt Roh- und Metadaten auf Grundlage der Prinzipien des Open Access, (Linked) Open Data sowie der FAIR-Prinzipien (findable, accessible, interoperable, reusable).
- Volltextindexierung und Volltexterkennung (HTR, OCR) für quantitative Volltextanalysen im *distant reading*, die ein qualitatives *close reading* auf eine neue Grundlage stellen.
- Publikation der Digitalisate im Open Access oder unter möglichst offenen Lizenzen zur Nachnutzbarkeit in Open Science-Szenarien.
- Referenzierbarkeit der Digitalisate durch stabilen Uniform Resource Name (URN) oder Uniform Resource Identifier (URI).
- Möglichkeit der Einbindung in virtuelle Forschungsumgebungen und (kollaborativen) Weiterverarbeitung der Digitalisate mit quantitativen und qualitativen Forschungsansätzen durch Tools und Methoden der DH (z.B. Annotationen, Visualisierungen).
- Möglichkeit vergleichenden Arbeitens, z. B. durch Verwendung des *International Image Interoperability Framework (IIIF)*, um Digitalisate von Texten und Objekten institutionsübergreifend austauschen und standortunabhängig in unterschiedlichen Viewern präsentieren sowie mit Bildbearbeitungs- und Annotationstools bearbeiten zu können.
- Auffindbarkeit durch bibliothekarisch-formale und wissenschaftlich-intellektuelle Erschließung mit geeigneten normierten Metadaten.
- Vernetzung von Quellen und Forschungsliteratur miteinander.

Speziell für Volltextdatenbanken für *historische Originalquellen* wie Texte oder kulturelle Artefakte empfiehlt sich:

- Erschließung durch wissenschaftliche Beschreibungen (z.B. Kataloge).

- Verknüpfung mit Editionen (gedruckt, digital).
- Verknüpfung mit Forschungsliteratur (gedruckt, digital).
- Kontextualisierung mit weiteren relevanten Texten und Objekten (gedruckt, digital) in Fach- und Kulturportalen.

Speziell für Volltextdatenbanken für *digitalisierte und genuin digitale Forschungsliteratur* empfiehlt sich:

- Erschließung durch bibliothekarische Metadaten und Normdaten.
- Erschließung durch Thesauri, Fachklassifikationen oder intellektuelle Inhaltserschließung.
- Verknüpfung über Linkresolver mit lokalen Bibliothekskatalogen.
- Verknüpfung mit digitalisierten historischen Originalquellen.

Welche dieser Nutzererwartungen historische Volltextdatenbanken bereits erfüllen und wo weiterhin Entwicklungsbedarf besteht, wird im folgenden Kapitel exemplarisch gezeigt.

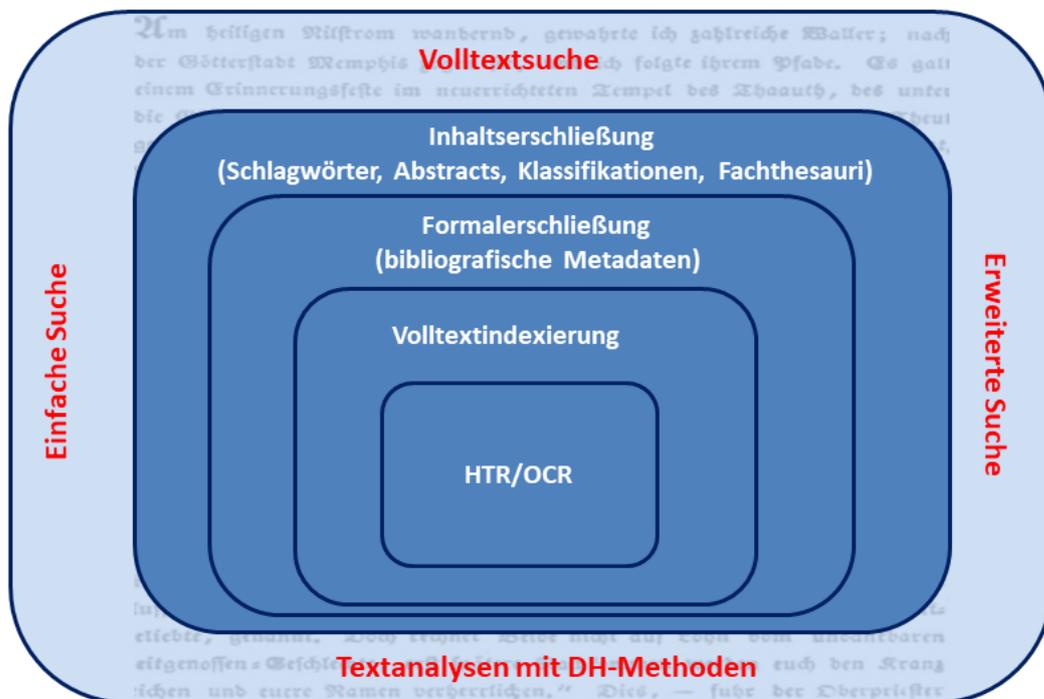


Abbildung 6. Möglichkeiten der Erschließung und Nutzung von digitalisierten Texten und Quellen in Volltextdatenbanken

## 2. Historische Volltextdatenbanken und digitale Ressourcen

### 2.1 Infrastrukturen und Zugänglichkeit

Eine auch nur annähernd vollständige Bestandsaufnahme historischer Volltextdatenbanken kann im Rahmen dieses Guides nicht geleistet werden. Der praxisorientierte Überblick ist vor dem Hintergrund der Konzepte der Verteilten deutschen Nationalbibliothek und der Verteilten nationalen Forschungsbibliothek (Kap. 1.2) untergliedert nach Volltextdatenbanken für Forschungsliteratur (Kap. 2.5) einerseits, für Quellen und Quelleneditionen (Kap. 2.6) andererseits. Hinter beiden Konzepten stehen Institutionen, die mit konkreten historischen Volltextdatenbanken verbunden sind. Unterschieden wird ferner zwischen umfassenden Portalen und einzelnen Volltextdatenbanken. Die folgende Übersicht präsentiert schwerpunktmäßig Volltextdatenbanken für Quellen nach historischen Epochen, Regionen und Themen. Die diesen Guide ergänzende Linkliste beinhaltet weitere ausgewählte Ressourcen.

Wie in der Einführung (Kap. 1.1) beschrieben, sind historische Volltextdatenbanken in unterschiedlichen Kontexten entstanden und durch öffentliche Wissenschaftseinrichtungen und kommerzielle Verlage gleichermaßen geprägt: Bibliotheken beispielsweise bieten einerseits in Volltextdatenbanken digitalisierte oder genuin digitale Texte und Digitalisate von Quellen mitsamt Metadaten im *Open Access* an, sie finanzieren andererseits durch unterschiedliche Lizenzierungsmodelle den Zugang zu kommerziellen digitalen Produkten von Verlagen im *Closed Access*. Während Universitätsbibliotheken diese Ressourcen allein oder konsortial ausschließlich für ihre eigenen Forschenden, Lehrenden und Studierenden lizenzieren, ermöglichen Landes- und Staatsbibliotheken diese Zugänge für eine wissenschaftlich interessierte Öffentlichkeit außerhalb von Wissenschaftsinstitutionen.

Sowohl wissenschaftliche wie auch private Nutzerinnen und Nutzer in Deutschland profitieren von den Nationallizenzen. Die DFG finanzierte 2004 – 2010 den Erwerb von Lizenzen, um Studierenden, Wissenschaftlern und der interessierten Öffentlichkeit den freien Zugriff auf kostenpflichtige elektronische Verlagsprodukte

---

zu ermöglichen. Die Nationallizenzen wurden 2011 durch Allianz-Lizenzen abgelöst, für die sich Bibliotheken in Konsortien zusammenfanden, um ausgewählte Datenbanken zu lizenzieren.

Etwas komplexer ist die Ablösung der Sondersammelgebiete (SSGs) durch Fachinformationsdienste (FIDs): Während auf die gedruckte Literatur der SSGs jedermann zugreifen konnte, sind die FID-Lizenzen nur für fachlich definierte Communities zugänglich (Kap. 1.2). Für die Geschichtswissenschaften sind mehrere Bibliotheken als Träger von Fachinformationsdiensten relevant, beispielsweise die Bayerische Staatsbibliothek für die Alte Geschichte, die Geschichte Deutschlands, Österreichs, der Schweiz, Frankreichs, Italiens und die Technikgeschichte in Kooperation mit dem Deutschen Museum (Abb. 5).

Die Situation, dass digitale Forschungsressourcen wie historische Volltextdatenbanken unterschiedlichen Zugangsmodalitäten unterliegen, bestimmt ihre Nutzung grundlegend und erfordert eine souveräne Orientierung auf diesem Feld. Da seit Beginn der Digitalisierung und der digitalen Transformation die gesellschaftliche Relevanz, die der freie Zugang zu wissenschaftlichen Informationen besitzt, immer deutlicher wurde, liegt in der weiteren Transformation hin zum Open Access eine der großen Herausforderungen der Zukunft.

### 2.2 Verständnis und Definitionen

So heterogen die analogen Vorläufer der historischen Volltextdatenbanken sind, so wenig eindeutig, ist eine Definition des Begriffs „Volltextdatenbank“. Aus geschichtswissenschaftlicher Praxis und informationstechnologischer sowie bibliothekarischer Theorie lässt sich dennoch ein gemeinsames Kernverständnis von Volltextdatenbanken herleiten. Im Allgemeinen sind Volltextdatenbanken Sammlungen elektronischer Volltexte mit bibliografischen und weiteren Metadaten. Historische Volltextdatenbanken im Speziellen sind Datenbanken, in denen für die Geschichtswissenschaften relevante einzelne Texte oder Sammlungen von Quellentexten, Quelleneditionen und Forschungsliteratur präsentiert werden. Als Volltextdatenbanken können aber auch Datenbanken für nicht-tex-

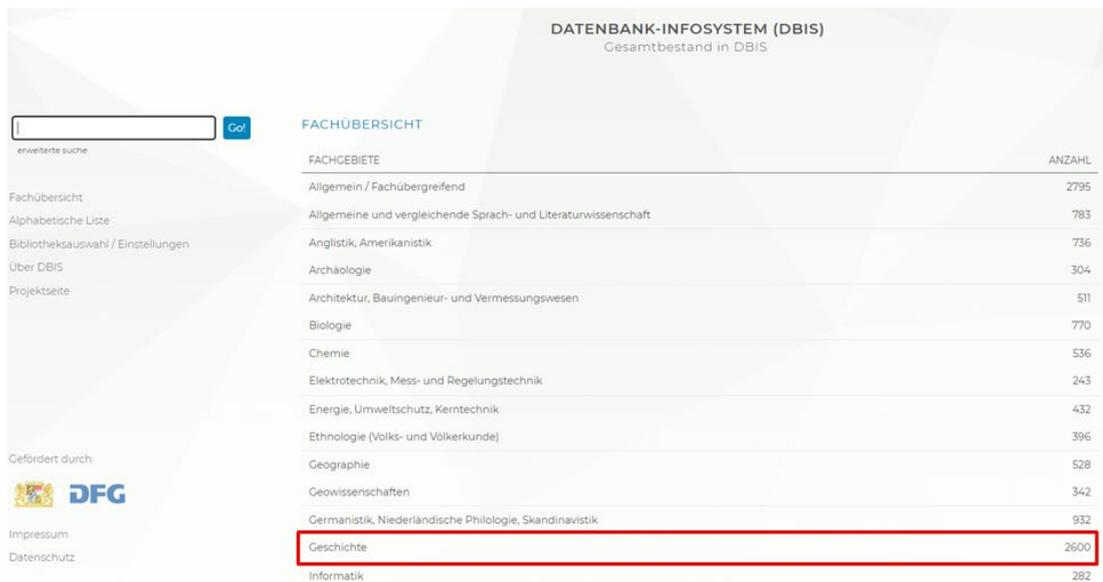
tuelle Quellen, Objekte und kulturelle Artefakte verstanden werden, sofern diese Textträger sind und von Transkriptionen begleitet werden. Beispiele hierfür sind Texte auf unterschiedlichen Trägern wie Stein, Papyrus, Pergament oder historische Karten. Dabei ist es unerheblich, ob die Texte lediglich als Images digitalisiert sind oder ob durch eine implementierte OCR-Erkennung eine Volltextsuche möglich ist.

Diese weite Definition möchte dem *material turn* innerhalb der Geschichtswissenschaften gerecht werden: (Text-) Objekte sind genauso wichtig wie reine Texte, denn auch sie vermögen durch ihr narratives Potenzial Geschichte zu erzählen – wenn sie von Historikerinnen und Historikern angemessen entziffert werden.

Objekte, bei denen Texte keinerlei Rolle spielen, werden hingegen in der Regel in Bilddatenbanken erfasst und spielen hier keine Rolle.

## 2.3 Das Datenbank-Infosystem (DBIS)

Das *Datenbank-Infosystem (DBIS)*<sup>100</sup> ist das wichtigste Verzeichnis wissenschaftlicher Datenbanken im deutschen Sprachraum.



DATENBANK-INFOSYSTEM (DBIS)  
Gesamtbestand in DBIS

erweiterte suche

FACHÜBERSICHT

FACHGEBIETE	ANZAHL
Allgemein / Fachübergreifend	2795
Allgemeine und vergleichende Sprach- und Literaturwissenschaft	783
Anglistik, Amerikanistik	736
Archäologie	304
Architektur, Bauingenieur- und Vermessungswesen	511
Biologie	770
Chemie	536
Elektrotechnik, Mess- und Regelungstechnik	243
Energie, Umweltschutz, Kerntechnik	432
Ethnologie (Volks- und Völkerkunde)	396
Geographie	528
Geowissenschaften	342
Germanistik, Niederländische Philologie, Skandinavistik	932
<b>Geschichte</b>	<b>2600</b>
Informatik	282

Fachübersicht  
Alphabetische Liste  
Bibliotheksauswahl / Einstellungen  
Über DBIS  
Projektseite

Gefördert durch  


Impressum  
Datenschutz

Abbildung 7. Datenbank-Infosystem (DBIS) – Fachübersicht (<https://dbis.ur.de/fachliste.php?lett=l>, 16.07.2024)

<sup>100</sup> <https://dbis.ur.de/fachliste.php?lett=l>

So heterogen die in DBIS verzeichneten Datenbanken sind, so gibt es doch formale Kriterien für die Aufnahme: Umfang der Datenmenge, langfristige Verfügbarkeit, kontinuierliche Pflege und Aktualisierung, Wissenschaftlichkeit, Seriosität der Inhalte und Herausgeber. In DBIS nicht aufgenommen werden Linklisten, Literaturlisten im HTML- oder PDF-Format, einzelne E-Books oder E-Journals sowie Bibliotheks-OPACs. DBIS ermöglicht sowohl eine bibliotheksübergreifende als auch eine lokale Sicht auf die Datenbanken sämtlicher Wissenschaftsdisziplinen und ihre Zugangsmodalitäten. Unter den etwa 2.600 Datenbanken allein für das Fach Geschichte finden sich knapp 1.060 historische Volltextdatenbanken – ein deutlicher Beleg für die kontinuierlich steigende Bedeutung von Volltextdatenbanken für das geschichtswissenschaftliche Arbeiten.

Unter die in DBIS als „Volltextdatenbank“ klassifizierten Produkten werden inhaltlich und formal heterogene Angebote subsumiert. DBIS versteht als „Volltextdatenbank“ eine „Datenbank jeglicher Art mit direkten Zugriffen auf Volltexte.“ Entsprechend dieser weiten Definition werden auch Portale, in denen Volltexte nur einen Teil des Angebotes darstellen, oder Datenbanken, die keine Texte, sondern audiovisuelle Medien beinhalten, subsumiert. Des Weiteren sind Vollständigkeit der Texte sowie ihre Erschließung durch Abstracts oder Schlagwörter zentrale Kriterien. In DBIS firmieren als „Volltextdatenbanken“ sowohl Datenbanken mit durch Metadaten erschlossenen Texten, die lediglich als Image-Digitalisate vorliegen, als auch Datenbanken mit digitalisierten Texten, die mit OCR-Software bearbeitet wurden und echte Volltextsuchen ermöglichen.

Ob in den von DBIS als Volltextdatenbanken für Quellen, Quelleneditionen und Forschungsliteratur klassifizierten Produkten eine Volltextsuche in den mit HTR/OCR bearbeiteten Bilddigitalisaten von Originalquellen (Inschriften, Handschriften, Inkunabeln, Drucke) oder digitalisierter Forschungsliteratur möglich ist, ist produktabhängig. Beispielsweise werden die Verzeichnisse der im deutschen Sprachraum erschienenen Drucke (VD16, VD17 und VD18) als „National- bzw. Regionalbibliografie“ klassifiziert, im

Fall des VD17 jedoch zusätzlich als „Volltextdatenbank“. Eine Volltextsuche ist im VD17 jedoch in Zukunft erst dann möglich, wenn die Volltexttransformation der VD's im Rahmen des Projektes OCR-D insgesamt realisiert ist. Ein anderes Beispiel ist die Datenbank *IKAR-Landkartendrucke vor 1850*<sup>101</sup>, die als Fachbibliografie und zugleich als Volltextdatenbank klassifiziert ist, obwohl keine Volltextsuche in den Karten selbst möglich ist. Diese Beispiele zeigen, dass die Klassifizierung „Volltextdatenbank“ in DBIS uneinheitlich ist und keinen Rückschluss erlaubt, ob im Einzelfall tatsächlich eine Volltexterkennung durchgeführt wurde.

Mit der „erweiterten Suche“ lassen sich in DBIS gezielt Datenbank-Typen für unterschiedliche Fächer auswählen. Aus der Produktübersicht wird deutlich, unter welchem Lizenztyp eine Datenbank zur Verfügung steht: eine Ampelsymbolik kennzeichnet die unterschiedlichen Zugangsarten.

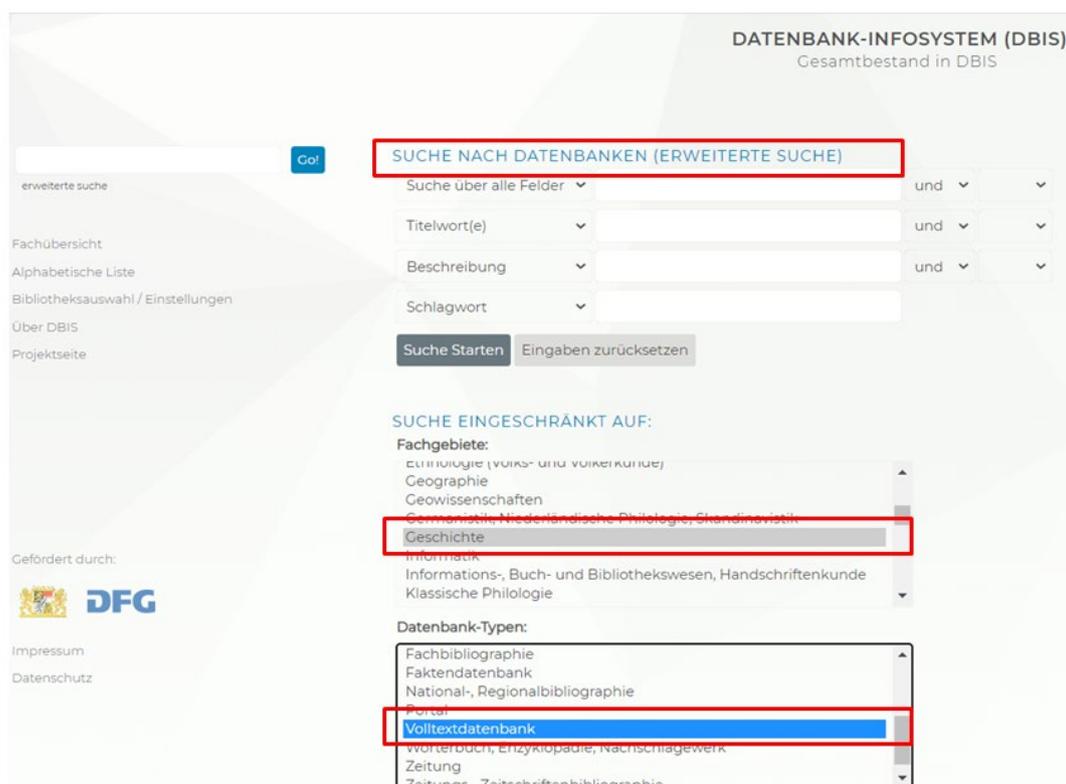


Abbildung 8. Datenbank-Infosystem (DBIS) – erweiterte Suche ([https://dbis.ur.de//suche.php?bib\\_id=alle&colors=3&ocolors=40](https://dbis.ur.de//suche.php?bib_id=alle&colors=3&ocolors=40), 16.07.2024)

<sup>101</sup> [https://dbis.ur.de//detail.php?bib\\_id=alle&colors=&ocolors=&lett=fs&tid=0&titel\\_id=1549](https://dbis.ur.de//detail.php?bib_id=alle&colors=&ocolors=&lett=fs&tid=0&titel_id=1549)



Abbildung 9. Datenbank-Infosystem (DBIS) – Zugangsarten und Ampelsymbolik  
([https://dbis.ur.de/dbliste.php?bib\\_id=ubfre&colors=31&ocolors=40&lett=a](https://dbis.ur.de/dbliste.php?bib_id=ubfre&colors=31&ocolors=40&lett=a), 16.07.2024)

Nahezu sämtliche der im Folgenden besprochenen Produkte sind in DBIS verzeichnet, inhaltlich beschrieben und im Hinblick auf die institutionelle Lizenzierung definiert. Daher empfiehlt sich DBIS als erster Zugriff auf Volltextdatenbanken im Recherchealltag. Um die Lektüre dieses Überblicks zu erleichtern, wird sowohl in den Fußnoten als auch in der separaten Liste relevanter Volltextdatenbanken jedoch die direkte Verlinkung auf die Ressource präferiert.

Neben der Fachsicht innerhalb von DBIS empfiehlt sich für Historikerinnen und Historiker der Einstieg in die Fachinformationsrecherche über die jeweiligen FIDs als Teile der Verteilten nationalen Forschungsbibliothek. Diese tragen dafür Sorge, dass die relevanten Forschungsressourcen, zu denen auch die Volltextdatenbanken gehören, deutschlandweit zur Verfügung stehen. Hierdurch ergänzen sie das System der Nationallizenzen. Auch wenn die FID-Lizenzen für die Geschichtswissenschaften in DBIS erfasst sind, finden sich auf den Portalen der einzelnen FIDs vielfältige weitere Fachinformationen.

### 2.4 Kulturportale und digitale Bibliotheken

Für die Geschichtswissenschaften spielen Kulturportale sowie digitale Bibliotheken, die das kulturelle und wissenschaftliche Erbe öffentlich zugänglich machen, eine wichtige Rolle. Eine digitale Bibliothek verfolgt grundsätzlich dieselben Ziele wie eine papierene Bibliothek: Sammlung, Strukturierung, Erschließung und Archivierung von digital(isiert)en textuellen und kulturellen Inhalten.

Digitale Bibliotheken können zugleich Virtuelle Bibliotheken sein, die über keine eigenen Bestände verfügen, sondern als elektronisches Informationssystem Metadaten und Volltexte aus unterschiedlichen Quellen unter einer einheitlichen Oberfläche zusammenführen und recherchierbar machen. Im Folgenden werden die wichtigsten regionalen, nationalen und internationalen Kulturportale und digitale Bibliotheken mit dem Schwerpunkt auf historischen Schriftquellen vorgestellt.

*Europeana*<sup>102</sup>, das Kulturportal Europas, setzt auf den nationalen Kultur- und Wissenschaftsinstitutionen mit ihren digitalisierten Sammlungen auf. Da die Erschließung heterogener Quellen unterschiedlicher Institutionen nicht immer homogen ist und nicht sämtliche Metadaten aus den Ursprungsdatenbanken in *Europeana* abgebildet werden, empfiehlt sich für qualitativ hochwertige Rechercheergebnisse immer auch eine differenzierte Suche in den Nachweissystemen der bestandshaltenden Institutionen selbst.

Die *Deutsche Digitale Bibliothek (DDB)*<sup>103</sup> ist bei der *Deutschen Nationalbibliothek (DNB)*<sup>104</sup> und der *Stiftung Preußischer Kulturbesitz*<sup>105</sup> angesiedelt. Als nationales Portal weist sie das kulturelle Erbe deutscher Gedächtnisinstitutionen – beispielsweise Bücher, Archivalien, Bilder, Skulpturen, Tondokumente, Filme, Noten – zentral nach und stellt diese als nationaler Aggregator der *Europeana* zur Verfügung. Wie in der *Europeana* stehen in der DDB neben bibliografischen Metadaten auch Volltexte und Digitalisate zur Verfügung, die jedoch nicht immer für eine Volltexterkennung aufbereitet sind. Innerhalb der DDB weist das *Archivportal-D*<sup>106</sup> digitalisiertes Archivgut, Findbücher und Informationen zu deutschen Archiven zentral nach.

Die DDB baut ihrerseits auf den Kulturportalen der Bundesländer auf, beispielsweise:

---

<sup>102</sup> <https://www.europeana.eu/de>

<sup>103</sup> <https://www.deutsche-digitale-bibliothek.de/>

<sup>104</sup> [https://www.dnb.de/DE/Home/home\\_node.html](https://www.dnb.de/DE/Home/home_node.html)

<sup>105</sup> <https://staatsbibliothek-berlin.de/>

<sup>106</sup> <https://www.archivportal-d.de/?>

- Baden-Württemberg: *LEO-BW – Landeskunde entdecken online*<sup>107</sup>.
- Bayern: *bavarikon – Kultur und Wissensschätze Bayerns*<sup>108</sup>.
- Hessen: *LAGIS – Landesgeschichtliches Informationssystem*<sup>109</sup>.
- Niedersachsen: *Kulturerbe Niedersachsen*<sup>110</sup>.
- Sachsen: *Sachsen.digital*<sup>111</sup>.

Insbesondere für die Regionalgeschichte sind diese Landesportale von großer Bedeutung, da sie häufig auch über unterschiedliche Themenportale verfügen.

Wie die DNB bieten auch Nationalbibliotheken anderer Länder übergreifende Portale und spezifische Angebote, in denen für die eigene nationale Geschichte relevante Quellen und Forschungsliteratur digitalisiert und in Volltextdatenbanken veröffentlicht werden. Für die historische Forschung lassen sich hier vielfältige Entdeckungen machen.

Nach dem Vorbild der *Europeana* entstand in den USA die *Digital Public Library of America (DPLA)*<sup>112</sup> mit Digitalisaten von Kulturgütern aus Bibliotheken, Archiven und Museen der USA unter freien Lizenzen.

Die größte und bedeutendste Bibliothek der USA, die *Library of Congress (LoC)*<sup>113</sup> in Washington, präsentiert in ihren *Digital Collections*<sup>114</sup> zentrale Dokumente der US-amerikanischen Geschichte, darunter Handschriften, seltene historische Drucke und Bücher, Zeitungen, Karten, Noten, Ton- und Filmdokumente aus sämtlichen Bundesstaaten.

---

<sup>107</sup> <https://www.leo-bw.de/>

<sup>108</sup> <https://www.bavarikon.de/>

<sup>109</sup> <https://www.lagis-hessen.de/>

<sup>110</sup> <https://kulturerbe.niedersachsen.de/start/>

<sup>111</sup> <https://sachsen.digital/>

<sup>112</sup> <https://dp.la/>

<sup>113</sup> <https://www.loc.gov/>

<sup>114</sup> <https://www.loc.gov/collections/>

---

In Europa digitalisiert die *British Library* (BL) in London als größte Bibliothek der Welt nach der *Library of Congress* (LoC) ihre reichen Sammlungen und präsentiert diese in ihren *Digital Collections*<sup>115</sup> – ein exzellenter Einstieg in die Recherche nach Schriftquellen wie Handschriften, Autographen, historischen Drucken, Büchern, Zeitungen, Karten, audiovisuellen Dokumenten und Forschungsliteratur zur Geschichte Englands und des Commonwealth.

Stellvertretend für die übrigen Nationalbibliotheken Europas sei die *Bibliothèque nationale de France* (BnF) in Paris genannt, die im Projekt *Gallica*<sup>116</sup> gemeinsam mit zahlreichen Partnerinstitutionen eine digitale Bibliothek für das Kulturerbe Frankreichs aufbaut, das zu großen Teilen bereits im Volltext durchsuchbar ist – ein ausgezeichnete Einstieg in die französische Geschichte.

Neben diesen regionalen, nationalen und internationalen Kulturportalen gibt es zahlreiche digitale Bibliotheken mit speziellen Volltextdatenbanken, die für historische Forschungen von Interesse sein können. Als älteste digitale Bibliothek der Welt gilt das 1971 begründete *Project Gutenberg*<sup>117</sup>, das zum Ziel hat, urheberrechtsfreie Bücher mit literarischen Texten weltweit zugänglich zu machen. Wurden am Beginn des Projekts ausgewählte Bücher von Freiwilligen manuell abgetippt und die Transkriptionen korrigiert, bevor diese im Internet veröffentlicht wurden, ermöglichte die kontinuierliche Verbesserung von Scan-Technik und Texterkennungsoftware ein immer schnelleres Wachstum dieser digitalen Bibliothek. Während im *Project Gutenberg* vorwiegend englischsprachige Bücher enthalten sind, bietet das Projekt *Gutenberg-DE*<sup>118</sup> gemeinfreie deutschsprachige literarische Werke.<sup>119</sup>

---

<sup>115</sup> <https://www.bl.uk/>

<sup>116</sup> <https://gallica.bnf.fr/accueil/fr/content/accueil-fr?mode=desktop>

<sup>117</sup> <https://www.gutenberg.org/>

<sup>118</sup> <https://www.projekt-gutenberg.org/>

<sup>119</sup> Zwischen 2018 und 2021 war *Project Gutenberg* in Deutschland gesperrt, bis die gerichtliche Auseinandersetzung mit dem S. Fischer Verlag bezüglich der Werke Thomas und Heinrich Manns sowie Alfred Döblins beigelegt wurde: Im Unterschied zu den urheberrechtlichen Bestimmungen der USA sind in

Unter den kommerziellen Anbietern digitaler Bibliotheken auch historischer Bücher, in denen Volltextsuchen möglich sind, ist *Google Books*<sup>120</sup> am bekanntesten. Mit seiner 1997 online gegangenen Suchmaschine wollte Google nicht nur die Informationen des World Wide Web erschließen. Die Idee des parallel entwickelten *Library Projects* war, das weltweite gedruckte kulturelle Erbe der „Gutenberg-Galaxis“ mit seinem darin niedergelegten menschlichen Wissen systematisch zu digitalisieren, mit OCR zu bearbeiten und eine globale Volltextsuche über sämtliche Textgattungen und Epochen zu spannen. Im Unterschied zum Ziel der Arbeitsgemeinschaft Sammlung Deutscher Drucke (AG SDD), das schriftliche Kulturerbe an deutschen Bibliotheken systematisch zu erfassen und zu digitalisieren, war Googles Ziel ein globales. Daher kooperierte der Internetkonzern einerseits mit kommerziellen Verlagen, andererseits mit wissenschaftlichen Bibliotheken weltweit – im deutschen Sprachraum mit der Österreichischen Nationalbibliothek Wien, mit der Bayerischen Staatsbibliothek München und mit der Staatlichen Bibliothek Regensburg. Google stellte die Digitalisate her – die Bibliotheken erhielten von Google eine digitale Kopie eines jeden gescannten Buches aus ihren Sammlungen.

Das von Google von Anfang an verfolgte Ziel ist jedoch nicht ohne Hindernisse zu erreichen – zu komplex ist die urheberrechtliche Situation im Falle der jüngeren Texte. Hierin liegt der Grund, dass lediglich ältere Bücher, deren urheberrechtlicher Schutz erloschen ist, frei zugänglich sind, während in allen anderen Fällen ein stark eingeschränkter Zugriff auf die Volltexte möglich ist: *Google Books* unterscheidet daher zwischen „Vollansicht“, „Eingeschränkte Vorschau“, „Auszugsansicht“ und „Keine Vorschau verfügbar“. Im Jahr 2019 gab *Google Books* anlässlich seines 15-jährigen Bestehens bekannt, mehr als 40 Millionen Bücher in mehr als 400 Sprachen digitalisiert zu haben.

Auf Grundlage dieses monumentalen Volltextcorpus lassen sich vielfältige quantitative Analysen durchführen. Mit dem *Google*

---

Deutschland Werke eines Autors 70 Jahre nach seinem Tod urheberrechtlich geschützt, bevor sie gemeinfrei werden.

<sup>120</sup> <https://books.google.de/?hl=de>

*Books Ngram Viewer*<sup>121</sup> beispielsweise lässt sich erkennen, wann ein bestimmter Begriff innerhalb von *Google Books* erstmals fassbar wird und wie sich seine Verwendung im Laufe der Zeit verändert hat.

Am Beispiel von *Google Books* werden zugleich die Grenzen einer wissenschaftlichen Recherche lediglich auf Grundlage einer Volltextsuche in Texten mit Volltextindexierung deutlich. Google hatte seine Suchmaschinentechnologie, die auf der Indexierung des World Wide Web beruhte, auf das digitalisierte gedruckte schriftliche Kulturerbe übertragen. Da eine systematische, bibliothekarische und wissenschaftliche Erschließung der einzelnen, von Google digitalisierten Bücher fehlt, muss eine Recherche in *Google Books* mit einer Stichwortsuche auskommen. Die gewählten Suchbegriffe sollten in unterschiedlichen Sprachen formuliert werden, um Zugriff auf die jeweils nationale Literatur zu erhalten.

Hier bieten Bibliotheken noch immer einen deutlichen Mehrwert: Sie verfügen über fachnahe Konzepte der qualitativen Strukturierung von Informationen und erschließen ihre Texte mit differenzierten Metadaten wie Schlagwörtern oder Fachthesauri, die für wissenschaftliche Recherchen notwendig sind. Die Verknüpfung mit Normdaten ermöglicht eine gezielte Referenzierbarkeit und Vernetzung von Texten, Objekten oder Daten. Insofern bietet *Google Books* zwar eine attraktive Ergänzung zu geschichtswissenschaftlichen Volltextdatenbanken, doch sollte man sich auch der Grenzen bewusst sein. Unter dem Aspekt der Sichtbarkeit des globalen kulturellen und wissenschaftlichen gedruckten Erbes innerhalb digitaler Bibliotheken ist zu beachten, dass eine Recherche mit der Suchmaschine Google unter Berücksichtigung von *Google Books* ein Ranking der Ergebnisse zeigt, in dem englischsprachige Literatur dominiert. Daher empfiehlt sich die gleichzeitige Konsultation von Kulturportalen und digitalen Bibliotheken wie der *Europeana* oder der *DDB*.

*Wikisource*<sup>122</sup>, eine Sammlung gemeinfreier oder unter einer freien respektive Creative-Commons-Lizenz (CC-BY bzw. CC-BY-

---

<sup>121</sup> <https://books.google.com/ngrams/>

<sup>122</sup> <https://de.wikisource.org/wiki/Hauptseite>

SA) stehender Texte, wird häufig unbewusst genutzt, da diese mit *Wikipedia* verbunden ist. Für die Geschichtswissenschaft ist *Wikisource* von besonderer Bedeutung, da insbesondere ältere Texte mit dem Status von historischen Quellen Berücksichtigung finden. Grundlage der in *Wikisource* präsentierten Texte sind Erstausgaben, Ausgaben letzter Hand oder kritische Editionen, so dass wissenschaftliche Qualität gewährleistet ist. Aber auch digitalisierte geschichtswissenschaftliche Zeitschriften und zahlreiche Volltexte gehören zu den Ressourcen für Historikerinnen und Historiker.

Nicht nur eine digitale Bibliothek für Volltexte von Büchern, Musik, Filmen, Software oder Bildern, sondern zugleich ein Dienst, der Webseiten in unterschiedlichen Versionen speichert, ist das 1996 gegründete gemeinnützige *Internet Archive*<sup>123</sup>. Da es nicht möglich ist, das gesamte World Wide Web in allen Zuständen dauerhaft zu archivieren, bietet die Speicherung von „Momentaufnahmen“ dennoch einen breiten Zugriff auf Webseiten, die zeitweise im Internet verfügbar waren – dieser erfolgt mit Hilfe der *Wayback Machine*. Ein Spiegelserver des in San Francisco ansässigen *Internet Archive* findet sich übrigens in der *Bibliotheca Alexandrina* in Ägypten, am Ort der größten Bibliothek der antiken Welt.

Innerhalb des *Million Book Project* als Teilprojekt des *Internet Archive* werden gemeinfrei gewordene Bücher digitalisiert und in der *Open Library*<sup>124</sup> publiziert. Hier soll in einem kollaborativen Ansatz jedes jemals publizierte Buch auf einer eigenen Webseite dokumentiert werden. Dabei kann zum gemeinfreien bibliografischen Nachweis auch der direkte Zugriff auf das Digitalisat mit dem Volltext treten. Im Unterschied zu *Google Books* mit der Digitalisierung auch urheberrechtlich geschützter Literatur konzentriert sich die *Open Library* auf gemeinfreie Bücher.

Ein gemeinsames Projekt von zahlreichen US-amerikanischen Universitätsbibliotheken sowie Forschungseinrichtungen aus der ganzen Welt ist die *HathiTrust Digital Library*<sup>125</sup>. Bibliografische Datenbank und Repositorium digitalisierter Bücher zugleich, ist eine

---

<sup>123</sup> <https://archive.org/>

<sup>124</sup> <https://openlibrary.org/about>

<sup>125</sup> <https://www.hathitrust.org/>

Volltextsuche in Millionen Dokumenten unterschiedlicher Fachbereiche möglich – die Geschichtswissenschaft bildet einen Schwerpunkt. Der aus dem Hindi und Urdu stammende Name „Hathi“ bedeutet übrigens „Elefant“: Diesem wird ein besonderes Gedächtnis nachgesagt – insofern gleicht er den Archiven und Bibliotheken als Gedächtnis der Menschheit.

Eine wichtige deutschsprachige Volltextbibliothek ist *Zeno.org*<sup>126</sup>. Sie basiert auf der kommerziellen Reihe *Digitale Bibliothek* (CD's, DVD's) und umfasst Texte vom Anfang des Buchdrucks bis zum Beginn des 20. Jahrhunderts. Seit 2009 ist die Volltextsammlung Teil des Repositoriums von *TextGrid*<sup>127</sup>, der Virtuellen Forschungsumgebung für die Geisteswissenschaften, und kann zur wissenschaftlichen Bearbeitung, beispielsweise in digitalen Editionen, genutzt werden.

Insbesondere für kulturgeschichtliche Fragestellungen relevant ist das *Deutsche Textarchiv (DTA)*<sup>128</sup> als Referenzcorpus der neuhochdeutschen Sprache. Es umfasst etwa 1.500 sorgfältig ausgewählte und nach Erstausgaben digitalisierte Texte unterschiedlicher Disziplinen aus dem 17.–20. Jahrhundert. Im Konsortium *Text+*<sup>129</sup> der *Nationalen Forschungsdateninfrastruktur (NFDI)*<sup>130</sup> ist das DTA als Repository und strukturiertes, linguistisch annotiertes Volltextcorpus historischer Texte eingebunden.

Ein Blick in die Welt der Archive und Museen rundet den Blick auf die für die Geschichtswissenschaften relevante Anbieter digitaler Bibliotheken und Volltextdatenbanken ab. Von nationaler Bedeutung für das deutsche Archivwesen ist das *Bundesarchiv*<sup>131</sup>, die für Sicherung des Archivgutes der Bundesrepublik Deutschland und seiner Vorgängerstaaten zuständige Institution. Über seine (Volltext-) Datenbanken für Bilder, Filme, Töne und Karten ist die

---

<sup>126</sup> <http://www.zeno.org/>

<sup>127</sup> <https://www.textgrid.de/>

<sup>128</sup> <https://www.deutschestextarchiv.de/>

<sup>129</sup> <https://www.nfdi.de/textplus/>

<sup>130</sup> <https://www.nfdi.de/>

<sup>131</sup> <https://www.bundesarchiv.de/DE/Navigation/Home/home.html>

---

Recherche innerhalb der allgemeinen wie auch der speziellen Archivbestände möglich. Die beiden zentralen historischen Museen der Bundesrepublik Deutschland – das *Deutsche Historische Museum*<sup>132</sup> (Berlin) und das *Haus der Geschichte*<sup>133</sup> (Bonn) – präsentieren nicht nur Dauerausstellungen aus eigenen Beständen und zahlreiche Sonderausstellungen mit Leihgaben zu Themen der deutschen Geschichte. Ein gemeinsames Projekt beider Museen mit dem Bundesarchiv ist das *Lebendige Museum Online (LeMO)*<sup>134</sup>, in dem auch (Volltext-) Datenbanken eine zentrale Rolle spielen.

Lag der Fokus bis hierher auf retrodigitalisierten Texten, die für die Geschichtswissenschaft als ältere, gemeinfreie Forschungsbeiträge oder als Quellen interessant sein können, so wird im folgenden Kapitel ein kurzer Überblick über Volltextdatenbanken aktueller geschichtswissenschaftlicher Forschungsliteratur gegeben.



Abbildung 10. Kulturportale und digitale Bibliotheken (Auswahl)

<sup>132</sup> <https://www.dhm.de/>

<sup>133</sup> <https://www.hdg.de/stiftung>

<sup>134</sup> <https://www.dhm.de/lemo>

## 2.5 Volltextdatenbanken für Forschungsliteratur

Portale und Volltextdatenbanken für Forschungsliteratur können hier nicht in der notwendigen Breite thematisiert werden. Hinweise zu relevanten Volltextdatenbanken finden sich in den Guides zu Epochen oder Regionen. Daher sei nur cursorisch auf wenige große fachübergreifende Ressourcen hingewiesen.

Eine der wichtigsten Volltextdatenbanken für Zeitschriften aus dem gesamten Spektrum der Geistes-, Kultur- und Sozialwissenschaften aus dem Zeitraum 1800–2000 ist das *Periodicals Archive Online (PAO)*<sup>135</sup>. Es bietet über eine differenzierte Recherchemaske den Zugriff auf 3 Millionen durch Abstracts erschlossene Artikel aus 700 Zeitschriften.

Unverzichtbar für die Recherche nach Fachartikeln internationaler Zeitschriften nicht nur aus den Geschichtswissenschaften ist die Volltextdatenbank *Journal Storage (JSTOR)*<sup>136</sup>. Enthalten sind internationale wissenschaftliche Zeitschriften vom ersten Jahrgang an – je nach Titel und lokaler Lizenz ist der Zugriff auf die aktuellen Hefte aufgrund einer *moving wall* nicht möglich. Das deutsche Pendant zu JSTOR ist das Volltextarchiv *DigiZeitschriften*<sup>137</sup> mit deutschsprachigen Fachzeitschriften, unter denen die Geschichtswissenschaften mit etwa 350 Traditionszeitschriften vertreten sind.

Zu fachspezifischen Volltextdatenbanken bieten die einschlägigen FIDs, wie *Propylaeum*<sup>138</sup> für die Alte Geschichte oder *historicum.net*<sup>139</sup> für die mittlere und neuere Geschichte, weitere Hinweise. Auf fachübergreifende Ressourcen wie Google Books, das Internet Archive oder Hathi Trust und andere wurde bereits hingewiesen (Kap. 2.4). Weitere Hinweise finden sich in der kommentierten *Linkliste Volltextdatenbanken*<sup>140</sup>.

---

<sup>135</sup> [https://about.proquest.com/en/products-services/periodicals\\_archive/](https://about.proquest.com/en/products-services/periodicals_archive/)

<sup>136</sup> <https://www.jstor.org/>

<sup>137</sup> <https://www.digizeitschriften.de/>

<sup>138</sup> <https://www.propylaeum.de/>

<sup>139</sup> <https://www.historicum.net/>

<sup>140</sup> <https://guides.clio-online.de/guides/sammlungen/volltextdatenbanken/2024/linkliste>

## 2.6 Volltextdatenbanken für Quellen und Quelleneditionen

Neben der Verteilten nationalen Forschungsbibliothek ist die Verteilte deutsche Nationalbibliothek Anbieter und Plattform historischer Volltextdatenbanken für Quellen und Quelleneditionen. Die nationalbibliografischen Verzeichnisse VD16, VD17 und VD18 sowie das zvdd sind vor allem für die deutsche Geschichte interessant – insbesondere durch die Perspektive ihrer geplanten Volltexttransformation im Rahmen des Projektes OCR-D. Die im Folgenden vorgestellten Volltextdatenbanken für historische Quellen berücksichtigen Produkte im *Open* wie im *Closed Access*.

Volltextdatenbanken für Quellen entstanden und entstehen dadurch, dass schriftbasierte Originalquellen wie Inschriften, Papyri, Handschriften oder Drucke digitalisiert und als Bilder in Datenbanken veröffentlicht werden. Im strengen Sinn handelt es sich in diesen Fällen aber zunächst (noch) nicht um „Volltextdatenbanken“ – die Quellen werden lediglich durch detaillierte Metadaten beschrieben. Erst die Entwicklung der Texterkennung für Handschriften (HTR) oder Drucke (OCR) ermöglicht im nächsten Schritt, aus den reinen Bilddigitalisaten echte Volltexte zu generieren, in denen recherchiert werden kann und deren Weiterverarbeitung innerhalb von DH-Szenarien möglich ist.

Mit Volltextdatenbanken digitalisierter Originalquellen sind Volltextdatenbanken für Quelleneditionen verbunden – vielfach werden diese nicht nur untereinander, sondern auch mit bibliografischen Datenbanken oder Volltextdatenbanken für Forschungsliteratur vernetzt.

### Alte Geschichte

Antike Textquellen können auf unterschiedlichen Trägern überliefert sein, beispielsweise in Stein gehauen als Inschriften oder handgeschrieben auf Papyri oder Pergament in Form von Rotuli oder Codices. Relevante Quellen für die Alte Geschichte sind jedoch nicht allein Textzeugnisse, sondern auch kulturelle Artefakte wie archäologische Funde.

Zentrales Rechercheportal für bibliografische Informationen, Beschreibungen, Digitalisate, Transkriptionen und Übersetzungen

von Papyri, Ostraka oder Holztafeln aus bedeutenden internationalen Sammlungen ist *papyri.info*<sup>141</sup>. Der *Papyrological Navigator (PN)*<sup>142</sup> ermöglicht eine integrierte Recherche innerhalb unterschiedlicher Datenbanken, beispielsweise im *Heidelberger Gesamtverzeichnis der griechischen Papyrusurkunden Ägyptens*<sup>143</sup>, in der *Duke Databank of Documentary Papyri (DDBDP)*<sup>144</sup> und im *Advanced Papyrological Information System (APIS)*<sup>145</sup>. Letzteres weist Informationen (Bibliografien, Beschreibungen), Abbildungen (Digitalisate) oder Übersetzungen zu papyrologischen Materialien (Papyri, Ostraka, Holztafeln) aus internationalen Sammlungen nach.

Einen Überblick über die wichtigsten Papyrussammlungen Deutschlands bietet das *Papyrus Portal*<sup>146</sup>. Es ermöglicht sowohl eine parallele Suche über einzelne Papyrus-Datenbanken als auch den Wechsel in lokale Präsentationen mit häufig differenzierteren Rechercheoptionen.

Epigraphische Textzeugnisse der Antike werden als archäologische Objekte ebenfalls in Volltextdatenbanken publiziert, die neben Beschreibungen der Inschriften mit spezifischen Metadaten auch Transkriptionen, Übersetzungen und Digitalisate umfassen. Ein Beispiel ist die *Epigraphische Datenbank Heidelberg (EDH)*<sup>147</sup>, die auf eine umfassende Dokumentation lateinischer und bilinguer (z.B. lateinisch-griechischer) Inschriften des Römischen Reiches zielt. Berücksichtigt werden insbesondere die außerhalb der großen Corpora publizierten Inschriften. Die Suchmaske ermöglicht eine differenzierte Recherche innerhalb des umfangreichen Materials – Inschriftentexte, Fotodokumentationen, Bibliografie und Geografie. Partner der EDH sind unter anderen die *Epigraphik-Datenbank*

---

<sup>141</sup> <https://papyri.info/>

<sup>142</sup> <https://papyri.info/search>

<sup>143</sup> <https://aquila.zaw.uni-heidelberg.de/start>

<sup>144</sup> <https://papyri.info/browse/ddbdp/>

<sup>145</sup> <https://papyri.info/browse/apis/>

<sup>146</sup> <https://www.papyrusportal.de/content/start.xml>

<sup>147</sup> <https://edh.ub.uni-heidelberg.de/home>

---

*Claus/Slaby (EDCS)*<sup>148</sup>, das *Corpus Inscriptionum Latinarum (CIL)*<sup>149, 150</sup> *Inscriptiones Graecae (IG)*<sup>151</sup>, *Searchable Greek Inscriptions (PHI Greek Inscriptions)*<sup>152</sup>, *Trismegistos (TM)*<sup>153</sup>, *Europeana EAGLE*<sup>154</sup> oder *Ubi Erat Lupa (Lupa)*<sup>155</sup>. Es ist zu beachten, dass papyrologische und epigraphische Quellen häufig in gemeinsamen Volltextdatenbanken erfasst sind.

*Perseus Digital Library (PDL)*<sup>156</sup> ist eine der ältesten online verfügbaren geisteswissenschaftlichen Textsammlungen – mit Schwerpunkt auf der antiken Überlieferung. In der Kollektion *Greek and Roman Materials*<sup>157</sup> bietet sie auf Grundlage zitierfähiger Editionen Volltexte klassischer griechischer und lateinischer Literatur, teilweise mit (englischen) Übersetzungen, die im *Scaife Viewer*<sup>158</sup> miteinander verglichen werden können. Insbesondere Analysen historischer Begrifflichkeiten oder philologische Fragestellungen zum Wortgebrauch sind möglich.

Der *Thesaurus Linguae Graecae (TLG)*<sup>159</sup> enthält klassische griechische Texte aus der Zeit zwischen etwa 800 v. Chr. und 600 n. Chr. sowie mittelalterliche historiographische, lexikographische und scholastische griechische Texte aus der Zeit zwischen etwa 600 n.

---

<sup>148</sup> [https://db.edcs.eu/epigr/epi.php?s\\_sprache=de](https://db.edcs.eu/epigr/epi.php?s_sprache=de)

<sup>149</sup> <https://cil.bbaw.de/ace/search?page=1>

<sup>150</sup> In der Datenbank sind die Archivalien des CIL erschlossen – die Abklatsche, Fotos und Scheden zu den im CIL edierten Inschriften. Da die Einträge mit der Epigraphik-Datenbank Clauss – Slaby (EDCS) verknüpft sind, steht hierüber auch der Inschriftentext zur Verfügung.

<sup>151</sup> <http://telota.bbaw.de/ig/>

<sup>152</sup> <https://epigraphy.packhum.org/>

<sup>153</sup> <https://www.trismegistos.org/>

<sup>154</sup> <https://www.eagle-network.eu/>

<sup>155</sup> <http://lupa.at/>

<sup>156</sup> <http://www.perseus.tufts.edu/hopper/>

<sup>157</sup> <http://www.perseus.tufts.edu/hopper/collection?collection=Perseus:collection:Greco-Roman>

<sup>158</sup> <https://scaife.perseus.org/>

<sup>159</sup> <http://stephanus.tlg.uci.edu/>

Chr. und 1453 n. Chr. Die Suchmaske erlaubt komplexe Recherchen nach Autor, Werk, Datierung, Gattung, aber auch eine Volltextsuche innerhalb der Werke eines oder mehrerer Autoren. Die Darstellung der Texte erfolgt wahlweise in griechischen oder transliteriert in lateinischen Buchstaben. Die Verknüpfung der Texte des TLG mit weiteren Volltextdatenbanken für Quellen (PDL) oder für Forschungsliteratur (JSTOR) ermöglicht vernetztes digitales Arbeiten mit griechischen Textquellen für die Alte Geschichte.

Die *Library of Latin Texts Complete Plus*<sup>160</sup> umfasst die beiden zuvor separaten Datenbanken *Library of Latin Texts – Series A* (LLT-A) und *Series B* (LLT-B). Inkludiert sind inzwischen mehr als 5.400 Werke von etwa 1.300 Autoren von den Anfängen der lateinischen Literatur im 3. Jh. v. Chr. bis zum 2. Vatikanischen Konzil (1962–1965): römische Klassiker, Kirchenväter, mittelalterliche lateinische Literatur sowie Texte der Reformation und Gegenreformation. Volltextsuche und Textanalyse ermöglichen vielfältige philologische Fragestellungen.

Die *Sammlung Tusculum Online*<sup>161</sup> basiert auf der traditionsreichen Buchreihe mit inzwischen mehreren Hundert Bänden. Sie umfasst Editionen, Übersetzungen und Kommentare der griechischen und lateinischen Klassiker der Antike, künftig auch spätantiker, christlicher, byzantinischer und neulateinischer Literatur, basierend auf den teilweise vergriffenen Druckausgaben. Im Gegensatz zur *Library of Latin Texts Complete Plus* ist in der *Sammlung Tusculum Online* jedoch keine textübergreifende Volltextsuche möglich, sondern nur in den einzelnen Texten.

Ein umfassendes Informationssystem für die interdisziplinären altertumswissenschaftlichen Disziplinen ist das Portal *iDAI.objects (Arachne)*<sup>162</sup> des Deutschen Archäologischen Instituts (DAI). Es ist eingebettet in eine modulare Forschungsinfrastruktur, die Objekte, Bücher, Bilder, bibliografische Daten und Forschungsdaten sowie Digitalisate nach einem einheitlichen Datenmodell verwaltet. Für

---

<sup>160</sup> <https://www.brepols.net/products/IS-9782503603216-1>

<sup>161</sup> <https://www.degruyter.com/serial/tusc-b/html>

<sup>162</sup> <https://arachne.dainst.org/>

geschichtswissenschaftliches Arbeiten, das sich auf eine breite Vielfalt altertumswissenschaftlicher Quellen stützen möchte, bieten sich hier herausragende Voraussetzungen.

Monumentale Editionen des quellenverliebten 19. Jahrhunderts zur Patrologie- und Mittelalterforschung sind die *Patrologia Graeca (PG)*<sup>163</sup>, die *Patrologia Latina (PL)*<sup>164</sup> sowie die im folgenden Abschnitt zum Mittelalter besprochenen *Monumenta Germaniae Historica* (MGH). Die PG basiert auf der *Patrologia Graeco-Latina*, die von Jacques-Paul Migne zwischen 1857 und 1866 in 161 Bänden herausgegeben wurde.<sup>165</sup> Sie enthält zentrale Werke der christlich-griechischen Kirchenliteratur spätantiker und mittelalterlicher Theologie, Philosophie und Geschichte aus der Zeit zwischen 100 n. Chr. und 1478 und ist grundlegend für historische Forschungen zum frühen Christentum. Die Volltexterschließung erfolgt durch ein lateinisches und griechisches Inhaltsverzeichnis, einen Autoren-, Werktitel- und Sachindex.

In Ergänzung zur PG steht die PL, die Edition des lateinischen Schrifttums der Kirche von den Anfängen bis ins Hochmittelalter in insgesamt 221 Bänden, die Jacques-Paul Migne in zwei Reihen zwischen 1844 und 1855 publiziert hat.<sup>166</sup> Auf dieser Ausgabe beruht die Datenbank, die differenzierte Volltextsuchen ermöglicht. Texte der PL sind auch Bestandteil der *Library of Latin Texts*.<sup>167</sup>

---

<sup>163</sup> <https://classiques-garnier.com/patrologia-graeca-1.html>

<sup>164</sup> [https://about.proquest.com/en/products-services/patrologia\\_latina/](https://about.proquest.com/en/products-services/patrologia_latina/)

<sup>165</sup> [Patrologia cursus completus / Series Graeca] Patrologiae cursus completus: in qua prodeunt patres, doctores scriptoresque ecclesiae Graecae [...] accurante J. P. Migne. Paris, 1857–1866.

<sup>166</sup> [Patrologia cursus completus / Series Latina] Patrologiae cursus completus: seu bibliotheca universalis, integra, uniformis, commoda, oeconomica, omnium ss. Patrum [...] ecclesiasticorum, sive Latinorum, sive Graecorum [...] accurante J. P. Migne. Paris, 1844–1855.

<sup>167</sup> Neben den lizenzpflichtigen Verlags-Editionen stehen PG und PL auch im Internet Archive oder Google Books im Open Access zur Verfügung: Patristica <http://patristica.net/>.

## Geschichte des Mittelalters

Zu den zentralen Quellen für die Mediävistik gehören handschriftliche Urkunden, historiographische oder literarische Texte, Inschriften sowie kulturelle Artefakte. Da diese in Archiven, Bibliotheken oder Museen aufbewahrt werden, ist die Kenntnis institutioneller Infrastrukturen hilfreich. Ihre Sammlungen sind häufig Grundlage von Digitalisierungsprojekten und (Volltext-) Datenbanken, die digital vernetztes Arbeiten mit Methoden und Tools der Digital Humanities ermöglichen.

Einen hervorragenden Einstieg in die Überlieferung urkundlicher Quellen bietet *Monasterium (MOM)*<sup>168</sup>, das virtuelle Urkundenarchiv Europas. Es präsentiert etwa 500.000 digitalisierte Dokumente aus mehr als 60 europäischen Archiven – beispielsweise Bilder, Regesten, ältere gedruckte sowie neue Editionen. Eine Volltextsuche innerhalb der Metadaten, aber auch in transkribierten Urkunden, ist möglich.

Die seit dem 19. Jahrhundert entstandenen Handschriftenkataloge einzelner Bibliotheken sind unverzichtbare Instrumente der Text- und Überlieferungsgeschichte. Seit Beginn der Digitalisierung wurden Handschriftenkataloge und Handschriftenoriginale schrittweise digitalisiert und sind Grundlagen von Volltextdatenbanken.

Einen Überblick über das Handschriftenerbe im deutschsprachigen Raum bietet die Recherche in digitalen Handschriftenbibliotheken Deutschlands, Österreichs und der Schweiz. In Deutschland ist das *Handschriftenportal (HSP)*<sup>169</sup> zentrale Informationsinfrastruktur für europäische Buchhandschriften in deutschen Sammlungen. Recherchierbar sind neben den bibliothekarischen Metadaten der Originale die wissenschaftlichen Beschreibungen der maßgeblichen Handschriftenkataloge. Zugleich ist das HSP zentrales Portal der Handschriftendigitalisate, die sich in den digitalen Bibliotheken der jeweiligen besitzenden Einrichtungen befinden. Durch Technologien wie IIF erfüllt das HSP internationale Standards der

---

<sup>168</sup> <https://www.icar-us.eu/cooperation/online-portals/monasterium-net/>

<sup>169</sup> <https://handschriftenportal.de/>

wissenschaftlichen Arbeit mit Digitalisaten. Die dynamische Entwicklung automatischer Volltexterkennung mit HTR lässt Volltextsuchen innerhalb des Handschriftenerbes im HSP vorstellbar werden.

Für die Erforschung der Handschriftenüberlieferung in Österreich koordiniert das Institut für Mittelalterforschung der Österreichischen Akademie der Wissenschaften den Aufbau des Portals *manuscripta.at – mittelalterliche Handschriften in Österreich*<sup>170</sup>. Sein Ziel, verstreute, schwer zugängliche Daten zu österreichischen Handschriften als Verweise, Links, Images oder Volltexte gebündelt zu präsentieren, macht *manuscripta.at* dem deutschen Handschriftenportal vergleichbar.

Das schweizerische Pendant *e-codices – virtuelle Handschriftenbibliothek der Schweiz*<sup>171</sup> erschließt die mittelalterlichen und neuzeitlichen Handschriften aus öffentlichen, kirchlichen und privaten Sammlungen der Schweiz. Als Nationalbibliografie und Bestandsverzeichnis zugleich bietet *e-codices* den Zugriff auf wissenschaftliche Beschreibungen und digitale Reproduktionen der Handschriften. Es besteht die Möglichkeit, kollaborativ Annotationen oder bibliografische Angaben den einzelnen Handschriften hinzuzufügen.

Neben *e-codices* weist *e-manuscripta.ch*<sup>172</sup> digitalisierte handschriftliche Quellen aus Schweizer Bibliotheken und Archiven nach: Texthandschriften (Einzel- und Sammelhandschriften), Briefe, Musikalien, Karten und Bilder. Gemeinsame Transkriptionsarbeit durch moderiertes Crowdsourcing ermöglicht schrittweise eine intellektuelle Volltexttransformation des Handschriftenerbes in der Schweiz – Grundsätze des Open Access und einer Citizen Science verbinden sich harmonisch.

Im Unterschied zu den genannten Portalen, die insbesondere das in Gedächtnisinstitutionen der Schweiz aufbewahrte Handschrif-

---

<sup>170</sup> <https://manuscripta.at/>

<sup>171</sup> <https://www.e-codices.unifr.ch/de>

<sup>172</sup> <https://www.e-manuscripta.ch/>

---

tenerbe nachweisen, unterstützt *Fragmentarium – Laboratory for Medieval Manuscript Fragments*<sup>173</sup> die praktische Forschung und Arbeit mit Handschriftenfragmenten, die internationale „digitale Fragmentologie“.

Ergänzend zu den nationalen Handschriftenportalen im deutschsprachigen Raum bietet der *Handschriftencensus (HSC)*<sup>174</sup> einen Überblick über sämtliche deutschsprachige Handschriften des Mittelalters aus dem Zeitraum 750–1520 in internationalen Sammlungen. Das umfassende Bestandsverzeichnis bietet differenzierte Informationen und Metadaten zu mittelalterlichen Autoren, Werken und ihrer Überlieferung. Zu jedem Textzeugen ist nicht nur die relevante Literatur bibliografisch erfasst, sondern auch das Digitalisat verlinkt.

Für die Geschichte des Mittelalters bieten klassische Printeditionen die Grundlage von Volltextdatenbanken und komfortable Möglichkeiten des Zugriffs auf Quellen. Unentbehrlich sind die *Monumenta Germaniae Historica (MGH)*<sup>175</sup> als grundlegende Sammlung von Quelleneditionen (Historiographie, Rechtstexte, Urkunden, Briefe, Dichtung) aus dem Zeitraum 500 – 1500, in der seit 1819 in mehr als 300 Bänden etwa 1.300 Texte erschienen sind. Die aktuellen gedruckten Editionen werden mit einer *moving wall* von drei Jahren digitalisiert und online publiziert. Differenzierte Suchoptionen ermöglichen einen umfassenden Volltextzugriff auf diese wichtigste Editionsreihe von Quellen zur mittelalterlichen Geschichte Deutschlands und Europas.

Ursprünglich als Vorarbeit zu den MGH konzipiert, entstand mit den *Regesta Imperii (RI)*<sup>176</sup> ein weiteres Monument der Wissenschaftsgeschichte, das seit seiner Begründung 1839 für Historikerinnen und Historiker unverzichtbar ist. Chronologisch geordnet werden sämtliche urkundlichen und historiographischen Quellen

---

<sup>173</sup> <https://fragmentarium.ms/>

<sup>174</sup> <https://handschriftencensus.de/>

<sup>175</sup> <https://www.dmgh.de/>

<sup>176</sup> <http://www.regesta-imperii.de/unternehmen.html>

---

der römisch-deutschen Herrscher von den Karolingern bis zu Maximilian I. (751–1519) sowie der Päpste des frühen und hohen Mittelalters verzeichnet. Bei der Textsorte *Regest* handelt es sich um keine historisch-kritische Edition, sondern um eine Zusammenfassung des Inhalts der jeweiligen überlieferten Quelle. Sämtliche gedruckten Regestenbände wurden digitalisiert und ihre Inhalte zusätzlich in die Regestendatenbank übernommen. Diese bietet differenzierte Rechercheoptionen in den Volltexten und Vernetzungen beispielsweise mit der *Regesta Imperii-Literaturdatenbank (RI-OPAC)*<sup>177</sup> als zentraler Bibliografie für die mediävistische Forschung. Sofern eine in den RI erfasste Urkunde innerhalb der MGH ediert wurde, findet sich eine entsprechende Verlinkung.

Eine weitere Verknüpfung der RI besteht mit dem *Lichtbildarchiv älterer Originalurkunden Marburg (LBA)*<sup>178</sup>, das seit 1928 die original überlieferten Urkunden des römisch-deutschen Reiches aus der Zeit vor 1250 sammelt und diese digitalisiert in einer Datenbank zugänglich macht. Die differenzierte Recherchemaske erlaubt beispielsweise die Suche nach Ausstellern, Empfängern, Mitsiegler, Datierungen oder den gegenwärtigen Aufbewahrungsorten von Urkunden. Sind diese innerhalb der RI ediert, gelangt man direkt in die Regesten und kann die Digitalisate unmittelbar vergleichen.

Neben Urkunden sind weitere Textquellen für die Geschichte des Mittelalters relevant. Das Repertorium *Geschichtsquellen des deutschen Mittelalters*<sup>179</sup> ist ein bibliografisches und quellenkundliches Verzeichnis erzählender Geschichtsquellen wie Chroniken, Annalen oder Briefen aus der Zeit Karls des Großen bis zu Maximilian I. (ca. 750 – 1500). Es ist hervorgegangen aus dem traditionsreichen *Repertorium Fontium Historiae Medii Aevi*<sup>180</sup>. Neben der Beschrei-

---

<sup>177</sup> [http://opac.regesta-imperii.de/lang\\_de/](http://opac.regesta-imperii.de/lang_de/)

<sup>178</sup> <https://www.uni-marburg.de/de/fb06/mag/lichtbildarchiv-aelterer-originalurkunden>

<sup>179</sup> <http://www.geschichtsquellen.de/start>

<sup>180</sup> Repertorium fontium historiae Medii Aevi. Hrsg. Istituto Storico Italiano per il Medio Evo [...] Primum ab Augusto Potthast digestum, nunc cura collegii

---

bung des Inhaltes der verzeichneten Quellen finden sich Nachweise der handschriftlichen Überlieferung, der relevanten Editionen, der Übersetzungen und Forschungsbeiträge.

Nicht unerwähnt bleiben soll schließlich eine für die Erforschung des Mittelalters wichtige Volltextdatenbank für epigraphische Zeugnisse, die *Deutschen Inschriften Online (DIO)*<sup>181</sup>. Auf Grundlage der gedruckten Edition *Die deutschen Inschriften des Mittelalters und der Frühen Neuzeit*<sup>182</sup> hat sie das Ziel, sämtliche lateinischen und deutschen Inschriften bis zum Jahr 1650 aus Deutschland, Österreich und Südtirol zu sammeln und zu edieren. Aufgenommen werden sowohl erhaltene Originalinschriften als auch kopial überlieferte Dokumente. Die Datenbank geht weit über die Digitalisierung der gedruckten Edition hinaus, indem sie weiteres Material integriert und differenzierte Recherchen sowohl in den Einzelbänden, als auch im Gesamtbestand und innerhalb der transkribierten Inschriften ermöglicht.

### Geschichte der Frühen Neuzeit

Bevor sich die Frühe Neuzeit als eigenständige Epoche innerhalb der Geschichtswissenschaft fest etabliert hat, war sie bereits prominenter Gegenstand bibliothekarischer Forschung und bibliografischer Dokumentation.

Insbesondere das als Zeitalter der Inkunabeln, der Wiegendrucke, bezeichnete erste halbe Jahrhundert des Buchdrucks zwischen 1450 und 1500 beschäftigte bereits in der Frühen Neuzeit Bibliothekare und Bibliophile. Im 19. Jahrhundert formierte sich auf Grundlage erster Inkunabelverzeichnisse die Inkunabelkunde als bibliothekswissenschaftliche Disziplin. Diese Verzeichnisse wurden immer vollständiger und immer weiter verbessert, so dass sie inzwi-

---

historicorum e pluribus nationibus emendatum et auctum. 11 Bände, Rom: 1962–2010.

<sup>181</sup> <https://www.inschriften.net/>

<sup>182</sup> Eine Übersicht über die Editionsreihe findet sich unter:  
<https://www.inschriften.net/projekt/publikationen/editionsreihe.html>.

schen – wie auch die Handschriftenkataloge – als (Volltext-)Datenbanken mit anderen Forschungsressourcen vernetzt werden können.

Der auf älteren Katalogen<sup>183</sup> aufbauende, an der Königlichen Bibliothek zu Berlin vor dem Ersten Weltkrieg begonnene und heute von der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz fortgeführte *Gesamtkatalog der Wiegendrucke (GW)*<sup>184</sup> verzeichnet alphabetisch sämtliche Drucke des 15. Jahrhunderts.<sup>185</sup> Da er bis heute nur zu Teilen vollendet ist, wird er parallel als Datenbank aufgebaut, um sämtliche in den gedruckten Bänden sowie im Manuskript nachgewiesenen Inkunabelausgaben in unterschiedlicher Erschließungstiefe zu verzeichnen. Über unterschiedliche Suchmasken und Register lassen sich differenziert Druckernamen, Druckorte, Autoren, Werke oder die GW-Nummern recherchieren.

Wie der deutsche GW verfolgt der von der British Library seit 1980 herausgegebene *Incunabula Short Title Catalogue – the international database of 15th-century European printing (ISTC)*<sup>186</sup> das Ziel, einen Überblick über die weltweit erhaltenen Inkunabeln zu bieten.<sup>187</sup> Dieser internationale Kurztitelkatalog umfasst als Meta-Katalog die Kerndaten aus nationalen Inkunabelkatalogen mit dem Schwerpunkt auf Europa und Nordamerika. Mit 30.518 Ausgaben (Stand 2016) bietet der ISTC eine hervorragende Abdeckung. Verknüpfungen gibt es mit dem GW und mit vorhandenen Digitalisaten weltweit.

---

<sup>183</sup> Hierzu zählen Ludwig Hains Inkunabelverzeichnis *Repertorium bibliographicum* (1825–1838), das von Walter Arthur Copingers *Supplement to Hain's Repertorium bibliographicum* (1895–1902) und von Walter Reichlings *Appendices ad Hainii-Coperingi Repertorium bibliographicum* (1905–1911) zwar vervollständigt wurde, aber wissenschaftlichen Anforderungen nicht mehr genügte.

<sup>184</sup> <https://www.gesamtkatalogderwiegendrucke.de/>

<sup>185</sup> Nach dem 1925 erschienen ersten Band liegen inzwischen 11 Bände (A-H) mit der Beschreibung von etwa 36.000 Inkunabeln vor. Parallel wird eine Datenbank aufgebaut.

<sup>186</sup> [https://data.cerl.org/istc/\\_search](https://data.cerl.org/istc/_search)

<sup>187</sup> An der Konzeption war maßgeblich die Inkunabelforscherin Lotte Hellinga beteiligt, die den Katalog *Incunabula in American libraries* (1964) von Frederick R. Goff (1916–1982) als Grundlage für den neuen Katalog wählte.

---

Für Forschungen zur konkreten Materialität von Inkunabeln sind zwei Datenbanken relevant, die mit GW, ISTC und weiteren Inkunabelverzeichnissen verknüpft sind: *Material Evidence in Incunabula (MEI)*<sup>188</sup> und der von Paul Needham erstellte *Index Possessorum Incunabulorum (IPI)*<sup>189</sup>. In MEI werden sämtliche bekannten materialspezifischen Besonderheiten wie Buchschmuck, Einband, Stempel, Annotationen, Signaturen und Provenienzen erfasst. IPI basiert auf der Auswertung von etwa 200 publizierten Inkunabelkatalogen sowie Needhams eigenen Forschungen zu den früheren Besitzern von Inkunabeln. Insbesondere die Möglichkeit, nach vielfältigsten Provenienzen (zum Beispiel Personen, Institutionen, Monogramme) zu recherchieren, ist eine hervorragende Ergänzung zur Recherche in MEI.

Die größte Inkunabelsammlung des deutschen Sprach- und Kulturraums mit 20.337 Exemplaren von 9.782 Ausgaben besitzt die *Bayerische Staatsbibliothek München*<sup>190</sup>. Nachgewiesen wird diese im Inkunabelkatalog *BSB-Ink online*<sup>191</sup>, der auf der zwischen 1988 und 2021 erschienenen Printausgabe<sup>192</sup> beruht und Teil des *Inkunabelkatalogs deutscher Bibliotheken (INKA)*<sup>193</sup> ist. Die genauen Exemplarbeschreibungen in INKA bieten ein hervorragendes Werkzeug für die Provenienzforschung und referenzieren die maßgeblichen Verzeichnisse und Bibliografien.

Für die Recherche nach Inkunabeln in deutschen und internationalen Sammlungen bieten die genannten Verzeichnisse, auf die vielfach auch der Volltextzugriff möglich ist, eine hervorragende Ausgangsbasis. Wie für Handschriften dürfte auch für die digitalisierten Inkunabeln in absehbarer Zeit eine Volltexterkennung möglich sein, so dass künftig nicht nur in den Inkunabelkatalogen mit

---

<sup>188</sup> [https://data.cerl.org/mei/\\_search](https://data.cerl.org/mei/_search)

<sup>189</sup> [https://data.cerl.org/ipi/\\_search](https://data.cerl.org/ipi/_search)

<sup>190</sup> <https://www.bsb-muenchen.de/sammlungen/historische-drucke/bestaende/inkunabeln/>

<sup>191</sup> <https://inkunabeln.digitale-sammlungen.de/start.html>

<sup>192</sup> Inkunabelkatalog. 8 Bände, Wiesbaden: Reichert, 1988–2021.

<sup>193</sup> <http://www.inka.uni-tuebingen.de/>

sämtlichen Metadaten, sondern auch in den Volltexten der Inkunabeln selbst recherchiert werden kann.

Eine Normierung der Drucktypen und des Textlayouts begann erst im 16. Jahrhundert, in dem mediengeschichtlich die Epoche der Alten Drucke ihren Anfang nahm. Aus der Katalogisierung der deutschen und auf deutschem Territorium erschienenen Drucke des 16., 17. und 18. Jahrhunderts entwickelte sich die retrospektive Nationalbibliografie, die im Rahmen der Arbeitsgemeinschaft Sammlung Deutscher Drucke (AG SDD) zur Verteilten deutschen Nationalbibliothek wurde (vgl. Kap. 1.2). Im Unterschied zu dem zunächst als Druckausgabe erschienenen, erst später in eine Datenbank überführten *VD16*<sup>194</sup> wurden *VD17*<sup>195</sup> und *VD18*<sup>196</sup> von Beginn an als Datenbanken konzipiert. Da zeitgleich um die Jahrtausendwende mit den Digitalisierungszentren in München und Göttingen die Digitalisierung des schriftlichen Kulturerbes begann, konnten mit der autoptischen Katalogisierung der Texte zunächst ausgewählte Schlüsselseiten, später die Werke vollständig digitalisiert werden. Auf Grundlage der retrospektiven deutschen Nationalbibliografie für das 16.–18. Jahrhundert begann die virtuelle digitale deutsche Nationalbibliothek Gestalt anzunehmen. Ihre Volltexttransformation durch Einsatz von OCR-Technologien wird in Zukunft eine Volltextsuche über das gesamte schriftliche Kulturerbe der Frühen Neuzeit ermöglichen.

Da sich die VDs auf die deutschsprachigen oder auf deutschem Territorium gedruckten Werke konzentrieren, ist für die Recherche historischer Drucke auch anderer Sprachen und Herkunft ein weiteres Portal relevant. Das *Zentrale Verzeichnis Digitalisierter Drucke (zvdd)*<sup>197</sup> ist das Nachweisportal für die in Deutschland erstellten Digitalisate von Druckwerken vom 15. Jahrhundert bis heute – insofern sind auch das 19. und 20. Jahrhundert inkludiert, auch wenn für diesen Bereich angesichts der großen Menge der Drucke sowie

---

<sup>194</sup> <https://bvbat01.bib-bvb.de/TP61/start.do?View=vd16&SearchType=2>

<sup>195</sup> <https://kxp.k10plus.de/DB=1.28/>

<sup>196</sup> <https://kxp.k10plus.de/DB=1.65/DB=1.65/>

<sup>197</sup> <https://www.zvdd.de/startseite/>

ihrer Katalogsituation Vollständigkeit nur schwer zu erzielen ist (vgl. Kap. 1.2). Über eine einheitliche Recherchemaske lässt sich nach Titeln, Autoren, Druckern, Druckorten oder Erscheinungszeiträumen suchen.

Für den deutschsprachigen Raum ist ferner *e-rara*<sup>198</sup> als Plattform für digitalisierte Drucke aus Schweizer Bibliotheken zu nennen. Nachgewiesen werden Bücher, Karten und illustrierte Materialien von den Anfängen des Buchdrucks bis ins 20. Jahrhundert.

Auch die englische Buchproduktion der Frühen Neuzeit ist in unterschiedlichen Volltextdatenbanken inzwischen ausgezeichnet dokumentiert. *Early English Books Online (EEBO)*<sup>199</sup> umfasst Bücher aus der Zeit vom Beginn der englischen Buchproduktion bis zu Shakespeare und ist vergleichbar mit VD16 und VD17. Sämtliche Werke sind im Volltext durchsuchbar und können heruntergeladen oder ausgedruckt werden.

Die frühneuzeitliche englische Textproduktion des 18. Jahrhunderts wird außerdem durch die *Eighteenth Century Collections Online (ECCO)*<sup>200</sup> repräsentiert.<sup>201</sup> Enthalten sind zwischen 1701 und 1800 vorwiegend in England erschienene Bücher, Pamphlete, Essays und Einblattdrucke sämtlicher Fachgebiete. Insbesondere für vergleichende Forschungen zwischen dem deutschen und englischen Kulturraum, beispielsweise zu Fragen der Verbreitung von Autoren, Texten und damit Wissen in der Frühen Neuzeit, bieten EEBO und ECCO einen reichen Schatz an Quellen, dessen systematische Erforschung durch die Verbindung von *close reading* und *distant reading* und den Einsatz der Möglichkeiten der Digital Humanities erst am Anfang steht.

Zentrale Gegenstände der Frühneuzeitforschung betreffen die religiösen und sozialen Diskurse im Spannungsfeld von Reforma-

---

<sup>198</sup> <https://www.e-rara.ch/>

<sup>199</sup> <https://about.proquest.com/en/products-services/eebo/>

<sup>200</sup> <https://www.gale.com/intl/primary-sources/eighteenth-century-collections-online>

<sup>201</sup> Die verlagsseitige Suchmaske wird nach Übernahme der Daten durch die SUB Göttingen leider nicht mehr angeboten.

---

tion, Gegenreformation und Dreißigjährigem Krieg. Für den Protestantismus bietet die *Digital Library of Classic Protestant Texts (DLCP)*<sup>202</sup> Volltexte von etwa 1.500 Werken von mehr als 325 protestantischen Autoren des 16. und 17. Jahrhunderts wie Martin Luther, Johannes Calvin oder Ulrich Zwingli. Neben theologischen Schriften beinhaltet DLCP auch Beichtschriften, Bibelkommentare, Streitschriften, Katechismen und liturgischen Schriften.

Das Pendant für den Katholizismus ist die *Digital Library of the Catholic Reformation (DLCR)*<sup>203</sup> mit etwa 2.000 Texten katholischer Autoren desselben Zeitraumes. Theologische Schriften sind genauso inkludiert wie Papsturkunden, Synodalbeschlüsse, Katechismen, Beichthandbücher, Bibelkommentare, religiöse Dramen, liturgische Schriften, Inquisitionshandbücher oder Andachtsbücher. In beiden Fällen handelt es sich um kombinierte Volltext- und Faksimiledatenbanken, deren Suchmasken differenzierte Recherchen innerhalb der Volltexte ermöglichen.

In der Frühen Neuzeit gewannen nach der Entdeckung des amerikanischen Kontinents die USA eine immer wichtigere Bedeutung innerhalb der transatlantischen Geschichte. Eine ausgezeichnete Quellengrundlage bietet die Volltextdatenbank *Early American Imprints 1639–1819*<sup>204</sup>. Innerhalb der *Series I: Evans, 1639–1800 (EAI I)*<sup>205</sup>, die auf der Bibliografie von Charles Evans<sup>206</sup> beruht, finden sich mehr als 36.000 digitale Ausgaben der zwischen 1639 und 1800 in Nordamerika publizierten Bücher.<sup>207</sup> Die *Series II: Shaw-*

---

<sup>202</sup> <https://search.alexanderstreet.com/tcpt>

<sup>203</sup> <https://search.alexanderstreet.com/dlcr>

<sup>204</sup> <https://www.readex.com/products/early-american-imprints1639-1819>

<sup>205</sup> <https://www.readex.com/products/early-american-imprints-series-i-evans-1639-1800>

<sup>206</sup> Evans, Charles, *American bibliography: a chronological dictionary of all books, pamphlets and periodical publications printed in the United States of America; with bibliographical and biographical notes*, New York 1903–1971.

<sup>207</sup> Wie im Falle von ECCO wird die verlagsseitige Suchmaske nach Übernahme der Daten durch die SUB Göttingen leider nicht mehr angeboten.

---

*Shoemaker, 1801–1819 (EAI II)*<sup>208</sup>, basierend auf der Bibliografie von Ralph R. Shaw und Richard H. Shoemaker<sup>209</sup>, schließt daran an und umfasst 37.000 digitalisierte Werke, die zwischen 1801 und 1819 in Nordamerika erschienen sind.

Wie diese Beispiele zeigen, ist die Epoche der Frühen Neuzeit ein „goldenes Zeitalter für die Retrodigitalisierung“<sup>210</sup>: Die durch Gutenberg eingeläutete Medienrevolution steigerte die Produktion von Schriftlichkeit in zuvor nicht gekanntem Umfang. Insofern bietet sich der Frühneuzeitforschung ein im Vergleich zu vorausgehenden Epochen außerordentlich vielfältiges Portfolio an schriftlichen Quellen. Der in der Regel gute konservatorische Zustand der Bücher dieser Jahrhunderte durch die Verwendung von Hadernpapier sowie die längst erloschenen Urheberrechte geben den Gedächtnisinstitutionen große Freiheiten in ihrer Digitalisierungsentscheidung. Die Volltextdatenbanken, in die die digitalisierten Sammlungen aufgenommen werden, werden zum wichtigsten Instrument frühneuzeitlicher Quellenkunde.

### Neuere Geschichte und Zeitgeschichte

Angesichts der seit dem 19. Jahrhundert weiter expandierenden Schriftlichkeit durch technologische Innovationen der Buchherstellung, neue Textsorten und neue soziale Gruppen von Leserinnen und Lesern stehen der Erforschung der jüngsten Epochen der Geschichte immer neue und immer mehr Quellen zur Verfügung. Begünstigend wirkt, dass OCR-Erkennung moderner Typographien weniger aufwändig ist als im Falle von Inkunabeln, frühen Drucken oder handschriftlichen Dokumenten älterer Epochen.

---

<sup>208</sup> <https://www.readex.com/products/early-american-imprints-series-ii-shaw-shoemaker-1801-1819>

<sup>209</sup> Shaw, Ralph R., Shoemaker, Richard H., *American bibliography: a preliminary checklist*, New York 1958–1965.

<sup>210</sup> Enderle, Wilfried, *Frühe Neuzeit*, in: *Clio Guide – Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften*, hrsg. von Silvia Daniel, Wilfried Enderle, Rüdiger Hohls, Thomas Meyer, Jens Prellwitz, Claudia Prinz, Annette Schuhmann, Silke Schwandt, 3. erw. und aktualisierte Aufl., Berlin 2023, <https://doi.org/10.60693/yqjz-7f44>.

Ein ähnliches Konzept wie ECCO für englische Quellen des 18. Jahrhunderts verfolgt für das 19. Jahrhundert die Volltextdatenbank *Nineteenth Century Collections Online (NCCO)*<sup>211</sup>. Im Unterschied zu ECCO umfasst sie jedoch ausgewählte internationale Sammlungen aus den Bereichen Geschichte, Politik, Literatur und darstellende Kunst. Im Einzelnen: *Asia and the West: Diplomacy and Cultural Exchange*, *British Politics and Society*, *British Theatre, Music and the Arts: High and Popular Culture*, *Europe and Africa: Commerce, Christianity, Civilization, and Conquest*, *European Literature, 1790–1840: The Corvey Collection*<sup>212</sup>, *Photography: The World Through the Lens* und *Science, Technology and Medicine, 1780–1925*. Die Quellen sind in ihren Metadaten sowie im Volltext differenziert durchsuchbar, ermöglichen einen Download und die Weiterbearbeitung.

Eine wichtige Quellengattung für das 19. und 20. Jahrhundert (und partiell auch für die Frühe Neuzeit), die digitalisiert und in Volltextdatenbanken zur Verfügung steht, sind Zeitungen, für die hier nur einige herausragende Beispiele genannt werden. Eine umfassende Übersicht dazu gibt der *Clio-Guide Zeitungen*<sup>213</sup>. Grundsätzlich ist zu beachten, dass historische Zeitungen aus der Zeit vor 1945 in der Regel ohne urheberrechtliche Schranken digitalisiert und im Open Access zugänglich gemacht werden können, während für Zeitungen aus der zweiten Hälfte des 20. Jahrhunderts Urheberrechte bestehen. Daher hängt die Nutzungsmöglichkeit von Lizenzen ab – sei es durch die lokale Bibliothek, durch den FID oder als Nationallizenz. Einen jeweils aktuellen Überblick bieten DBIS sowie die jeweiligen Lizenzinformationen der zuständigen FIDs.<sup>214</sup>

---

<sup>211</sup> <https://www.gale.com/intl/primary-sources/nineteenth-century-collections-online>

<sup>212</sup> Volltexte englischer, französischer und deutscher Literatur aus der Fürstlichen Bibliothek Corvey.

<sup>213</sup> <https://guides.clio-online.de/guides/sammlungen/zeitungen/2023>

<sup>214</sup> [Historicum.net](https://www.historicum.net/services/fid-lizenzen) bietet aktuell beispielsweise folgende Lizenzen zu Volltextdatenbanken und Zeitungen: <https://www.historicum.net/services/fid-lizenzen>.

Eine breite Titelvielfalt seit dem 16. Jahrhundert bietet der digitale Zeitungs- und Zeitschriftenlesesaal der Österreichischen Nationalbibliothek mit *ANNO – Austrian Newspapers Online*<sup>215</sup>. Das Portal *digiPress*<sup>216</sup> präsentiert den historischen Zeitungsbestand der Bayerischen Staatsbibliothek vom 17. Jahrhundert bis ins frühe 20. Jahrhundert mit regionalem Schwerpunkt in Süddeutschland. Der Zugriff auf die Quellen ist über die Titelliste, den Kalender sowie die Volltextsuche möglich. *ZEFYS - Zeitungsinformationssystem der Staatsbibliothek zu Berlin*<sup>217</sup> besitzt einen regionalen Schwerpunkt im Norden und Nordosten des deutschen Sprachraumes: Die *Amtspresse Preußens* beispielsweise umfasst zentrale preußische Tageszeitungen der Zeit zwischen 1856 und 1944, das *DDR-Pressportal* die politisch maßgeblichen Zeitungen der Deutschen Demokratischen Republik der Jahre zwischen 1945 und 1994. Im Hinblick auf eine regionale Abdeckung Deutschlands ist schließlich das Zeitungsportal Nordrhein-Westfalen, *zeit.punktNRW*<sup>218</sup>, zu nennen, das regionale Zeitungen aus den Bibliotheken und Archiven des Landes mit Volltextsuche digital bereitstellt.

Der Nachweis der Digitalisierungsprojekte zu historischen Zeitungen in Deutschland erfolgt künftig durch das *Deutsche Zeitungsportal*<sup>219</sup> als Teil der *Deutschen Digitalen Bibliothek (DDB)*.<sup>220</sup> Es wird von mehreren großen Bibliotheken getragen und bietet historische Zeitungen des 17. bis 20. Jahrhunderts mit der Möglichkeit einer Volltextsuche an.

---

<sup>215</sup> <https://anno.onb.ac.at/>

<sup>216</sup> <https://digipress.digitale-sammlungen.de/>

<sup>217</sup> <https://zefys.staatsbibliothek-berlin.de/>

<sup>218</sup> <https://zeitpunkt.nrw/>

<sup>219</sup> <https://www.deutsche-digitale-bibliothek.de/newspaper>

<sup>220</sup> Aus dem DFG-Projekt „Digitalisierung historischer Zeitungen“ (Pilotphase 2013–2015) entstanden „Empfehlungen zur Digitalisierung historischer Zeitungen in Deutschland (Masterplan Zeitungsdigitalisierung)“; dieser – wie auch weitere Praxisempfehlungen – sind auf der Seite der ZDB dokumentiert: <https://zeitschriftendatenbank.de//zeitungsdigitalisierung/>.

Für Großbritannien sind zwei Volltextdatenbanken für Zeitungen exemplarisch zu nennen: *19th Century British Library Newspapers*<sup>221</sup> mit mehr als 160 lokalen und regionalen Zeitungen aus dem Bestand der British Library aus der Zeit von 1732–1950 sowie die *Burney collection*<sup>222</sup> mit insgesamt 1.270 englischen Zeitungen und Flugschriften des 17. und 18. Jahrhunderts aus der Sammlung des Charles Burney (1757–1817). Mit enthalten sind Parlamentsveröffentlichungen, tagesaktuelle Nachrichten aus London, dem Britischen Empire, den Nachbarländern und deren Königshäusern oder Regierungen.

Für Frankreich steht beispielsweise die einflussreiche Tageszeitung *Le Monde* als Volltextdatenbank für die Jahre 1944–2000 als *FID-Lizenz*<sup>223</sup> zur Verfügung.

Vor dem Hintergrund dieser ausgewählten Volltextdatenbanken für Zeitungen einzelner Länder Europas entsteht mit *Europeana Newspapers*<sup>224</sup> ein gemeinsames Angebot, in dem europäische Zeitungen als historische Quellen künftig vergleichend erforscht werden können.

Erweitert man den Blick von Europa auf die Vereinigten Staaten von Amerika, so stehen auch hier zahlreiche Volltextdatenbanken für Zeitungen zur Verfügung. *America's Historical Newspapers*<sup>225</sup> ist eine umfassende Volltextdatenbank für digitalisierte Zeitungen der USA aus der Zeit zwischen 1690 und 1922. Sie beinhaltet neben überregional bedeutenden Titeln auch Zeitungen aus 50 Bundesstaaten sowie interessante deutsch-amerikanische Periodika. Die Volltextsuche ermöglicht die Analyse politischer und gesellschaftlicher Diskurse auf dem nordamerikanischen Kontinent in einer diachronen Perspektive.

---

<sup>221</sup> <https://nl.sub.uni-goettingen.de/collection/nlh-blh>

<sup>222</sup> <https://nl.sub.uni-goettingen.de/collection/nlh-bcn>

<sup>223</sup> <https://www.historicum.net/services/leid-lizenzen/le-monde-proquest>

<sup>224</sup> <http://www.europeana-newspapers.eu/>

<sup>225</sup> <https://www.readex.com/products/americas-historical-newspapers>

---

Die großen US-amerikanischen Tageszeitungen, *New York Times* für die Jahre 1851–2020<sup>226</sup> und *Washington Post* für die Jahre 1877–2007<sup>227</sup> stehen als *FID-Lizenzen*<sup>228</sup> zur Verfügung. Die Volltextsuche ermöglicht die Recherche innerhalb der einzeln indexierten Artikel, Kommentare, Rezensionen, Fotos, Karikaturen, Werbeanzeigen, Leserbriefe und Familienanzeigen.

Auch das Regierungshandeln des US-Kongresses ist in digitalisierten Quellen ausgezeichnet dokumentiert. Die Volltextdatenbank *U.S. Congressional Serial Set 1817–1980/1994*<sup>229</sup> enthält etwa 350.000 Publikationen aus Gremien und Arbeitsgruppen des US-Kongresses sowie von Regierungsbehörden. Teil der Datenbank sind die *American State Papers 1789–1838*, die Dokumente aus der Periode der ersten 14 US-Kongresse vor 1817 beinhalten. Die Datenbank bietet eine Fülle an Materialien zu allen Bereichen der Geschichte, Politik, Wirtschaft und Kultur der Vereinigten Staaten, aber vielfach auch darüber hinaus, sofern die Beziehungen zu anderen Ländern betroffen sind.

Trotz vielfältiger Schwierigkeiten und Restriktionen bei der Erschließung und Digitalisierung der in osteuropäischen Bibliotheken und Archiven aufbewahrten Quellen existieren inzwischen Volltextdatenbanken zur Geschichte Osteuropas. *Integrum World Wide*<sup>230</sup> ist die umfangreichste Volltextdatenbank Russlands und der GUS mit etwa 360 Millionen Quellendokumenten aus den Bereichen Politik, Kultur, Wirtschaft und Gesellschaft. Enthalten sind beispielsweise Texte aus der russischen und englischen Presse (regionale und überregionale Periodika, Monitor-Dienste von Fernsehen und Radio, Presseagenturen), Statistiken (Goskomstat), Geset-

---

<sup>226</sup> <https://pq-static-content.proquest.com/collateral/media2/documents/brochure-hnp-newyorktimes.pdf>

<sup>227</sup> <https://pq-static-content.proquest.com/collateral/media2/documents/brochure-hnp-washingtonpost.pdf>

<sup>228</sup> <https://www.historicum.net/services/fid-lizenzen/the-new-york-times-with-index-the-washington-post-proquest>

<sup>229</sup> <https://www.readex.com/products/us-congressional-serial-set-1817-1994>

<sup>230</sup> <http://www.integrumworld.com/>

zestexte, Regierungsveröffentlichungen, Patentschriften (Rospatent), bibliografische Datenbanken der Russischen Akademie der Wissenschaften (INION) oder Internetquellen.

Eine besondere Art Volltextdatenbank stellt das vom *Deutschen Historischen Museum* (DHM) und *Haus der Geschichte* (HdG) betriebene *Lebendige Museum online (LeMO)*<sup>231</sup> dar: Hier werden (Volltext-) Datenbanken musealer Objekte mit Texten, Karten, Statistiken, Bildern sowie Film- und Tondokumenten in einem attraktiven Portal präsentiert.

Das *Georg-Eckert-Institut – Leibniz-Institut für Bildungsmedien (GEI)*<sup>232</sup>, das auch für seine Forschungsaktivitäten im Bereich Volltexterkennung bekannt ist, bietet in seiner Volltextdatenbank *GEI-Digital – Die digitale Schulbuchbibliothek*<sup>233</sup> eine Volltextdatenbank für digitalisierte historische Schulbücher, die als Quellen für bildungswissenschaftliche Fragen dienen können.

### 3. Resümee und Ausblick

Volltextdatenbanken oder Bibliotheks- und Archivreisen?

Welche konkreten Nutzungsszenarien bieten historische Volltextdatenbanken für Historikerinnen und Historiker? Statt zeitintensiver Bibliotheks- und Archivreisen zu den Originalquellen, statt aufwändiger Suche nach relevanten Textstellen in Quelleneditionen oder Forschungsliteratur ermöglichen Volltextdatenbanken eine zielgenaue Identifizierung relevanter Quellen und Forschungsliteratur sowie den unmittelbaren Zugriff hierauf. Auf Grundlage digitaler Bibliotheken und vernetzter Volltextdatenbanken lassen sich quantitative Analysen historischer Kulturen und Gesellschaften in einem zuvor unbekanntem Umfang vornehmen. Beispielsweise ist denkbar, durch automatische Analyse großer Textcorpora von Primärquellen größerer Zeiträume Thesauri historischer Begriffe zu generieren und ihre Semantik differenzierter

---

<sup>231</sup> <https://www.dhm.de/lemo>

<sup>232</sup> <https://www.gei.de/>

<sup>233</sup> <https://gei-digital.gei.de/>

---

und umfassender zu analysieren, als es beispielsweise in den *Geschichtlichen Grundbegriffen* in vordigitaler Zeit möglich war.<sup>234</sup> Künftig wird nicht nur das schriftliche Kulturerbe in einem System vernetzter historischer Volltextdatenbanken zur Verfügung stehen. Es wird möglich sein, Forschungsfragen in einer Qualität und Quantität zu bearbeiten, die in analoger Zeit kaum in einer einzigen Forschergeneration hätten bearbeitet werden können.

#### Von der «Bibliothek der Bücher» zur «Bibliothek der Texte»

Die digitale Transformation verwandelt die traditionelle „Bibliothek der Bücher“ in eine „Bibliothek der Texte“<sup>235</sup>, die als Volltexte beliebig kontextualisiert werden können. Text, Mensch und Welt verschmelzen in einem prinzipiell grenzenlosen Hypertext. Texte benötigen keine Bücher mehr – sie haben sich von Büchern und Bibliotheken emanzipiert.<sup>236</sup>

Der Bibliothekskatalog der Zukunft wird nicht mehr nur bibliografische Metadaten von Forschungsliteratur und Standorte gedruckter Bücher nachweisen, sondern als Suchmaschine eine kombinierte Recherche in Metadaten und Volltexten im Textraum sämtlicher digital(isiert)en Bücher und Forschungsressourcen der jeweiligen Bibliothek ermöglichen und den direkten Zugriff auf die Quellen und Volltexte bieten. Doch nicht genug: Lokale Texträume verschmelzen mit globalen Textwelten, in denen geschichtswissenschaftliche Quellen und Forschungsliteratur eine disziplinäre Textsphäre bilden, die verwoben ist mit Textsphären anderer Disziplinen: Interdisziplinäre Volltextrecherchen im universellen

---

<sup>234</sup> Brunner, Otto u.a. (Hrsg.), *Geschichtliche Grundbegriffe: Historisches Lexikon zur politisch-sozialen Sprache in Deutschland*, Stuttgart 1972–1997. Hier werden insgesamt 122 Begriffe wie Adel, Demokratie, Herrschaft, Staat in einzelnen Artikeln analysiert. Da der Fokus auf der Zeit von 1700 bis zur Gegenwart liegt, bieten sich künftig entsprechende Analysen für Antike und Mittelalter an. Da in den Literaturwissenschaften Volltextdatenbanken ebenfalls eine zentrale Rolle spielen, lassen sich Analysen historischer Quellen vor dem Hintergrund literarischer Texte der jeweiligen Epoche durchführen.

<sup>235</sup> Stäcker, *Konversion des kulturellen Erbes für die Forschung*, S. 1ff.

<sup>236</sup> Ebd.

Text- und Datenraum global vernetzter digitaler Bibliotheken werden möglich. Bibliotheken haben aber auch in Zukunft die Aufgabe, Bücher, Texte, Informationen, Daten und Wissen für den lesenden Menschen zu strukturieren.

#### Von der «Bibliothek der Texte» zur «Bibliothek der Daten»

Die digitale Transformation verwandelt die traditionelle „Bibliothek der Bücher“ aber nicht allein in eine „Bibliothek der Texte“, sondern zugleich in eine „Bibliothek der Informationen“ und in eine „Bibliothek der Daten“. Nachdem angesichts der wachsenden Bedeutung von Forschungsdaten innerhalb digital arbeitender Wissenschaften der *Rat für Informationsinfrastrukturen (RFII)*<sup>237</sup> die Gründung der *Nationalen Forschungsdateninfrastruktur (NFDI)*<sup>238</sup> angestoßen hat, befinden sich mehrere geisteswissenschaftliche Konsortien im Aufbau.<sup>239</sup> Vertreten durch den Verband der Historiker und Historikerinnen Deutschlands (VHD) sind die Geschichtswissenschaften Teil des *NFDI4Memory-Konsortiums*<sup>240</sup>, das Forschungseinrichtungen, Gedächtnisinstitutionen und Informationsinfrastrukturen aus den historisch arbeitenden Geisteswissenschaften vereint.<sup>241</sup>

Historische Forschungsdaten („historical research data“) werden sehr breit verstanden: „Under the broad label of ‚data‘ one finds texts (handwritten or printed), images, objects, statistical databases, maps, and films, among others, each with their own specificities and challenges. Different types of data and different re-

---

<sup>237</sup> <https://rfii.de/de/start/>

<sup>238</sup> <https://www.nfdi.de/>

<sup>239</sup> Der 2020 gegründete NFDI-Verein ist eine wissenschaftsgetriebene Initiative mit dem Ziel des Aufbaus fachspezifischer Forschungsdateninfrastrukturen und wird von Bund und Ländern getragen. Mit Beginn der dritten Förderphase 2023 befinden sich 26 Konsortien aus den unterschiedlichen Wissenschaftsdisziplinen im Aufbau.

<sup>240</sup> <https://4memory.de/>

<sup>241</sup> Vgl. <https://www.nfdi.de/konsortien-nfdi4memory/>.

search methods require modular and flexible solutions that currently only exist in limited contexts.”<sup>242</sup> Kernziele von *NFDI4Memory* sind: „Linking Research, Memory Institutions and Infrastructures“, „Integrating Historical Source Criticism into Data Services“, „Network of Historically Oriented Research Communities“, „Knowledge Order for the Digital Future of the Past“, „Advancing the Analog / Digital Interface of Historical Source Material and Data“, „Generating Standards for Historical Research Data and Sustainability“, „Education and Citizen Participation“.<sup>243</sup> Ein zentrales Anliegen von *NFDI4Memory* ist es, die spezifische Tradition der Quellenkritik für das digitale Zeitalter im Sinne einer digitalen Quellenkritik weiterzuentwickeln. Auf diese Weise können aus geschichtswissenschaftlicher Perspektive im Zeitalter der Digital Humanities die „Bibliothek der Bücher“, die „Bibliothek der Texte“ und die „Bibliothek der Daten“ durchaus zusammengedacht werden.<sup>244</sup>

Neben *NFDI4Memory* sind für geschichtswissenschaftliches Arbeiten weitere NFDI-Konsortien relevant: *Text+*<sup>245</sup> als Konsortium für Sprach- und textbasierte Forschungsdateninfrastruktur, *NFDI4Culture*<sup>246</sup> als Konsortium für Forschungsdaten zu materiellen und immateriellen Kulturgütern und *NFDI4Objects*<sup>247</sup> – Forschungsdateninfrastruktur für die materiellen Hinterlassenschaften der Menschheitsgeschichte. Vergleichbar mit den früheren SSGs oder FIDs der Bibliotheken entstehen innerhalb der NFDI Informationsinfrastrukturen für Forschungsdaten, bei denen nicht nur Bibliotheken, sondern zahlreiche weitere Partner eng zusammenwirken.

---

<sup>242</sup> Vgl. <https://4memory.de/historical-research-data/>.

<sup>243</sup> Vgl. <https://4memory.de/linkage/>.

<sup>244</sup> Dass hierin kein Widerspruch besteht, zeigt auch Lauer, der Buchwissenschaft als „datenintensive Wissenschaft“ bezeichnet (Lauer, Was ist Buchwissenschaft, S. 176).

<sup>245</sup> <https://www.nfdi.de/textplus/>

<sup>246</sup> <https://www.nfdi.de/nfdi4culture/>

<sup>247</sup> <https://www.nfdi4objects.net/>

---

### Volltextdatenbanken, Volltextsuche und historisches Arbeiten

Volltextdatenbanken zählen zu den wichtigsten Instrumenten historischen Arbeitens, sei es für Quellen, Quelleneditionen oder Forschungsliteratur. Die technisch realisierte Volltexterkennung wird zum „Qualitätskriterium von digitalen Sammlungen“<sup>248</sup>. Da geschichtswissenschaftliches Arbeiten in der Regel mit einer Recherche nach Fachliteratur zu einem Thema in einer bibliografischen Datenbank beginnt, bieten historische Volltextdatenbanken die komfortable Möglichkeit, aus dieser Recherche heraus über Linkresolver direkt zum Volltext innerhalb einer Volltextdatenbank zu gelangen. Immer häufiger entwickeln sich bibliografische Datenbanken auch selbst zu Volltextdatenbanken, so dass die bibliografische Recherche unmittelbar zum Volltext der Forschungsliteratur führt. Geschichtswissenschaftliches Arbeiten bedeutet aber auch die intensive Arbeit mit Quellen. Wie Forschungsliteratur werden diese ebenfalls nicht nur als digitale Faksimiles reproduziert, sondern durch Anwendung der HTR/OCR-Technologien in Volltextdatenbanken aufbereitet.

Das Ziel, das gesamte handschriftliche und gedruckte Kulturerbe aus Antike, Mittelalter und Neuzeit digital zu faksimilieren und im Volltext zu erkennen, liegt für den deutschen Sprachraum inzwischen in greifbarer Nähe. Die für die Geschichtswissenschaft zentralen Schriftquellen werden sich künftig als fortlaufender und vernetzter Text vor den Augen der Lesenden entfalten. Es wird möglich sein, nicht nur innerhalb der Kontinente digital verfügbarer Forschungsliteratur und Forschungsdaten, sondern zugleich innerhalb des digitalisierten schriftlichen Kulturerbes zu „googeln“.

In beiden Fällen – Volltextdatenbanken für Forschungsliteratur und Volltextdatenbanken für Quellen – ermöglicht eine Volltextsuche die Formulierung von Fragen und die Gewinnung von Erkenntnissen, für die eine automatisierte Analyse großer Textmengen durch corpusbasierte Forschungsfragen grundlegende Voraussetzung ist. Volltextdatenbanken sind somit weit mehr als nur eine

---

<sup>248</sup> Hertling, Anke; Klaes, Sebastian, Volltexte für die Forschung: OCR partizipativ, iterativ und on Demand, in: O-Bib. Das offene Bibliotheksjournal 9 (2022) 3, S. 1, <https://doi.org/10.5282/o-bib/5832>.

neue Form der Publikation und Archivierung zuvor gedruckten oder genuin digital repräsentierten Wissens, indem sie differenzierte Instrumente der Erschließung und Analyse von Texten bereitstellen und neue Fragestellungen und Antworten ermöglichen. Auch in den nächsten Jahren ist mit einer kontinuierlichen Zunahme von Volltextdatenbanken zu rechnen, da nicht nur immer größere Textmengen erstmals digitalisiert sind, sondern auch viele der bisher lediglich in Image-Digitalisaten vorliegenden Sammlungen mit automatischen Texterkennungsverfahren aufbereitet werden.

Vor dem Hintergrund der skizzierten Entstehung und Charakterisierung historischer Volltextdatenbanken lassen sich bereits jetzt Szenarien der Zukunft erkennen, in denen historische Volltextdatenbanken ihre Potenziale vollends entfalten können. Insbesondere ist zu erwarten, dass die beiden Säulen geschichtswissenschaftlichen Arbeitens mit Texten – das Studium der Quellen und der Forschungsliteratur – durch immer umfassendere Volltextdatenbanken ganz neue Qualitäten ermöglichen. Zugleich werden diese Textwelten immer differenzierter miteinander verwoben – Quellen mit Quellen, Quellen mit Forschungsliteratur und Forschungsliteratur mit Forschungsliteratur. In Bezug auf den Umgang mit historischen Quellen tritt neben die Kompetenzen in traditionellen historischen Grundwissenschaften wie Paläographie oder Kodikologie die digitale Quellenkritik. Insbesondere für Forschungsprojekte im Kontext historischer Volltextdatenbanken spielen die Fächer Informatik, Informationswissenschaft und Bibliothekswissenschaft eine gleichermaßen zentrale Rolle.

#### Aufbruch aus dem Elfenbeinturm

Künftig werden sich nicht nur menschliche Historikerinnen und Historiker in diese Textwelten von Quellen und Forschungsliteratur begeben, um in individueller oder kollaborativer Forschung zu neuen Erkenntnissen zu gelangen. Je mehr Texte digitalisiert, volltexterkant und prozessierbar für weitere Bearbeitung gemacht werden, je mehr Texte als genuine Open-Access-Publikationen vorliegen, desto größere Potenziale gewinnen nicht nur Algorithmus-

basierte Analyseverfahren wie Text Mining, sondern auch KI-Technologien wie Large Language Models (LLMs).<sup>249</sup>

Bereits jetzt ist überdeutlich, dass sich (geschichts-)wissenschaftliches Arbeiten in Zukunft fundamental ändern wird. Historikerinnen und Historiker werden weiterhin forschungsrelevante Fragen auf der Basis intensiver Lektüre von Quellen und Forschungsliteratur bearbeiten. Historische Kompetenz wird zunehmend jedoch auch darin bestehen, KI-basierte oder durch den Einsatz von KI ergänzte Analysen zu prüfen und fachlich zu bewerten. Beschleunigtes *distant reading*, durch das Quellen und Forschungsliteratur schneller aufgefunden oder analysiert werden können, ermöglicht mehr Zeit für interpretatorische Arbeit im *close reading*. Geschichtswissenschaftliches Arbeiten basiert künftig nicht nur auf einer fundierten Fachkompetenz, die sich durch Anwendung klassischer hermeneutischer Verfahren auszeichnet, sondern gleichermaßen auf einer fundierten Informations-, Medien- und Datenkompetenz, die einer stetig wachsenden Menge verfügbarer wissenschaftlicher Informationen souverän begegnet, und schließlich auf einer fundierten Datenkompetenz, die mit unterschiedlichsten historischen Daten umzugehen weiß. Hierzu gehören beispielsweise Metadaten von Digitalisaten historischer Quellen oder TEI/XML-Code, mit dem Texte ausgezeichnet werden, um beispielsweise digitale Editionen zu erstellen.

### Von der traditionellen zur digitalen Quellenkritik

Innerhalb des Prozesses, den die traditionelle, analoge Geschichtswissenschaft über die Digitalisierung hin zur digitalen Transformation durchläuft, nehmen historische Volltextdatenbanken eine Schlüsselrolle ein: An ihnen werden zahlreiche Aspekte dieses Wandels exemplarisch deutlich. Selbstverständlich untersuchen auch heute noch Historikerinnen und Historiker originale historische Quellen. Sofern von diesen analogen Originalquellen jedoch

---

<sup>249</sup> Ein aktuell verbreitetes Modell ist ChatGPT: <https://chat.openai.com/>. Das Dossier "Generative KI" des Hochschulforums Digitalisierung gibt einen breiten aktuellen Einstieg in die Thematik: <https://hochschulforumdigitalisierung.de/dossier/generative-ki/>.

digitale Repräsentationen wie digitale Faksimiles oder Transkriptionen hergestellt werden, ist auch mit diesen im Forschungsprozess umzugehen. Ein weiterer Quellentypus, der für die Zeitgeschichte immer wichtiger wird, sind Quellen, die „digital born“ sind, wie beispielsweise elektronische Akten oder Postings in Sozialen Netzwerken. Da sich digitale Quellen beliebig kopieren lassen, stellt sich nicht nur die Frage nach möglichen Unterscheidungskriterien zwischen Original und Kopie im Digitalen, sondern auch die grundsätzliche Frage nach einer zeitgemäßen Quellenkritik. Eine solche „digitale Quellenkritik“ sollte eingebettet sein in eine „digitale Hermeneutik“, wie sie Andreas Fickers vorschlägt: „Es ist höchste Zeit, die digitale Hermeneutik zu entwickeln und sie zum Standard für die Ausbildung zukünftiger Historiker zu machen [...] Die Sprachen des Historikers waren lange Zeit tote Sprachen wie Latein oder Altgriechisch. Der Geschichtswissenschaftler der Zukunft muss neben diesen auch Programmiersprachen verstehen.“<sup>250</sup>

Geschichtswissenschaftliche Informations- und Medienkompetenz, Datenkompetenz sowie digitale Quellenkritik sind künftig die Säulen, auf denen die Grundsätze guter wissenschaftlicher Praxis ruhen – gerade auch im Hinblick auf die Einbindung von Methoden künstlicher Intelligenz. Geschichtswissenschaftliches Arbeiten in der digitalen Welt bietet eine Vielzahl von Perspektiven und Potenzialen, die in Gestalt historischer Volltextdatenbanken *in nuce* erkennbar sind.

### *Literaturhinweise*

Adams, Thomas R.; Barker, Nicolas, A New Model for the Study of the Book, in: A Potencie of Life. Books in Society. The Clark Lectures 1986–1987, Hrsg. von Nicolas Barker, London 1993, S. 5–43.

Blickle, Peter, Die Revolution von 1525, 4., durchgesehene und bibliografisch erweiterte Auflage, München 2004.

---

<sup>250</sup> Fickers, Andreas: Digitale Hermeneutik – Ein Update für das kritische Denken in den Geisteswissenschaften,  
<https://www.science.lu/de/forschungstrends/digitale-hermeneutik-ein-update-fuer-das-kritische-denken-den-geisteswissenschaften>.

- Bösch, Frank; Schlotheuber, Eva, Quellenkritik im digitalen Zeitalter: Die Historischen Grundwissenschaften als zentrale Kompetenz der Geschichtswissenschaft und benachbarter Fächer, in: Historische Grundwissenschaften und die digitale Herausforderung, Hrsg. für H-Soz-Kult von Rüdiger Hohls, Claudia Prinz und Eva Schlotheuber, Historisches Forum 18 (2016), S. 16–20, <https://edoc.hu-berlin.de/handle/18452/19491>.
- Darnton, Robert, What is the History of Books?, in: Daedalus 111 (1982), S. 65–83.
- Darnton, Robert, What is the History of Books?, in: Modern Intellectual History 4 (2007), S. 495–508.
- Deck, Klaus-Georg, Digital Humanities – Eine Herausforderung an die Informatik und an die Geisteswissenschaften, in: Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden, hrsg. von Martin Huber und Sybille Krämer, 2018 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 3), [https://doi.org/10.17175/sb003\\_002](https://doi.org/10.17175/sb003_002).
- DFG-Praxisregeln Digitalisierung, DFG-Vordruck 12.151, [http://www.dfg.de/formulare/12\\_151/12\\_151\\_de.pdf](http://www.dfg.de/formulare/12_151/12_151_de.pdf); aktualisierte Neufassung 2022 <https://zenodo.org/record/7435724>.
- Digital Humanities in den Geschichtswissenschaften, hrsg. von Christina Antenhofer, Christoph Kühberger und Arno Strohmeier, Wien 2024.
- Enderle, Wilfried, Frühe Neuzeit, in: Clio Guide – Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften, hrsg. von Silvia Daniel, Wilfried Enderle, Rüdiger Hohls, Thomas Meyer, Jens Prellwitz, Claudia Prinz, Annette Schuhmann, Silke Schwandt, 3. erw. und aktualisierte Aufl., Berlin 2023, <https://doi.org/10.60693/yqjz-7f44>.
- Eisermann, Falk, The Gutenberg Galaxy’s Dark Matter: Lost Incunabula, and Ways to Retrieve Them, in: Flavia Bruni und Andrew Pettegree (Hrg.), Lost Books. Reconstructing the Print World of Pre-Industrial Europe. Leiden/Boston: Brill 2016, S. 31–54.
- Engl, Elisabeth, OCR-D kompakt. Ergebnisse und Stand der Forschung in der Förderinitiative, in: Bibliothek – Forschung und Praxis 44 (2020), 2, S. 218–230, <https://doi.org/10.1515/bfp-2020-0024>.
- Ernst, Michael, Rechtliche Rahmenbedingungen der Digitalisierung kulturellen Erbes. Legal framework for digitising cultural heritage, in: Bibliotheksdienst 52 (2018) 9, S. 687–697; <http://doi.org/10.1515/bd-2018-0082>.
- Euler, Ellen, Open Access, Open Data und Open Science als wesentliche Pfeiler einer (nachhaltig) erfolgreichen digitalen Transformation der Kulturerbeinstitutionen und des Kulturbetriebes, 2018, <https://doi.org/10.11588/artdok.00006135>.

- Fabian, Bernhard, Buch, Bibliothek und geisteswissenschaftliche Forschung. Zu Problemen der Literaturversorgung und Literaturproduktion in der Bundesrepublik Deutschland, Göttingen 1983.
- Federbusch, Maria; Polzin, Christian, Volltext via OCR – Möglichkeiten und Grenzen. Beiträge aus der Staatsbibliothek zu Berlin - Preußischer Kulturbesitz 43. Berlin: Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, 2013, [http://staatsbibliothek-berlin.de/fileadmin/user\\_upload/zentrale\\_Seiten/historische\\_drucke/pdf/SBB\\_OCR\\_STUDIE\\_WEBVERSION\\_Final.pdf](http://staatsbibliothek-berlin.de/fileadmin/user_upload/zentrale_Seiten/historische_drucke/pdf/SBB_OCR_STUDIE_WEBVERSION_Final.pdf).
- Fickers, Andreas: Digitale Hermeneutik - Ein Update für das kritische Denken in den Geisteswissenschaften, <https://www.science.lu/de/forschungstrends/digitale-hermeneutik-ein-update-fuer-das-kritische-denken-den-geisteswissenschaften>, 27.04.2018.
- Gasser, Sonja, Das Digitalisat als Objekt der Begierde. Anforderungen an digitale Sammlungen für Forschung in der Digitalen Kunstgeschichte, in: Objekte im Netz. Wissenschaftliche Sammlungen im digitalen Wandel, Hrsg. von Udo Andraschke, Sarah Wagner, Bielefeld 2020, S. 261–276, <https://doi.org/10.1515/9783839455715-009>.
- Haber, Peter, Digital Past. Geschichtswissenschaft im digitalen Zeitalter, München 2011.
- Hartwig, Uwe, Evaluation von Volltextdaten mit Open-Source-Komponenten, in: O-Bib. Das offene Bibliotheksjournal 9 (2022) 4, S. 1–21, <https://doi.org/10.5282/o-bib/5888>.
- Hertling, Anke; Klaes, Sebastian, Volltexte für die Forschung: OCR partizipativ, iterativ und on Demand, in: O-Bib. Das offene Bibliotheksjournal 9 (2022) 3, S.1–11, <https://doi.org/10.5282/o-bib/5832>.
- Horstmann, Jan, Digitale Manuskriptanalyse, in: forTEXT. Literatur digital erforschen, 2018, <https://fortext.net/routinen/methoden/digitale-manuskriptanalyse>.
- Horstmann, Jan, Topic Modeling, in: forTEXT. Literatur digital erforschen, 2018, <https://fortext.net/routinen/methoden/topic-modeling>.
- Horstmann, Jan; Jan-Erik Stange, Textvisualisierung, in: forTEXT. Literatur digital erforschen, 2018, <https://fortext.net/routinen/methoden/textvisualisierung>.
- Horstmann, Jan, Stilometrie, in: forTEXT. Literatur digital erforschen, <https://fortext.net/routinen/methoden/stilometrie>.
- Huff, Dorothee; Stöbener, Kristina, Projekt OCR-BW: Automatische Texterkennung von Handschriften, in: O-Bib. Das Offene Bibliotheksjournal 9 (2022) 4, S. 1–19, <https://doi.org/10.5282/o-bib/5885>.

- Johrendt, Jochen, Digitalisierung als Chance, in: Historische Grundwissenschaften und die digitale Herausforderung, Hrsg. für H-Soz-Kult von Rüdiger Hohls, Claudia Prinz und Eva Schlotheuber, Historisches Forum 18 (2016), S. 41–42, <https://edoc.hu-berlin.de/handle/18452/19491>.
- Kann, Bettina, Hintersonleitner, Michael, Volltextsuche in historischen Texten – Erfahrungen aus den Projekten der Österreichischen Nationalbibliothek, in: Bibliothek – Forschung und Praxis 39 (2015) 1, S. 73–79, <http://doi.org/10.1515/bfp-2015-0004>.
- Klimpel, Paul, Kulturelles Erbe digital. Eine kleine Rechtsfibel, Berlin 2020, [https://www.digis-berlin.de/wp-content/uploads/2020/09/digiS\\_PKlimpel\\_Rechtsfibel.pdf](https://www.digis-berlin.de/wp-content/uploads/2020/09/digiS_PKlimpel_Rechtsfibel.pdf).
- Klimpel, Paul, In Bewegung. Die Rechtsfibel für Digitalisierungsprojekte in Kulturerbe-Einrichtungen, Berlin 2022, <https://nbn-resolving.org/urn:nbn:de:0297-zib-86485>.
- Kobel, Esther; Volp, Ulrich, Distant reading – Perspektiven einer digitalen Zeit. Eine Einführung, in: Journal of Ethics in Antiquity and Christianity 4 (2022), S. 5–10.
- Lauer, Gerhard, Was ist Buchwissenschaft, wenn sie eine Disziplin ist?, in: Archiv für Geschichte des Buchwesens 77 (2022), S. 173–178.
- McLuhan, Marshall, Die Gutenberg-Galaxis. Das Ende des Buchzeitalters, München 1968.
- McLuhan, Marshall; Powers, Bruce R., The global village: der Weg der Mediengesellschaft in das 21. Jahrhundert, Paderborn 1995.
- Mühlberger, Günter, Digitalisierung historischer Zeitungen aus dem Blickwinkel der automatisierten Text- und Strukturerkennung (OCR), in: Zeitschrift für Bibliothekswesen und Bibliographie 58 (2011), 1, S. 10–18, <http://doi.org/10.3196/186429501158135>.
- Otlet, Paul, Traité de documentation. Le livre sur le livre. Théorie et pratique, Brüssel 1934.
- Rautenberg, Ursula (Hrsg.), Buchwissenschaft in Deutschland. Ein Handbuch, Berlin - Boston 2010.
- Rehbein, Malte, Digitalisierung braucht Historiker/innen, die sie beherrschen, nicht beherrscht, in: Historische Grundwissenschaften und die digitale Herausforderung, Hrsg. für H-Soz-Kult von Rüdiger Hohls, Claudia Prinz und Eva Schlotheuber, Historisches Forum 18 (2016), S. 45–52, <https://edoc.hu-berlin.de/handle/18452/19491>.

- Renggli, Alexa, e-manuscripta.ch: Volltext – Crowdsourcing mit Qualitätssicherung, in: O-Bib. Das offene Bibliotheksjournal 5 (2018) 4, S. 58–70, <https://doi.org/10.5282/o-bib/2018H4S58-70>.
- Sahle, Patrick, Digitale Editionsformen: zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels, 3 Bände, Norderstedt 2013.
- Sahle, Patrick, Digitale Edition, in: Jannidis, Fotis; Kohle, Hubertus; Rehbein Malte (Hrsg.): Digital Humanities: Eine Einführung, Stuttgart 2017, S. 234–249.
- Schaab, Rupert, Die Flut der Texte. Notwendigkeit und Perspektiven, die Bibliothek digital zu erweitern, in: WLB-Forum 1 (2022) 1, S. 4–7, <https://doi.org/10.53458/wissteilen.vi1.2923>.
- Schmitz, Wolfgang, Grundriss der Inkunabelkunde. Das gedruckte Buch im Zeitalter des Medienwechsels, Stuttgart 2023.
- Schnitzlein, Peter, Volltextsuche, stufenloser Zoom, Open Access – Bayerische Staatsbibliothek stellt ihre neuen Digitalen Sammlungen mit rund 2,7 Millionen digitalisierten Titeln vor, in: Pressemitteilung. Bayerischen Staatsbibliothek, 21. April 2021, <https://www.bsb-muenchen.de/article/volltextsuche-stufenloser-zoom-open-access-bayerische-staatsbibliothek-stellt-ihre-neuen-digitalen-sammlungen-mit-rund-2-7-millionen-digitalisierten-titeln-vor0-4096/>.
- Schumacher, Mareike, Netzwerkanalyse, in: forTEXT. Literatur digital erforschen, 2018, <https://fortext.net/routinen/methoden/netzwerkanalyse>.
- Schumacher Mareike, Abbyy FineReader, in: forTEXT. Literatur digital erforschen, 2019; <https://fortext.net/tools/tools/abbyy-finereader>.
- Stäcker, Thomas, Konversion des kulturellen Erbes für die Forschung: Volltextbeschaffung und -bereitstellung als Aufgabe der Bibliotheken, in: O-Bib 1 (2014), S. 220–237. <http://doi.org/10.5282/o-bib/2014H1S220-237>.
- Steinhauer, Eric W., Die Bibliothek 2040 – eine Einrichtung der digitalen Transformation mit vielen Büchern?, in: Bibliothek: Forschung und Praxis 47 (2023) 1, S. 29–32, <https://doi.org/10.1515/bfp-2023-0014>.
- Weil, Stefan, Neue Frakturmodelle für Tesseract. Präsentation auf dem Kitodo Anwendertreffen 18.–19. November 2019, <https://madoc.bib.uni-mannheim.de/53748/1/2019-11-18.pdf>.
- Zotter, Hans, Erlebnisräume, gebaut aus Erinnerungen. Die Sondersammlung als Teaching Library, in: Sondersammlungen im 21. Jahrhundert: Organisation, Dienstleistungen, Ressourcen, Hrsg. von Graham Jefcoate, Jürgen Weber, Wiesbaden 2008, S. 136–144.

*Dr. Marcus Schröter ist Leiter des Dezernats Historische Sammlungen, Digitalisierung, Bestandserhaltung und Fachreferent für Geschichtswissenschaften und Buchwesen an der Universitätsbibliothek Freiburg im Breisgau. Neben seiner Lehrtätigkeit in Freiburg unterrichtet er an der Bibliotheksakademie Bayern. Seine Arbeitsschwerpunkte sind Buch- und Bibliotheksgeschichte, Editionswissenschaft, Digitalisierung und Digital Humanities sowie Didaktik der Informations- und Medienkompetenz.*

Zitation: Marcus Schröter, Clio-Guide: Volltextdatenbanken, in: Clio Guide – Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften, hrsg. von Silvia Daniel, Wilfried Enderle, Rüdiger Hohls, Thomas Meyer, Jens Prellwitz, Claudia Prinz, Annette Schuhmann, Silke Schwandt, 3. erw. und aktualisierte Aufl., Berlin 2023–2024, <https://doi.org/10.60693/6fa2-2567>