

Digital History

Studies in Digital History and Hermeneutics



Herausgegeben von
Andreas Fickers, Valérie Schafer, Sean Takats
und Gerben Zaagsma

Band 6

Digital History

Konzepte, Methoden und Kritiken Digitaler
Geschichtswissenschaft

Herausgegeben von
Karoline Döring, Stefan Haas, Mareike König
und Jörg Wettlaufer

DE GRUYTER
OLDENBOURG

Gefördert durch das Centre for Contemporary and Digital History (C²DH) der Universität Luxembourg, die Fritz Thyssen Stiftung und das Deutsche Historische Institut Paris (DHIP).



Fritz Thyssen Stiftung
für Wissenschaftsförderung

ISBN 978-3-11-075693-7

e-ISBN (PDF) 978-3-11-075710-1

e-ISBN (EPUB) 978-3-11-075714-9

ISSN 2629-4540

DOI <https://doi.org/10.1515/9783110757101>



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz. Weitere Informationen finden Sie unter <http://creativecommons.org/licenses/by/4.0>.

Library of Congress Control Number: 2022936745

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

© 2022 bei den Autorinnen und Autoren, Zusammenstellung © 2022 Karoline Döring, Stefan Haas, Mareike König und Jörg Wettlaufer, publiziert von Walter de Gruyter GmbH, Berlin/Boston. Dieses Buch ist als Open-Access-Publikation verfügbar über www.degruyter.com.

Einbandabbildung: © Stefan Haas

Satz: bsix information exchange GmbH, Braunschweig

Druck und Bindung: CPI books GmbH, Leck

www.degruyter.com

Inhaltsverzeichnis

Einleitung

Karoline Döring, Stefan Haas, Mareike König und Jörg Wettlaufer
Von Lille nach Göttingen und am Ende hinein in den digitalen (Tagungs-) Raum — 3

Digitalisierung der Geschichtswissenschaften

Torsten Hiltmann

Vom Medienwandel zum Methodenwandel

Die fortschreitende Digitalisierung und ihre Konsequenzen für die Geschichtswissenschaften in historischer Perspektive — **13**

Andreas Fickers

What the D does to history

Das digitale Zeitalter als neues historisches Zeitregime? — **45**

Tobias Hodel

Die Maschine und die Geschichtswissenschaft

Der Einfluss von *deep learning* auf eine Disziplin — **65**

Neue Methoden der Digital History

Katrin Moeller

Die Modellierung des zeitlichen Vergleichs als Kernkompetenz von Digital History?

Fachliche Voraussetzungen und digitale Konzepte — **83**

Leif Scheuermann

Über die Rolle computerbasierter Modellrechnungen und Simulationen für eine digitale Geschichte — 107

Christopher Banditt

Die Rekonstruktion der materiellen Ausstattung von Haushalten in Ostdeutschland 1980–2000

Ein Beispiel für quantitative und digitale Methoden in den Geschichtswissenschaften — 119

Florian Windhager, Eva Mayr, Matthias Schlögl und Maximilian Kaiser

Visuelle Analyse und Kuratierung von Biographiedaten — 137

Clemens Beckstein, Robert Gramsch-Stehfest, Clemens Beck, Jan Engelhardt, Christian Knüpfer, Georg Zwillling

Digitale Prosopographie

Die automatisierte Auswertung des *Repertorium Germanicum*, eines Quellenkorpus zur Geschichte geistlicher Eliten des 15. Jahrhunderts — 151

Digitale Praktiken in der Reflexion

Georg Vogeler, Christopher Pollin, Roman Bleier

„Ich glaube, Fakt ist...“:

Der geschichtswissenschaftliche Zugang zum digitalen Edieren — 171

Silke Schwandt

Geschichte visualisieren: Digitale Praktiken in der Geschichtswissenschaft als Praktiken der Wissenschaftsreflexion — 191

Stefan Haas und Christian Wachter

Visual Heuristics — 213

Moritz Feichtinger

Materialität und Praxis

Zur Analyse historischer Datenverarbeitung am Beispiel des Vietnamkrieges — 229

Melanie Althage

Potenziale und Grenzen der Topic-Modellierung mit *Latent Dirichlet Allocation* für die Digital History — 255

Digitale Lehre, Wissenschaftskommunikation und Public History

Lara Mührenberg, Ute Verstegen

Christliche Archäologie im *Inverted Classroom* und auf virtueller Exkursion — 281

René Smolarski, Hendrikje Carius, Marcus Plaul

Perspectives and Challenges of Historical Research with Citizen Participation: A Critical Reflection on the Example of “Cinema in the GDR” — 303

Mareike König und Paul Ramisch

Die twitternde Zunft

Historikertage auf Twitter (2012–2018) — 319

Oliver Kiechle

Noisy digitalia – Kommunikationslärm in Born Digital-Quellen am Beispiel von Spam im Usenet — 347

Einleitung

Karoline Döring, Stefan Haas, Mareike König und
Jörg Wettlaufer

Von Lille nach Göttingen und am Ende hinein in den digitalen (Tagungs-)Raum

Die Idee zur Tagung „Digital History – Konzepte, Methoden und Kritiken digitaler Geschichtswissenschaften“, die von der Arbeitsgruppe Digitale Geschichtswissenschaft im Verband der Historiker und Historikerinnen Deutschlands im März 2021 veranstaltet wurde und aus der dieser Sammelband hervorgegangen ist,¹ hat einen Vorlauf, der bis in das Jahr 2017 zurückreicht – nämlich zur damaligen DH-Nord Tagung mit dem Thema „(De)constructing Digital History“. Diese Tagungsserie, die vom Maison européenne des sciences de l’homme et de la société in Lille veranstaltet wird, inspirierte Jörg Wettlaufer und Mareike König dazu, eine ähnliche Veranstaltung zur Digitalen Geschichtswissenschaft in Deutschland durchzuführen. Die Idee war geboren, und schnell fanden sich in Karoline Döring und Stefan Haas Mitstreiter:innen, die halfen, die Sache gemeinsam auf den Weg zu bringen. Auch die AG Digitale Geschichtswissenschaft war gerne bereit, die Tagung in ihr Veranstaltungsprogramm aufzunehmen. Im März 2020 wäre es fast soweit gewesen: Insgesamt 21 Vortragende und 122 eingeschriebene Zuhörer:innen wollten sich in Göttingen im Historischen Gebäude der Staats- und Universitätsbibliothek drei Tage lang zu Vorträgen, Austausch und Diskussionen treffen. Doch es kam anders: Vier Tage vor Veranstaltungsbeginn erhielt das Organisationsteam von der Universität die Nachricht, dass aufgrund des damals neuartigen COVID-19-Virus die Veranstaltung nicht wie geplant stattfinden könne. Die Enttäuschung bei den Organisator:innen und allen Beteiligten war groß, die Einsicht in die Notwendigkeit einer Verschiebung angesichts der Pandemie aber ebenfalls. Da die epidemische Lage auch Monate später noch angespannt war, entschloss sich das Organisationsteam, auf ein Online-Format zu wechseln – ein Format, das neben einigen Einschränkungen zugleich viele Vorteile mit sich brachte. Für die Planung und Vorbereitung bedeutete der Wechsel allerdings, vieles wieder auf Anfang zu stellen und von Neuem zu beginnen. Der für alle Beteiligten arbeitsintensive Neubeginn hat sich indes gelohnt.

¹ Wir danken Andreas Fickers, dem Direktor des Luxembourg Centre for Contemporary and Digital History (C²DH) für die Aufnahme dieses Bandes in die von ihm mitverantwortete Reihe sowie dem Deutschen Historischen Institut Paris und der Fritz Thyssen Stiftung, die diese Publikation mit finanziellen Mitteln ermöglicht haben.

1 Konzept des Bandes

Mit der Änderung des Tagungsformats ging eine Neukonzeption der Veranstaltung einher, die sich auch sehr positiv auf die Beiträge zu diesem Sammelband ausgewirkt hat. Orientierte sich die in Göttingen geplante Präsenztagung noch an den üblichen 20 Minuten Vortrags- und zehn Minuten Diskussionszeit, so lag der Fokus der Durchführung online ganz auf Diskussion und Austausch. Das Organisationsteam hatte dazu die Beitragenden gebeten, im Vorfeld ausführliche Tagungsbeiträge allen angemeldeten Teilnehmer:innen in einem nicht-öffentlichen Bereich des Tagungsblogs² in einem frei wählbaren Format zur Verfügung zu stellen. Besonders die Respondent:innen, deren Mitwirkung ins Online-Format übernommen wurde, konnten sich damit intensiv auf die Inhalte der Sektionen vorbereiten, die nun als diskussionsintensive Roundtables und nicht mehr als vortragslastige Panels durchgeführt wurden. Vom ausführlichen schriftlichen Beitrag über Screencasts bis zu Videos auf YouTube reichte schließlich die Bandbreite des vorher verfügbar gemachten Materials.

Die Beiträge wie auch der gesamte Sammelband spiegeln wider, wie sehr beides vom veränderten Format und Konzept der Online-Tagung profitiert hat. Die Beiträger:innen wurden für die geplante Präsenzveranstaltung in einem offenen Call for Papers ausgewählt und die Einreichungen gemeinsam vom Organisationsteam und einem profilierten wissenschaftlichen Beirat³ begutachtet. Ihre Überlegungen und Ergebnisse präsentierten die Beiträger:innen auf der Online-Tagung einem zahlenmäßig gegenüber der Präsenzveranstaltung stark angewachsenen, internationalen Fachpublikum von über 450 angemeldeten Teilnehmer:innen aus Deutschland, Österreich, der Schweiz und Luxemburg. Dieses Fachpublikum ging zudem durch das vorher zur Verfügung gestellte Material gut vorbereitet und konstruktiv-kritisch in die Diskussionen. Zuletzt arbeitete das Organisationsteam noch einmal intensiv mit den Beiträger:innen an den schriftlichen Ausarbeitungen für die Publikation der Tagungsergebnisse. Die Rückmeldungen auf diesen verschiedenen Ebenen sind in die hier versammelten Beiträge eingeflossen und haben die adressierten Fragestellungen, Methoden und Projekte weiterentwickelt. Auch zwei Jahre nach der Tagung ist mit dem Erscheinen des Sammelbandes somit eine aktuelle Bilanz in den oft

² Tagungsblog Digitale Geschichte, <https://digitalhist.hypotheses.org/>.

³ Wir danken Andreas Fickers (Luxemburg), Gudrun Gersmann (Köln), Simone Lässig (Washington), Malte Rehbein (Passau), Patrick Sahle (Wuppertal), Charlotte Schubert (Leipzig) und Georg Vogeler (Graz) für ihr kritisches Auge bei der Auswahl der Beitragenden und für ihre Beteiligung in verschiedenen Rollen an der Tagung.

schnelllebigen Forschungsfeldern der digitalen Geschichtswissenschaft möglich. Wovon geht diese Bilanz aus und welche Fragen adressiert sie?

2 Digitalisierung – Digital History – Digitale Geschichtswissenschaft

Die Meinungen über die Erfolge der Digital Humanities und speziell der digitalen Geschichtswissenschaft gehen auseinander.⁴ Der vorliegende Band ist daher wie die Tagung mit dem englischen Begriff „Digital History“ überschrieben, der die größte Schnittmenge verspricht. Es handelt sich dabei um einen jungen Begriff, der erst seit etwas mehr als eine Dekade im deutschen Sprachraum verwendet wird, zunächst von Peter Haber 2009,⁵ dann auch von Wolfgang Schmale 2010⁶. Die Übernahme in den deutschsprachigen Raum zu diesem Zeitpunkt und in dieser Diktion ist kein Zufall, da nach der Etablierung der Digital Humanities einige Jahre zuvor ein Vakuum zwischen der vergleichsweise bescheiden auftretenden Historischen Fachinformatik und den scheinbar forschender daher kommenden Digital Humanities entstanden war und zugleich mit „Digital History“ ein anglophones Vorbild bereitstand, digitale Geschichtswissenschaft im deutschsprachigen Raum im Sinne der Digital Humanities zu betreiben.⁷

Im Gegensatz zur „Digital History“ hat es die „Digitale Geschichtswissenschaft“ noch nicht als Lemma in die deutschsprachige Wikipedia geschafft. Dort ist (noch) als deutsche Sprachvariante die Historische Fachinformatik verknüpft und der entsprechende Eintrag fehlt. Dafür gibt es einen Eintrag zur „Historischen Digital Literacy“,⁸ worunter die Kompetenz verstanden wird, mit Hilfe digitaler Technologien historische Informationen zu finden und Geschich-

⁴ Vgl. zum Beispiel Manfred Thaller, „Controversies around the Digital Humanities. An Agenda,“ *Historical Social Research* 37/3 (2012): 7–29. doi: 10.12759/hsr.37.2012.3.7-23.

⁵ Zunächst in Peter Haber, „Sechs Stichworte für eine Digitale Geschichtswissenschaft,“ *hist.net. Plattform für Geschichte und digitale Medien*. 24.09.2009. <https://www.hist.net/archives/3131>. Zugriff am 18.01.2022. Dann in seiner Habilitationsschrift, die 2011 veröffentlicht worden ist: Peter Haber, *Digital Past. Geschichtswissenschaft im digitalen Zeitalter* (München: Oldenbourg Verlag 2011). Allerdings schrieb er schon 2006 von „Geschichtswissenschaften im digitalen Zeitalter“, siehe Peter Haber, „Geschichtswissenschaften im digitalen Zeitalter. Eine Zwischenbilanz,“ *Schweizerische Zeitschrift für Geschichte* 56/2 (2006): 168–183. doi: 10.5169/seals-1679.

⁶ Wolfgang Schmale, *Digitale Geschichtswissenschaft* (Wien: Böhlau 2010).

⁷ Edward L. Ayers, „The Past and Futures of Digital History,“ *History News* 56/4 (2001): 5–9.

⁸ https://de.wikipedia.org/wiki/Historische_Digital_Literacy. Zugriff am 29.11.2021.

te zu schreiben. Gleichzeitig entsteht mit neuen Ausdrücken wie „computergestützte Geschichte“ eine weitere Differenzierung. Der Selbstfindungsprozess ist folglich noch nicht abgeschlossen.

Sichtbarer Meilenstein der deutschsprachigen Entwicklung war die Gründung der Arbeitsgruppe „Digitale Geschichtswissenschaft“ im Rahmen des Verbands deutscher Historiker und Historikerinnen im Jahre 2012.⁹ Digitale Geschichtswissenschaft grenzte sich damit zur älteren Fachinformatik ab und öffnete sich zugleich – mit einigen Einschränkungen – den Digital Humanities und ihren Methoden und Fragestellungen. Ein neues, weites und an den Rändern offenes Forschungs-, Praxis- und Methodenfeld bildete sich, das allein schon durch technische Neuerungen weiterhin im Fluss ist und sein wird.

Die historische Forschung und Lehre haben sich in den letzten Jahren durch die Digitalisierung von Quellen, Methoden und Forschungsumgebungen, innerhalb derer Geschichtswissenschaft durchgeführt, produziert, diskutiert und verbreitet wird, tiefgreifend verändert.¹⁰ Massendigitalisierungsprojekte ermöglichen einen zeit- und ortsunabhängigen Zugang zu Quellen und Literatur. Kommerzielle und OpenSource-Programme stehen bereit, um mittels qualitativer und/oder quantitativer Datenanalyse verschiedene methodische Verfahren zur Analyse und Interpretation dieser Quellen anzuwenden. Die Fachinformation und -kommunikation hat sich überwiegend ins Netz verlagert und schließt partizipative Medien ein. Die Bandbreite an digitalen Lehrmethoden hat stark zugenommen, während die Online-Präsentation von Forschungsergebnissen und Public-Science-Projekten den Dialog und das aktive Einbinden der breiten Öffentlichkeit in den Forschungsprozess ermöglichen. Lehrstühle zur Digital History werden eingerichtet (zum Beispiel in Berlin, Bamberg und Bielefeld), die Historikertage haben selbstverständlich digitale Sektionen und fast jedes neue Forschungsprojekt hat einen (wenn auch manchmal kleinen) digitalen Anteil. Auch die Narrative unterliegen zunehmend weitreichenden Veränderungen und Modifikationen. Alles zusammengenommen hat erkennbar weitreichende Auswirkungen auf Konzeption, Theorie und Praxis der Geschichtswissenschaft,

9 Für allgemeine Informationen über die Arbeitsgruppe, ihr Profil, ihre Ziele, Mitglieder und Arbeitsweise siehe <https://www.historikerverband.de//arbeitsgruppen/ag-digitale-gw/die-ag-digitale-gw.html> sowie das Blog der AG <https://digigw.hypotheses.org/>. Zugriff am 29.11.2021.

10 Vgl. Mareike König, „Digitale Methoden in der Geschichtswissenschaft. Definitionen, Anwendungen, Herausforderungen,“ *BIOS – Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen* 1/2 (2017): 7–21. doi: 10.3224/bois.v30i1-2.02; Guido Koller, *Geschichte digital. Historische Welten neu vermessen* (Stuttgart: Kohlhammer 2016); Manfred Thaller, *From History to Applied Science in the Humanities* (= Historical Social Research. Supplement 29 (2017) und Gerben Zaagsma, „On Digital History,“ *BMGN – Low Countries Historical Review* 128 (2013): 3–29.

wobei der soziale Sammelpunkt zunehmend der letztgenannte Faktor wird: In der digitalen Geschichtswissenschaft kommt heute, und das zeigt auch das Spektrum dieses Sammelbandes, ähnlich zu den Digital Humanities eine *Community of Practice* zusammen.

Auch wenn sich die Digital History inzwischen einen zunehmend respektablen Platz im Fach erworben hat, bleibt sie in den Augen vieler Kritiker:innen ein „ewiges Versprechen“¹¹. In den USA ist dieser Eindruck, dass digitale Geschichtswissenschaft zu viel versprochen und zu wenig geliefert habe, 2017 im Whitepaper „History and Argument“¹² offen diskutiert worden. Für den deutschsprachigen Raum zieht dieser Sammelband nun eine Bilanz, indem er Veränderungen in der Art und Weise, wie Geschichtsforschung durchgeführt und kommuniziert wird, auf der einen Seite und neue Objekte, Methoden und Werkzeuge der Analyse und ihre Auswirkungen auf der anderen Seite thematisiert. Der Sammelband ist daher themen- und epochenübergreifend angelegt und behandelt drei größere Fragenkomplexe:

Konzepte: Zum ersten soll das Potential digitaler Methoden ausgelotet und kritisch beleuchtet werden. Der Fokus liegt dabei auf dem Beitrag, den digitale Methoden für die Entwicklung von historischen Argumenten und Interpretationen leisten können.

Methoden: Zum zweiten wird danach gefragt, wie sich geschichtswissenschaftliche Forschung, ihre Themen, Praktiken sowie Kommunikations- und Publikationskulturen im digitalen Zeitalter verändern und wie das Verhältnis von digitalen und klassisch hermeneutischen Methoden neu zu bestimmen ist.

Kritiken: Zum dritten werden Auswirkungen des digitalen Wandels auf die Lehre und auf eine Geschichtswissenschaft, die sich als Public History verstehen möchte, thematisiert.

Wie schon mit der Tagung selbst strebt das Organisationsteam, das in den vergangenen Monaten zum Herausgeber:innenteam des Sammelbandes geworden ist, einen offenen und kritischen Dialog von digital arbeitenden und nicht-digital arbeitenden Wissenschaftler:innen an. Oft forschen diese zum selben Thema und richten den Blick aus unterschiedlichen Perspektiven auf die gleichen Fragen: Können digitale Methoden bei der Beantwortung der Kernfragen

11 Cameron Blevins, „Digital History’s Perpetual Future Tense,“ in: *Debates in the Digital Humanities 2016*, hg. v. Matthew K. Gold und Lauren F. Klein (Minneapolis: University of Minnesota 2016), 308–324, <http://dhdebates.gc.cuny.edu/debates/text/77>. Zugriff am 29.11.2021.

12 Stephen Robertson und Lincoln A. Mullen, „Digital History and Argument“, (white paper presented at the workshop „Arguing with Digital History“ at James Mason University. Roy Rosenzweig Center for History and New Media, 13.11.2017), <http://rrchnm.org/wordpress/wp-content/uploads/2017/11/digital-history-and-argument.RRCHNM.pdf>. Zugriff am 29.11.2021.

des historischen Forschens einen Beitrag leisten oder bestätigen sie (bislang) nur schon Bekanntes und analog gewonnene Erkenntnis, wenn auch auf einer anderen empirischen Basis? Ist digitale Geschichtswissenschaft die Antwort auf die Frage nach der Beziehung traditioneller Geschichtswissenschaft (und ihrer Methoden) zu den Digital Humanities? Gibt es einen qualitativen Unterschied zwischen analogen und digitalen Methoden in Bezug auf Interpretation und Einordnung von Ergebnissen oder lösen sich diese schließlich doch im Diskurs oder gar in den Praktiken auf? Die Beiträge des Sammelbandes zeigen Konzepte zum Umgang mit Digitalität und der digitalen Transformation, bestimmen den Ort digitaler Methoden innerhalb der Geschichtswissenschaften neu und erkunden kritisch aktuelle wie zukünftige Arbeitsfelder digitaler Geschichte.¹³ Bei aller fachlich notwendigen Differenz ist ihnen gemeinsam, dass sie die Auswirkungen dieses zum Teil gar nicht mehr so neuen Zusammentreffens von Geschichtswissenschaft und Informatik auf die gegenwärtige Arbeit von Historiker:innen zu erfassen, zu bewerten und kritisch zu begleiten suchen. Die Diskussion, die darüber im Tagungsraum online begonnen wurde, soll mit den publizierten Beiträgen weitergeführt werden. Wir wünschen eine anregende Lektüre!

Bibliographie

- Ayers, Edward L. „The Pasts and Futures of Digital History.“ *History News* 56/4 (2001): 5–9.
- Blevins, Cameron. „Digital History’s Perpetual Future Tense“. In: *Debates in the Digital Humanities 2016*, hg. v. Matthew K. Gold und Lauren F. Klein, 308–324. Minneapolis: University of Minnesota 2016. <http://dhdebates.gc.cuny.edu/debates/text/77>. Zugriff am 29.11.2021.
- Haber, Peter. „Geschichtswissenschaften im digitalen Zeitalter. Eine Zwischenbilanz“. *Schweizerische Zeitschrift für Geschichte* 56/2 (2006): 168–183, doi: 10.5169/seals-1679.
- Haber, Peter. „Sechs Stichworte für eine Digitale Geschichtswissenschaft“. *hist.net. Plattform für Geschichte und digitale Medien*. 24.09.2009. <https://www.hist.net/archives/3131>. Zugriff am 18.01.2022.
- Haber, Peter. *Digital Past. Geschichtswissenschaft im digitalen Zeitalter*. München: Oldenbourg Verlag 2011.
- König, Mareike. „Digitale Methoden in der Geschichtswissenschaft. Definitionen, Anwendungen, Herausforderungen.“ *BIOS – Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen* 1/2 (2017): 7–21, doi: 10.3224/bois.v30i1-2.02.
- Koller, Guido. *Geschichte digital. Historische Welten neu vermessen*. Stuttgart: Kohlhammer 2016.

13 Haber, *Digital Past*.

- Robertson, Stephen und Lincoln A. Mullen. „Digital History and Argument.“ White Paper presented at the workshop „Arguing with Digital History“ at James Mason University. Roy Rosenzweig Center for History and New Media, 13.11.2017, <http://rrchnm.org/wordpress/wp-content/uploads/2017/11/digital-history-and-argument.RRCHNM.pdf>. Zugriff am 29.11.2021.
- Schmale, Wolfgang. *Digitale Geschichtswissenschaft*. Wien: Böhlau 2010.
- Thaller, Manfred. „Controversies around the Digital Humanities. An Agenda.“ *Historical Social Research* 37/3 (2012): 7–29, doi: 10.12759/hsr.37.2012.3.7-23.
- Thaller, Manfred. *From History to Applied Science in the Humanities* (= Historical Social Research. Supplement 29 (2017)).
- Zaagsma, Gerben. „On Digital History“. *BMGN – Low Countries Historical Review* 128 (2013): 3–29.

Digitalisierung der Geschichtswissenschaften

Torsten Hiltmann

Vom Medienwandel zum Methodenwandel

Die fortschreitende Digitalisierung und ihre Konsequenzen für die Geschichtswissenschaften in historischer Perspektive

Abstract: This article assesses the current process of digitalisation by comparing it with previous media changes in order to better understand its practical implications for historical scholarship. In doing so, it shows that we are not just dealing with another turn, but with a fundamental change in the very nature of communication itself. By distinguishing between data and their representation and, building on this, between medial and conceptual digitality, the article demonstrates that digitalisation proceeds in different phases and discusses the resulting consequences for historical scholarship, which are reflected in the need for a strong data culture, the critical appropriation of new methods, and the further refinement of our theories.

Keywords: digital history, media change, methodological change, epistemology, digitization, data culture, theory

Zusammenfassung: Ziel des Beitrags ist es, den aktuellen Prozess der Digitalisierung durch den Vergleich mit früheren medialen Entwicklungen und Umbrüchen in seiner Tragweite und Ausprägung einzuordnen und daraus Schlussfolgerungen für die Geschichtswissenschaften zu ziehen. Der Beitrag zeichnet nach, dass wir es dabei nicht nur mit einem weiteren Turn zu tun haben, sondern mit einer grundlegenden Veränderung unserer Kommunikation insgesamt. Durch die Unterscheidung zwischen den Daten und ihrer Repräsentation und darauf aufbauend zwischen medialer und konzeptioneller Digitalität, macht der Beitrag deutlich, dass sich auch die Digitalisierung in unterschiedlichen Adaptionsstufen vollzieht. Er diskutiert die sich daraus ergebenden Konsequenzen für die Geschichtswissenschaften und fasst diese in der Notwendigkeit einer starken Datenkultur, der kritischen Aneignung neuer Methoden und der Weiterentwicklung unserer Theorien zusammen.

Schlagwörter: Digital History, Medienwandel, Methodenwandel, Epistemologie, Digitalisierung, Datenkultur, Theorie

Die Geschichtswissenschaften sind einem konstanten Wandel unterworfen. Dies gilt auch für den Grad ihrer Digitalisierung und die Art und Weise, mit der sie digitale Medien und Methoden in ihre Arbeitsweisen und ihre Fachkultur integrieren. In einer kurzfristigeren Perspektive von nur einigen Jahren mögen diese Entwicklungen kaum sichtbar sein. Weitert man den Blick jedoch auf die letzten 20 bis 30 Jahre, so erscheint der bereits zurückgelegte Weg erheblich und nicht von der Hand zu weisen. Während frühe Projekte schon in den 1970er Jahren für ihre Forschungen auf Großrechenanlagen zurückgriffen, hat sich der Computer dann in den 1980er und vor allem in den 1990er Jahren auch in der Breite durchgesetzt und ist zu einem alltäglichen Arbeitsinstrument geworden.¹ Seitdem müssen Texte nicht erst genauer vorgeplant werden, bevor man sie an der Schreibmaschine tippt, sondern werden direkt am Computer entworfen.² Der direkte Austausch zwischen den Forschenden erfolgt nicht mehr per Brief, sondern mindestens per Email, und statt die zentralen Nachschlagewerke, Enzyklopädien, Inventare und Kataloge als gedruckte Werke in der Bibliothek zu konsultieren, geschieht dies mittlerweile in den meisten Fällen über entsprechende digitale Angebote im World Wide Web. Ähnliches gilt für die Fachkommunikation und das Rezensionswesen. Dabei wurde Diensten wie H-Soz-Kult und Clio-online, die unsere Fachkultur heute erheblich mitprägen, anfänglich mit viel Skepsis begegnet.³ Zu erwähnen ist schließlich ebenso die zunehmende Rolle von Open-Access-Publikationen⁴ und die Umstellung des Erwerbs fachhis-

¹ Zu den verschiedenen Stufen der frühen Aneignung computerbasierter Technologien durch die Geschichtswissenschaften vgl. Manfred Thaller, „Entzauberungen: Die Entwicklung einer fachspezifischen historischen Datenverarbeitung in der Bundesrepublik [1990],“ *Historical Social Research/Historische Sozialforschung. Supplement* 29 (2017): 178–192.

² Zur empfundenen Erleichterung der ersten Nutzer:innen siehe Bernd Gregor, „Computergestützter Satz: Protokoll der Entstehung dieses Buches,“ in *Computerfibel für die Geisteswissenschaften: Einsatzmöglichkeiten des Personal Computers und Beispiele aus der Praxis*, hg. v. Manfred Krifka (München: C. H. Beck, 1986), 265–282.

³ Vgl. Thomas Meyer, „H-Soz-Kult und Clio-online: Von der Mailingliste zur Online-Community,“ *Bibliotheksdienst* 52/3–4 (2018): 185–195, hier: 186, doi: 10.1515/bd-2018-0026; Rüdiger Hohls, „Historische Fachkommunikation im Umbruch: Ein Rückblick auf die Anfänge des H-Net und von H-Soz-u-Kult,“ in *20 Jahre Arbeitsgemeinschaft Geschichte und EDV*, hg. v. Jörn Kobes, Kai Ruffing und Wolfgang Spickermann (Gutenberg: Computus Druck Satz & Verlag, 2013), 147–166, hier: 156–157.

⁴ Diese Entwicklung wird nicht zuletzt durch die Politik der Drittmittelgeber befördert. Die Deutsche Forschungsgemeinschaft informiert über ihre diesbezügliche Position auf einer neu eingerichteten Website, siehe „Open Access“, Deutsche Forschungsgemeinschaft, https://www.dfg.de/foerderung/programme/infrastruktur/lis/open_access/index.html. Zugriff am 14.12.2021.

torischer Zeitschriften durch die Universitätsbibliotheken und Institute von Print-Ausgaben auf Online-Publikationen.

Die angeführten Entwicklungen haben das historische Arbeiten und die historische Wissensproduktion in ihren Bedingungen bereits grundlegend verändert, ohne dass dies in der Breite des Faches wirklich tiefgreifend reflektiert wurde.⁵ Doch scheint die Entwicklung in den letzten zwei bis drei Jahren noch einmal deutlich an Dynamik gewonnen zu haben, was den Eindruck eines allgemeinen Umbruchs weiter stärkt. Dieser Eindruck resultiert nicht zuletzt aus der Einrichtung gleich mehrerer Professuren mit einem Schwerpunkt in den digitalen Geschichtswissenschaften (Berlin, Bielefeld, Bamberg, Gießen, Trier)⁶, denen sicherlich noch weitere folgen dürften. Im Rahmen des Aufbaus einer Nationalen Forschungsdateninfrastruktur wurde auf Initiative des Verbands der Historikerinnen und Historiker Deutschlands eine eigene Konsortialinitiative für die historisch arbeitenden Geisteswissenschaften aufgebaut, an der mehr als 80 Institutionen beteiligt sind, um den Weg der Geschichtswissenschaften hin zu einer stärker datengetriebenen Forschung in einem digitalen Forschungsumfeld zu begleiten und voranzubringen.⁷ Im Oktober 2021 erschienen darüber hinaus die ersten Beiträge der neuen und auch in ihren Präsentationsmöglichkeiten hochinnovativen Zeitschrift „Journal of Digital History“, die mit ihrer Einbindung von Daten und Methoden auf unterschiedlichen auswählbaren und ausführbaren Text- und Codelayern für die gesamten Digital Humanities wie für die zukünftigen Geschichtswissenschaften wegweisend ist.⁸ Und nicht zuletzt sei hier auch das erhebliche Tagungsaufkommen in den letzten beiden Jahren genannt, das nicht allein mit der Erfahrung der Pandemie zu be-

5 In den Digitalen Geschichtswissenschaften hat sich die Auseinandersetzung lange vor allem auf die Möglichkeiten des Internet bzw. des World Wide Web und der Hypertextualität konzentriert, vgl. Peter Haber, *Digital Past: Geschichtswissenschaft im digitalen Zeitalter* (München: Oldenbourg Verlag, 2011), doi: 10.1515/9783486712339; Wolfgang Schmale, *Digitale Geschichtswissenschaft* (Wien: Böhlau, 2010); sowie zuletzt Christian Wachter, *Geschichte digital schreiben: Hypertext als non-lineare Wissensrepräsentation in der Digital History* (Bielefeld: transcript, 2021), doi: 10.1515/9783839458013. Allgemeinere Perspektiven öffnete zuletzt Adam Crymble, *Technology and the Historian: Transformations in the Digital Age* (Urbana: University of Illinois Press, 2021). Vgl. zudem den Beitrag von Andreas Fickers in diesem Band.

6 Die meisten dieser neu geschaffenen Professuren sind in einer Liste von Patrick Sahle zu den im Bereich der Digital Humanities ausgeschriebenen Professuren einsehbar: „Professuren für Digital Humanities“, *DHdBlog* (06.02.2019), <https://dhd-blog.org/?p=11018>. Zugriff am 15.12.2021.

7 „NFDI4Memory“, *NFDI4Memory*, <https://4memory.de/>. Zugriff am 15.12.2021.

8 „Journal of Digital History“, University of Luxembourg, C2DH, <https://journalofdigitalhistory.org/en>. Zugriff am 15.12.2021.

gründen ist. Neben der schon für 2020 geplanten „Digital-History“-Tagung⁹, für die dieser Beitrag entstand, und der ebenfalls ursprünglich für 2020 geplanten „Data for History“-Konferenz¹⁰ fanden im deutschsprachigen Raum allein im Jahr 2021 mindestens zwölf weitere geschichtswissenschaftliche Tagungen und Workshops mit einem digitalen Schwerpunkt statt.¹¹

Man könnte auch sagen, dass wir uns gerade in einer sehr dynamischen Phase in der Entwicklung der digitalen Geschichtswissenschaften und der (weiteren) Digitalisierung unseres Faches als solches befinden. Eine ähnliche Wahrnehmung von digitalem Aufbruch gab es in den Geschichtswissenschaften zwar immer wieder; dies tut der derzeitigen Entwicklung aber keinen Abbruch. Vielmehr stellt sich die Frage, wie man diese aktuelle Entwicklung genauer einordnen und auf dieser Grundlage proaktiv mitgestalten kann. Genau darum soll es in diesem Beitrag gehen.

Dafür soll im Folgenden die Perspektive weiter geöffnet und der zugrundeliegende Prozess der Digitalisierung als solcher historisiert werden. Aus der Perspektive eines Mediävisten sollen hierfür zum Vergleich zwei grundlegende mediale Umbrüche herangezogen werden, die die mittelalterliche Kultur und Gesellschaft und damit auch die unsere nachhaltig beeinflussten und mit denen die Digitalisierung immer wieder verglichen wird: der Übergang von der Mündlichkeit zur Schriftlichkeit sowie die Entstehung des Buchdrucks. Ziel ist es, auf der Basis dieses Vergleichs die Grundlagen der Digitalisierung und die damit verbundenen Konsequenzen zu identifizieren und am Beispiel der Geschichtswissenschaften die sich daraus ergebenden Entwicklungslinien nachzuvollziehen, um schließlich zu prüfen, welche Schlussfolgerungen sich konkret für die digitalen Geschichtswissenschaften selbst ableiten lassen. Der Beitrag soll damit zeigen, dass wir es bei der Digitalisierung nicht einfach nur mit einem wei-

⁹ „Digital History: Konzepte, Methoden und Kritiken digitaler Geschichtswissenschaften“, <https://digitalhist.hypotheses.org/>. Zugriff am 15.12.2021.

¹⁰ „Data for History: Modeling Time, Places, Agents“, Data for History Consortium, <https://d4h2020.sciencesconf.org/>. Zugriff am 15.12.2021.

¹¹ Dabei ging es bei den meisten dieser Veranstaltungen um ganz spezifische Anwendungsfelder wie Gender History (<https://www.gw.uni-jena.de/digitalgenderhistory>), Rechtsgeschichte (<https://www.lhlt.mpg.de/2255460/event-21-03-04-digital-methods-and-resources-in-legal-history>), Baltische Geschichtsforschung (<https://www.balt-hiko.de/baltische-historikertreffen/>), Historische Bildungsforschung (https://www.uni-muenster.de/EW/ife/arbeitsbereiche/hist_bildf/digi-hbf.html), historische Kartographie (https://kbl.badw.de/fileadmin/user_upload/Files/KBL/Aktuelles/KBL_Flyer_WEB.pdf), zeithistorische Portale (<https://www.zlb.uni-jena.de/veranstaltungen/zeithistorische+portale+und+digitale+sammlungen>) oder auch frühneuzeitliche Universitäts- und Gelehrtengeschichte (<https://pcp-on-web.htwk-leipzig.de/project/pcp-reloaded/de/>). Zugriff jeweils am 15.12.2021.

teren Turn zu tun haben, sondern mit einer ganz grundlegenden Veränderung unserer gesamten Kommunikations- und Arbeitsweisen als Historiker:innen, aber auch als Gesellschaft insgesamt, die weitreichende Konsequenzen hat. Unter näherer Betrachtung lassen sich in diesem Prozess, was die Reichweite und Nutzung der damit einhergehenden Veränderungen und Potentiale angeht, zwei verschiedene mögliche Adaptionstufen unterscheiden, von denen wir, wie im Folgenden gezeigt werden soll, die zweite erst allmählich zu beschreiten beginnen.

1 Struktureller Vergleich von Verschriftlichung, Erfindung des Buchdrucks und Digitalisierung

Es fehlt natürlich nicht an Beiträgen, die die Digitalisierung in die Medien- und Kommunikationsgeschichte¹² und übergreifender in die Geschichte von Wahrnehmung und Denken einordnen.¹³ Dabei wird sie immer wieder, ganz zu Recht, nach Sprache, Schrift und Buchdruck als vierte „kognitive Revolution“ bezeichnet, als technische Innovation, die nicht nur quantitative Auswirkungen hatte, sondern die Art des Denkens selbst veränderte.¹⁴ Gerade um das Jahr 2000 wurde hierzu häufig der Vergleich mit der Einführung des Buchdrucks herangezogen.¹⁵ Der Bezug zum Buchdruck scheint dabei ein Allgemeinplatz ge-

12 Für einen Überblick der Literatur vgl. zum Beispiel Ramón Reichert, „Theorien digitaler Medien,“ in *Digital Humanities: Eine Einführung*, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein (Stuttgart: J.B. Metzler, 2017), 19–34, doi: 10.1007/978-3-476-05446-3_3. Als Klassiker auf diesem Feld der Mediengeschichte ist hier zu nennen: Marshall McLuhan, *The Gutenberg Galaxy: The Making of Typographic Man* (Toronto: University of Toronto Press, 1962), der das Ende der Gutenberg-Galaxis zwar in den elektronischen Medien wie dem Fernsehen sah, in seiner grundsätzlichen Einordnung des medialen Wandels und der Buchkultur (und ihrer Ablösung) jedoch immer wieder als grundlegende Referenz dient.

13 In den letzten Jahren stark diskutiert: Armin Nassehi, *Muster: Theorien der digitalen Gesellschaft* (München: C. H. Beck, 2019). Siehe ferner Florian Süßenguth (Hg.), *Die Gesellschaft der Daten: Über die digitale Transformation der sozialen Ordnung* (Bielefeld: transcript, 2015); Michael Giesecke, *Von den Mythen der Buchkultur zu den Visionen der Informationsgesellschaft* (Frankfurt am Main: Suhrkamp, 2002).

14 Stevan Harnad, „Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge,“ *Public-Access Computer Systems Review* 2/1 (1991): 39–53; Luciana Floridi, *The 4th Revolution. How the Infosphere is Reshaping Human Reality* (Oxford: Oxford University Press, 2014).

15 Wolfgang Frühwald, „Das Ende der Gutenberg-Galaxis: Über den Einfluß des Mediums auf den Inhalt wissenschaftlicher Publikationen,“ *Leviathan* 26/3 (1998): 305–318; Jürgen Fröhlich,

worden zu sein, so dass 2015 in der Zusammenfassung des von Wolfgang Schmale herausgegebenen Bandes „Digital Humanities“ von dem „in diesem Kontext sonst obligatorischen Satz, die Digitalisierung sei in ihrer Bedeutung nur mit der Erfindung des Buchdrucks vergleichbar,“¹⁶ gesprochen wird.

2 Erfindung des Buchdrucks mit beweglichen Lettern

Tatsächlich ist dieser Vergleich naheliegend. Durch den Buchdruck¹⁷ konnten Texte mechanisch in deutlich größerer Zahl und damit wesentlich günstiger produziert werden als durch ihre handschriftliche Vervielfältigung. Zugleich war damit eine erhebliche Qualitätssteigerung verbunden, da der Text nun weitgehend identisch reproduziert werden konnte.¹⁸ Dies führte dazu, dass Wissen und Informationen deutlich schneller und weiter kommuniziert und geteilt werden konnten und damit letztlich eine neue „Sphäre öffentlicher Kommunikation (ob als Literatur, Poesie, Aufklärungsschriften oder Manifeste)“ entstand.¹⁹ Spezifische Texte und Informationen konnten nun erheblich größere Gruppen erreichen, was im 16. Jahrhundert zu einschneidenden politischen (Bauernkriege) und religiösen Auswirkungen (Reformation) führte.²⁰ Aber auch neue Formen von religiöser und herrschaftlicher Propaganda wurden auf diese Weise er-

„Was aber soelchs himmelkorn bedüt / ist Gott allein zewüssen“: Phänomene struktureller Amnesie in medialen Umbruchzeiten (Flugblatt und Internet),“ in *Mediaevistik und neue Medien*, hg. v. Klaus van Eickels, Ruth Wesselbaumer und Ingrid Bennewitz (Ostfildern: Thorbecke, 2004), 49–66. Für eine breitere Einordnung siehe vor allem Horst Wenzel, *Mediengeschichte vor und nach Gutenberg* (Darmstadt: Wissenschaftliche Buchgesellschaft, 2007).

16 Stefan Zahlmann, „Die Digital Humanities und der Mensch: Ein Kommentar,“ in *Digital Humanities: Praktiken der Digitalisierung, der Dissemination und der Selbstreflexivität*, hg. v. Wolfgang Schmale, *Historische Mitteilungen – Beihefte*, Bd. 91 (Stuttgart: Steiner, 2015), 175–180, hier: 175.

17 Bei der weiteren Verwendung des Begriffs „Buchdruck“ ist stets der Buchdruck mit beweglichen Lettern gemeint.

18 Die voluminöse Studie von Michael Giesecke, *Der Buchdruck in der frühen Neuzeit: Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien* (Frankfurt am Main: Suhrkamp, 1991) gilt weiterhin als Standardwerk zur Geschichte des Buchdrucks.

19 Manfred Faßler und Wulf Halbach, *Geschichte der Medien* (München: Fink, 1998), 10.

20 Elizabeth L. Eisenstein, *The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early Modern Europe. Volumes I and II* (Cambridge [England], New York: Cambridge University Press, 1979).

probt.²¹ Um diese neue Technologie herum entstanden mit den Druckereien und vor allem mit dem Verlagswesen ganz neue Strukturen, die die Verbreitung von Texten, Informationen und Wissen über die nächsten Jahrhunderte grundlegend prägten.²²

Die Analogien zu den Entwicklungen seit den 1990er Jahren liegen auf der Hand. Mit der Etablierung des Internets hat die Möglichkeit zur Veröffentlichung und Verbreitung von Texten, wie allgemein von Informationen und Wissen, einen vergleichbaren Entwicklungssprung erlebt, nicht zuletzt durch Formate wie Blogs und die später aufgekommenen sozialen Medien.²³ Texte und andere Medien können nun innerhalb einer gegebenen Infrastruktur kostenlos oder mit geringen Kosten ein potentiell unbegrenztes Publikum erreichen, was zu einer grundlegenden Veränderung der Kommunikations- und Publikationsstrukturen und der mit diesen verbundenen Institutionen wie Verlage, Fernsehanstalten und so weiter führte.²⁴ Wie durch die Erfindung und Verbreitung des Buchdrucks in der Frühneuzeit haben die neuen Medien auf Grundlage des Internets zu tiefgreifenden politischen und gesellschaftlichen Umbrüchen geführt (oder sind noch dabei).²⁵

Dennoch beschreiben diese Beschleunigung der Kommunikation, die Vermischung von Produzent:innen und Konsument:innen sowie die Vergrößerung der Gruppe der potentiellen Kommunikationsteilnehmer:innen, die mit dem Internet und den sozialen Medien assoziiert werden, nur einen Effekt der Digitalisierung, jedoch nicht die Digitalisierung selbst. Deren eigentlicher Kern und deren eigentliche Tragweite liegen woanders – ähnlich wie dies beim Buchdruck und bei der Verschriftlichung der Fall ist. Denn die Art und Weise, wie Informa-

21 Vgl. u. a. Robert W. Scribner, „Flugblatt und Analphabetentum: Wie kam der gemeine Mann zu reformatorischen Ideen?“ in *Flugschriften als Massenmedium der Reformationszeit*, hg. v. Hans-Joachim Köhler (Stuttgart: Klett-Cotta, 1981), 65–76; Jan-Dirk Müller, „Publizistik unter Maximilian I.: Zwischen Buchdruck und mündlicher Verkündigung,“ in *Sprachen des Politischen: Medien und Medialität in der Geschichte*, hg. v. Ute Frevert und Wolfgang Braungart (Göttingen: Vandenhoeck und Ruprecht, 2004), 95–122.

22 Andreas Würzler, *Medien in der Frühen Neuzeit* (München: De Gruyter Oldenbourg, 2009); Hans Erich Bödeker, „Aufklärung als Kommunikationsprozeß,“ *Aufklärung* 2/2 (1987): 89–111.

23 Vgl. Newton Key, „Crowdsourcing the Early Modern Blogosphere,“ in *Historyblogosphere: Bloggen in den Geschichtswissenschaften*, hg. v. Peter Haber und Eva Pfanzelter (München: Oldenbourg Wissenschaftsverlag, 2013), 101–118, doi: 10.1524/9783486755732.101 sowie zu den sozialen Medien allgemein: Jan-Hinrik Schmidt und Monika Taddicken (Hg.), *Handbuch Soziale Medien* (Wiesbaden: Springer, 2017).

24 Siehe etwa für den Bereich der Politik und der politischen Institutionen Martin Emmer, „Soziale Medien in der politischen Kommunikation,“ in *Handbuch Soziale Medien*, hg. v. Jan-Hinrik Schmidt und Monika Taddicken (Wiesbaden: Springer, 2017), 81–99.

25 Vgl. Felix Stalder, *Kultur der Digitalität* (Berlin: Suhrkamp, 2016), 203–278.

tionen gespeichert und vermittelt werden, bleiben jeweils gleich. Beim Buchdruck geschieht dies, wie schon zuvor in den Handschriften, über die Schrift. Im Internet wiederum, wie schon zuvor in der Elektronischen Datenverarbeitung und beim Home-PC, über die Repräsentation von Informationen in Form von Daten. Kurz, das Internet und die Vernetzung unserer Gesellschaft sind nicht die Digitalisierung, sondern bauen auf dieser auf, ganz so, wie auch der Buchdruck auf dem Prozess der Verschriftlichung aufbaut. Die Wurzeln des eigentlichen Medienwandels liegen also tiefer.

3 Verschriftlichung

Möchte man den tiefgreifenden Wandel, der mit der Digitalisierung verbunden ist, wirklich fassen, muss man meines Erachtens noch ein Stück weiter bis zum Prozess der Verschriftlichung selbst zurückgehen, wie er sich im Mittelalter besonders im 12. und 13. Jahrhundert vollzogen hat.²⁶ Zwar hatte der Buchdruck mit der mechanischen Vervielfältigung von Texten in Hinblick auf deren Verbreitung und Nutzung weitreichende Folgen, die Grundlagen der Kommunikation blieben jedoch die gleichen wie zuvor bei der Handschriftenproduktion: Die Texte wurden schriftlich vermittelt. Der mediale Bruch von der Oralität zur Schriftlichkeit hingegen betraf die Art und Weise selbst, wie Texte, Aussagen und Informationen repräsentiert, gespeichert und vermittelt werden. Hier änderte sich der grundlegende Modus der Kommunikation und führte zu deutlich weitreichenderen gesellschaftlichen und kulturellen Veränderungen.

So ist die mündliche Kommunikation dadurch geprägt, dass der Akt der Kommunikation an eine bestimmte Situation, das heißt an einen bestimmten Ort, eine bestimmte Zeit und bestimmte Teilnehmer:innen gebunden ist. Mit anderen Worten, Sprechende und Hörende müssen an der Interaktion direkt und zur gleichen Zeit teilnehmen. Die mündlich übertragenen Inhalte werden dabei allein als Sprache im Gedächtnis der Teilnehmenden gespeichert und sind an

²⁶ Zum Verschriftlichungsprozess im Mittelalter siehe die Arbeiten der beiden Sonderforschungsbereiche SFB 231 „Träger, Felder, Formen pragmatischer Schriftlichkeit im Mittelalter“ in Münster und SFB 321 „Übergänge und Spannungsfelder zwischen Mündlichkeit und Schriftlichkeit“ in Freiburg, sowie einführend Hagen Keller, „Vom ‚heiligen Buch‘ zur ‚Buchführung‘: Lebensfunktionen der Schrift im Mittelalter,“ *Frühmittelalterliche Studien* 26 (1992): 1–31; Michael T. Clanchy, *From Memory to Written Record: England 1066–1307*, 3. Aufl. (Malden: Wiley-Blackwell, 2013). Prozesse der Verschriftlichung gab es natürlich auch schon in den frühen Hochkulturen, im antiken Griechenland und in anderen Weltregionen, vgl. für einen Überblick Peter Stein, *Schriftkultur: Eine Geschichte des Schreibens und Lesens* (Darmstadt: Primus, 2006).

diese gebunden, die sie memorieren. Wenn diese Inhalte dann wieder aufgerufen werden, geschieht dies wiederum in einem spezifischen Kontext und – wenn nicht durch besondere Techniken abgefedert (Reime, Bewegungen)²⁷ – werden die so gespeicherten Inhalte immer wieder (wenn auch nicht unbedingt bewusst) dem neuen Kontext angepasst, das heißt den Umständen entsprechend aktualisiert. Mündlich kodierte und übertragenes Wissen wird damit meist nicht „wortwörtlich“ vermittelt, sondern verändert sich. Unwichtig Gewordenes oder für unwichtig Gehaltenes wird ausgelassen und „vergessen“, Wichtiges wird tradiert und gegebenenfalls angepasst und erweitert. Dies hat zu nicht weniger komplexen Formen von Gedächtniskulturen mit eigenen kulturellen Praktiken geführt, die jedoch aufgrund ihrer Kommunikationsweise grundsätzlich anders funktionieren.²⁸

Bei der schriftlichen Kommunikation hingegen löst sich die Kommunikationssituation auf, indem die Aussagen und Informationen und damit auch das darin vermittelte Wissen in ein Objekt transferiert und dadurch materialisiert und verstetigt werden.²⁹ Das können Inschriften in Steinen sein, aber auch Schrift auf Pergament oder Papier, ob sie dort nun per Hand oder mit Drucklettern aufgebracht wurde. Dadurch werden die Inhalte der Kommunikation, das heißt zumeist die jeweiligen Texte und das damit verbundene Wissen, fixiert und können in genau dieser Art und Weise, also in dem gleichen Wortlaut, über Raum und Zeit hinweg vermittelt und in neuen Kontexten identisch wieder aufgerufen werden. Auch diese Form der Kodierung, Speicherung und Vermittlung von Wissen ist eine eigene Kulturtechnik oder auch Technologie, die zuerst erworben und erlernt werden muss und mit Schreibmaterialien und -utensilien, aber auch mit Archiven und Bibliotheken ihre ganz eigene Infrastruktur verlangt, damit das so kodierte Wissen gesammelt und gepflegt werden kann.

Die Konsequenzen dieses Umbruchs können nicht hoch genug eingeschätzt werden. Er veränderte nicht nur Umfang und Reichweite des einzelnen Kommunikationsaktes. Die damit abgelegten Inhalte konnten so zugleich gesammelt, verglichen, neu strukturiert und schriftlich erweitert werden. Mit der Schrift

27 Walter Ong, *Orality and Literacy: The Technologizing of the Word* (London: Methuen, 1982), 33–36, 56–66.

28 Jan Assmann, *Das kulturelle Gedächtnis: Schrift, Erinnerung und politische Identität in frühen Hochkulturen* (München: C. H. Beck, 1992); Harald Weinrich, „Gedächtniskultur – Kulturgedächtnis,“ in *Mediengebrauch und Erfahrungswandel: Beiträge zur Kommunikationsgeschichte*, hg. v. Detlev Schöttker (Göttingen: Vandenhoeck und Ruprecht, 2003), 37–52.

29 Konrad Ehlich, „Funktion und Struktur schriftlicher Kommunikation,“ in *Schrift und Schriftlichkeit: Ein interdisziplinäres Handbuch internationaler Forschung*, hg. v. Hartmut Günther und Otto Ludwig, Bd. 1. (Berlin, New York: De Gruyter Mouton, 2008), 18–41, doi: 10.1515/9783110111293.1.1.18.

konnten Texte und Wissen gespeichert und immer wieder neu abgerufen werden, was schließlich zu zahlreichen gesellschaftlichen und kulturellen Innovationen führte, sei es in der Wissensorganisation, im Rechtswesen (prospektives Recht), in der Verwaltung, in der Wirtschaft oder in der Organisation des politischen Lebens sowie politischer Institutionen. Durch diese neue Form der Speicherung von Text und Sprache haben sich letztlich auch die Sprache und das Denken selbst verändert.³⁰ Dies alles, indem Sprache und Text durch Schrift in Objekte ausgelagert und darin fixiert wurden.

4 Digitalisierung

Der aktuelle Prozess der Digitalisierung ist in seiner Fundamentalität wie in seinen Konsequenzen dem der Verschriftlichung vergleichbar. Auch hier verändert sich ganz grundlegend die Art und Weise, wie Aussagen, Informationen und Wissen kommunikativ vermittelt und gespeichert werden. Denn die Schrift, um bei diesem Beispiel zu bleiben, wird hier nicht mehr in einem Objekt fixiert, sondern in digitalen Daten kodiert. Dabei handelt es sich wiederum um eine eigene Kulturtechnik und Technologie, die noch einmal völlig anders funktioniert als die Schrift selbst. Auch hier braucht es spezifische Utensilien und Infrastrukturen, was in erster Linie eine elektronische Rechenmaschine, also den Computer meint, der diese digitalen Daten erzeugen und bearbeiten kann. Ohne diesen kann Digitalität nicht funktionieren.

Unter digitalen Daten kann man dabei im weitesten Sinne eine spezifische Repräsentation von Informationen verstehen, die dafür in ihre kleinsten Teile zerlegt und in einzelnen Zeichen repräsentiert werden. Dabei geht die zeichenhafte Repräsentation der Informationen deutlich weiter als bei Schriftsystemen, bei denen ebenso jeder Buchstabe einen bestimmten Laut beziehungsweise jede als Wort zusammengefügte Zeichenfolge oder jedes Schriftzeichen eine mehr oder weniger bestimmte Bedeutung kodieren. Denn während Texte in der Schrift als Zeichensystem materiell festgeschrieben, vermittelt und durch Gelesenwerden wieder dekodiert werden, löst das Digitale durch eine weitere zeichenhafte Repräsentationsstufe die Bindung an eine spezifische materielle Fixierung auf und virtualisiert diese. Dafür wird zum Beispiel die Schrift nach be-

30 Zu den Konsequenzen der Schriftlichkeit vgl. Jack Goody und Ian Watt, „The Consequences of Literacy,“ *Comparative Studies in Society and History* 5 (1963): 304–345, doi: 10.1017/S0010417500001730; Ong, *Orality and Literacy*; Jack Goody, *The Logic of Writing and the Organization of Society* (Cambridge: Cambridge University Press, 1987).

stimmten formalen Regeln weiter kodiert, etwa in Zeichensätzen wie UTF-8, wo beispielsweise der Buchstabe A durch die Hexadezimalzahl 41 wiedergegeben wird, die dann zur konkreten Speicherung und Verarbeitung innerhalb des Rechners wiederum mit 01000001 in einer spezifischen Abfolge von Nullen und Einsen, genauer gesagt von Zuständen als Binärcode repräsentiert wird. Das heißt, im Digitalen ist die Schrift, die wir auf dem Monitor oder auf Ausdrucken sehen, nicht identisch mit der Information, die am Ende abgespeichert wird, sondern nur *eine* mögliche Darstellungsform und Interpretation der letztlich in Binärcode hinterlegten Daten.³¹

Anders als Schrift können die in dieser Weise kodierten Daten zwar materiell fixiert werden, müssen es aber nicht. So kann ich meine als Binärcode abgelegten Daten zum Beispiel als Abfolge unterschiedlicher Vertiefungen in CD-Roms einbrennen oder auch, dann als Schriftzeichen repräsentiert, auf Papier ausdrucken. Sie können aber auch so gespeichert werden, dass sie an gleicher Stelle immer wieder verändert und neu abgespeichert werden können. Auf der Festplatte zum Beispiel werden die unterschiedlichen Zustände 0 und 1 als Magnetisierungen gespeichert, auf einer SSD (*Solid State Disk*) als elektrische Ladungen (die übrigens nur ungefähr zehn Jahre vorhalten). Im Rechenwerk des Computers selbst werden die Nullen und Einsen bzw. deren Abfolge als unterschiedliche Ladungen verarbeitet und neu zusammengesetzt und können so im Rechner oder über Netzwerke zwischen Rechnern in Form unterschiedlicher Spannungen transportiert werden. Kurz gesagt, Informationen werden im Digitalen als distinkte Daten repräsentiert, welche wiederum in einer Abfolge von zwei unterschiedlichen Zuständen kodiert werden, die ebenso materiell fixiert als auch einfach nur als Sequenz flüchtiger elektrischer Spannungen existieren können.³²

Die Speicherung, Verarbeitung und Anzeige von Texten in Form von Schrift ist nur eine mögliche Nutzung der digitalen Repräsentationsform. Schon daher ist die Digitalisierung nicht einfach nur eine Erweiterung der Schriftlichkeit. Denn im Rahmen der Digitalität kann letztlich alles, was in diskreten Werten als Daten repräsentiert werden kann, auch als Daten gespeichert und verarbeitet werden. Es spielt also keine Rolle, ob es sich dabei um einen in Schrift kodierten Text handelt, um Bilder, Töne, Videos oder materielle Objekte. Am Ende

³¹ Fotis Jannidis, „Zahlen und Zeichen,“ in *Digital Humanities: Eine Einführung*, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein (Stuttgart: J. B. Metzler, 2017), 59–67, doi: 10.1007/978-3-476-05446-3_5.

³² Harald Klink, „Aufbau des Computers und Vernetzung,“ in *Digital Humanities: Eine Einführung*, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein (Stuttgart: J. B. Metzler, 2017), 35–58, doi: 10.1007/978-3-476-05446-3_4.

werden all diese unterschiedlichen Modi nach dem gleichen Prinzip und in der gleichen Form als Daten repräsentiert und damit gespeichert, berechnet, transportiert oder auf dieser Grundlage wieder ausgegeben.³³ So werden zum Beispiel bei Bildern, um ein weiteres Beispiel zu nennen, diese bei der Erfassung als Rastergrafiken in einzelne Bildpunkte oder Pixel zerlegt und je nach Farbraum (Graustufen, RGB, CMYK etc.) jedem dieser Pixel einer oder mehrere (Farb-)Werte zugeordnet, die dann wiederum als Zahlen und am Ende als Binär-code kodiert und gespeichert werden.³⁴

Während Schrift also in einer ganz konkreten Form (zum Beispiel Drucktypen oder Schreiberhand) auf ein Objekt angebracht bzw. darin eingeschrieben wird und von dort wiederum vom Menschen als Schriftcode erkannt und verarbeitet, das heißt gelesen wird, werden im Digitalen bestimmte Informationen (wie der Text, aber auch Angaben zu dessen Darstellung wie Schriftart oder Schriftgröße) in mehreren Bearbeitungsstufen als Zeichen kodiert, die vom Menschen nur noch mit Hilfe einer Maschine und entsprechenden Programmen nach bestimmten Regeln (Formalisierung) interpretiert werden können.



Abb. 1: Die Stufen der Kodierung einer Webseite von ihrer spezifischen Ausgabe für Desktoprechner bis zum Binärcode am Beispiel der Webseite <https://de.wikipedia.org/wiki/Krim> (Zugriff am 23.03.2022). Abbildung von Torsten Hiltmann.

³³ In Medientheorien wird dieses Phänomen mit dem Begriff der Medienkonvergenz beschrieben; vgl. Friedrich A. Kittler, *Grammophon, Film, Typewriter* (Berlin: Brinkmann & Bose, 1986), 7; Reichert, „Theorien digitaler Medien“, 23–24.

³⁴ Malte Rehbein, „Digitalisierung“, in *Digital Humanities: Eine Einführung*, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein (Stuttgart: J. B. Metzler, 2017), 179–189, doi: 10.1007/978-3-476-05446-3_12.

Einen in ein materielles Objekt eingeschriebenen Text kann man daher kaum ändern, ohne in diesem Objekt Spuren zu hinterlassen (zum Beispiel Streichungen, herausgerissene Seiten und so weiter). Im Digitalen hingegen ist die spezifische Verbindung zwischen Information und Materialität, zwischen dem Inhalt und der Form seiner Darstellung aufgehoben und wird bei jeder Darstellung neu berechnet, so dass der Inhalt hier ohne weitere Spuren immer wieder verändert und neu arrangiert werden kann.

Aus dem Gesagten, das die Konsequenzen der Digitalisierung für unseren Umgang mit Informationen nur anreißen kann, sollte deutlich geworden sein, dass wir es im Rahmen des Digitalen mit einer ganz neuen Form von Kommunikation zu tun haben, bei der sich die Grundlagen der Vermittlung und Speicherung von Informationen noch einmal ganz grundsätzlich von denen der Mündlichkeit und der Schriftlichkeit unterscheiden – und damit auch unsere Möglichkeiten, mit diesen Informationen umzugehen. Dabei zeichnen sich digitale Daten als digitale Repräsentation von Informationen zusammenfassend durch folgende Eigenschaften aus:

- Informationen werden nicht in einzelne materielle Objekte eingeschrieben, sondern in Informationsinfrastrukturen als Binärcode (0, 1) repräsentiert; sie werden elektronisch verarbeitet und oft nur flüchtig als unterschiedliche Spannungen bzw. Ladungen gespeichert, wodurch sie rasch verändert, aber auch schnell verloren gehen oder zerstört werden können.
- Die Speicherung von Information und die konkrete Form ihrer Darstellung, anders gesagt: die Ebene der Daten (Datenebene) und die Ebene ihrer Präsentation (Präsentationsebene), sind distinkt, da jede menschenlesbare Darstellung der Informationen erst durch eine maschinengestützte Interpretation der Daten erzeugt werden muss, wozu es stets einer spezifischen Infrastruktur bedarf, womit die Bedingungen der Darstellung und damit auch deren jeweilige Form veränderlich sind.
- Durch diese besondere Repräsentationsform der Informationen verschwimmen im Digitalen auf der Datenebene die Grenzen zwischen unterschiedlichen medialen Formen (Text, Bilder, Video, Ton, materielles Objekt, und andere) beziehungsweise werden aufgehoben.
- Digitale Daten können durch das Kopieren des Codes im Prinzip beliebig häufig identisch und verlustfrei reproduziert, übertragen und an anderer Stelle gespeichert werden.
- Die digitalen Daten können darüber hinaus in einem vernetzten System wie dem Internet beliebig oft miteinander verbunden, gespeichert, bereitgestellt, kopiert, aggregiert, getrennt und in unterschiedlicher und immer wieder neuer Form und Zusammensetzung dargestellt werden.

- Die Bearbeitung erfolgt dabei maschinell auf der Grundlage von Algorithmen, womit auch sehr umfangreiche Datenmengen unterschiedlicher Modalität und Provenienz automatisch verarbeitet, verändert oder ausgewertet werden können.

Entscheidend für uns ist, dass damit auf der Datenebene anders mit Informationen umgegangen werden kann als in der analogen Schriftlichkeit, wo für jede Änderung, Umstrukturierung oder Erweiterung der materielle Datenträger modifiziert oder sogar ein neuer geschaffen werden muss. Im Digitalen hingegen werden Informationen flexibel gespeichert und können grundsätzlich ohne großen Aufwand verändert werden. Anders ausgedrückt: In einer Bibliothek sind alle Bücher (aber auch Videos, Schallplatten und CDs) als einzelne singuläre Objekte erfasst und gespeichert, zu denen man meist nur über einen Katalog und entsprechende Signaturen Zugang findet. Jeder dieser Datenträger ist nur für sich zugänglich, jeder Eingriff in die dort hinterlegten Informationen hinterlässt Spuren. Im Digitalen hingegen sind all diese Informationen (ob als Texte, Bilder oder Musik) letztlich in der gleichen Weise als digitale Daten kodiert und können damit nicht nur auf der Ebene der Objekte, sondern auch auf der Ebene der Inhalte selbst potentiell beliebig ausgelesen, verknüpft und in unterschiedlichsten Kombinationen abgefragt, analysiert und verändert werden.

5 Medienwandel zwischen medialer Aneignung und konzeptioneller Nutzung

Bevor wir jedoch danach fragen, was sich daraus für die Digitalisierung in den Geschichtswissenschaften ableiten lässt, müssen wir zunächst noch einen weiteren Aspekt des medialen Wandels näher betrachten. Denn dieser vollzieht sich selbstverständlich nicht auf einmal, sondern in verschiedenen Etappen. Dabei lassen sich anscheinend sowohl hinsichtlich des Schriftgebrauchs wie bei der Nutzung des Buchdrucks zwei ganz ähnliche Adaptionsschritte oder auch Verwendungsformen unterscheiden, die auch in Hinblick auf die Digitalisierung eine Rolle zu spielen scheinen.

So differenziert Wulf Oesterreicher zwischen einer *Verschriftung* und einer *Verschriftlichung* und meint damit, kurz zusammengefasst, zum einen eine me-

diale und zum anderen eine konzeptionelle Verwendung der Schriftlichkeit.³⁵ Unter dem Prozess der Verschriftung wird dabei schlicht die Transkodierung von Äußerungen vom phonischen ins graphische Medium verstanden, ohne dass sich dabei in Hinblick auf Text und Sprache konzeptionell etwas ändert. Gemeint sind damit beispielsweise Schwörformeln oder die Aufzeichnung von Zeugenaussagen, die ganz im Duktus der Mündlichkeit festgehalten und in diesem Sinne auch gebraucht werden. Man könnte ebenso frühmittelalterliche Verträge nennen, in denen der Rechtsakt nur schriftlich festgehalten, nicht aber auch vollzogen wurde.³⁶

Die Verschriftlichung hingegen beschreibt Prozesse, bei denen sich durch den Wechsel des Mediums zugleich konzeptionell etwas ändert, wenn dadurch zum Beispiel neue Sprachmittel und Gattungsformen entstehen, sich das verwendete Vokabular ausdifferenziert und erweitert und zunehmend komplexe Ausdrucksformen mit spezifischeren, eindeutigeren Bedeutungen, gebraucht werden (zum Beispiel bei Gesetzestexten).³⁷ Damit sind vor allem Texte gemeint, die gar nicht mehr zum mündlichen Vortrag, sondern zum Lesen gedacht sind. Das heißt, hier geht es um die Entwicklung der eigentlichen Schriftsprache sowie ganz neuer, aus dem Medium heraus entstehender Kommunikationspraktiken und Kulturtechniken wie zum Beispiel der Verwendung von Kapitelüberschriften, Inhaltsverzeichnissen und Registern bei der Sammlung und Organisation größerer Informationsbestände.³⁸ Schaut man sich die Innovationen und veränderten sozialen Praktiken an, die durch die erweiterte Verwendung von Schriftlichkeit entstanden sind, so sind diese vor allem mit dieser zweiten Adaptionstufe der Verschriftlichung verbunden. Erst durch die schriftliche Ausarbeitung komplexer Verträge und Gesetze, auf deren wortgetreue Einhaltung nun geachtet und auf die immer wieder zurückverwiesen werden kann, konnten komplexe Rechtssysteme mit weiteren neuen Textgattungen entstehen (zum Beispiel Verfahrensordnungen, Rechtskommentare). In der Wirtschaft waren nun neue Entwicklungen möglich, da mit der Buchführung neue Steuerungssysteme und über Schuldbriefe und Schecks neue Finanzflüsse eingeführt werden konnten, die wiederum zur Grundlage neuer Institutionen wie den Ban-

35 Wulf Oesterreicher, „Verschriftung und Verschriftlichung im Kontext medialer und konzeptioneller Schriftlichkeit,“ in *Schriftlichkeit im frühen Mittelalter*, hg. v. Ursula Schäfer (Tübingen: Narr, 1993), 267–292.

36 Josef Hartmann, „Urkunden,“ in *Die archivalischen Quellen: Mit einer Einführung in die historischen Hilfswissenschaften*, hg. v. Friedrich Beck und Eckart Henning, 3. Aufl. (Köln, Weimar und Wien: Böhlau, 2004), 9–39, hier: 28.

37 Oesterreicher, „Verschriftung und Verschriftlichung,“ 272.

38 Ivan Illich, *Im Weinberg des Textes: Als das Schriftbild der Moderne entstand* (München: C. H. Beck, 2010), 99–120.

ken wurden. In Konsequenz dieser konzeptionellen Schriftlichkeit sind zudem neue Formen der Literatur (zum Beispiel Romane), die Briefkultur oder auch das Konzept von Autor:innenschaft möglich geworden. All das ist erst im Rahmen einer konzeptionellen Schriftlichkeit entstanden, die der Logik des neuen Mediums folgt, sich dessen Eigenschaften zunutze macht und das Medium auch in diesem Sinne verwendet.

Ähnliches lässt sich beim Buchdruck beobachten, der seinerseits den Zwischenschritt über die Inkunabeln nahm, womit die Drucke vor dem Jahr 1500 bezeichnet werden. Denn das Ziel der frühen Buchdrucker war zunächst, mit der neuen Technik die gebräuchlichen Handschriften nachzuahmen, nur eben effizienter. Medienbedingte Innovationen erfolgten erst später. Es ging also auch hier anfangs darum, die neue Technologie in die hergebrachte Verwendungslogik, also das Reproduzieren von Handschriften zu integrieren. Erst zum Beginn des 16. Jahrhunderts verselbstständigte sich der Buchdruck, als aus dem Medium selbst heraus neue Nutzungsformen und Praktiken entstanden, wie etwa neue, ganz eigene Schrifttypen, neue Formate und vor allem das Frontispiz als Titelseite³⁹, das über Buchinhalt, Drucker und Verleger informiert. Auch neue Publikationsformate wie die Einblattdrucke für die politische Kommunikation haben sich erst in dieser Zeit entwickelt. Wir können auch hier in einem ersten Schritt eine Aneignung des neuen Mediums oder einer neuen Technologie feststellen, welche diese zunächst in die bekannten Gebrauchslogiken integriert, bevor es in einem weiteren Schritt tatsächlich zu konzeptionell neuen Verwendungsweisen kommt. Und es scheint ganz so, dass sich ein ähnliches Muster auch in Bezug auf die Digitalisierung beobachten lässt. Diese Beobachtungen sollen im Folgenden dazu dienen, die Entwicklungen in den digitalen Geschichtswissenschaften einzuordnen.

6 Die Digitalisierung der Geschichtswissenschaften

Wenn wir uns mit der Geschichte der Digitalen Geschichtswissenschaften befassen, wird schnell deutlich, dass wir hier in einer langen Tradition stehen. Denn

³⁹ Zur Geschichte der Frontispize siehe Margaret M. Smith, *The Title-Page: Its Early Development 1460–1510* (London: The British Library & Oak Knoll Press, 2000) sowie im Detail Ursula Rautenberg, „Die Entstehung und Entwicklung des Buchtitelblatts in der Inkunabelzeit in Deutschland, den Niederlanden und Venedig – Quantitative und qualitative Studien,“ *Archiv für Geschichte des Buchwesens* 62 (2008): 1–105.

was wir heute als Digitalisierung verstehen, hat bereits einen Vorlauf, der bis in die 1960er Jahre zurückreicht. Dabei ist die Entwicklung der Digitalen Geschichtswissenschaften stets eng mit der jeweiligen technologischen und gesellschaftlichen Entwicklung verbunden. Zwar lassen sich schon für die Zeit vor 1990 international über 700 Aufsätze und Publikationen zu der Frage nachweisen, wie man Computer und computerbasierte Methoden in den Geschichtswissenschaften nutzen kann.⁴⁰ Jedoch mussten hierfür alle Quellen und Informationen zunächst aufwendig in digitale Daten umgewandelt werden, um diese dann tatsächlich auch mit einem Rechner prozessieren zu können. Mit der Verbreitung des Internets und der immer tiefergreifenden Digitalisierung von Kultur und Gesellschaft hat sich auch für die Geschichtswissenschaften der Rahmen noch einmal grundlegend geändert. Entsprechend soll es im Folgenden nur um die Zeit nach 1990 gehen, für die sich analog zum oben gesagten zwei unterschiedliche Phasen der Digitalisierung unterscheiden lassen.

6.1 Medialer Wandel und das Primat der Präsentation

Die erste Phase lässt sich als eine Phase der medialen Digitalisierung beschreiben. So sind am Ende der 1990er Jahre und zu Beginn der 2000er Jahre gleich mehrere Projekte und Initiativen entstanden, deren Ziel es war und zum Teil auch heute noch ist, historisches Kulturgut wie auch eigene Text- und Datensammlungen zu digitalisieren und Nutzer:innen (womit zumeist auch Historiker:innen gemeint sind) über das Internet zugänglich zu machen.

Bereits 1997 und damit zeitgleich mit dem Start der digitalen Bibliothek Gallica der Bibliothèque nationale de France⁴¹ und ganze acht Jahre vor der Veröffentlichung von Google Books begann das Münchener Digitalisierungszentrum der Bayerischen Staatsbibliothek mit der Digitalisierung von Handschriften.⁴²

⁴⁰ Für frühe Beiträge vgl. etwa Jerome M. Clubb und Howard Allen, „Computers and Historical Studies,“ *Journal of American History* 54/3 (1967): 599–607, doi: 10.2307/2937409; Vern L. Bullough, „The Computer and the Historian: Some Tentative Beginnings,“ *Computers and the Humanities* 1/3 (1967): 61–64, doi: 10.1007/BF00119888; oder auch Carl August Lückerrath, „Prolegomena zur elektronischen Datenverarbeitung im Bereich der Geschichtswissenschaft,“ *Historische Zeitschrift* 207 (1968): 265–296, doi: 10.1524/hzhz.1968.207.jg.265.

⁴¹ Emmanuelle Bermès, „La BNF face au numérique: De nouveaux objets patrimoniaux“, *Revue de la Bibliothèque nationale de France* 61 (2020): 163–167.

⁴² Markus Brantl und Astrid Schoger, „Das Münchener Digitalisierungszentrum zwischen Produktion und Innovation,“ in *Information, Innovation, Inspiration: 450 Jahre Bayerische Staatsbibliothek*, hg. v. Rolf Griebel und Klaus Ceynowa (München: Saur, 2008), 253–280, doi: 10.1515/9783598440892.4.253.

Ein Jahr später publizierten die Regesta Imperii ihre ersten Regesten digital auf CD-Rom und begannen 2001 damit, alle bereits erschienenen Bände zu retrodigitalisieren und in digitaler Form über das Internet bereitzustellen.⁴³ 2001 war zugleich der Beginn des sehr ambitionierten und weitreichenden Projekts der Manuscripta Mediaevalia, das Handschriftendigitalisate und vor allem Handschriftenkataloge aus deutschen Bibliotheken erschloss.⁴⁴ Ebenfalls im Jahr 2001 wurde das Prometheus-Bildarchiv gegründet⁴⁵. 2004 begannen die Monumenta Germaniae Historica mit der Digitalisierung und Bereitstellung ihrer Editionen als dMGH über das Internet.⁴⁶ Diese Entwicklung setzte sich in den letzten Jahren mit immer weiteren Angeboten fort, begleitet von zentralen Portalen wie der Deutschen Digitalen Bibliothek⁴⁷, der Europeana⁴⁸, dem Archivportal-D⁴⁹, dem Deutschen Zeitungsportal⁵⁰ oder dem Handschriftenportal⁵¹, die die von verschiedenen Institutionen bereitgestellten Daten und Digitalisate sammeln und gemeinsam durchsuchbar zugänglich machen.

Auch wenn hier viele mediävistische Großprojekte genannt wurden, mag diese Zeitreihe und vor allem die damit beschriebene Entwicklung in ähnlicher Form auch für die anderen geschichtswissenschaftlichen Epochen gelten. Nach mittlerweile 20 Jahren stehen dem Fach, so lässt sich festhalten, Millionen Datensätze, digitalisierte Bilder, Handschriften und Objekte, Editionen, Texte und Regesten, Katalogeinträge und Findmittel digital über das Internet zur Verfü-

43 Paul-Joachim Heinig, „Die Herausforderung der ‚Neuen Medien‘ (CD-Rom, Bildplatte und Internet): Zukünftige Gestaltungsfragen und Publikationsformen am Beispiel der Regesta Imperii,“ in *Die Regesta Imperii im Fortschreiten und Fortschritt*, hg. v. Harald Zimmermann (Köln, Weimar und Wien: Böhlau, 2000), 129–148.

44 Robert Giel, „Manuscripta mediaevalia: Handschriften aus deutschen Bibliotheken im Internet,“ *Gazette du livre médiéval* 39 (2001): 34–40.

45 „Vom Projekt zum Verein – Die Geschichte von prometheus | Das prometheus-Bildarchiv: Hochwertige Bilder zu Kunst, Kultur und Geschichte,“ *Prometheus. Das verteilte digitale Bildarchiv für Forschung und Lehre*, https://www.prometheus-bildarchiv.de/about/project_history. Zugriff am 15.12.2021.

46 Bernhard Assmann und Patrick Sahle, *Digital ist besser: Die Monumenta Germaniae Historica mit den dMGH auf dem Weg in die Zukunft – eine Momentaufnahme* (Köln: Universität zu Köln, 2008), urn:nbn:de:hbz:38-23179.

47 „Deutsche Digitale Bibliothek – Kultur und Wissen online“, <https://www.deutsche-digitale-bibliothek.de/>. Zugriff am 15.12.2021.

48 „Europeana – Discover Inspiring European Cultural Heritage“, <https://www.europeana.eu/en>. Zugriff am 15.12.2021.

49 „Archivportal-D“, <https://www.archivportal-d.de/>. Zugriff am 15.12.2021.

50 „Deutsches Zeitungsportal – Deutsche Digitale Bibliothek“, <https://www.deutsche-digitale-bibliothek.de/newspaper>. Zugriff am 02.12.2021.

51 „Handschriftenportal“, <https://handschriftenportal.de/>. Zugriff am 15.12.2021.

gung (schon allein das Archivportal-D umfasst die Beschreibungen von 24 Mio. Archivstücken, davon 1,3 Mio. mit Digitalisat⁵²). Blickt man noch etwas allgemeiner auf die Frage der Digitalisierung der Geschichtswissenschaften, sind natürlich auch die zahlreichen genuin digitalen Quellen hinzuzufügen, die spätestens seit den 1990er Jahren immer wichtiger werden und die die historische Arbeit in der Zeitgeschichte mehr und mehr prägen, angefangen bei digitalem Verwaltungsarchivgut und offiziell bereitgestellten Daten (zum Beispiel govData⁵³), wie sie aktuell zum Beispiel in der Corona-Krise eine Rolle spielen, über vielfältige Quellen wie Zeitungen, Fernsehsendungen, Internetportale, Blogs und Podcasts, E-Mails und Kurzmitteilungen (wie die berühmten SMS der früheren Bundeskanzlerin Angela Merkel) bis hin zu den wohlbekannten Social-Media-Kanälen wie jenen auf Twitter, die mittlerweile selbst Teil des Regierungshandelns sind.⁵⁴ Damit steht eine fast unüberschaubare und zum Teil äußerst prekäre Menge an digitalen Quellen – und damit letztlich auch an Daten – für die Auswertung zur Verfügung.

6.2 Persistenz der analogen Gebrauchslogik

Bislang wird für die Auswertung dieser Quellen jedoch meist nur die *Präsentationsebene* genutzt. Diese wandelt die in den Daten repräsentierten Informationen wieder in eine Form um, die sie uns – nach dem Vorbild ihrer materiellen Vorlagen, oder zumindest deren Gebrauchslogik folgend – am Rechner darstellt oder abspielt. Durch die Anzeige am Bildschirm, das Abspielen als Ton oder als Ausdruck auf Papier werden diese Informationen, oder zumindest ein Teil davon, wieder in analoge Signale umgewandelt und bereitgestellt. Digitalität und Internet werden vor allem als bequeme Speicher und schnelle Kommunikationsmittel verstanden, die es uns erlauben, Texte und Handschriften auch außerhalb der Bibliotheken und Archive zu konsultieren, ohne dabei deren Öffnungszeiten und andere Limitierungen wie die Anzahl der täglich erlaubten Dokumente berücksichtigen zu müssen. Dabei werden die digitalen Medien weitgehend in einen analogen Gebrauchszusammenhang integriert und die so repräsentierten Inhalte in der gleichen Weise verwendet wie vor ihrer Digitali-

52 „Archivportal-D“, <https://www.archivportal-d.de/>. Zugriff am 15.12.2021.

53 „GovData | Datenportal für Deutschland“, GovData, <https://www.govdata.de>. Zugriff am 15.12.2021.

54 Frédéric Clavert, „History in the Era of Massive Data: Online Social Media as Primary Sources for Historians,“ *Geschichte und Gesellschaft* 47/1 (2021): 175–194, doi: 10.13109/gege.2021.47.1.175.

sierung. Das heißt, selbst wenn sie als Daten vorliegen und übermittelt werden, werden Texte weiterhin als Texte gelesen, Bilder weiterhin als Bilder gesehen, Audioaufzeichnungen als Audioaufzeichnungen gehört und Filme als Filme geschaut. Für die Bearbeitung dieser Quellen und Informationen wird weiterhin auf die gleichen Analysemethoden zurückgegriffen wie zuvor im Analogen. Erweiterung erfährt die Bearbeitung höchstens in der Form von Volltextsuchen, die eine schnellere Orientierung im Text erlauben. Der damit einhergehende Bruch zwischen Original und Digitalisat wird jedoch nur selten reflektiert.⁵⁵ Dabei ist dieser – wie weiter oben beschrieben – erheblich. Ein Digitalisat ist keine Kopie und kein Faksimile; es ist ein Modell, das nur ausgewählte Eigenschaften der Vorlage in Daten überträgt, aus denen dann für jede Ausgabe nach bestimmten Vorgaben und abhängig vom jeweiligen System, dem konkreten Ausgabegerät und der Software eine neue (Teil-)Darstellung produziert beziehungsweise simuliert wird. Dafür greifen wir auf Programme wie Adobe Acrobat (PDF), Media Player (Video, Audio), DFG-Viewer (Handschriftendigitalisate), Blender (3D-Modelle) oder Browser wie Chrome und Firefox (unter anderem Textdaten und Bilder) zurück, die die in den Daten enthaltenen Informationen aufnehmen, verarbeiten und in einer jeweils spezifischen Art und Weise in jene analoge Darstellung umwandeln, auf die wir dann für unsere Analysen zurückgreifen.

Einen ähnlichen Filter und eine erhebliche Einschränkung im Zugang zu den Daten bilden auch die verschiedenen graphischen Benutzeroberflächen, die den Nutzer:innen auf der Präsentationsebene Zugang zu den in den Daten gespeicherten Informationen gewähren und einen Teil davon für den analogen Zugang (lesen, sehen, hören) aufbereiten. Dabei vereinfachen sie einerseits die Suche nach bestimmten Informationen und strukturieren den Nutzer:innen die Ergebnisse vor, womit sie ihnen Orientierung in der Menge an potentiellen Informationen bieten. Andererseits schränken sie damit aber auch den Zugang zu den in den Daten repräsentierten Informationen erheblich ein und ermöglichen

55 Siehe hierzu Michael Bender, Thomas Kollatz und Andrea Rapp, „Objekte im digitalen Diskurs: Epistemologische Zugänge zu Objekten durch Digitalisierung und diskursive Einbindung in virtuelle Forschungsumgebungen und -infrastrukturen,“ in *Objektepistemologien: Zur Vermessung eines transdisziplinären Forschungsraums*, hg. v. Markus Hilgert, Henrike Simon und Kerstin P. Hofmann (Berlin: Edition Topoi, 2018), 107–132, doi: 10.17171/3-59; Andreas Fickers, „Digitale Metaquellen und doppelte Reflexivität,“ in *Historische Grundwissenschaften und die digitale Herausforderung*, hg. v. Rüdiger Hohls, Claudia Prinz und Eva Schlotheuber (Berlin: Clio-Online und Humboldt-Universität zu Berlin, 2016), 52–57, doi: 10.18452/18771, sowie Pascal Föhr, *Historische Quellenkritik im digitalen Zeitalter*, E-humanities (Glückstadt: VWH Verlag Werner Hülsbusch, 2019), doi: 10.5451/unibas-006805169.

nur jene Nutzungsszenarien, für die die Benutzeroberfläche konzipiert wurde beziehungsweise umgesetzt werden konnte.

Diese Zugänge auf der Präsentationsebene waren lange Zeit die einzige Möglichkeit, die die Anbieter entsprechender Daten bereitgestellt haben, um mit den Daten zu interagieren. So war es in den verschiedenen Zeitungsportalen lange Zeit nur möglich, die verschiedenen digitalisierten Ausgaben online zu durchblättern und die jeweils interessierenden Seiten anzuschauen und im Bedarfsfall einzeln herunterzuladen, was sich in den letzten Jahren glücklicherweise grundlegend geändert hat.⁵⁶ Mittlerweile liegen die Zeitungen meist per OCR bearbeitet auch als Volltext vor und können – mit Einschränkungen – über den gesamten Bestand hin nach bestimmten Begriffen durchsucht werden. Die Ergebnisse verlinken meist auf die jeweiligen Seiten, die dann wiederum als Bild ausgegeben werden, um die jeweiligen Fundstellen im Kontext betrachten zu können. Zwar kann man sich den erkannten Volltext anzeigen lassen und die Fundstelle entweder als Volltext oder als bildbasierte PDF herunterladen. Auch kann man die Wiedergabe der Digitalisate auf unterschiedliche Weise verändern, um sie am Bildschirm lesbarer zu machen. Was man bislang jedoch nicht herunterladen konnte, waren die Text- oder Bilddaten selbst, die diesen Präsentationen zugrunde liegen, um diese eigenständig mit entsprechenden Analysemethoden bearbeiten zu können – und sei es nur für bestimmte Titel oder Zeitabschnitte.⁵⁷ Die Präsentationsebene beziehungsweise graphische Benutzeroberfläche bleibt damit die einzige Ebene, auf der man mit den digital repräsentierten Quellen interagieren kann.

Noch deutlicher kann man das bei der wichtigen und schon etwas älteren Publikation der Verhandlungen des Deutschen Reichstags und seiner Vorläufer beobachten.⁵⁸ Auch hier können die Nutzer:innen allein über das graphische Benutzerinterface der Webseite mit den Daten interagieren. Man kann die er-

⁵⁶ Vgl. zum Beispiel die Zeitungsportale „ANNO – Austrian newspaper Online“ der Österreichischen Nationalbibliothek (<https://anno.onb.ac.at/>) und „Zeitungportal NRW“ der Universitäts- und Landesbibliotheken in Bonn und Münster (<https://zeitpunkt.nrw/>). Zugriff jeweils am 15.12.2021.

⁵⁷ Seit Oktober 2021 bietet die Deutsche Digitale Bibliothek mit dem „Deutschen Zeitungsportal“ (<https://ddb.de/newspaper>) jedoch einen Aggregierungsservice an, der die über verschiedene Institutionen und Webangebote verteilten digitalisierten historischen Zeitungsbestände gemeinsam durchsuchbar und abrufbar macht. Hier können die so aggregierten Daten seit kurzem auch über die Schnittstelle (API) der DDB (<https://labs.deutsche-digitale-bibliothek.de/app/ddbapi/>) direkt auf der Datenebene angesprochen werden (Dokumentation: <https://api.deutsche-digitale-bibliothek.de/OpenAPI>). Zugriff jeweils am 15.12.2021.

⁵⁸ „Datenbank der deutschen Parlamentsabgeordneten 1867–1938“, Bayerische Staatsbibliothek, <https://www.reichstag-abgeordnetendatenbank.de/>. Zugriff am 15.12.2021.

schlossenen Texte zwar im Volltext durchsuchen, doch wird für die Fundstellen nur das Bild der jeweiligen Seite angezeigt. Gleiches geschieht, wenn man die Datenbank mit den detaillierten Einträgen zu den 5.211 Abgeordneten aus der Zeit zwischen 1867 und 1942 konsultiert. Hier kann man mit einer sehr praktischen und detaillierten Abfrage die gesamte Datensammlung durchsuchen, die daraus resultierenden Ergebnisse jedoch nur über die Ergebnisanzeige nutzen. Diese verweist entweder via Link auf die nicht durchsuchbaren Digitalisate des Parlamentsalmanachs beziehungsweise der Parlamentsprotokolle oder aber auf deren jeweilige Transkription. Die umfangreichen strukturierten Daten, die der Datenbank zugrunde liegen (unter anderem Berufsfeld, Konfession, Wahlperiode, Partei), hingegen sind bislang nicht abrufbar und damit auch nicht eigenständig auswertbar. An eine weitere Verwendung dieser umfangreichen Daten durch die Nutzer:innen selbst ist nicht gedacht. Sie dienen nur als Findmittel für eine dynamische Zusammenstellung von Texten und Digitalisaten, damit diese einzeln am Monitor gelesen werden können.

Dies lässt sich schließlich auch bei digitalen Editionen beobachten, wo die Texte zwar für die Präsentation über entsprechende Portale aufwendig aufbereitet und für diese verschiedenen Darstellungs- und damit Nutzungsszenarien bereitgestellt werden, von denen man ausgeht, dass sie für die Nutzer:innen nützlich sind. Doch wird hier ebenfalls meist davon ausgegangen, dass diese den Text, wie bei herkömmlichen Editionen als Buch, jetzt eben am Bildschirm lesen werden. Die mühsam aufbereiteten Daten der Edition, die ähnlich wie die Daten zu den Reichstagsabgeordneten von den Nutzenden selbst noch einmal auf ganz eigene und kreative Art mit digitalen Methoden und Tools ausgewertet werden könnten, werden jedoch selbst nicht bereitgestellt. In dem von Greta Franzini kuratierten Katalog digitaler Editionen⁵⁹ kann man sehen, dass gerade einmal für 70 der 179 in TEI/XML ausgeführten Editionen auch die zugrundeliegenden Daten zum Download angeboten werden. Für alle anderen bleibt die Präsentationsebene die einzige Möglichkeit zur Interaktion mit den bereitgestellten Informationen.

6.3 Konzeptioneller Wandel und das Primat der Daten

Damit bleiben die Möglichkeiten, mit den umfangreich digitalisierten Quellen und digital verfügbaren Daten umzugehen, bei den meisten Angeboten auf eine Art der Interaktion begrenzt, die der gewohnten analogen Gebrauchslöge folgt.

⁵⁹ „Catalogue Digital Editions“, Greta Franzini, <https://dig-ed-cat.acdh.oeaw.ac.at/>. Zugriff am 15.12.2021.

Informatik und den Digital Humanities für verschiedene Analyseschritte bereits etabliert sind.

Ich möchte das abschließend an drei kleinen Beispielen demonstrieren. Das erste zeigt dabei geradezu idealtypisch den Weg von der analogen über die digitalisierte zur digitalen Forschung in den Geschichtswissenschaften. Ausgangspunkt sind die *Germania Sacra*, ein Verzeichnis von kirchlichen Personen und Institutionen im Alten Reich, das seit 1929 als Publikationsreihe in gedruckter Form erschien. Diese Bände wurden seit 2008 retrodigitalisiert und als PDF über das Internet bereitgestellt, außerdem wurden die Personenregister der einzelnen Bände in einer Datenbank erfasst und damit über die verschiedenen Bände hinweg gemeinsam abfragbar gemacht.⁶⁰ Die Datenbank kann über die Website des Projekts genutzt und auf diese Weise zum Beispiel eine Liste mit allen Domherren erstellt werden, die zwischen 1350 und 1520 am Domstift in Münster nachweisbar sind – was am Ende eine Liste von 196 Personen ergibt. Möchte man nun herausfinden, wie die Vernetzung der Münsteraner Domherren zu kirchlichen Institutionen in anderen Städten aussah, indem man untersucht, an welchen anderen Institutionen außerhalb Münsters sie in ihrer Karriere Ämter innehatten und wie sich diese über die Zeit entwickelte, kann man sich für jeden einzelnen dieser Domherren dessen Ämterlaufbahn anzeigen lassen und diese 196 Einzeleinträge mit ihren Listen auswerten. Oder aber man kann das Angebot des Projekts annehmen und diese Abfrage über die API (Programmierschnittstelle) stellen⁶¹ beziehungsweise die Ergebnisliste der Abfrage über die graphische Benutzeroberfläche als Daten herunterladen. Diese kann man daraufhin weiter bearbeiten und so umstrukturieren, dass man sich die in den Daten enthaltenen Informationen in einer raumzeitlichen Visualisierung ausgeben lässt. Durch die Bearbeitung und Umstrukturierung der Daten mit ein paar einfachen Operationen lassen sich auch größere Datenbestände rasch auswerten und mögliche Muster und Entwicklungen aufzeigen, die dann am Material selbst näher geprüft werden können. Auf der Grundlage der zur Verfügung

60 Bärbel Kröger und Christian Popp, „Das Forschungsportal Germania Sacra Online: Der Weg in die digitale Zukunft,“ in *100 Jahre Germania Sacra: Kirchengeschichte schreiben vom 16. bis zum 21. Jahrhundert*, hg. v. Hedwig Röckelein (Berlin, Boston: De Gruyter Akademie Forschung, 2018), 135–148, doi: 10.1515/9783110619584-007.

61 Die Filterkriterien können dabei einfach als Parameter über die URL übergeben werden, wie das Beispiel zeigt: [http://personendatenbank.germania-sacra.de/api/v1.0/person?query\[0\]\[field\]=person.belegdaten&query\[0\]\[value\]=1521&query\[0\]\[operator\]=lower&query\[0\]\[connector\]=and&query\[1\]\[field\]=person.belegdaten&query\[1\]\[value\]=1349&query\[1\]\[operator\]=greater&query\[1\]\[connector\]=and&query\[2\]\[field\]=amt.institution&query\[2\]\[value\]=Domstift%20M%C3%BCnster&query\[2\]\[operator\]=like&query\[2\]\[connector\]=and&query\[3\]\[field\]=bezeichnung&query\[3\]\[value\]=Domherr&query\[3\]\[operator\]=like&offset=0&limit=500&format=turtle](http://personendatenbank.germania-sacra.de/api/v1.0/person?query[0][field]=person.belegdaten&query[0][value]=1521&query[0][operator]=lower&query[0][connector]=and&query[1][field]=person.belegdaten&query[1][value]=1349&query[1][operator]=greater&query[1][connector]=and&query[2][field]=amt.institution&query[2][value]=Domstift%20M%C3%BCnster&query[2][operator]=like&query[2][connector]=and&query[3][field]=bezeichnung&query[3][value]=Domherr&query[3][operator]=like&offset=0&limit=500&format=turtle).

stehenden Daten lässt sich hier mit ein paar Handgriffen zeigen, dass das Ämternetzwerk der Münsteraner Stiftsherren im untersuchten Zeitraum zunächst deutlich überregional ausgeprägt war, was sich jedoch spätestens ab 1450 änderte, wonach nur noch regionale Vernetzungen sichtbar sind. Die Veränderungen scheinen im Netzwerk mit der Münsteraner Stiftsfehde und demnach mit einem einschneidenden Ereignis in der Münsteraner Geschichte zusammenzuhängen. Damit haben wir statt auf der Präsentationsebene zu bleiben mit den über die Schnittstelle des Datenbankportals bereitgestellten Daten selbst gearbeitet und diese zur Grundlage unserer Untersuchung gemacht.

Vergleichbare Beispiele lassen sich ebenso für die direkte Arbeit mit Textquellen sowie mit Bildern finden. So könnte man beispielsweise für die Frage, in welcher Weise Raimund von Aguilers in seiner *Historia Francorum* bei der Schilderung der Ereignisse auf dem Ersten Kreuzzug auf die Bibel zurückgriff, den Quellentext einzeln durcharbeiten und sich dabei auf die eigene Kenntnis der Bibel stützen, die – wie Robert Huygens anmahnte – ein guter Editor mehr als einmal komplett gelesen haben sollte.⁶² Man kann aber statt mit der Anzeige als PDF oder als Onlinepräsentation auch auf der Datenebene mit diesen Texten arbeiten und diese Forschungsfrage unter Rückgriff auf die Methoden der *text reuse analysis* bearbeiten.⁶³ Gleiches gilt beispielsweise auch für die Suche nach Wappendarstellungen in mittelalterlichen Handschriften. Auch hier könnte man jedes Manuskript als Digitalisat aufrufen und auf der Suche nach entsprechenden Abbildungen virtuell durchblättern. Oder man kann auf die zugrunde liegenden Daten zurückgreifen und mit Hilfe von Verfahren des maschinellen Lernens aus der *computer vision* ein Modell trainieren, mit dem sich die Wappendarstellungen automatisiert über die Bilddaten eines größeren Korpus hinweg auffinden lassen.⁶⁴

62 Robert Burchard Constantijn Huygens, *Ars edendi: A Practical Introduction to Editing Medieval Latin Texts*, (Turnhout: Brepols, 2000), 11.

63 Torsten Hiltmann et al., „Digital Methods in Practice: The Epistemological Implications of Applying Text Re-Use Analysis to the Bloody Accounts of the Conquest of Jerusalem (1099),“ *Geschichte und Gesellschaft* 47/1 (2021): 122–156, doi: 10.13109/gege.2021.47.1.122.

64 Torsten Hiltmann, Benjamin Risse und Sebastian Thiele, „Friends with Benefits. Wie Deep-Learning basierte Bildanalyse und kulturhistorische Heraldik voneinander profitieren,“ in *DHd 2020 Spielräume: Digital Humanities zwischen Modellierung und Interpretation. 7. Tagung des Verbands „Digital Humanities im deutschsprachigen Raum“ (DHd 2020)*, hg. v. Christof Schöch und Patrick Helling, 135–138. Paderborn, 2020. doi: 10.5281/zenodo.4621890.

7 Von einer digitalisierten zu einer digitalen Geschichtswissenschaft

Vergleicht man also die Entwicklungen in den Geschichtswissenschaften mit dem, was wir sowohl beim Prozess der Verschriftlichung als auch bei der Einführung des Buchdrucks mit beweglichen Lettern gesehen haben, kann man auch hier davon sprechen, dass dieser Prozess in zwei Etappen verläuft. Zunächst die Aneignung des neuen Mediums (mediale Digitalität), bei der anfänglich nur die mediale Ebene gewechselt wird, der konkrete Gebrauch sich aber weiter an den hergebrachten, das heißt hier: analogen Gebrauchslogiken orientiert – ganz wie bei der Verschriftung. Das neue digitale Medium wird zunächst einmal in die gewohnten Gebrauchszusammenhänge integriert. Dem folgt eine weitere Stufe, in deren Rahmen sich der Gebrauch dieser neuen Technologien tatsächlich an den Bedingungen der Digitalität und damit an den oben aufgeführten Eigenschaften digitaler Daten selbst orientiert (konzeptionelle Digitalität). Erst hieraus entsteht eine neue, spezifisch digitale Gebrauchslogik. Auch wenn sich die Geschichtswissenschaften, wie am Anfang demonstriert, im Rahmen der Digitalisierung schon erheblich gewandelt haben, können durch die Verwendung digitaler Methoden, die dieser neuen datenbasierten Gebrauchslogik folgen, den herkömmlichen Methoden noch einmal neue Perspektiven hinzugefügt werden. Ein Prozess, bei dem wir als Historiker:innen jedoch erst am Anfang stehen.

Aus dieser Beobachtung lassen sich meines Erachtens mehrere Konsequenzen und Aufgaben für die Digitalen Geschichtswissenschaften ableiten. Denn die unhintergehbare Grundlage dieser Entwicklung sind digitale Daten oder, genauer gesagt, digitale Forschungsdaten – die wir noch deutlich besser verstehen müssen als bisher und die wir neben den Quellen selbst in den Mittelpunkt unserer Arbeit stellen sollten.

1. **Data First:** Das bedeutet, dass wir darüber nachdenken müssen, inwieweit die Geschichtswissenschaften oder mindestens die digitalen Geschichtswissenschaften eine starke und klare Datenkultur herausbilden müssen. Zentral hierfür wäre das Data-First-Prinzip, das meint, dass man die Daten in den Mittelpunkt stellt und zur Grundlage des digitalen Arbeitens macht, anstelle der verschiedenen Applikationen und unterschiedlichen Repräsentationsformen, die auf diesen Daten aufbauen. Ganz ähnlich wie in anderen gesellschaftlichen Bereichen sollten im Zeitalter der Digitalität auch in den Geschichtswissenschaften die Daten im Zentrum stehen. Erst dadurch können wir die Erkenntnispotentiale nutzen, die uns über dieses neue Medium zur Verfügung stehen und die wir brauchen, um die großen Mengen an di-

gitalisierten oder genuin digitalen Daten handhabbar zu halten und produktiv zu nutzen. Dafür aber müssten, und das wäre die fundamentale Voraussetzung, die verschiedenen Projekte und Institutionen die bei ihnen produzierten Daten nicht nur auf der in ihrer Nutzbarkeit begrenzten Präsentationsebene publizieren, sondern zum Beispiel über Schnittstellen oder in Form von *data dumps* zugleich unter Befolgung der FAIR-Prinzipien als (Forschungs-)Daten bereitstellen.⁶⁵

2. **Methodenentwicklung und Methodenreflexion:** Wie wir gesehen haben, basiert die Priorisierung der Daten auf einer neuen digitalen Gebrauchslogik, die sich aus den Eigenschaften der Daten und deren Prozessierung ergibt, die wir noch viel besser verstehen müssen. Dies gilt vor allem für die neuen methodischen Möglichkeiten, die damit verbunden sind. Denn um die Daten angemessen nutzen zu können, brauchen wir eine entsprechende Methodenentwicklung und eine kritische Methodenreflexion aus der spezifischen Perspektive der Geschichtswissenschaften. Es geht darum, zu verstehen, wie wir in der historischen Forschung die uns zur Verfügung stehenden Daten nutzen können, gleichzeitig jedoch immer aus der Perspektive unseres Faches reflektieren, was bei der Anwendung der Methoden in den einzelnen Schritten mit den Daten und den darin enthaltenen Informationen geschieht, um sicherzustellen, dass diese Methoden und die Art und Weise, wie sie eingesetzt werden, auch aus geschichtswissenschaftlicher Perspektive zu sauberen und validen Ergebnissen führen.
3. **Theorieentwicklung:** Was wir drittens brauchen, um im Rahmen dieses Wandels tatsächlich gute Methoden zu entwickeln, ist eine vertiefte Auseinandersetzung mit Digitalität und Epistemologie und darauf aufbauend eine entsprechende Theorieentwicklung. Denn eine reine Übernahme fachfremder Methoden reicht nicht aus; sie müssen auch in Hinblick auf die theoretischen und methodischen Grundlagen der Geschichtswissenschaften geprüft und an diese angepasst werden. Dafür aber müssen wir die Grundlagen unseres Faches gut kennen, auch unter den Bedingungen der Digitalität. Das bedeutet, wir brauchen ein noch klareres Verständnis dafür, wie historische Forschung funktioniert und wie sie sich durch Digitalität verändert. Wir müssen verstehen, wie historisches Wissen produziert wird, wie wir historische Daten repräsentieren und analysieren, wie wir

⁶⁵ Vgl. hierzu zuletzt Christof Schöch, „Open Access für die Maschinen,“ in *Die Zukunft des kunsthistorischen Publizierens*, hg. v. Maria Effinger und Hubertus Kohle (Heidelberg: arthistoricum.net, 2021), doi: 10.11588/ARTHISTORICUM.663.C9210. Zu den wichtigen FAIR-Prinzipien siehe: „FAIR Principles“, GO FAIR, <https://www.go-fair.org/fair-principles/>. Zugriff am 15.12.2021.

aber auch mit Unsicherheiten und Ambiguitäten umgehen, um am Ende eines digitalen Prozesses verlässliche und vor allem kritisch überprüfbare und belastbare Ergebnisse zu erhalten. Das bedeutet ferner, dass wir noch genauer reflektieren müssen, was Digitalität selbst meint und wie diese Art der Speicherung und Kommunikation von Wissen und Information in Form von Daten unsere Gesellschaft und letztlich auch unsere Geschichte ganz grundsätzlich beeinflusst und verändert.

Es ist schon ein Paradox. Auf der einen Seite schaut die digitale Geschichtswissenschaft bereits auf eine fast 60-jährige Geschichte zurück. Zugleich ist aber immer wieder, auch heute, das Gefühl verbreitet, erst am Anfang einer Entwicklung zu stehen. Der Blick in die Geschichte der Verschriftlichung und in die Entstehung des Buchdrucks konnte hoffentlich zeigen, dass es sich bei den aktuellen Entwicklungen jedoch nicht um etwas Neues, sondern nur um eine weitere Etappe in einem längeren Prozess handelt, den wir auf Grundlage dieses Wissens aktiv mitgestalten können.

Die Feststellung, dass sich die tatsächlichen Potentiale neuer Medien und Technologien ihren Nutzer:innen erst allmählich erschließen, lässt sich letztendlich auch auf die mit der Digitalisierung der Geschichtswissenschaften einhergehenden neuen Techniken und Methoden übertragen. So ließe sich auch erklären, warum die gern gestellte Frage nach besonders herausragenden Ergebnissen, die mit den neuen digitalen Methoden erbracht wurden, um diese zu überhaupt zu legitimieren,⁶⁶ in dieser Weise gar nicht zu beantworten ist. Denn auch hier gilt, dass man sich diese Methoden zunächst einmal Stück für Stück aneignen, sie erforschen und vor allem anwenden muss, um erkennen zu können, welche tatsächlichen Potentiale damit verbunden sind.

Bibliographie

Assmann, Jan. *Das kulturelle Gedächtnis: Schrift, Erinnerung und politische Identität in frühen Hochkulturen*. München: C. H. Beck, 1992.

Assmann, Bernhard und Patrick Sahle. *Digital ist besser: Die Monumenta Germaniae Historica mit den dMGH auf dem Weg in die Zukunft – eine Momentaufnahme*. Köln: Universität zu Köln, 2008. urn:nbn:de:hbz:38-23179.

⁶⁶ Vgl. Timothy Brennan, „The Digital-Humanities Bust,“ *The Chronicle of Higher Education* 64, 8 (15.10.2017), <https://www.chronicle.com/article/the-digital-humanities-bust/>. Zugriff am 15.12.2021.

- Bender, Michael, Thomas Kollatz und Andrea Rapp. „Objekte im digitalen Diskurs: Epistemologische Zugänge zu Objekten durch Digitalisierung und diskursive Einbindung in virtuelle Forschungsumgebungen und -infrastrukturen.“ In *Objektepistemologien: Zur Vermessung eines transdisziplinären Forschungsraums*, hg. v. Markus Hilgert, Henrike Simon und Kerstin P. Hofmann, 107–132. Berlin: Edition Topoi, 2018. doi: 10.17171/3-59.
- Bermès, Emmanuelle. „La BNF face au numérique: De nouveaux objets patrimoniaux.“ *Revue de la Bibliothèque nationale de France* 61 (2020): 163–167.
- Bödeker, Hans Erich. „Aufklärung als Kommunikationsprozeß.“ *Aufklärung* 2/2 (1987): 89–111.
- Brantl, Markus und Astrid Schoger. „Das Münchener Digitalisierungszentrum zwischen Produktion und Innovation.“ In *Information, Innovation, Inspiration: 450 Jahre Bayerische Staatsbibliothek*, hg. v. Rolf Griebel und Klaus Ceynowa, 253–280. München: Saur, 2008. doi: 10.1515/9783598440892.4.253.
- Brennan, Timothy. „The Digital-Humanities Bust.“ *The Chronicle of Higher Education* 64, 8 (15.10.2017), <https://www.chronicle.com/article/the-digital-humanities-bust/>. Zugriff am 15.12.2021.
- Bullough, Vern L. „The Computer and the Historian: Some Tentative Beginnings.“ *Computers and the Humanities* 1/3 (1967): 61–64. doi: 10.1007/BF00119888.
- Clanchy, Michael T. *From Memory to Written Record: England 1066–1307*. 3. Aufl. Malden: Wiley-Blackwell, 2013.
- Clavert, Frédéric. „History in the Era of Massive Data: Online Social Media as Primary Sources for Historians.“ *Geschichte und Gesellschaft* 47/1 (2021): 175–194. doi: 10.13109/gege.2021.47.1.175.
- Clubb, Jerome M. und Howard Allen. „Computers and Historical Studies.“ *Journal of American History* 54/3 (1967): 599–607. doi: 10.2307/2937409.
- Ehlich, Konrad. „Funktion und Struktur schriftlicher Kommunikation.“ In *Schrift und Schriftlichkeit: Ein interdisziplinäres Handbuch internationaler Forschung*. Bd. 1, hg. v. Hartmut Günther und Otto Ludwig, 18–41. Berlin, New York: De Gruyter Mouton, 2008. doi: 10.1515/9783110111293.1.1.18.
- Eisenstein, Elizabeth L. *The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early Modern Europe. Volumes I and II*. Cambridge [England], New York: Cambridge University Press, 1979.
- Emmer, Martin. „Soziale Medien in der politischen Kommunikation.“ In *Handbuch Soziale Medien*, hg. v. Jan-Hinrik Schmidt und Monika Taddicken, 81–99. Wiesbaden: Springer, 2017.
- Faßler, Manfred und Wulf Halbach. *Geschichte der Medien*. München: Fink, 1998.
- Fickers, Andreas. „Digitale Metaquellen und doppelte Reflexivität.“ In *Historische Grundwissenschaften und die digitale Herausforderung*, hg. v. Rüdiger Hohls, Claudia Prinz und Eva Schlotheuber, 52–57. Berlin: Clio-Online und Humboldt-Universität zu Berlin, 2016. doi: 10.18452/18771.
- Floridi, Luciana. *The 4th Revolution. How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press, 2014.
- Föhr, Pascal. *Historische Quellenkritik im digitalen Zeitalter*. E-humanities. Glückstadt: VWH Verlag Werner Hülsbusch, 2019. doi: 10.5451/unibas-006805169.
- Fröhlich, Jürgen. „Was aber soelchs himmelkorn bedüt / ist Gott allein zewüssen“: Phänomene struktureller Amnesie in medialen Umbruchzeiten (Flugblatt und Internet).“ In *Mediaevistik und neue Medien*, hg. v. Klaus van Eickels, Ruth Wesselbaumer und Ingrid Bennewitz, 49–66. Ostfildern: Thorbecke, 2004.

- Frühwald, Wolfgang. „Das Ende der Gutenberg-Galaxis: Über den Einfluß des Mediums auf den Inhalt wissenschaftlicher Publikationen.“ *Leviathan* 26/3 (1998): 305–318.
- Giel, Robert. „‚Manuscripta mediaevalia‘: Handschriften aus deutschen Bibliotheken im Internet.“ *Gazette du livre médiéval* 39 (2001): 34–40.
- Giesecke, Michael. *Der Buchdruck in der frühen Neuzeit: Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien*. Frankfurt am Main: Suhrkamp, 1991.
- Giesecke, Michael. *Von den Mythen der Buchkultur zu den Visionen der Informationsgesellschaft*. Frankfurt am Main: Suhrkamp, 2002.
- Goody, Jack und Ian Watt. „The Consequences of Literacy.“ *Comparative Studies in Society and History* 5 (1963): 304–345. doi: 10.1017/S0010417500001730.
- Goody, Jack. *The Logic of Writing and the Organization of Society*. Cambridge: Cambridge University Press, 1987.
- Gregor, Bernd. „Computergestützter Satz: Protokoll der Entstehung dieses Buches.“ In *Computerfibel für die Geisteswissenschaften: Einsatzmöglichkeiten des Personal Computers und Beispiele aus der Praxis*, hg. v. Manfred Krifka, 265–282. München: C. H. Beck, 1986.
- Haber, Peter. *Digital Past: Geschichtswissenschaft im digitalen Zeitalter*. München: Oldenbourg Verlag, 2011. doi: 10.1515/9783486712339.
- Harnad, Stevan. „Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge.“ *Public-Access Computer Systems Review* 2/1 (1991): 39–53.
- Hartmann, Josef. „Urkunden.“ In *Die archivalischen Quellen: Mit einer Einführung in die historischen Hilfswissenschaften*, hg. v. Friedrich Beck und Eckart Henning, 9–39. 3. Aufl. Köln, Weimar und Wien: Böhlau, 2004.
- Heinig, Paul-Joachim. „Die Herausforderung der ‚Neuen Medien‘ (CD-Rom, Bildplatte und Internet): Zukünftige Gestaltungsfragen und Publikationsformen am Beispiel der Regesta Imperii.“ In *Die Regesta Imperii im Fortschreiten und Fortschritt*, hg. v. Harald Zimmermann, 129–148. Köln, Weimar und Wien: Böhlau, 2000.
- Hiltmann, Torsten, Benjamin Risse und Sebastian Thiele. „Friends with Benefits. Wie Deep-Learning basierte Bildanalyse und kulturhistorische Heraldik voneinander profitieren.“ In *DHd 2020 Spielräume: Digital Humanities zwischen Modellierung und Interpretation. 7. Tagung des Verbands „Digital Humanities im deutschsprachigen Raum“ (DHd 2020)*, hg. v. Christof Schöch und Patrick Helling, 135–138. Paderborn, 2020. doi: 10.5281/zenodo.4621890.
- Hiltmann, Torsten, Jan Keupp, Melanie Althage und Philipp Schneider. „Digital Methods in Practice: The Epistemological Implications of Applying Text Re-Use Analysis to the Bloody Accounts of the Conquest of Jerusalem (1099).“ *Geschichte und Gesellschaft* 47/1 (2021): 122–156. doi: 10.13109/gege.2021.47.1.122.
- Hohls, Rüdiger. „Historische Fachkommunikation im Umbruch: Ein Rückblick auf die Anfänge des H-Net und von H-Soz-u-Kult.“ In *20 Jahre Arbeitsgemeinschaft Geschichte und EDV*, hg. v. Jörn Kobes, Kai Ruffing und Wolfgang Spickermann, 147–166. Gutenberg: Computus Druck Satz & Verlag, 2013.
- Huygens, Robert Burchard Constantijn. *Ars edendi: A Practical Introduction to Editing Medieval Latin Texts*. Turnhout: Brepols, 2000.
- Illich, Ivan. *Im Weinberg des Textes: Als das Schriftbild der Moderne entstand*. München: C. H. Beck, 2010.

- Jannidis, Fotis. „Zahlen und Zeichen.“ In *Digital Humanities: Eine Einführung*, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein, 59–67. Stuttgart: J. B. Metzler, 2017. doi: 10.1007/978-3-476-05446-3_5.
- Kaden, Ben und Michael Kleineberg. „Zur Situation des digitalen geisteswissenschaftlichen Publizierens: Erfahrungen aus dem DFG-Projekt ‚Future Publications in den Humanities‘.“ *Bibliothek Forschung und Praxis* 41, 1 (2017): 7–14. doi: 10.1515/bfp-2017-0009.
- Keller, Hagen. „Vom ‚heiligen Buch‘ zur ‚Buchführung‘: Lebensfunktionen der Schrift im Mittelalter.“ *Frühmittelalterliche Studien* 26 (1992): 1–31.
- Key, Newton. „Crowdsourcing the Early Modern Blogosphere.“ In *Historyblogosphere: Bloggen in den Geschichtswissenschaften*, hg. v. Peter Haber und Eva Pfanzer, 101–118. München: Oldenbourg Wissenschaftsverlag, 2013. doi: 10.1524/9783486755732.101.
- Kittler, Friedrich A. *Grammophon, Film, Typewriter*. Berlin: Brinkmann & Bose, 1986.
- Klinke, Harald. „Aufbau des Computers und Vernetzung.“ In *Digital Humanities: Eine Einführung*, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein, 35–58. Stuttgart: J. B. Metzler, 2017. doi: 10.1007/978-3-476-05446-3_4.
- Kröger, Bärbel und Christian Popp. „Das Forschungsportal Germania Sacra Online: Der Weg in die digitale Zukunft.“ In *100 Jahre Germania Sacra: Kirchengeschichte schreiben vom 16. bis zum 21. Jahrhundert*, hg. v. Hedwig Röckelein, 135–148. Berlin, Boston: De Gruyter Akademie Forschung, 2018. doi: 10.1515/9783110619584-007.
- Lückerath, Carl August. „Prolegomena zur elektronischen Datenverarbeitung im Bereich der Geschichtswissenschaft.“ *Historische Zeitschrift* 207 (1968): 265–296. doi: 10.1524/hzhz.1968.207.jg.265.
- McLuhan, Marshall. *The Gutenberg Galaxy: The Making of Typographic Man*. Toronto: University of Toronto Press, 1962.
- Meyer, Thomas. „H-Soz-Kult und Clio-online: Von der Mailingliste zur Online-Community.“ *Bibliotheksdienst* 52/3–4 (2018): 185–195. doi: 10.1515/bd-2018-0026.
- Müller, Jan-Dirk. „Publizistik unter Maximilian I.: Zwischen Buchdruck und mündlicher Verkündigung.“ In *Sprachen des Politischen: Medien und Medialität in der Geschichte*, hg. v. Ute Frevert und Wolfgang Braungart, 95–122. Göttingen: Vandenhoeck und Ruprecht, 2004.
- Nassehi, Armin. *Muster: Theorien der digitalen Gesellschaft*. München: C. H. Beck, 2019.
- Oesterreicher, Wulf. „Verschriftung und Verschriftlichung im Kontext medialer und konzeptioneller Schriftlichkeit.“ In *Schriftlichkeit im frühen Mittelalter*, hg. v. Ursula Schäfer, 267–292. Tübingen: Narr, 1993.
- Ong, Walter. *Orality and Literacy: The Technologizing of the Word*. London: Methuen, 1982.
- Pinfield, Stephen, Simon Wakeling, David Bawden und Lyn Robinson. *Open Access in Theory and Practice: The Theory-Practice Relationship and Openness*. London, New York: Routledge, 2020. doi: 10.4324/9780429276842.
- Rautenberg, Ursula. „Die Entstehung und Entwicklung des Buchtitelblatts in der Inkunabelzeit in Deutschland, den Niederlanden und Venedig – Quantitative und qualitative Studien.“ *Archiv für Geschichte des Buchwesens* 62 (2008): 1–105.
- Rehbein, Malte. „Digitalisierung.“ In *Digital Humanities: Eine Einführung*, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein, 179–189. Stuttgart: J. B. Metzler, 2017. doi: 10.1007/978-3-476-05446-3_12.
- Reichert, Ramón. „Theorien digitaler Medien.“ In *Digital Humanities: Eine Einführung*, hg. v. Fotis Jannidis, Hubertus Kohle und Malte Rehbein, 19–34. Stuttgart: J. B. Metzler, 2017. doi: 10.1007/978-3-476-05446-3_3.
- Schmale, Wolfgang. *Digitale Geschichtswissenschaft*. Wien: Böhlau, 2010.

- Schmidt, Jan-Hinrik und Monika Taddicken (Hg.). *Handbuch Soziale Medien*. Wiesbaden: Springer, 2017.
- Schöch, Christof. „Open Access für die Maschinen.“ In *Die Zukunft des kunsthistorischen Publizierens*, hg. v. Maria Effinger und Hubertus Kohle. Heidelberg: arthistoricum.net, 2021. doi: 10.11588/ARTHISTORICUM.663.C9210.
- Scribner, Robert W. „Flugblatt und Analphabetentum: Wie kam der gemeine Mann zu reformatorischen Ideen?“ In *Flugschriften als Massenmedium der Reformationszeit*, hg. v. Hans-Joachim Köhler, 65–76. Stuttgart: Klett-Cotta, 1981.
- Smith, Margaret M. *The Title-Page: Its Early Development 1460–1510*. London: The British Library & Oak Knoll Press, 2000.
- Stalder, Felix. *Kultur der Digitalität*. Berlin: Suhrkamp, 2016.
- Stein, Peter. *Schriftkultur: Eine Geschichte des Schreibens und Lesens*. Darmstadt: Primus, 2006.
- Süssenguth, Florian (Hg.). *Die Gesellschaft der Daten: Über die digitale Transformation der sozialen Ordnung*. Bielefeld: transcript, 2015.
- Thaller, Manfred. „Entzauberungen: Die Entwicklung einer fachspezifischen historischen Datenverarbeitung in der Bundesrepublik [1990].“ *Historical Social Research / Historische Sozialforschung. Supplement* 29 (2017): 178–192.
- Wachter, Christian. *Geschichte digital schreiben: Hypertext als non-lineare Wissensrepräsentation in der Digital History*. Bielefeld: transcript, 2021. doi: 10.1515/9783839458013.
- Weinrich, Harald. „Gedächtniskultur – Kulturgedächtnis.“ In *Mediengebrauch und Erfahrungswandel: Beiträge zur Kommunikationsgeschichte*, hg. v. Detlev Schöttker, 37–52. Göttingen: Vandenhoeck und Ruprecht, 2003.
- Wenzel, Horst. *Mediengeschichte vor und nach Gutenberg*. Darmstadt: Wissenschaftliche Buchgesellschaft, 2007.
- Würgler, Andreas. *Medien in der Frühen Neuzeit*. München: De Gruyter Oldenbourg, 2009.
- Zahlmann, Stefan. „Die Digital Humanities und der Mensch: Ein Kommentar.“ In *Digital Humanities: Praktiken der Digitalisierung, der Dissemination und der Selbstreflexivität*, hg. v. Wolfgang Schmale, Historische Mitteilungen – Beihefte, Bd. 91, 175–180. Stuttgart: Steiner, 2015.

Andreas Fickers

What the D does to history

Das digitale Zeitalter als neues historisches Zeitregime?

Abstract: The digital interferes in multiple ways in our current day practice of history. This article argues that it not only impacts the way we search, store, analyze, and visualize historical sources and how we tell our stories, but the dynamic, real-time, and connected nature of digital research infrastructures and the Internet has a deep influence on how we think about history. As a new temporal regime, the digital age shapes our memory practices and changes the way we imagine and experience the past. By mobilizing the concept of digital hermeneutics, the chapter proposes a conceptual framework that helps to understand the various interferences of the „D“ and to use the critical potential of humanities to deconstruct and contextualize our data-driven present.

Keywords: digital hermeneutics, time regime, hyperconnectivity, data visualization, real time, aura

Zusammenfassung: Das Digitale greift auf vielfältige Weise in unsere heutige Geschichtspraxis ein. In diesem Artikel wird argumentiert, dass es sich nicht nur auf die Art und Weise auswirkt, wie wir historische Quellen suchen, speichern, analysieren und visualisieren und wie wir unsere Geschichten erzählen, sondern dass die dynamische, in Echtzeit ablaufende und vernetzte Natur der digitalen Forschungsinfrastrukturen und des Internets einen tiefgreifenden Einfluss darauf hat, wie wir über Geschichte denken. Als neues zeitliches Regime prägt das digitale Zeitalter unsere Erinnerungspraktiken und verändert die Art und Weise, wie wir uns die Vergangenheit vorstellen und wie wir Geschichte erleben. Unter Verwendung des Konzepts der digitalen Hermeneutik schlägt das Kapitel einen konzeptionellen Rahmen vor, der hilft, die verschiedenen Interferenzen des „D“ zu verstehen und das kritische Potenzial der Geisteswissenschaften zu nutzen, um unsere datengesteuerte Gegenwart zu dekonstruieren und zu kontextualisieren.

Schlagwörter: Digitale Hermeneutik, Zeitregime, Hyperkonnektivität, Datenvisualisierung, Echtzeit, Aura

1 Der digitale Historiker als „randonneur“

Im Oktober 2019 fand im Deutschen Historischen Institut in Washington die Tagung „Digital Hermeneutics: From Research to Dissemination“ statt, welche das GHI in Zusammenarbeit mit dem Roy Rosenzweig Centre for History and New Media und dem Luxembourg Centre for Contemporary and Digital History organisiert hatte.¹ In der Abschlussdiskussion, an der unter anderem Alan Liu, Tim Hitchcock und Jessica Owens zu resümierenden Kommentaren aufgefordert wurden, entspann sich eine interessante Diskussion darüber, wie sich die mannigfaltigen methodologischen und epistemologischen Implikationen des Digitalen für die Geschichtswissenschaften beschreiben beziehungsweise fassen ließen. Mit anderen Worten: What does the „D“ to history? In ihrer Direktheit und vordergründigen Einfachheit verbirgt sich in dieser Frage eine Komplexität, deren Vielschichtigkeit und Verästelungen mich bis heute umtreiben. Das, was in diesem Aufsatz präsentiert wird, spiegelt demnach einen zeit- und standortgebundenen Einblick in einen Denkprozess wider, der sich durch vorsichtiges „heuristisches Tasten“² auszeichnet und sich der Serres'schen Analogie des „randonneurs“ verpflichtet fühlt. „Schlagen Sie sich seitwärts“ – „débrouillez-vous“, so lautet Serres' Ratschlag in dem Kapitel „Methode und Wanderung“ in *Die fünf Sinne. Eine Philosophie der Gemenge und Gemische*.³

Sich dem Digitalen als großes „D“ anzunähern, fühlt sich in der Tat wie der von Serres beschriebene Exodus des Odysseus an: jenen Weg, der sich durch Abweichungen, Fluktuationen und Zerstreuungen kennzeichnet und nicht durch cartesianische Linearität. Die Semantik des Begriffs „randonnée“, der im französischen „courir à randon“ das Verfolgen des Wildes und das Spurenlesen meint, im Englischen aber als „random“ die Erinnerung an den unregelmäßigen, unvorhergesehenen Fluchtweg des Wildes behielt und damit auf die Zufälligkeit der zurückgelegten Strecke verweist, bringt genau jenes Gefühl der Ambivalenz zum Ausdruck, welches mich beim Versuch beschleicht, die multiplen Interferenzen des Digitalen in der Praxis historischen Arbeitens zu „erfassen“. Erfassen ist hier im Cassirer'schen Sinne als doppelter Akt des „Begreifens“ ge-

¹ Für einen Konferenzbericht siehe Tim van der Heijden, Juliane Tatarinow und Gerben Zaagsma, „Transatlantic Perspectives on Digital Hermeneutics“, 05.11.2019, <https://www.c2dh.uni.lu/de/thinking/transatlantic-perspectives-digital-hermeneutics>. Zugriff am 16.12.2021.

² Zum Begriff des „heuristischen Tastens“ als experimenteller Erkenntnismodus siehe Olaf Breidbach, Peter Heering, Matthias Müller und Heiko Weber (Hg.), *Experimentelle Wissenschaftsgeschichte* (München: Brill, 2010), 13–72.

³ Michel Serres, *Die fünf Sinne. Eine Philosophie der Gemenge und Gemische* (Frankfurt a. M.: Suhrkamp, 1998), 366.

meint: „Alle geistige Bewältigung der Wirklichkeit“, so Cassirer in seiner Schrift *Form und Technik* aus dem Jahr 1930, „ist an diesen doppelten Akt des ‚Fassens‘ gebunden: an das ‚Begreifen‘ der Wirklichkeit im sprachlich-theoretischen Denken und an ihr ‚Erfassen‘ durch das Medium des Wirkens; an die gedankliche wie an die technische Formgebung.“⁴

Genau an dieser Schnittstelle zwischen Theorie und Praxis, konzeptioneller Durchdringung und methodologischer Reflektion von Geschichtswissenschaft im digitalen Zeitalter setzen meine Überlegungen zu einer digitalen Hermeneutik der Geschichtswissenschaft an.⁵ Es ist dieses Spannungsfeld zwischen einer auf methodologischen Grundlagen des 19. Jahrhunderts beruhenden Disziplin und der sich radikal wandelnden Wissensökonomie des 21. Jahrhunderts, welches zum Überdenken und zur Anpassung zentraler Konzepte, Praxen und damit des Relevanz- und Wissenschaftsanspruches des Faches zwingt. Digitale Forschungsinfrastrukturen, Datenbestände und Werkzeuge beeinflussen heute die Art und Weise wie wir Geschichte denken (Fragestellungen und Forschungsperspektiven), machen (Methoden und Interpretation) und erzählen (Argumentationen und narrative Formate); kurzum: wie wir historischen Sinn produzieren und versuchen, das Vergangene deutend zu verstehen.

Als praktische Erkenntnistheorie der Geschichtswissenschaft interessiert sich digitale Hermeneutik also für die Spannung zwischen „Fremdheit“ und „Vertrautheit“ der historischen Überlieferung, diesem „Zwischen“, das laut Hans-Georg Gadamer der „wahre Ort der Hermeneutik“ ist.⁶ Wenn wir akzeptieren, dass Hybridität „the new normal“ ist,⁷ brauchen wir ein Update der historischen Hermeneutik, um die methodologischen und epistemologischen Spannungen in aktuellen Geschichtspraktiken explizit zu machen. Angewandte digitale Hermeneutik ist ebenso sehr „Theorie der Praxis“ wie „Praxis der Theorie“. Der konzeptionelle Rahmen der digitalen Hermeneutik lädt dazu ein, systematisch über die digitalen „Interferenzen“ im iterativen Prozess historischer Sinnbildung nachzudenken. Das heißt konkret:

⁴ Ernst Cassirer, „Form und Technik,“ in Ernst Cassirer, *Symbol, Technik, Sprache. Aufsätze aus den Jahren 1927–1933* (Hamburg: Meiner, 1995), 52.

⁵ Siehe hierzu Andreas Fickers, „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?“ *Zeithistorische Forschungen* 17 (2020) 1: 157–168; Andreas Fickers und Juliane Tatarinow (Hg.), *Digital Hermeneutics. Between Theory and Practice* (Berlin: DeGruyter, 2022), im Druck.

⁶ Hans-Georg Gadamer, *Wahrheit und Methode. Grundzüge einer philosophischen Hermeneutik* (Tübingen: Mohr Siebeck, 2010), 300.

⁷ Gerben Zaagsma, „On Digital History,“ *BMGN – Low Countries Historical Review*, 128 (2013) 4: 3–29.

- die Black Boxes von algorithmusgesteuerten Suchmaschinen zu öffnen und über die Heuristik der Suche in Online-Katalogen und Repositorien nachzudenken;⁸
- über die 6 V's der Datenintegrität nachzudenken (volume, velocity, variety, validity, veracity, value) und uns in historischer Datenkritik zu schulen;⁹
- zu verstehen, wie digitale Werkzeuge die epistemischen Objekte unserer Untersuchungen ko-konstruieren und Benutzer:innen zu Manipulator:innen hochspezifischer Forschungsinstrumente machen;¹⁰
- den „look of certainty“¹¹ der Datenvisualisierung zu dekonstruieren, indem wir die indexikalische aber dynamische und relationale Beziehung zwischen „Backend“ und „Frontend“ von dynamischen Interfaces erforschen;
- eine multimodale Kompetenz zu entwickeln, um die narrativen Konventionen des transmedialen Geschichtenerzählens und die relationale Logik von Web-Applikationen und Archiven zu entschlüsseln.¹²

Über die mannigfaltigen Herausforderungen eines „Updates“ der klassischen Hermeneutik im Bereich der historischen Quellen- und Werkzeugkritik finden sich in dem Blog des Doktorandenkollegs „Digital History & Hermeneutics“ zahlreiche reflexive Notizen und „Hands-on“-Erfahrungsberichte.¹³ An dieser Stelle möchte ich mich deshalb mit zwei anderen Dimensionen des Digitalen beschäftigen, welche einen vielleicht subtileren Einfluss auf die geschichtswissenschaftliche Forschung haben, da sie eher geschichtsphilosophischer und erkenntnistheoretischer Natur sind. Zum einen die Dimension des Symbolischen digitaler Werkzeuge und Infrastrukturen, also das, was Sheila Jasanoff und Sang-Hyun Kim als „sociotechnical imaginaries“ der „dreamscapes of modernity“ bezeichnen;¹⁴ zum anderen die Dimension des Imaginären, das heißt den

8 David Gugerli, *Suchmaschinen. Die Welt als Datenbank* (Frankfurt a. M.: Suhrkamp, 2009).

9 Carl Lagoze, „Big Data, Data Integrity, and the Fracturing of the Control Zone,“ *Big Data & Society* 1 (2014): 1–11.

10 Marijn Koolen, Jasmijn van Gorp und Jacco van Ossenbruggen, „Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice,“ *Digital Scholarship in the Humanities* 34 (2019) 2: 368–385.

11 Johanna Drucker, *Visualization and Interpretation. Humanistic Approaches to Display* (Cambridge, Massachusetts: MIT Press, 2020).

12 Steve F. Anderson, *Technologies of History. Visual Media and the Eccentricity of the Past* (New Hampshire: Dartmouth College Press, 2011); Tracy Bowen und Carl Withaus (Hg.), *Multimodal Literacies and Emerging Genres* (Pittsburgh: University of Pittsburgh Press, 2013).

13 Siehe Digital History & Hermeneutics, <https://dhh.uni.lu/category/blog>.

14 Sheila Jasanoff und Sang-Hyun Kim, *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power* (Chicago: University of Chicago Press, 2015).

Einfluss digitaler Quellenbestände, Archive und anderer Repräsentationen von Vergangenheit auf unsere historische Imagination und das, was Frank Ankersmit als „sublime historische Erfahrung“ bezeichnet.¹⁵

2 Mythos des Digitalen und neues Zeitregime?

Nicht nur in der populärwissenschaftlichen Literatur ist die Ära der Digitalität eng mit dem Nimbus des Revolutionären und der disruptiven Kraft radikaler Innovationen verwoben. Wie zuvor bei allen neuen Massenmedien geschehen, wurde auch der Erwartungshorizont des Digitalen durch soziotechnische Imaginationen geprägt, die auf relativ stabilen Tropen modernistischer und futuristischer Sprach- und Bildfiguren basieren.¹⁶ In seinem Buch *The Digital Sublime* aus dem Jahre 2005 hat der Kommunikationswissenschaftler Vincent Mosco die mythologische Dimension der Cyberspace- und Internetnarrative dekonstruiert, und auf die subtile Anziehungskraft jenes Diskurses hingewiesen, der das digitale Zeitalter als radikal gesellschaftstransformierenden Einschnitt in der Geschichte der Menschheit konstruiert.¹⁷

Auch in der Geschichtswissenschaft postulierte der Pionier der digitalen Geschichtswissenschaft Roy Rosenzweig im Jahre 2003 den Übergang von einem „age of scarcity“ in ein „age of abundance“, welches sich durch ein nie gekanntes Angebot an online zugänglichen Zeugnissen aus der nahen und fernen Vergangenheit kennzeichnet.¹⁸ Während sich der Topos der Informationsüberflutung wie ein roter Faden durch die Geschichte des Wissens zieht, hat das Internet und das World Wide Web in den Worten eines anderen Pioniers der digitalen Geschichtswissenschaften im deutschsprachigen Raum, Peter Haber, ein weiteres Phantasma wiederbelebt: das „Phantasma des universell verfügbaren Wissens“, das er in seinem Buch „Digital Past“ als „Google Syndrom“ beschrieben hat.¹⁹ Zehn Jahre nach Habers und Rosenzweigs Publikationen Anfang der 2000er Jahre wurden die Analogien von Wikipedia als Bibliothek Alex-

¹⁵ Frank Ankersmit, *Sublime Historical Experience* (Stanford: Stanford University Press, 2005).

¹⁶ Siehe beispielhaft Patrice Flichy, *The Internet Imaginaire* (Cambridge, Massachusetts: MIT Press, 2007).

¹⁷ Vincent Mosco, *The Digital Sublime. Myth, Power, and Cyberspace* (Cambridge, Massachusetts: MIT Press, 2004).

¹⁸ Roy Rosenzweig, „Scarcity or Abundance? Preserving the Past in a Digital Era,“ *The American Historical Review* 108/3 (2003): 735–762.

¹⁹ Peter Haber, *Digital Past. Geschichtswissenschaft im digitalen Zeitalter* (München: Oldenbourg Verlag, 2011).

andriens unserer Zeit und der Wikipedianer als Diderots legitime Erben von der „big data“-Metapher überrollt, die sich in Windeseile zum neuen Mantra gesellschafts- wie wissenschaftspolitischer Phantasien entwickelte.²⁰ Dies nur, um dann ihrerseits in den letzten zwei, drei Jahren von der x-ten Renaissance der Künstlichen Intelligenz als zentraler soziotechnischer Imagination überlagert zu werden.

Die in diesen Diskursen zum Ausdruck kommenden Erwartungshorizonte ließen sich problemlos in die lange Dauer stabiler narrativer Muster von technischen Fortschrittserzählungen einbetten.²¹ Doch dies ist nicht mein Anliegen. Vielmehr stelle ich mir die Frage, welche Vorstellungen von Zeit und historischer Imagination und Erfahrung in diesen Phantasmen, Metaphern und Mythen zum Ausdruck kommen, und welche Bedeutung diese Vorstellungen oder Erwartungshaltungen für die geschichtswissenschaftliche Forschung haben könnten. Seit der sozialkonstruktivistischen Wende der Geistes- und Sozialwissenschaften wissen wir, dass Diskurse keine passiven sprachlichen Abbildungen von Realität, sondern aktiver Teil der Gestaltung von Gesellschaft und Politik, Wissenschaft und Technik sind. Sie beeinflussen die Art und Weise, wie wir über Geschichte denken. Meine Ausgangsfrage lässt sich also folgendermaßen reformulieren: Haben wir es beim Zeitalter der Digitalität mit einem neuen Zeitregime zu tun, das unser Verhältnis zur Vergangenheit, Gegenwart und Zukunft auf andere, neue Weise strukturiert?

„We’ve become victims of our ever-increasing capacity to store, organise, instantly access, and share vast amounts of cultural data,“ so Simon Reynolds in seinem Buch *Retromania*. „Not only has there never before been a society so obsessed with the cultural artifacts of its immediate past, but there has never before been a society that is able to access the immediate past so easily and so copiously.“²² Der rapide technologische Wandel erzeugt laut Reynolds Gefühle der Nostalgie und habe neben der Vergangenheit auch die Gegenwart zum „foreign country“ gemacht. Diese Feststellung deckt sich mit den Befunden von Historiker:innen und Kulturwissenschaftler:innen wie François Hartog, Andreas Huyssen oder Aleida Assmann, die unserer Gegenwart ein kulturelles Zeitre-

20 Siehe beispielhaft Viktor Mayer-Schönberger und Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (London: John Murray, 2013).

21 Andreas Fickers, „Neither Good, nor Bad; nor Neutral: The Historical Dispositif of Communication Technologies,“ in *Journalism and Technological Change. Historical Perspectives, Contemporary Trends*, hg. v. Martin Schreiber und Clemens Zimmermann (Frankfurt a. M.: Campus, 2014), 30–52.

22 Simon Reynolds, *Retromania. Pop Culture’s Addiction to its Own Past* (London: Faber & Faber, 2011), 21.

gime attestieren, welches durch eine ausgedehnte, ja „gefäßige“ Gleichzeitigkeit gekennzeichnet ist „die alles in ihren Schlund zieht [...] und nicht nur die Differenz der Zeiten, sondern auch das historische Bewusstsein zerstört.“²³ „Un présent monstre“, so Hartog.²⁴ Gleichzeitig kennzeichnet sich dieser Päsentismus laut Aleida Assmann durch eine nie gekannte „Reaktualisierung der Vergangenheit“, einen neuen „Chronotopos“ aus, in dem sich die „Zugänge zur Vergangenheit vervielfältigt haben.“²⁵

Es überrascht, dass weder Assmann, Hartog noch Huyssen – zumindest nicht in systematischer oder argumentativ zwingender Weise – die digitale Wende als eine der möglichen Ursachen dieses neuen Zeitregimes anführen. Machen Assmann und Huyssen den „memory boom“ seit den 1980er Jahren als Hauptursache für das neue Zeitregime aus, verweist Hartog auf die Krise des modernen Zeitregimes, welches sich seit der Sattelzeit durch seine systematische Zukunftsorientierung auszeichnete. Dagegen machen medienwissenschaftliche Studien²⁶ sowie zahlreiche Autor:innen aus dem Feld der Digital Humanities die „digitale Revolution“ und den „big data turn“ für die „breite Gegenwart“²⁷ verantwortlich. Die Kultur der Digitalität zeichnet sich aus dieser Perspektive durch eine sozio-technische Imagination aus, die von der Idee der Konnektivität bestimmt wird. Laut Andrew Hoskins, einem der führenden Autor:innen im Bereich der *Digital Memory Studies*, ermöglicht der „connective turn“ einen bislang ungekannten und privilegierten Zugang zur Vergangenheit: „The networked self and society foster a view that collapses past and present into an orgy of hyperconnectivity: an impossible fantasy of prior generations with their now forgotten closed and contained media imaginaries.“²⁸

23 Andreas Huyssen, „The Crisis of Success: What Next in Memory Studies?“ *Memory Studies* 5/2 (2012): 226–228.

24 François Hartog, *Régimes d'historicité. Présentisme et expériences du temps* (Paris: Éditions du Seuil, 2012), 270.

25 Aleida Assmann, *Ist die Zeit aus den Fugen? Aufstieg und Fall des Zeitregimes in der Moderne* (München: Carl Hanser, 2013), 277.

26 Josée Van Dijk, *The Culture of Connectivity. A Critical History of Social Media* (Oxford: Oxford University Press, 2013).

27 Hans Ulrich Gumbrecht, *Unsere breite Gegenwart* (Frankfurt a. M.: Suhrkamp, 2010).

28 Andrew Hoskins (Hg.), *Digital Memory Studies. Media Pasts in Transition* (London: Routledge, 2018), 2.

3 Hyperkonnektivität und digitales Vergessen

Während Hoskins in der Hyperkonnektivität eine neue Erinnerungskultur in Echtzeit erkennt, deuten andere das digitale Zeitalter als Zeitregime, das sich durch massenhaftes Vergessen auszeichnet: „In the digital age, in what is perhaps the most fundamental change for humans since our humble beginnings, that balance of remembering and forgetting has become inverted. Committing information to digital memory has become the default, and forgetting the exception,“²⁹ so Viktor Mayer-Schönberger in seinem Essay *Delete. The Virtue of Forgetting in the Digital Age*. Mayer-Schönberger macht hierfür eine simple Rechnung auf: Waren der Aufwand für die Speicherung und Inszenierung analoger Zeugnisse im Kontext einer antizipativen Erinnerungspraxis immens hoch, sind die Kosten für digitale Speicherung so dramatisch gesunken, dass es sich aus ökonomischer Hinsicht nicht mehr lohnt, Zeit in die Selektionsarbeit zu investieren.³⁰ Diese neue Form des digitalen Vergessens bezeichnet Aleida Assmann als „Verwahrensvergessen“³¹: Wir speichern Tausende von digitalen Photos, Videos und E-Mails auf unseren Smartphones oder Laptops in der naiven Hoffnung, sie irgendwann einmal wieder hervorzukramen!

Noch pertinenter ist die Frage des digitalen Vergessens beim Leitmedium des digitalen Zeitalters: dem Internet. Entgegen der weit verbreiteten Auffassung, das Internet vergesse nichts, warnte bereits die UNESCO in ihrer 2003 veröffentlichten *Charter on the Preservation of Digital Heritage* vor einem zukünftigen „digital dark age“³². Studien zur Halbwertszeit von Hyperlinks im Internet zeigen, dass von 100 Links nach zwei Jahren nur noch 50 % funktionieren. Zu diesem als „link rot“ bezeichneten Phänomen gesellt sich die Tatsache, dass die durchschnittliche Lebensdauer einer Webseite nur acht Monate beträgt. Durch den exponentiellen Anstieg neuer Datenmengen wird die dem dynamischen Charakter des Netzes eingeschriebene Instabilität kaschiert. Dieser ephemere Charakter des World Wide Web stellt zukünftige Historiker:innengenerationen vor enorme methodologische Herausforderungen. Das Internet wird bisher nur extrem lückenhaft archiviert, und selbst dort, wo nationale oder private Initiativen existieren, haben wir es mit speziellen digitalen Konglomeraten zu tun, die

²⁹ Viktor Mayer-Schönberger, *Delete. The Virtue of Forgetting in the Digital Age* (Princeton: Princeton University Press, 2009), 196.

³⁰ Ibid., 68.

³¹ Aleida Assmann, *Formen des Vergessens* (Göttingen: Wallstein, 2016), 36–41.

³² *Charter on the Preservation of the Digital Heritage* (2009) <https://unesdoc.unesco.org/ark:/48223/pf0000179529.page=2>. Zugriff am 10.12.2021.

Niels Brügger als „reborn digital medium“ beschreibt.³³ Das World Wide Web ist ein Medium, welches sichtbare und unsichtbare Textebenen, Hyperlinks und fragmentierte Informationseinheiten in einer komplexen Netzwerkstruktur miteinander verbindet. Diese dynamische und relationale Architektur ist dafür verantwortlich, dass wir es beim Wiederaufrufen archivierter Webpages mit „digitalen Wiedergeburten“ zu tun haben, deren informationstechnische, inhaltliche wie darstellerische Integrität auf multiple Weise korrumpiert sind. Eine archivierte und „wiedergeborene“ Webpage ist immer das Resultat einer doppelten Transformation: Sowohl die Archivierung (etwa durch die Benutzung von „Webcrawling Software“³⁴) als auch die Wiederaktivierung (zum Beispiel durch die „Wayback Machine“³⁵) greifen aktiv in die semantische Ordnung und visuelle Repräsentation der relationalen Datenbestände ein und machen so klassische Prinzipien der Archivwissenschaft wie „respect des fonds“ und „respect de l'ordre“ obsolet.³⁶

Es geht mir an dieser Stelle aber nicht um das Problem der historischen Datenintegrität von Webseiten, sondern um die dem Web eingeschriebene doppelte Temporalität als „passive storage“ und „active memory“-Medium. Während Historiker:innen beim Besuch klassischer Archive nicht verändernd in die Organisation und Logik des „Systems“ eingreifen, werden Nutzer:innen der „Wayback Machine“ zu Ko-Konstrukteur:innen eines neuen epistemischen Objekts, da es die „wiederbelebte“ Webseite so in der Vergangenheit nie gab. Wendy Chun, Professorin für Neue Medien an der Simon Fraser University in Vancouver/Canada spricht deshalb vom World Wide Web als „enduring ephemeral [...] creating unforeseen degenerative links between humans and machines.“³⁷ An-

33 Niels Brügger, *The Archived Web. Doing History in the Digital Age* (Cambridge, Massachusetts: MIT Press, 2018), 6.

34 Eine praxisorientierte Beschreibung dieser Technik und ihrer Implikationen für die digitale Quellenkritik findet sich bei Ian Milligan, *History in the Age of Abundance? How the Web is Transforming Historical Research* (Montreal: McGill-Queen's University Press 2019), besonders Kapitel 6: The (Practical) Historian in the Age of Big Data, 213–235.

35 Valérie Schafer, „De la Wayback Machine à la bibliothèque: les différents saveurs des archives du Web,“ *La Gazette des archives* 253/1 (2019): 95–107. Zur Wayback-Machine siehe: <https://archive.org/web/>.

36 Jefferson Bailey, „Disrespect des Fonds: Rethinking Arrangement and Description in Born-Digital Archives,“ *Archive Journal* (June 2013), <https://www.archivejournal.net/essays/disrespect-des-fonds-rethinking-arrangement-and-description-in-born-digital-archives/>. Zugriff am 16.12.2021; Anat Ben-David und Adam Amram, „The Internet Archive and the socio-technical construction of historical facts,“ *Internet Histories. Digital Technology, Culture & Society* 2 (2018) 1/2: 179–201.

37 Wendy Hui Kyong Chung, „The Enduring Ephemeral, or the Future Is a Memory,“ *Critical Inquiry* 35 (2008) 1: 148–171, hier 148.

ders als die auch anderen Quellengattungen innewohnende Spannung zwischen „evidence“ und „erasure“ – man denke hier beispielsweise an Photographien oder Tonbandaufzeichnungen, haben wir es bei Digitalisaten und besonders dem World Wide Web mit einer Gleichzeitigkeit des Ungleichzeitigen zu tun, dessen Zeitlichkeit sich mit den Worten der brasilianischen Historiker Mateus Pereira und Valdei Araujo als „updatism“ beschrieben ließe: „a specific temporalization of time that connects the past with the present and the future.“³⁸ Entgegen der pessimistischen Note, welche in Hartogs Konzept des Präsentismus mitschwingt und in dem der Sieg der Gedächtnispolitik über die Geschichte bedauert wird, oder Gumbrechts Idee der „breiten Gegenwart“, in der das definitive Ende des Prozesses der Moderne beklagt wird, bedeutet „updatism“ den kontinuierlichen Akt der Reaktualisierung vergangener, gegenwärtiger wie zukünftiger Zeitbezüge in „Echtzeit“. Und dies eben nicht als individuellen Akt historischer Vergegenwärtigung, sondern als Resultat eines kollektiven und vernetzten Prozesses medialer Teilhabe und Partizipation.

4 Dynamische Datenvisualisierungen und Hermeneutik des Dazwischenseins

Doch bedeutet diese digitale Verfügbarkeit, dass sich mit dem mobilen Zuhandensein von Vergangenem auf Tablets und Smartphones auch die historische Erfahrung und Imagination verändern? Durchaus! Hierzu zwei Beispiele. Wie Claudio Fogu, Wulf Kansteiner und Todd Presner am Beispiel der Erinnerung an den Holocaust gezeigt haben, stellt die virtuelle Inszenierung von Zeitzeugen im *Visual History Archive* der Shoah Foundation eine ästhetische und moralische Herausforderung für die Geschichtspädagogik wie Gedächtnispolitik dar, da sie direkt auf die Wahrnehmung und Erfahrung von authentischen Zeugnisaussagen oder Orten – etwa während eines Besuches des Konzentrationslagers Auschwitz-Birkenau – einwirken können.³⁹ Die „Selfie-Debatte“⁴⁰ von 2014 hat

38 Mateus H. F. Pereira und Valdei Araujo, „Updatism: Gumbrecht’s broad present, Hartog’s Presentism and beyond“ *Diacronie. Studi di Storia Contemporanea: Miraggi e realtà: spazi e tempi della storia*, 43 (2020) 3. http://www.studistorici.com/2020/10/29/pereira-araujo_numero_43/. Zugriff am 16.12.2021.

39 Claudio Fogu, Wulf Kansteiner und Todd Presner (Hg.), *Probing the Ethics of Holocaust Culture* (Harvard: Harvard University Press, 2016).

40 Im Juni 2014 hatte die 18-jährige US-Amerikanerin Breanna Mitchel ein Selfie von ihrem Besuch in Auschwitz-Birkenau gepostet, was international für Entrüstung gesorgt hatte. Zu dieser Debatte siehe Maria Zalewska, „Selfies from Auschwitz: Rethinking the Relationship Bet-

die komplexe Überlagerung von institutionalisierter Gedenkkultur und digitaler Medienpraxis beispielhaft zum Vorschein gebracht: Authentisches Erleben vor Ort und synchrones Teilen der eigenen „dark tourism“-Erfahrungen in sozialen Medien verschmelzen in der digitalen Erinnerungskultur zu einem historischen Erlebnis. Erfolgreiche Inszenierung des Authentischen erfordere im digitalen Zeitalter geradezu den Einsatz digitaler Medien, so Kansteiner.⁴¹

Am Beispiel der Erinnerung an den Ersten Weltkrieg auf Twitter (zum 100-jährigen Gedenken) hat Frédéric Clavert gezeigt, dass die Zirkulation von Nachrichten in den sozialen Medien ein eigenes Zeitregime bilden, das sich radikal von der „longue durée“ des kulturellen Gedächtnisses unterscheidet, welches die Erinnerung an den Ersten Weltkrieg prägt.⁴² Auf Basis der Analyse von über drei Millionen Tweets, die er zwischen dem 1. April 2014 und dem 22. November 2016 auf Basis von deutschen, französischen und englischen Hashtags in einer Datenbank archiviert hat, stellt Clavert die Hypothese auf, dass das Erinnern im Hochfrequenztakt der sozialen Medien eine neue Qualität kommunikativer Gedächtnisbildung bedeutet, die sich durch verteilte Autorität und einen kollektiven Bezug zu den Spuren der Vergangenheit auszeichnet.

Akzeptiert man den von Andreas Hepp postulierten kommunikationswissenschaftlichen Befund der „deep mediatization“ der besagt, dass sich unsere Gegenwart durch Rhythmussteigerung und erhöhtes Tempo der Ko-Konstruktion medialer Realität auszeichnet, so scheint es folgerichtig, dass auch unsere Erfahrung und Imagination der Vergangenheit – die im Wesentlichen medial vermittelt sind – von dieser Entwicklung betroffen sind.⁴³ War die Geschichtskultur des 19. Jahrhunderts durch „sinnliche Anmutungskraft“⁴⁴ historischer Romane oder Nationalmuseen, die des 20. Jahrhunderts durch die auratische Wirkung von Ton- und Bildaufzeichnungen geprägt,⁴⁵ so regen im digitalen

ween Spaces of Memory and Places of Commemoration in the Digital Age,” *Studies in Russian, Eurasian and Central European New Media* 18 (2017): 95–116.

41 Wulf Kansteiner, „The Holocaust in the 21st Century: Digital Anxiety, Transnational Cosmopolitanism, and Never Again Genocide Without Memory,” in *Digital Memory Studies: Media Pasts in Transition*, hg. v. Andrew Hopkins (London: Routledge, 2018), 110–140.

42 Frédéric Clavert, „Face au passé: la Grande Guerre sur Twitter,” *Le Temps des médias* 31/2 (2018): 173–186.

43 Andreas Hepp, *Deep Mediatization* (London: Routledge, 2019).

44 Martin Sabrow und Achim Saupe, „Historische Authentizität. Zur Kartierung eines Forschungsfeldes,” in *Historische Authentizität*, hg. v. Martin Sabrow und Achim Saupe (Göttingen: Wallstein, 2016), 14.

45 Andreas Fickers, „Entre vérité et dire du vrai: Ein geschichtstheoretischer Grenzgang,” *Jeux sans Frontières? – Grenzgänge der Geschichtswissenschaft*, hg. v. Andreas Fickers, Rüdiger Hau-de, Stefan Krebs und Werner Tschacher (Bielefeld: transcript: 2018), 29–40.

Zeitalter dynamische historische Netzwerkvisualisierungen, animierte „deep mapping“-Technologien und interaktive multilineare Zeitachsen in virtuellen Ausstellungen die Phantasie und historische Imagination an.

Interaktive Interfaces und relationale Datenbanken, gefüllt mit Tausenden oder Millionen Quellen aus unterschiedlichen Genres, erzeugen einen neuen historischen Sinn, der sich laut Alan Liu der Ideologie der Linearität von historischem Denken verweigert.

We might say that the essential hermeneutic – or what we might today call algorithm – of *Historismus* was to interpret all the spatial (and political) barriers that impeded full-on human sociality as *temporal delay*. Civilization was the delayed action of sociality unfolding in historical time. [...] The temporality of shared culture is thus no longer experienced as unfolding narration but instead as ‘real time’ media.⁴⁶

Dynamische Visualisierungen der komplexen Beziehung von historischen Prozessen und Ereignissen erzeugen ein neues historisches Wissen, welches Liu als „hypergraphical knowledge“ bezeichnet:

The digital age promotes hypergraphical models of knowledge that conform to a world view in which knowledge is conceived by default to be multiperspectival and multiscalar, distributed in its foci and relations, and (connecting all the disparate nodes and levels) ultimately networked.⁴⁷

Wie Liu beschreibt auch der amerikanische Medientheoretiker Lev Manovich die digitalen Datenvisualisierungen kultureller Artefakte und Sammlungen als „neue Sprache“, die wir erlernen müssen, wenn wir in Zukunft „Sinn“ aus den „big data of the past“ generieren möchten. Denn bislang überlassen wir diese Sinnbildung – die sich meist auf die Sichtbarmachung von Mustern reduziert, die auf Basis statistischer Modellbildung erzeugt werden – den Computer- und Datenwissenschaftler:innen, deren epistemische Kultur sich deutlich von der hermeneutischen Tradition der Geisteswissenschaften unterscheidet. Zwar teilen sich beide Wissenschaftskulturen nun oftmals dieselben digitalen Werkzeuge und computer- bzw. softwarebasierten Methoden, aber die Problemstellungen und Untersuchungsobjekte divergieren erheblich:

Thus, we have two research universes that often use the same computational methods, but apply them to different ‘cultures’. On the humanities side, we have the past that stretches into hundreds or even thousands of years. On the computer science side, we

⁴⁶ Alan Liu, *Friending the Past. The Sense of History in the Digital Age* (Chicago: Chicago University Press, 2018), 19–20.

⁴⁷ Ibid., 73.

have the present that starts in the beginning of the 21st century. On the humanities side, we have artifacts created by professional elites. On the computer science side, we have artifacts and online behavior by everybody else.⁴⁸

Sind die meisten sozio-kulturellen Daten, welche anhand algorithmusbasierter machine-learning Methoden in den Computerwissenschaften analysiert werden, sogenannte „digital born“-Quellen (häufig Daten aus Sozialen Medien), haben wir es bei historischen Datenbeständen meist mit wesentlich kleineren Datenmengen zu tun, die sich generell durch größere Komplexität und Diversität auszeichnen.

Die meisten Werkzeuge, die in den digitalen Geschichtswissenschaften im Bereich des Text Mining zum Einsatz kommen – etwa zur Eigennamenerkennung oder Themenanalyse – sind darauf programmiert, Muster zu erkennen oder Häufigkeitsverteilungen zu berechnen, nicht aber Abweichungen und Partikularitäten zu identifizieren, die uns als Historiker:innen meist mehr interessieren als statistische Mittelwerte.⁴⁹ Diese Spannung zwischen der maschinenbetriebenen Suche nach Mustern und Trends und der problem- und frageorientierten Suche nach dem historisch Speziellen und Partikularen erfordert auch methodologisch eine Hermeneutik des Dazwischenseins, die Stephen Ramsey als „hermeneutics of screwing around“⁵⁰ beschrieben hat:

Trying to locate a hermeneutics at the boundary between mechanism and theory [...]. Algorithmic criticism proposes that we channel the heightened objectivity made possible by the machine into the cultivation of those heightened subjectivities necessary for critical work.⁵¹

Die Erkenntnis, dass „Daten“ in den Geisteswissenschaften komplex, unscharf und unvollständig sind, ist Allgemeingut. Es überrascht daher, dass viele Forschungsprojekte auch in den digitalen Geschichtswissenschaften sich auf relativ simple Fragestellungen reduzieren – etwa der Suche beziehungsweise Identifikation von sogenannten „named entities“ wie Personen-, Orts- und Zeitangaben. Fragen nach dem Wo, Wer und Wann könnte man boshaft als kleines Einmaleins der historischen Arbeit bezeichnen. Fragen nach dem Wie, Warum

⁴⁸ Lev Manovich, *Cultural Analytics* (Cambridge, Massachusetts: MIT Press, 2019), 46.

⁴⁹ Andreas Fickers, „Veins Filled with the Diluted Sap of Rationality: A Critical Reply to Jens Bod,“ *Low Countries Historical Review* 128/4 (2013): 155–163.

⁵⁰ Stephen Ramsey, „The Hermeneutics of Screwing Around; or What You Do with a Million Books,“ in *Pastplay. Teaching and Learning History with Technology*, hg. v. Kevin Kee (Ann Arbor: University of Michigan Press, 2014), 111–120.

⁵¹ Stephen Ramsey: *Reading Machines. Toward an Algorithmic Criticism* (Illinois: University of Illinois Press, 2011): x.

oder Wozu lassen sich mit den meisten digitalen Werkzeugen bis heute kaum beantworten. Viele textbasierte Projekte der ersten und zweiten Welle der Digital Humanities zeichnen sich durch eine eher positivistische Herangehensweise aus, was der Logik der Instrumente und Werkzeuge geschuldet war, mit denen die digitalen Korpora „befragt“ wurden. Das erkenntnistheoretische Potenzial solcher Arbeiten steigt meines Erachtens erheblich mit der Überwindung statistischer Modellbildung in Richtung frage- und problemgesteuerter Simulationen. Erst wenn die auf Mustererkennung und Häufigkeitsverteilungen spezialisierten Algorithmen zu dynamischen Simulationen historischen Wandels weiterentwickelt werden können, werden die digitalen Geschichtswissenschaften neue Impulse zur Theoriebildung komplexer historischer Prozesse beitragen.

Es war der überwiegend a-historische Ansatz der positivistischen und empiristischen Forschung im 19. Jahrhundert, der Wilhelm Dilthey dazu angespornt hat, seine hermeneutische Theorie der Geisteswissenschaften zu formulieren. Ein historisches Verständnis der Vergangenheit fußt laut Dilthey notwendigerweise auf der Erkenntnis der grundsätzlichen Geschichtlichkeit sowohl vergangener als auch gegenwärtiger Tatsachen und Wirkungen. Ohne die Möglichkeit struktureller Kausalitäten zu leugnen, betonte Dilthey die Individualität der Wahrnehmung, der Vorstellung und des Denkens, die nötig sei, um einen kritischen Zugang zur Geschichte zu entwickeln und sowohl strukturelle Kräfte als auch individuelles Handeln deutend verstehen zu können. Übertragen auf die digitalen Geschichtswissenschaften heißt dies in meinen Augen, dass erst die Kombination von maschinenbasierten Methoden des *distant reading* mit der hermeneutischen Methode des *close reading* dazu führen wird, generische Phänomene mit partikularen Erscheinungen in Verbindung zu setzen, und so statistische Beobachtungen und Korrelationen auf ihre historische Relevanz überprüfen zu können. Diese Fähigkeit des „scaleable reading“, des Ein- und Auszoomens vom Allgemeinen zum Spezifischen – also die Kombination der explorativen und interpretativen Analyse von Datenbeständen – ist eine neue Episode der Kulturtechnik des Lesens, die uns augenblicklich noch fremd ist.⁵² Es wird wohl eine Generation dauern, bis diese neue Form des skalenbasierten Lesens uns so vertraut ist, wie uns heute das Lesen am Bildschirm oder das vertikale Scrollen am Smartphone ist. Wie schon bei Dilthey gilt es, uns dem Fremden im hermeneutischen Prozess des Dazwischenseins anzunähern:

Die Auslegung wäre unmöglich, wenn die Lebensäußerungen gänzlich fremd wären. Sie wäre unnötig, wenn in ihnen nichts fremd wäre. Zwischen diesen beiden äußersten Ge-

52 Gerhard Lauer, *Lesen im digitalen Zeitalter* (Darmstadt: wbg Academic, 2020).

gensätzen liegt sie also. Sie wird überall erfordert, wo etwas fremd ist, das die Kunst des Verstehens zu eigen machen soll,

so Dilthey in seiner Vorlesung zum *Aufbau der geschichtlichen Welt in den Geisteswissenschaften* aus dem Jahre 1910.⁵³

5 Die Aura des Digitalen und deren Entzauberung

Wenn wir uns als Historiker:innen heute also computerbasierten Analysen und Interpretationen der Vergangenheit nähern, vollziehen wir genau jene hermeneutische Bewegung, die schon immer die Arbeit des fragend-forschenden Geschichtswissenschaftlers war: Wir wenden das von Carlo Ginzburg beschriebene diagnostische Verfahren der Spurensuche an und erzeugen gleichzeitig historischen Sinn durch die Produktion einer graphischen Erzählung – nunmehr im digitalen Raum „retrospektiver Wahrsagung“⁵⁴. Dass uns die informationstechnischen und datenbankstrukturellen Hintergründe (oder das „Backend“) solcher digitalen Inszenierungen von historischem Sinn oftmals verborgen bleiben, macht im Benjaminschen Sinne vielleicht gerade deren Aura aus:

Die Spur ist Erscheinung einer Nähe, so fern das sein mag, was sie hinterließ. Die Aura ist Erscheinung einer Ferne, so nah das sein mag, was sie hervorruft. In der Spur werden wir der Sache habhaft; in der Aura bemächtigt sie sich unser.⁵⁵

Auch wenn wohl nicht alle Historiker:innen von der Aura digitaler Repräsentationen und Visualisierungen gleichermaßen erfasst sein dürften, bin ich davon überzeugt, dass es für zukünftige Generationen von Historiker:innen von entscheidender Bedeutung sein wird, sich kritisch mit den Werkzeugen und Methoden der digitalen Datenanalyse und den Möglichkeiten datenbasierter Zeit- und Raumvisualisierungen auseinanderzusetzen. Sich als Historiker:innen in interdisziplinären und kollaborativen Projekten auf die Auseinandersetzung mit Methoden und Konzepten der Datenwissenschaften einzulassen, bedeutet nicht

⁵³ Wilhelm Dilthey, *Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften* (Frankfurt a. M.: Suhrkamp, 1981), 278.

⁵⁴ Carlo Ginzburg, *Spurensicherung. Die Wissenschaft auf der Suche nach sich selbst* (Berlin: Wagenbach, 2011); Patrick Boucheron, „La trace et l’aura. Un court-circuit de Walter Benjamin,“ in *Faire profession d’historien*, hg. v. Patrick Boucheron (Paris: Éditions de la Sorbonne, 2016), 9–35.

⁵⁵ Walter Benjamin, *Das Passagen-Werk*, in *W. Benjamin, Gesammelte Schriften Bd. V* (Frankfurt a. M.: Suhrkamp, 1982), 560.

nur klassische Kompetenzen wie die Quellenkritik oder Diskursanalyse auf digitale Daten und Korpora zu übertragen und den Spezifika digitaler Medien anzupassen, sondern auch einen Wissens- und Kompetenztransfer in Richtung der technischen Disziplinen. „If content is king, context is its crown“ – so brachte der schwedische Medienhistoriker Pelle Snickars die Bedeutung der hermeneutischen Methode für die Digital Humanities vor einigen Jahren auf den Punkt.⁵⁶

Genau diese Kontextualisierung von großen Datenmengen wird in interdisziplinären Forschungsfeldern wie der Klima- und Umweltforschung als *die* zentrale Herausforderung für die Zukunft gesehen – und beschäftigt auch Forschende der Geschichte, Archäologie und Anthropologie, die sich der Vergangenheit mit dem Konzept der „deep history“ annähern.⁵⁷ Wie aussagekräftig komplexe Modelle und Simulationen über vergangene Realitäten und zukünftige Szenarien sein werden, wird eben nicht nur von der Menge der vorhandenen Daten abhängen, sondern von deren Kontextualisierung und kritischen Analyse und Interpretation. „Big data is the answer... But what is the question?“ – fragten 2017 die beiden Wissenschafts- und Technikhistoriker Bruno Strasser und Paul Edwards in einem Artikel der Zeitschrift *Osiris*.⁵⁸ Als kritische Wissenschaft, die frage- und problemgesteuert ist, kann die Geschichtswissenschaft hier einen wichtigen Beitrag für eine Wissensökonomie des 21. Jahrhunderts leisten, was – auf dezidiert provokative Weise – 2014 von Jo Guldi und David Armitage in der Streitschrift *The History Manifesto* eingefordert wurde.⁵⁹ Es geht also nicht nur darum zu begreifen, „what the D does to history“, sondern auch danach zu fragen, „what history does to D“.

Bibliographie

- Anderson, Steve F. *Technologies of History. Visual Media and the Eccentricity of the Past*. New Hampshire: Dartmouth College Press, 2011.
- Ankersmit, Frank. *Sublime Historical Experience*. Stanford: Stanford University Press, 2005.
- Assmann, Aleida. *Ist die Zeit aus den Fugen? Aufstieg und Fall des Zeitregimes in der Moderne*. München: Carl Hanser, 2013.

⁵⁶ Pelle Snickars, „If Content is King, Context is its Crown,“ *VIEW. Journal of European Television History and Culture* 1/1 (2012): 34–39.

⁵⁷ Andrew Shryock und Daniel Lord Smail (Hg.), *Deep History. The Architecture of Past and Present* (Berkeley: University of California Press, 2011).

⁵⁸ Bruno J. Strasser und Paul N. Edwards, „Big Data Is the Answer... But What Is the Question?“ *Osiris* 32/1 (2017): 328–345.

⁵⁹ Jo Guldi und David Armitage, *The History Manifesto* (Cambridge: Cambridge University Press, 2014).

- Assmann, Aleida. *Formen des Vergessens*. Göttingen: Wallstein, 2016.
- Bailey, Jefferson. „Disrespect des Fonds: Rethinking Arrangement and Description in Born-Digital Archives.“ *Archive Journal* (June 2013), <https://www.archivejournal.net/essays/disrespect-des-fonds-rethinking-arrangement-and-description-in-born-digital-archives/>. Zugriff am 17.12.2021.
- Ben-David, Anat und Adam Amram. „The Internet Archive and the Socio-Technical Construction of Historical Facts.“ *Internet Histories. Digital Technology, Culture & Society* 2 (2018) 1/2: 179–201.
- Benjamin, Walter. *Das Passagen-Werk*. In *W. Benjamin, Gesammelte Schriften Bd. V*. Frankfurt a. M.: Suhrkamp, 1982.
- Boucheron, Patrick. „La trace et l'aura. Un court-circuit de Walter Benjamin.“ In *Faire profession d'historien*. Hg. v. Patrick Boucheron, 9–35. Paris: Éditions de la Sorbonne, 2016.
- Bowen, Tracy und Carl Withaus (Hg.). *Multimodal Literacies and Emerging Genres*. Pittsburgh: University of Pittsburgh Press, 2013.
- Breibach, Olaf, Peter Heering, Matthias Müller und Heiko Weber (Hg.). *Experimentelle Wissenschaftsgeschichte*. München: Brill, 2010.
- Brügger, Niels. *The Archived Web. Doing History in the Digital Age*. Cambridge, Massachusetts: MIT Press, 2018.
- Cassirer, Ernst. *Symbol, Technik, Sprache. Aufsätze aus den Jahren 1927–1933*. Hamburg: Meiner, 1995.
- Clavert, Frédéric. „Face au passé: la Grande Guerre sur Twitter.“ *Le Temps des médias* 31/2 (2018): 173–186.
- Dilthey, Wilhelm. *Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften*. Frankfurt a. M.: Suhrkamp, 1981.
- Drucker, Johanna. *Visualization and Interpretation. Humanistic Approaches to Display*. Cambridge, Massachusetts: MIT Press, 2020.
- van Dijk, Josée. *The Culture of Connectivity. A Critical History of Social Media*. Oxford: Oxford University Press, 2013.
- Fickers, Andreas. „Veins Filled with the Diluted Sap of Rationality: A Critical Reply to Jens Bod.“ *Low Countries Historical Review* 128/4 (2013): 155–163.
- Fickers, Andreas. „Neither Good, nor Bad; nor Neutral: The Historical Dispositif of Communication Technologies.“ In *Journalism and Technological Change. Historical Perspectives, Contemporary Trends*. Hg. v. Martin Schreiber und Clemens Zimmermann, 30–52. Frankfurt a. M.: Campus, 2014.
- Fickers, Andreas. „Entre vérité et dire du vrai: Ein geschichtstheoretischer Grenzgang.“ In *Jeux sans Frontières? – Grenzgänge der Geschichtswissenschaft*, hg. v. Andreas Fickers, Rüdiger Haude, Stefan Krebs und Werner Tschacher, 29–40. Bielefeld: transcript: 2018.
- Fickers, Andreas. „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?“ *Zeithistorische Forschungen* 17 (2020) 1: 157–168.
- Fickers, Andreas und Juliane Tatarinov (Hg.). *Digital Hermeneutics. Between Theory and Practice*. Berlin: DeGruyter, 2022 (im Druck).
- Fligny, Patrice. *The Internet Imaginaire*. Cambridge, Massachusetts: MIT Press, 2007.
- Fogu, Claudio, Wulf Kansteiner und Todd Presner (Hg.). *Probing the Ethics of Holocaust Culture*. Harvard: Harvard University Press, 2016.
- Gadamer, Hans-Georg. *Wahrheit und Methode. Grundzüge einer philosophischen Hermeneutik*. Tübingen: Mohr Siebeck, 2010.

- Ginzburg, Carlo. *Spurensicherung. Die Wissenschaft auf der Suche nach sich selbst*. Berlin: Wagenbach, 2011.
- Gugerli, David. *Suchmaschinen. Die Welt als Datenbank*. Frankfurt a. M.: Suhrkamp, 2009.
- Guldi, Jo und David Armitage. *The History Manifesto*. Cambridge: Cambridge University Press, 2014.
- Gumbrecht, Hans Ulrich. *Unsere breite Gegenwart*. Frankfurt a. M.: Suhrkamp, 2010.
- Haber, Peter. *Digital Past. Geschichtswissenschaft im digitalen Zeitalter*. München: Oldenbourg, 2011.
- Hartog, François. *Régimes d'historicité. Présentisme et expériences du temps*. Paris: Éditions du Seuil, 2012.
- van der Heijden, Tim, Juliane Tatarinov und Gerben Zaagsma, „Transatlantic Perspectives on Digital Hermeneutics“, 5.11.2019, <https://www.c2dh.uni.lu/de/thinking/transatlantic-perspectives-digital-hermeneutics>. Zugriff am 17.12.2021.
- Hepp, Andreas. *Deep Mediatization*. London: Routledge, 2019.
- Hoskins, Andrew (Hg.). *Digital Memory Studies. Media Pasts in Transition*. London: Routledge, 2018.
- Hui Kyong Chung, Wendy. „The Enduring Ephemeral, or the Future Is a Memory.“ *Critical Inquiry* 35 (2008) 1: 148–171.
- Huyssen, Andreas. „The Crisis of Success: What Next in Memory Studies?“ *Memory Studies* 5/2 (2012): 226–228.
- Jasanoff, Sheila und Sang-Hyun Kim. *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press, 2015.
- Kansteiner, Wulf. „The Holocaust in the 21st Century: Digital Anxiety, Transnational Cosmopolitanism, and Never Again Genocide Without Memory“. In *Digital Memory Studies: Media Pasts in Transition*, hg. v. Andrew Hopkins, 110–140. London: Routledge, 2018.
- Koolen Marijn, Jasmijn van Gorp und Jacco van Ossenbruggen, „Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice.“ *Digital Scholarship in the Humanities* 34 (2019) 2: 368–385.
- Lagoze, Carl. „Big Data, Data Integrity, and the Fracturing of the Control Zone.“ *Big Data & Society* 1 (2014): 1–11.
- Lauer, Gerhard. *Lesen im digitalen Zeitalter*. Darmstadt: wbv Academic, 2020.
- Liu, Alan. *Friending the Past. The Sense of History in the Digital Age*. Chicago: Chicago University Press, 2018.
- Mayer-Schönberger, Viktor. *Delete. The Virtue of Forgetting in the Digital Age*. Princeton: Princeton University Press, 2009.
- Mayer-Schönberger, Viktor und Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray, 2013.
- Manovich, Lev. *Cultural Analytics*. Cambridge, Massachusetts: MIT Press, 2019.
- Milligan, Ian. *History in the Age of Abundance? How the Web is Transforming Historical Research*. Montreal: McGill-Queen's University Press 2019.
- Mosco, Vincent. *The Digital Sublime. Myth, Power, and Cyberspace*. Cambridge, Massachusetts: MIT Press, 2004.
- Pereira, Mateus H. F. und Valdeir Araujo, „Updatism: Gumbrecht's broad present, Hartog's Presentism and beyond.“ *Diacronie. Studi di Storia Contemporanea: Miraggi e realtà: spazi e tempi della storia*, 43 (2020). http://www.studistorici.com/2020/10/29/pereira-araujo_numero_43/. Zugriff am 17.12.2021.

- Ramsey, Stephen. *Reading Machines. Toward an Algorithmic Criticism*. Illinois: University of Illinois Press, 2011.
- Ramsey, Stephen. „The Hermeneutics of Screwing Around; or What You Do with a Million Books.“ In *Pastplay. Teaching and Learning History with Technology*, hg. v. Kevin Kee, 111–120. Ann Arbor: University of Michigan Press, 2014.
- Reynolds, Simon. *Retromania. Pop Culture's Addiction to its Own Past*. London: Faber & Faber, 2011.
- Rosenzweig, Roy. „Scarcity or Abundance? Preserving the Past in a Digital Era.“ *The American Historical Review* 108/3 (2003): 735–762.
- Sabrow, Martin und Achim Saupe. „Historische Authentizität. Zur Kartierung eines Forschungsfeldes.“ In *Historische Authentizität*, hg. v. Martin Sabrow und Achim Saupe. 7–29. Göttingen: Wallstein, 2016.
- Schafer, Valérie. „De la Wayback Machine à la bibliothèque: les différents saveurs des archives du Web.“ *La Gazette des archives* 253/1 (2019): 95–107.
- Serres, Michel. *Die fünf Sinne. Eine Philosophie der Gemenge und Gemische*. Frankfurt a. M.: Suhrkamp, 1998.
- Shryock, Andrew und Daniel Lord Smail (Hg.). *Deep History. The Architecture of Past and Present*. Berkeley: University of California Press, 2011.
- Snickars, Pelle. „If Content is King, Context is its Crown.“ *VIEW. Journal of European Television History and Culture* 1/1 (2012): 34–39.
- Strasser, Bruno J. und Paul N. Edwards. „Big Data Is the Answer... But What Is the Question?“ *Osiris* 32/1 (2017): 328–345.
- Zaagsma, Gerben. „On Digital History.“ *BMGN – Low Countries Historical Review* 128 (2013) 4: 3–29.
- Zalewska Maria. „Selfies from Auschwitz: Rethinking the Relationship Between Spaces of Memory and Places of Commemoration in the Digital Age.“ *Studies in Russian, Eurasian and Central European New Media* 18 (2017): 95–116.

Tobias Hodel

Die Maschine und die Geschichtswissenschaft

Der Einfluss von *deep learning* auf eine Disziplin

Abstract: Deep learning is a method from the field of artificial intelligence that is currently being used in many disciplines to create appraisal decisions. The form of machine learning is also being used in history, for example for text recognition or the identification of named entities. Since deep learning will become a much stronger part of the methodological apparatus in the future, it is worth taking a critical look at what is happening. The moment of training plays a crucial role within the method. There, models are created and optimized. Based on the provided data, patterns can be recognized and imitated. Significantly, the created models are only verifiable in retrospect and with test procedures and are at most partially comprehensible. Thus, hermeneutic approaches are needed to understand and classify the models. Accordingly, the use of deep learning in history will entail a new reflection on methods, which must take into account technical circumstances on the one hand and disciplinary specifications on the other.

Keywords: maschine learning, model training, corpus, text recognition, named entity recognition, text analysis

Zusammenfassung: *Deep learning* ist ein Verfahren aus dem Bereich der künstlichen Intelligenz, das aktuell in vielen Disziplinen zur Erstellung von Bewertungsentscheiden genutzt wird. Auch in der Geschichtswissenschaft wird die Form des maschinellen Lernens bereits genutzt, etwa zur Texterkennung oder der Identifikation von benannten Entitäten. Da *deep learning* zukünftig noch viel stärker Teil des Methodenapparats werden wird, lohnt sich ein kritischer Blick auf die Vorgänge. Das Moment des Trainings spielt in der Methode eine entscheidende Rolle. Dort werden Modelle erstellt und optimiert. Aufgrund der vorgesetzten Daten können Muster erkannt und imitiert werden. Bezeichnenderweise sind die erstellten Modelle nur im Nachhinein und mit Testverfahren überprüfbar und höchstens bedingt nachvollziehbar. Damit braucht es hermeneutische Herangehensweisen, um die Modelle zu verstehen und einzuordnen. Die Nutzung von *deep learning* in der Geschichtswissenschaft wird entsprechend eine neue Methodenreflexion nach sich ziehen, die einerseits technische

Gegebenheiten und andererseits fachimmanente Spezifikationen berücksichtigen muss.

Schlagwörter: Maschinelles Lernen, Modelltraining, Korpus, Texterkennung, Named Entity Recognition, Textanalyse

Die Vision einer künstlichen Intelligenz (englisch *artificial intelligence* oder kurz AI) begleitet die Computerwissenschaft seit ihrem Entstehen. Seit wenigen Jahren fokussieren die Forschungen in diesem Bereich auf *deep learning*. Das Verfahren aus dem Bereich des maschinellen Lernens wird mittlerweile für zahlreiche Bewertungsentscheide eingesetzt, die vor wenigen Jahren noch als ungeeignet für die Bearbeitung durch Algorithmen oder allgemein „den Computer“ beurteilt wurden: Beispiele umfassen die Identifikation von Stimmen, selbstfahrende Autos und die Einschätzung der Rückfallgefahr bei Verurteilungen von Straftäter:innen.

Das Prinzip des *deep learning* ist an sich simpel: Neuronale Netze, dem menschlichen Gehirn nachempfundene vernetzte Speicherzellen, werden mit möglichst vielen Daten versorgt und in einem Trainingsprozess auf typischerweise eine zu lösende Aufgabe getrimmt. Von Spracherkennung über Bildanalyse zu Dokumentenauswertung – *deep learning* setzt sich als Technologie in unterschiedlichen Feldern durch. Sie wird insbesondere seit wenigen Jahren für naturwissenschaftliche Auswertungen benutzt, hält als Technologie entsprechend bereits Einzug in wissenschaftliche Disziplinen. Die Nutzung der Technologie ist aber nicht unproblematisch und in dem Kontext auftretende Probleme stehen im Zentrum dieser Seiten.¹

In den Geisteswissenschaften wird die Technologie aktuell erst in Ansätzen genutzt. Die Texterkennung von Drucken und Handschriften ist einer der Einsatzbereiche, die sich diesen Ansatz zu Nutze macht. Dies ist jedoch erst der Anfang, denn es ist absehbar, dass in naher Zukunft weit mehr (Be-)Wertungsentscheide manuell unterstützt oder gar autonom getroffen werden, die Beschäftigung mit der Methode ist entsprechend wichtig. *Named Entity Recognition*, aber auch visuelle und textuelle Strukturanalysen zeigen gemäß ersten Tests und *proof-of-concepts* bessere Resultate, als dies rein regelgeleitete Algorithmen ver-

¹ Der Aufsatz ist der Versuch einen technologischen Ansatz zu vermitteln und gleichzeitig zu problematisieren. Die angesprochenen Themen werden entsprechend nicht in aller (notwendiger) Tiefe diskutiert, sondern vielmehr angeschnitten. Der Autor dankt für die angeregte Diskussion mit vielen Anknüpfungspunkten an der virtuellen Tagung sowie Christa Schneider und David Schoch für Diskussionen und kritischen Kommentare bei der Erstellung des Textes.

mögen. Mit wenig Phantasie lassen sich gar die Einsatzmöglichkeiten noch erweitern und die Interpretation von Texten mit und dank *machine learning* modellieren.

Im Rahmen dieses Papers werden drei Themenblöcke angeschnitten, die unterschiedliche Anwendungen des maschinellen Lernens im Fokus haben. Erstens, und wohl am unproblematischsten, ist die Nutzung von *deep learning* zur Handschriftenerkennung. Problematischer ist zweitens, die Entitätenerkennung (*Named Entity Recognition*), die kulturwissenschaftliche Fragen zu Praktiken der Namensgebung und zum Individuum im Generellen aufwirft. Drittens kann schließlich mit *machine learning* Ansätzen Strukturerkennung betrieben werden – dies ist eine Vorgehensweise, die in analoger Form etwa aus der Urkundenlehre bereits bekannt ist. Um die Technologie in den Fokus zu stellen, ist es jedoch nötig, dass die drei Ansätze innerhalb des Arbeitens mit neuronalen Netzen verortet und vor allem die Resultate kritisch betrachtet werden. Dieses Paper orientiert sich daher an den drei Perspektiven *Training* von neuronalen Netzen, *Interpretation* von Input und Output, sowie *Konsequenzen* des Einsatzes maschineller Lernverfahren.

1 Trainieren: Die Induktion von *bias*

Training als Basis zur komplexen, statistisch unterstützten Wertung erweist sich als größte Stärke und gleichzeitig neuralgische Stelle der Aufbereitung von Quellenmaterial als Daten, da durch das Trainingsmaterial (Vor-)Urteile übernommen und verstärkt werden. Diese Effekte wurden etwa für Suchmaschinen oder bei Bewerbungsprozessen mehrfach nachgewiesen und problematisiert.² Je nach Form des maschinellen Lernens werden Trainingsmaterialien zum Erlernen, etwa von Annotationen, vorgegeben und überprüft (*supervised learning*) oder die Strukturen werden selbständig erlernt (*unsupervised learning*).

Im Rahmen der Handschriftenerkennung werden Bildausschnitte einem zu erkennenden Text gegenübergestellt. Die Aufgabe des neuronalen Netzes ist es, eine Entsprechung zwischen Anhäufungen von Pixeln und zu erkennenden Zeichen zu finden. Dabei agieren die meisten Systeme unabhängig vom Vorwissen und trainieren jeweils eigenständige Modelle. Damit gibt es keine natürliche

² Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York 2018); Aylin Caliskan, Bryson, Joanna J. und Arvind Narayanan, „Semantics derived automatically from language corpora contain human-like biases,“ *Science* 356 (6334), 183–186, doi: 10.1126/science.aal4230.

Verbindung zwischen Zeichen und Bild, es hängt vielmehr von den Vorgaben ab, die im Rahmen des Trainings gemacht werden. Die Distinktion von Zeichen liegt bei der trainierenden Person und auch im Umgang mit einem Zeichen in den vorhandenen Trainingsdaten. Der menschliche *bias* fließt zwar durch Transkriptionsentscheide in die Automatisierung ein, führt aber zu verhältnismäßig harmlosen Fehlern und Hyperkorrekturen.

Analog dazu verhält es sich bei der Erkennung von benannten Entitäten, einem Problem, das ebenfalls mit *supervised* Ansätzen bearbeitet wird und einen Algorithmus zur Nachahmung verleiten soll. Das System versucht *tokens* („Wörter“ im Satzkontext) als einer Einheit (Person, Ort, Organisation etc.) zugehörig zu bestimmen. Im Vergleich zur Erkennung von handschriftlichem Text basiert bei der *Named Entity Recognition* ein zentraler Schritt auf der Anwendung von Sprachmodellen. Solche Modelle können große Textmengen in hochdimensionalen Vektorräumen verorten und damit Ähnlichkeiten zwischen Wörtern aufzeigen, weil diese entweder häufig im selben Kontext auftauchen, synonym verwendet werden oder aus ähnlichen Zeichenfolgen bestehen. Beim Training der Entitätenerkennung wird einem System also entsprechend vermittelt, inwiefern Wörter im Umfeld eines Vektors zu einer gemeinsamen Gruppe gehören. Zudem sehen wir, wie die Aufbereitung von Trainingsmaterial zur Einspeisung von (Vor-)Urteilen führt, indem etwas als Person oder Ort verstanden wird. Als zusätzliches Problem stehen wir vor der Herausforderung, dass Sprache nicht statisch ist, was die Erzeugung von historischen oder domänenspezifischen Sprachmodellen erforderlich macht. Diese Modelle sind natürlich aufgrund ihrer Basis auch gefärbt (dafür möchte ich den Begriff der „Korpusfärbung“ beliebt machen)³ und bilden Sprache nur entsprechend dem zugrundeliegenden Korpus ab.

Insgesamt bewegen wir uns folglich in einem Bewertungszyklus, der einerseits als Arbeitserleichterung verstanden, aber gleichzeitig zum Verstärkungsmechanismus von (Vor-)Urteilen (*bias*) wird. Der *cycle of bias* (siehe Abbildung 1) entsteht dabei aus arbeitstechnisch sinnvollen und pragmatischen Abläufen, wobei Material aufbereitet (transkribiert/annotiert) und darauf aufbauend *machine learning* basierte Modelle trainiert werden. Auf der Grundlage dieser Modelle wird weiteres Material miteinbezogen. Dadurch erhalten implizite und explizite Vorstellungen eine Verstärkung. Wertungsentscheide wie, „was wird wie transkribiert“ oder „was ist ein Name“ verfestigen sich als Muster. Anhand der

³ Unter Korpusfärbung verstehe ich die thematische und sprachliche Ausrichtung, die durch ein Korpus eingeführt wird und Modelle, die auf maschinellen Lernverfahren basieren, dadurch nicht nur beeinflussen, sondern geradezu definieren. Das Korpus wird dadurch mitverantwortlich für die Bewertungen, die durch ein damit trainiertes Modell erstellt werden.

häufig manuellen Aufbereitung des Materials lassen sich immerhin noch Bevorzugungen identifizieren. Noch problematischer ist indes, wenn bereits vorliegende Modelle beziehungsweise neuronale Netze direkt übernommen und durch sogenanntes *fine-tuning* angepasst werden. Insbesondere bei Sprachmodellen ist diese Vorgehensweise etabliert (siehe dazu weiter unten). Dies führt zur Übernahme von Bewertungsentscheidungen, die häufig nicht direkt und erst aus hochproblematischen Resultaten ersichtlich werden.⁴ Insbesondere in den Bildwissenschaften wurde dieses Problem bereits erkannt, da aktuelle Bilderkennungsmodelle häufig auf einem kleinen Set an bereits bestehenden Modellen beruhen.

TRAININGSKREISLAUF

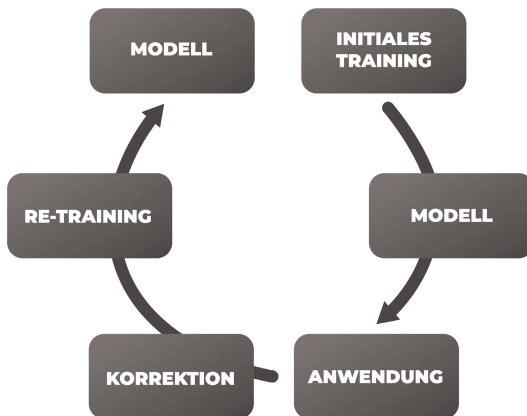


Abb. 1: „Cycle of bias.“ Abbildung von Viviane Blanchard und Tobias Hodel.

Wenn wir noch einen Schritt weitergehen und als Drittes versuchen Texteinheiten (Sätze, Absätze oder Sinneinheiten) einem Thema zuzuweisen oder nach semantischen Gesichtspunkten zu segmentieren, bewegen wir uns sowohl im Bereich des *supervised* als auch des *unsupervised learning*. Neben dem Trainieren

⁴ Siehe dazu aktuelle Eindrücke zum Sprachmodell GPT-3 Tom B. Brown et al., „Language Models are Few-Shot Learners,“ *arXiv:2005.14165 [cs]*, 22. Juli 2020, <http://arxiv.org/abs/2005.14165>, Zugriff am 16.03.2022 von OpenAI. Beispielsweise anhand des GPT-3 basierten Spiels AI Dungeon: Tom Simonite, „It Began as an AI-Fueled Dungeon Game. It Got Much Darker,“ *Wired*, 5. Mai 2021, <https://www.wired.com/story/ai-fueled-dungeon-game-got-much-darker/>, Zugriff am 28.02.2022.

von Annotationen (etwa für *sentiment analysis*), können auch Sätze (selten Satz-teile) aufgrund der vorkommenden Wörter zu Themenfeldern zusammengefügt werden (etwa mit *topic modeling*).⁵ Text wird in den meisten dieser Verfahren als ‚bag of words‘ verstanden, die Reihenfolge der Wörter also ignoriert und nur auf Frequenzen von Zeichenfolgen geachtet. Als zentrale Einheit gilt in diesen Verfahren der Satz, der durch einen Punkt (von einigen Ausnahmen abgesehen) von der nächsten Einheit abgetrennt wird. Entsprechend lässt sich an dieser Stelle ein neuralgischer Faktor identifizieren, da viele vormoderne Sprachen keine Entsprechung zum Satz kennen und die Interpunktion nicht existiert beziehungsweise nicht zwangsläufig Sinneinheiten abtrennt. Das zweite Problem ist die Aufbereitung der zu identifizierenden Teile oder Themen. Wenn wir dies anhand des Beispiels von Urkundenteilen durchdenken, werden je nach Ausgangsmaterial unterschiedliche Teile (etwa Protokoll, Kontext oder Eschatokoll) unterschiedlich stark gewichtet werden. Ein entsprechendes Training mit Übernahme der kanonisierten Wertung ist möglich, führt aber unweigerlich zur Verstärkung impliziter und expliziter Bevorzugungen. Im deutschsprachigen Raum wären solche Bevorzugungen etwa die Prägung der Diplomatik durch die Analyse ausgefertigter Königsurkunden im Gegensatz zu den zahlenmäßig massiv überwiegenden Urkunden, die unter dem Label „Privaturkunden“ zusammengefasst werden.

Die drei kurz skizzierten Themenbereiche stellen unterschiedliche Phasen im Prozess der Quellenaufbereitung dar. Dabei zeigt sich sowohl in relativ simplen Erkenn- oder Identifikationsprozessen als auch in komplexen Zuordnungen das Moment des Trainings als kritischer Vorgang, da die daraus generierten Modelle je nach Korpus (Ausgangsmaterial) in einen Modus des Nachahmens übergehen. Das Verständnis der Modelle ist in der Konsequenz ein hermeneutischer Prozess, der, wie bereits von Gadamer gefordert, eine Auseinandersetzung mit (eigenen) Urteilen und insbesondere Vorurteilen miteinschließt und folglich den Prozess des Trainings nur so nachvollziehbar macht.⁶

⁵ Zu Topic Modeling siehe auch David M. Blei, „Introduction to Probabilistic Topic Models,“ Communication of the ACM, 2011.

⁶ Die Rolle des (Vor-)Urteils wird bei Gadamer aufgedröselt (und weniger stark als Problem aufgefasst): Hans-Georg Gadamer, *Hermeneutik I: Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik*, 7. Aufl., Bd. 1, *Gesammelte Werke* (Tübingen: Mohr Siebeck, 2010), 270–290. (Seitenzahlen nach der 7. Auflage).

2 Interpretieren: Quellenkritik und Hermeneutik

Über den Prozess des Trainings hinaus stellt der Umgang mit Resultaten des *machine learning* insbesondere aus Sicht der *algorithm studies* eine Herausforderung dar. Die Algorithmen lassen sich zwar an unterschiedlichen Stellen zu Ausgaben zwingen (bekannt sind die Google Image-Traum Algorithmen),⁷ jedoch ist ein Nachvollzug der Entscheide innerhalb neuronaler Netze bislang nicht erfolgreich möglich. Die Kritik und die Auswertung der Resultate aus Vorgängen des maschinellen Lernens ähneln entsprechend hermeneutischen Interpretationen, die gerade durch den geschichtswissenschaftlichen Werkzeugapparat wie der Quellenkritik, aber auch andere Methoden analysiert werden müssen.⁸ Erst das wechselseitige *close-* und *distant-reading* der Quellen und der Resultate macht es möglich, die Belastbarkeit der maschinell gewonnenen Wertungen zu überprüfen.

Die Überprüfung der Fehler ist bei der Texterkennung auf den ersten Blick relativ simpel, da mehr oder minder standardisierte Transkriptionskonventionen existieren. Neue Erkennsalgorithmen erreichen dabei, je nach Anzahl der Trainingsseiten, unterschiedliche Resultate. Für Handschriften ist eine Erkennqualität mit Fehlerquoten im Bereich von 2,5 % technisch möglich, pro 1000 erkannten Zeichen muss entsprechend mit 25 Fehlern gerechnet werden. In dieser Fehlerquote ist die fehlerhafte Erkennung von Satzzeichen sowie Groß-/Kleinschreibung bereits enthalten. Bei regelmäßigen Schriften wird diese Fehlerquote durch das Training eines entsprechenden Modells mit ungefähr 50 000 Wörtern erreicht.⁹ Dieses Resultat lässt sich unter optimalen Bedingungen, das heißt unter Beizug von genügend Material von ähnlichen Schriften, auch für Modelle erreichen, die auf unterschiedlichen Händen basieren.¹⁰

7 Alexander Mordvintsev, Christopher Olah und Mike Tyka, „Inceptionism: Going Deeper into Neural Networks“, Google AI Blog, 17. Juni 2015, <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>. Zugriff am 28.02.2022.

8 Für die Literaturwissenschaften siehe als Beispiel: Ted Underwood, „Emerging conversations between literary history and sociology“, The Stone and the Shell, 02.12.2015, <https://tedunderwood.com/2015/12/02/emerging-conversations-between-literary-history-and-sociology/>, Zugriff am 16.03.2022.

9 Für dieses und andere Beispiele siehe Tobias Hodel, „Best-practices zur Erkennung alter Drucke und Handschriften – Die Nutzung von Transkribus large- und small-scale“, in *DHd 2020. Spielräume Digital Humanities zwischen Modellierung und Interpretation*, hg. v. Christof Schöch (dhd2020, Paderborn, 2020), 84–87, doi: 10.5281/zenodo.3666689.

10 Siehe dazu auch Tobias Hodel, David Schoch, Christa Schneider, Jake Purcell, „General Models for Handwritten Text Recognition: Feasibility and State-of-the Art. German Kurrent as an Example“, Journal of Open Humanities Data, 7(13) 2021, 1–10, doi: 10.5334/johd.46.

Es bleibt die Frage offen, inwiefern durch die Quantifizierung von Fehlern Aussagen zur Leistungsfähigkeit eines Erkennmodells gemacht werden können. Zentral sind aus historischer Perspektive schließlich die Fragestellung und die (digitale) Methode, die nach dem Erkennprozess zum Einsatz kommen sollen. Je nachdem fällt auch der Fehlertyp (Satzzeichen sind für *topic modeling* Algorithmen etwa unerheblich) oder die Art eines Fehlers (die Verwechslung von Stab-s mit „f“ führt im *close reading* zu keiner/wenig Verwirrung) ins Gewicht. Zukünftig wird es entsprechend wichtig sein, über quantifizierende Fehlerquoten hinaus, Angaben zur Fehleranfälligkeit eines Modells zu machen.

Der Einsatz von *Named Entity Recognition* verlangt anders gelagerte Diskussionen. Wie bereits oben angesprochen, wird dabei ebenfalls der Trainingsinput imitiert. Auch dies basiert auf Sprachmodellen, sodass die kritische Analyse eines solchen Modells Teil der Methodenkritik wird. Bei der Verwendung historischer Sprachformen entsteht jedoch zusätzlich das Problem, dass Sprachmodelle auf verhältnismäßig kleinen Datenmengen basieren.

Um die Leistungsfähigkeit bestehender Frameworks für nicht-standardisierte vormoderne Sprachen zu demonstrieren, wurde im Rahmen des Editionsprojekts Königsfelden ein Experiment zur Erkennung benannter Entitäten durchgeführt. Dabei wurde ein eigenes Sprachmodell (selbsttrainiert als FLAIR *embeddings*)¹¹ angelegt, das auf zeitlich nahen historischen Dokumenten aus dem 15. und 16. Jahrhundert basiert.¹² Das Training der benannten Entitäten basiert auf 645 Urkunden, für Verhältnisse des maschinellen Lernens also insgesamt eher wenig Material. Eine Besonderheit bildet das Tagging des Editionsprojekts, das die Strategie verfolgt, alle potentiell zugehörigen Informationen einem Namen zuzurechnen. Dadurch wurden auch Angaben, die heute nicht mehr als Namensteil verstanden würden, als solcher markiert und folglich auch fürs Training verwendet. Trotz des geringen Umfangs der Trainingsdaten konnten F-Scores im Bereich von 69–74 % erreicht werden.¹³

11 FLAIR ist ein open-source Framework für Natural Language Processing: <https://github.com/flairNLP/flair>. Alan Akbik et al., „FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP,“ in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)* (Minneapolis, Minnesota: Association for Computational Linguistics, 2019), 54–59, doi: 10.18653/v1/N19-4010.

12 Verwendet wurde das Bonner Frühneuhochdeutsch Korpus (<http://www.korpora.org/Fnh/dc/>), digital vorliegende Bände der Schweizerischen Rechtsquellen (<https://www.ssrq-sds-fds.ch/home/>) und Urkunden und Akten des Klosters Königsfelden (<https://www.hist.uzh.ch/de/fachbereiche/mittelalter/lehrstuehle/teuscher/forschung/projekte/koenigsfelden.html>, Zugriff am 16.03.2022).

13 F-Scores kombinieren Recall (Ausbeute) und Precision (Präzision) und sind ein häufig genutztes Mittel, um Klassifikatoren zu beurteilen.

Hanns	B-PER	B-PER	unser	B-PER	O	Erkenvid	O	B-PER
Zender	I-PER	I-PER	eidgnosschaft	I-PER	O	und	O	O
ouch	O	I-PER	vogt	I-PER	O	Hartman	O	B-PER
von	O	I-PER	zu	I-PER	O	ritt'	O	I-PER
Birmistorff	B-LOC	I-PER	Baden	I-PER	B-LOC	und	O	O
des	O	I-PER	des	I-PER	O	Peter	O	B-PER
gerichtz	O	I-PER	frommen	I-PER	B-PER	gebrud'	O	I-PER
daselbs	O	I-PER	vesten	I-PER	I-PER	truchsezen	O	I-PER
			Hans	I-PER	I-PER	*	O	I-PER
			Meisen	I-PER	I-PER	von	O	I-PER
			des	I-PER	I-PER	Habspg	O	I-PER
			räts	I-PER	I-PER			
			Zürich	I-PER	I-PER			

Abb. 2: Drei Beispiele für „Fehler“ im Tagging. Da die Definition einer benannten Entität sehr weit gefasst wurde, ist auch die maschinelle rechte Spalte sinnvoll. Im rechten Beispiel zeigt sich ein Annotationsfehler. Die Auswertung erfolgte durch Ismail Prada.

Dieser Wert ist für moderne Sprachen zwar nicht besonders hoch, zeigt aber das Potential des Ansatzes. Auch für dieses Verfahren lohnt sich ein Blick auf einzelne Resultate. Dadurch lässt sich eine Vielzahl von „Fehlern“ sichtbar machen, die korrekte Resultate widerspiegeln. Die „Fehler“ stammen in diesen Fällen von Annotator:innen, die inkorrekt auszeichneten oder aber von der Maschine selbst, die gar valable alternative Annotationen (Namen können teilweise Orts- oder Personennamen bezeichnen) liefert.¹⁴ Einschränkung muss erwähnt werden, dass die Transkription händisch erstellt und Eigennamen im Gegensatz zum restlichen Text großgeschrieben wurden. Die Algorithmen hatten entsprechend starke Indizien zur Identifikation von Entitäten.

Stärker noch als bei der Texterkennung zeigt sich für Annotationsaufgaben, wie sehr die unterschiedlichen Inputs (Sprachmodell, Transkriptionsvorgaben und Trainingsmaterial) das Resultat beeinflussen. Eine Analyse der Technologie und der Resultate muss die Komplexität und Unsicherheitsfaktoren beispielsweise von Annotationsaufgaben mitberücksichtigen, wobei aktuell der benötigte Werkzeugkasten dazu noch mehrheitlich fehlt und wiederum quantitative Angaben nur beschränkt Aussagen zur Fähigkeit eines Netzes erlauben.

¹⁴ Für diesen Versuch wurde nicht berücksichtigt, dass *pooling* von Wortvektoren (eigtl. eine Kontextualisierung innerhalb von Sätzen, indem auf vorangehende Wortvektoren zurückgegriffen wird) eine höhere Leistungsfähigkeit aufweisen. Alan Akbik, Tanja Bergmann und Roland Vollgraf, „Pooled Contextualized Embeddings for Named Entity Recognition,“ in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (NAACL-HLT 2019, Minneapolis, Minnesota: Association for Computational Linguistics, 2019), 724–28, doi: 10.18653/v1/N19-1078.

Bei der Analyse von Resultaten, die aus textbezogenen Annotationsaufgaben entstanden sind, stellt der Einbezug von Sprachmodellen eine weitere Hürde dar. Anhand von Vergleichen mit zur Verfügung gestellten Testdaten lässt sich aufzeigen, dass *language models*, die eine gewisse Größe aufweisen und mit den zu analysierenden Texten eng verwandt sind, bessere Resultate liefern. Es hängt bei dieser Aufgabe folglich nicht nur von der Menge an annotierten Daten ab, sondern auch vom Sprachmodell, welches als Ausgang gewählt wird.

Wie bereits oben angesprochen, wird aus pragmatischen Gründen und teilweise auch aus Gründen der Performanz auf vortrainierte Netze zurückgegriffen. Für moderne und von großen Firmen als zentral erachtete Sprachen (Englisch, Französisch, Deutsch etc.) ist dies etwa BERT¹⁵ beziehungsweise darauf aufbauende Varianten wie CamemBERT.¹⁶ Auch die bereits in den Fußnoten erwähnten Modelle *GPT-2* und *GPT-3* fallen in die Kategorie der vortrainierten Netze. Allen Sprachmodellen ist gemein, dass sie jeweils zwar eine erhöhte Leistungsfähigkeit in der Problemlösung (typischerweise ausfüllen von Lückentexten) und der Annotation aufweisen, jedoch auf (Trainings-)Materialien zurückgreifen, die entweder überhaupt nicht publiziert oder insofern nicht nachvollziehbar sind, als dass die Daten zwar zugänglich aber nicht mit Metadaten angereichert sind. Von Datenpublikation unter FAIR-Kriterien kann in keinem der Fälle gesprochen werden.¹⁷ Die Konsequenz ist, dass wir weder die Grundlagen beurteilen noch das Funktionieren der Modelle nachvollziehen können.¹⁸

In einem weit experimentelleren Stadium als die Identifikation von Entitäten befindet sich die Zuordnung von Annotationen, die Sinneinheiten klassifizieren. Bereits etwas etabliert, vor allem da kommerziell interessant, ist dabei die *sentiment analysis*, die Sätzen meistens positiven oder negativen Gefühlsausdrücken zuordnet. In der deutschsprachigen Digital Humanities Community

15 Jacob Devlin et al., „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,“ arXiv:1810.04805 [cs], 24. Mai 2019, <http://arxiv.org/abs/1810.04805>, Zugriff am 16.03.2022.

16 Louis Martin et al., „CamemBERT: A Tasty French Language Model,“ 9. Oktober 2019, HAL Id: hal-02445946.

17 Mark D. Wilkinson et al., „The FAIR Guiding Principles for Scientific Data Management and Stewardship,“ *Scientific Data* 3, Nr. 1 (Dezember 2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.

18 Ansätze in diese Richtung zum Verständnis von Sprachmodellen über Visualisierung werden aktuell getestet, siehe bspw. hier Challenge 4, https://www.cnd.philnat.unibe.ch/ueberuns/aktivitaeten/nlp_hackathon/ und das damit verbundene Poster: https://www.dh.unibe.ch/unibe/portal/fak_historisch/fsuf/d_dh/content/e330319/e336052/e1074527/PosterBDSGer.pdf, Zugriff am 16.03.2022.

werden entsprechende Ansätze bereits intensiv bearbeitet und auch für ältere Sprachstufen vorbereitet.¹⁹

Analog dazu können auch andere, etwa thematische Labels vergeben und trainiert werden. Die bereits oben beschriebenen Probleme werden dabei übernommen und die Komplexität nochmals um eine Stufe gesteigert, da die Sprachmodelle auf der Ebene „Satz“ angewandt werden. Das Verfahren wird dabei vom Wort (eigentlich *token*) auf eine Zeichenkette erweitert, die durch vordefinierte Stoppzeichen (Komma und [Doppel-]Punkte) abgetrennt werden. Das bedeutet, dass aus den hochdimensionalen Vektoren von Wörtern ein Vektor pro Satz errechnet wird, der gar auf ganze Texte erweitert werden kann. Dadurch lassen sich Textähnlichkeiten mathematisch über Cluster, also die Zusammenführung von nahen Vektoren, aufzeigen. Obwohl solche Verfahren bereits seit einigen Jahren verfügbar und auch die informatischen Anforderungen an die Infrastruktur nicht unermesslich sind, sind es tendenziell literarische Texte, die diesbezüglich mit solchen Verfahren behandelt wurden.²⁰ Mit spezifischen Datensets, die etwa historische Textgattungen abbilden und entsprechend das Erstellen von Sprachmodellen erleichtern, ist indes die Erweiterung der Einsatzgebiete in Sicht.

Da typischerweise auf den Satz als einfach zu segmentierende Einheit Bezug genommen wird, entstehen etwa für vormoderne Texte oder wenig gepflegte textuelle Formen (Stichwort: Kurznachrichten oder Social Media Posts) Her-

19 Siehe Thomas Schmidt, Manuel Burghardt und Katrin Dennerlein, „Kann man denn auch nicht lachend sehr ernsthaft sein?“ – Zum Einsatz von Sentiment Analyse-Verfahren für die quantitative Untersuchung von Lessings Dramen“ (Vortrag auf der DHd 2018 Kritik der digitalen Vernunft. 5. Tagung des Verbands „Digital Humanities im deutschsprachigen Raum“ (DHd 2018), Köln, Zenodo, 26. Februar 2018), doi: 10.5281/zenodo.4622557; David Wodasch et al., „Hinterlistig – schelmisch – treulos – Sentiment Analyse in Texten des 19. Jahrhunderts: Eine exemplarische Analyse für Länder und Ethnien“ (Vortrag auf der DHd 2018 Kritik der digitalen Vernunft. 5. Tagung des Verbands „Digital Humanities im deutschsprachigen Raum“ (DHd 2018), Köln, Zenodo, 26. Februar 2018), doi: 10.5281/zenodo.4622483. Siehe dazu auch die Resource S. Clematide und M. Klenner, „Evaluation and Extension of a Polarity Lexicon for German,“ in *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, hg. v. A. Montoyo et al. (Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA); Held in conjunction to ECAI 2010 Portugal, Lisbon, Portugal: University of Zurich, 2010), 7–13, doi: 10.5167/uzh-45506.

20 Ted Underwood, „The Literary Uses of High-Dimensional Space,“ *Big Data & Society* 2, Nr. 2 (1. Dezember 2015): 2053951715602494, doi: 10.1177/2053951715602494. Siehe auch einführend Ted Underwood und Matthew L. Jockers, „Text-Mining the Humanities,“ in *A New Companion to Digital Humanities*, hg. v. Susan Schreibman, Ray Siemens und John Unsworth (John Wiley & Sons, 2016), 291–306.

ausforderungen. Gerade für die angesprochenen Urkunden ist der Satz keine sinnvolle Einheit, um Zuordnungen zu erstellen.

Mit Blick auf diesen dritten Themenbereich (Textteile einem Thema zuordnen), stehen wir heute in einer initialen Findungsphase. Erste Modelle führen zu vielversprechenden Eindrücken, liefern jedoch noch zu wenig belastbare Resultate. Auch die Anwendung von *topic modeling* auf einzelne Sätze ist möglich, führt aber zu einem Clustering von ähnlichen Wortkonstruktionen und mahnt an die Auswertung von Kookkurrenzen. Aufschlüsse zu semantischen oder gar thematischen Feldern werden damit nur mittelbar gegeben.²¹

3 Konsequenzen: Von einer neuen Heuristik zu einer neuen Epistemologie?

Das Oszillieren zwischen praktischen Umsetzungen und theoretischen Überlegungen führt zu neuen Problemstellungen, die Epistemologie und heuristische Methoden der Geschichtswissenschaften betreffen. Maschinelles Lernen zeigt sich dabei bereits heute als nützliche Erweiterung der Disziplin an der Schwelle des Einsatzes von *big data*, die es kritisch zu betrachten und zu verfolgen gilt. Die Einsichten dienen dabei nicht nur der intradisziplinären Methodendiskussion, sondern führen darüber hinaus zu kritischen Positionen für den Einsatz von *deep learning* im alltäglichen Leben.

Die Nutzung von *deep learning* in einer hochgradig reflexiven Wissenschaft wie der Geschichtswissenschaft, bedeutet die Explizierung erkenntnistheoretischer Grundannahmen. Was etwa als „Text“ verstanden wird, muss offengelegt sein, wenn ein Algorithmus zur Erkennung von „Text“ gebracht wird.²² Dabei regen auch einzelne Vorstufen von Texten die Diskussion an, etwa wenn identifiziert werden muss, wo sich auf einem Artefakt Text befindet. Die Identifikation von Personen oder Orten in textuellen Strukturen greift ebenso auf Vorannah-

²¹ Siehe dazu auch Tobias Hodel, „Supervised and Unsupervised: Approaches to Machine Learning for Textual Entities,“ in *Archives, Access and AI: Working with Born-Digital and Digitised Archival Collections*, hg. v. Lise Jaillant (Transcript, 2021), 162–168.

²² Tobias Hodel, „Konsequenzen automatischer Texterkennung – Ein Aufriss zur Texterkennung mit Machine Learning,“ in *DHd 2018. Kritik der digitalen Vernunft* Konferenzabstracts. Universität zu Köln 26. Februar bis 2. März 2018, hg. v. Georg Vogeler (Köln, 2018), 249–251, <http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf>, Zugriff am 16.03.2022; Patrick Sahle, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 3: Textbegriffe und Recodierung.*, Bd. 3, 3 Bde., (Norderstedt: BoD, 2013), <http://kups.ub.uni-koeln.de/5352/>. Zugriff am 28.02.2022.

men zurück, indem Fragen nach bedeutungstragenden Namen gegenüber von Zuschreibungen abgewogen werden müssen. Zentral wird dabei die Dokumentation der Aufbereitung von Grundlagen, die Bewertungsentscheide nachvollziehbar macht. Indirekt lassen sich darauf aufbauend die Entscheide eines Modells nachvollziehen.

Gleichermaßen ist das Teilen der zugrunde liegenden Daten zentral, die eine Analyse der Annotations- und Bewertungsentscheide erlauben und somit das Problem des *bias* aktiv angehen. Die bereits existierenden und von diversen Forschungsförderanstalten eingeforderten FAIR-Richtlinien geben nicht nur diese Richtung vor, sondern lassen im akademischen Kontext fast keinen anderen Weg mehr zu. Anders sieht es bei der Nutzung kommerzieller Produkte oder Modelle aus, auf die weit weniger Zugriff besteht, die aber aufgrund ihrer Leistungsfähigkeit nicht komplett missachtet werden können. Der Umgang mit solchen Modellen muss im Kontext der *ethical AI* diskutiert und mit entsprechenden Ansätzen angegangen werden, etwa durch die Zertifizierung (*auditing*) von Algorithmen.²³

Die Anwendung von maschinellen Lernverfahren erfordert somit nicht eine komplett neue historische Methode, sondern eine Erweiterung des technischen Horizonts, indem zumindest im Grundsatz die Verfahren verstanden werden müssen. Überdies ist eine konsequente Erweiterung der Hermeneutik auf eingesetzte Methoden notwendig, da nicht mehr nur das erforschte Material, sondern auch die technischen Herangehensweisen nie vollständig überblickt und auch nur in (langsamer) Annäherung verstanden werden können.

Der *machine learning turn* führt nicht zu einer Abkehr von der historischen Methode, sondern vielmehr zu einer neuen Art der Beschäftigung mit Quellen, die nicht nur den Aussagewert beurteilt, sondern gleichzeitig auch die (automatisierte) Beschäftigung damit berücksichtigt.

Bibliographie

Akbik, Alan, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter und Roland Vollgraf. „FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP.“ In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, 54–59. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-4010.

²³ Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code*, 1. (Medford, MA: Polity, 2019).

- Akbik, Alan, Tanja Bergmann und Roland Vollgraf. „Pooled Contextualized Embeddings for Named Entity Recognition.“ In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 724–28. Minneapolis, Minnesota: Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-1078.
- Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. 1. Medford, MA: Polity, 2019.
- Blei, David M. „Introduction to Probabilistic Topic Models.“ *Communication of the ACM*, 2011.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, u. a. „Language Models are Few-Shot Learners.“ *arXiv:2005.14165 [cs]*, 22. Juli 2020. <http://arxiv.org/abs/2005.14165>, Zugriff am 16.03.2022.
- Caliskan, Aylin, Joanna J. Bryson und Arvind Narayanan. „Semantics Derived Automatically from Language Corpora Contain Human-like Biases.“ *Science* 356, Nr. 6334 (14. April 2017): 183–86. doi: 10.1126/science.aal4230.
- Clematide, S. und M. Klenner. „Evaluation and Extension of a Polarity Lexicon for German.“ In *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, hg. v. A. Montoyo, P. Martínez-Barco, A. Balahur und E. Boldrini, 7–13. Lisbon, Portugal: University of Zurich, 2010. doi: 10.5167/uzh-45506.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee und Kristina Toutanova. „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.“ *arXiv:1810.04805 [cs]*, 24. Mai 2019. <http://arxiv.org/abs/1810.04805>, Zugriff am 16.03.2022.
- Gadamer, Hans-Georg. *Hermeneutik I: Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik*. 7. Aufl. Bd. 1. Gesammelte Werke. Tübingen: Mohr Siebeck, 2010.
- Hodel, Tobias. „Konsequenzen automatischer Texterkennung – Ein Aufriss zur Texterkennung mit Machine Learning.“ In *DHd 2018. Kritik der digitalen Vernunft Konferenzabstracts. Universität zu Köln 26. Februar bis 2. März 2018*, hg. v. Georg Vogeler, 249–51. Köln, 2018. <http://dhd2018.uni-koeln.de/wp-content/uploads/boa-DHd2018-web-ISBN.pdf>. Zugriff am 16.03.2022.
- Hodel, Tobias. „Best-practices zur Erkennung alter Drucke und Handschriften – Die Nutzung von Transkribus large- und small-scale.“ In *DHd 2020. Spielräume Digital Humanities zwischen Modellierung und Interpretation*, hg. v. Christof Schöch, 84–87. Paderborn, 2020. doi: 10.5281/zenodo.3666689.
- Hodel, Tobias, David Schoch, Christa Schneider und Jake Purcell. „General Models for Handwritten Text Recognition: Feasibility and State-of-the Art. German Kurrent as an Example,“ in *Journal of Open Humanities Data* 7 (2021). doi: 10.5334/johd.46.
- Hodel, Tobias. „Supervised and Unsupervised: Approaches to Machine Learning for Textual Entities,“ in *Archives, Access and AI: Working with Born-Digital and Digitised Archival Collections*, hg. v. Lise Jaillant. Bielefeld: Transcript, 2021, 157–177.
- Martin, Louis, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric Villemonte de la Clergerie, Djamé Seddah und Benoît Sagot. „CamembERT: A Tasty French Language Model,“ 9. Oktober 2019. Hal Id: 02445946.
- Mordvintsev, Alexander, Christopher Olah und Mike Tyka. „Inceptionism: Going Deeper into Neural Networks.“ *Google AI Blog*, 17. Juni 2015. <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>, Zugriff am 16.03.2022.
- Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press, 2018.

- Sahle, Patrick. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 3: Textbegriffe und Recodierung*. Bd. 3. 3 Bde. Norderstedt: BoD, 2013. <http://kups.ub.uni-koeln.de/5352/>, Zugriff am 28.02.2022.
- Schmidt, Thomas, Manuel Burghardt und Katrin Dennerlein. „Kann man denn auch nicht lachend sehr ernsthaft sein?“ – Zum Einsatz von Sentiment Analyse-Verfahren für die quantitative Untersuchung von Lessings Dramen.“ Vortrag auf der DHd 2018 Kritik der digitalen Vernunft. 5. Tagung des „Verbands Digital Humanities im deutschsprachigen Raum“ (DHd 2018), Köln, 26. Februar 2018. doi: 10.5281/zenodo.4622557.
- Simonite, Tom. „It Began as an AI-Fueled Dungeon Game. It Got Much Darker.“ *Wired*, 5. Mai 2021. <https://www.wired.com/story/ai-fueled-dungeon-game-got-much-darker/>, Zugriff am 28.02.2022.
- Underwood, Ted. „Emerging Conversations between Literary History and Sociology.“ *The Stone and the Shell* (blog), 2. Dezember 2015. <https://tedunderwood.com/2015/12/02/emerging-conversations-between-literary-history-and-sociology/>, Zugriff am 16.03.2022.
- Underwood, Ted. „The Literary Uses of High-Dimensional Space.“ *Big Data & Society* 2, Nr. 2 (1. Dezember 2015): 2053951715602494. doi: 10.1177/2053951715602494.
- Underwood, Ted und Matthew L. Jockers. „Text-Mining the Humanities.“ In *A New Companion to Digital Humanities*, hg. v. Susan Schreibman, Ray Siemens und John Unsworth, 291–306. John Wiley & Sons, 2016.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, u. a. „The FAIR Guiding Principles for Scientific Data Management and Stewardship.“ *Scientific Data* 3, Nr. 1 (Dezember 2016): 160018. doi: 10.1038/sdata.2016.18.
- Wodasch, David, Maik Fiedler, Ben Heuwing und Thomas Mandl. „Hinterlistig – schelmisch – treulos – Sentiment Analyse in Texten des 19. Jahrhunderts: Eine exemplarische Analyse für Länder und Ethnien.“ Vortrag auf der DHd 2018 Kritik der digitalen Vernunft. 5. Tagung des Verbands „Digital Humanities im deutschsprachigen Raum“ (DHd 2018), Köln, 26. Februar 2018. doi: 10.5281/zenodo.4622483.

Neue Methoden der Digital History

Katrin Moeller

Die Modellierung des zeitlichen Vergleichs als Kernkompetenz von Digital History?

Fachliche Voraussetzungen und digitale Konzepte

Abstract: An overview of research into time perceptions, time references and the handling of longitudinal and cross-sectional data compares various procedures and methods of historical research. It asks about similarities and differences between classical and digital research topics and time-based methods and how both disciplines can meaningfully complement or enrich each other. In doing so, the hypothesis is put forward that digital methods can not only provide important impulses for longitudinal analyses across disciplines through mass data analyses, but that historical research can usefully complement the „fracture category“ of time („qualitative research“) by looking for similarities. In doing so, the author emphasises the necessity of common standards for methods and digital working procedures in order to actually generate secure foundations for scientific work.

Keywords: time perception, ages, employment research, entity recognition, methods, longitudinal analysis, digital history

Zusammenfassung: Ein Überblick zur Erforschung von Zeitwahrnehmungen, Zeitbezügen und des Umgangs mit längs- und querschnittsbezogenen Daten vergleicht verschiedene Verfahren und Methoden der historischen Forschung. Gefragt wird nach Gemeinsamkeiten und Unterschieden von klassischen und digitalen Forschungsthemen und zeitbasierten Methoden und wie sich beide Fachrichtungen sinnvoll ergänzen oder bereichern können. Dabei wird die These aufgestellt, dass digitale Methoden durch Massendatenanalysen nicht nur fachübergreifend wichtige Impulse zu Längsschnittanalysen setzen können, sondern geschichtswissenschaftliche Forschung die „Bruchkategorie“ der Zeit („qualitative Forschung“) durch Sichtungen nach Gemeinsamkeiten sinnvoll ergänzen. Dabei betont die Autorin die Notwendigkeit von gemeinsamen Standards für Methoden und digitalen Arbeitsverfahren, um tatsächlich gesicherte Grundlagen wissenschaftlichen Arbeitens zu erzeugen.

Schlagwörter: Zeitwahrnehmung, Alter, Berufsforschung, Entitätserkennung, Methoden, Längsschnittanalyse, Digital History

Zeit als Forschungsparadigma ist der Geschichtswissenschaft als ureigener innerer Kern und Gegenstand eingeschrieben: „Man kann nicht sagen, daß etwas ist, ohne zu sagen, was es ist. Indem man sie denkt, bezieht man Fakta schon auf Begriffe, und es ist doch wohl nicht einerlei, auf welche,“ zitiert Reinhart Koselleck in seiner Auseinandersetzung mit methodischen Fragen der Quellenkritik und Historizität bereits Friedrich Schlegel.¹ Der Wandel der Begrifflichkeiten, ihre Veränderlichkeit in der Zeit, ihre rezeptiven Nachwirkungen sowie die Ausgestaltung von Transformationsprozessen bilden eine der Hauptachsen historischer Forschung. Sogar die Kulturgeschichte der Zeitlichkeit selbst,² die Entstehung von Zeitbewusstsein, das Nebeneinander verschiedener Zeitordnungen, die Wahrnehmung und Strukturierung des Alltags und der Ökonomie durch Zeitregime standen und stehen seit einigen Jahren wieder verstärkt im Mittelpunkt einer umfassenden Analyse und von Diskussionen.³ Dabei repräsentiert die Zeit paradigmatisch die Veränderlichkeit von Begrifflichkeiten, den Wandel und den Neubeginn. Zeit ist damit eine Dimension der Transformation und des Bruchs, könnte man zahlreiche Forschungsansätze zusammenfassen. Zeit ist zudem kein gleichförmiger Fluss, sondern sie wird in der Moderne oder Postmoderne zunehmend als „Beschleunigung“ erlebt oder auch in Form von Zeitkonkurrenzen, wird als Abweichung von zeitlicher Ordnung im Sinne der Naturwissenschaft markiert.⁴ Diese Beschreibungen verweisen auf die verschiedenen Dimensionen von Zeit, die als subjektiv erlebte Prozesse auf das Individuum gerichtete Analysen beschreiben.⁵ Dabei scheinen auch das Wiederkeh-

1 Reinhart Koselleck, *Vergangene Zukunft: Zur Semantik geschichtlicher Zeiten* (Frankfurt a. M.: Suhrkamp, 2010), 349.

2 Albrecht Franz und Katja Patzel-Mattern, „Annäherungen an einen kulturwissenschaftlichen Zeitbegriff,“ in *Der Faktor Zeit: Perspektiven kulturwissenschaftlicher Zeitforschung*, hg. v. Katja Patzel-Mattern und Albrecht Franz (Stuttgart: Franz Steiner Verlag, 2015), 10.

3 Arndt Brendecke (Hg.), *Die Autorität der Zeit in der Frühen Neuzeit* (Berlin: LIT-Verl., 2007); Erhard Chvojka, Andreas Schwarcz und Klaus von Thienen (Hg.), *Zeit und Geschichte: Kulturgeschichtliche Perspektiven* (Wien, München: R. Oldenbourg Verlag, 2002); Ilse E. Plattner, *Zeitbewußtsein und Lebensgeschichte: Theoretische und methodische Überlegungen zur Erfassung des Zeitbewußtseins* (Heidelberg: Asanger, 1990); Elmar Schenkel und Kati Voigt (Hg.), *Verweile doch ...: Über die Erforschung der Zeit* (Leipzig: Edition Hamouda, Oktober 2015); Achim Landwehr (Hg.), *Frühe Neue Zeiten: Zeitwissen zwischen Reformation und Revolution* (Bielefeld: transcript, 2012); Gerald James Whitrow, *Time in History: Views of Time from Prehistory to the Present Day* (Oxford: Oxford Univ. Press, 1991).

4 Albrecht Franz und Katja Patzel-Mattern, „Annäherungen“, 11.

5 Plattner, *Zeitbewußtsein und Lebensgeschichte*.

rende und die rhythmische Ausgestaltung der Zeit zwar immer wieder auf,⁶ bleiben in vielen Darstellungen jedoch weitgehend im Hintergrund. Mit den Mitteln der hermeneutischen Analyse, des qualitativen Vergleichs und der historischen Quellenkritik – so kann man als Zwischenfazit festhalten – gelingt es uns hervorragend, zeitlichen Wandel inhaltlich mehrdimensional und quellenkritisch zu fassen, wenn die Zeitdimension damit einerseits den Bruch und die Abweichung vom Alltäglichen oder „Normalen“, andererseits vor allem die subjektiven und kulturell wandelbaren Wahrnehmungen neben der objektiv messbaren Zeit beschreibt.

Zeit ist damit notwendigerweise auch für die Digital Humanities und die Digital History ein zentrales Thema, vor allem, wenn sie wie häufig angemahnt, nicht einem bloßen Selbstzweck dienen soll, sondern produktive Ziele der historischen Erkenntnismehrung ihrer eigenen geisteswissenschaftlichen Herkunftsdisziplin verfolgt.⁷ Dennoch scheint gerade am Beispiel der Zeit kaum ein direkter Anknüpfungspunkt zwischen der aktuellen Diskussion der oben erwähnten und auf die „Zeit“ bezogenen historischen Debatte und den Digital Humanities zu bestehen. Dies zeigt bereits ein nur beispielhafter und damit unvollständiger Blick auf wichtige Projekte der Digital History, die sich mit dem zeitlichen Wandel beschäftigen.

Ein frühes, visuell ansprechendes, hochrangig in der Zeitschrift „Science“ publiziertes und in der Öffentlichkeit sehr positiv rezensiertes Projekt – und damit für viele bis heute impulsgebend – kartierte unter dem Titel „A network framework of cultural history“ Merkmale von Kultur in Raum und Zeit.⁸ Lebenszeiten wurden hier für das Entstehen und Vergehen von kulturellen Epochen sichtbargemacht. Entgegen der staunenden Öffentlichkeit kritisierten Historiker:innen und Expert:innen der Digital Humanities das Projekt jedoch grundlegend, weil es basale Prinzipien der Quellenkritik, vor allem jedoch der adäquaten fachwissenschaftlichen Operationalisierung von Fragestellungen, Methoden, Quellenanalyse und der damit gewonnenen Erkenntnisse verletzte, darunter fundamentale Prinzipien der Repräsentativität und Plausibilität von Daten. Als Fazit formulierte Rehbein sehr grundlegend: „dass wer die Entstehung und Verarbeitung von Daten nicht nachvollziehen kann, diese nicht in-

6 Alexander Deeg, „Advent ist immer,“ in *Verweile doch ...: Über die Erforschung der Zeit*, hg. v. Elmar Schenkel und Kati Voigt (Leipzig: Edition Hamouda, Oktober 2015), 99–106; Birgit Jochens, *Deutsche Weihnacht: Ein Familienalbum 1900–1945* (Berlin: Nicolai, 2011).

7 Malte Rehbein, „Digitalisierung braucht Historiker/innen, die sie beherrschen, nicht beherrscht,“ <https://www.hsozkult.de/debate/id/diskussionen-2905>. Zugriff am 19.02.2022.

8 Maximilian Schich et al., „Quantitative Social Science. A Network Framework of Cultural History,“ *Science* 345/6196 (2014): 558–562.

nerhalb von historischer Forschung verwenden *darf*.“⁹ Diese Kritik umfasst ein Grundproblem der Digital History, welche mit jedem einzelnen Projekt eine Vielzahl von fachspezifischen, technischen, informatischen und statistischen methodischen Kompetenzen bedürfte, die sich unterhalb von Leuchtturmprojekten oft aber gar nicht in dieser Interdisziplinarität bündeln lassen. Gleichzeitig braucht es im jungen Fach eben auch noch erst viele basale Voraussetzungen, um überhaupt an die komplexen Fragestellungen des eigenen Faches anknüpfen zu können. Gleiches gilt auch für andere beteiligte Fachdisziplinen, denn so ist bereits das Innenverhältnis von Statistik und Informatik häufig nicht konsolidiert. Viele Analysen, die Software und Programme heute ermöglichen, entsprechen nicht unbedingt den Standards oder den ausgewiesenen Anwendungsbereichen konsolidierter statistischer Verfahren. Vergleichbar ist dies vielleicht mit der Verwendung von Medikamenten zur Heilung von Krankheiten, die aber eben nicht geprüft und zugelassen sind. Es mag sein, dass solche Verwendungen zu einem guten Ergebnis führen. Der umgekehrte Fall ist jedoch ebenso möglich. Es braucht daher einen intensivierten Verständigungsprozess nicht nur über Formen der digitalen Quellenkritik, sondern auch zu verlässlich anzuwendenden Methoden und Verfahren.

Diese Problematik erweist sich auch für die Projekte noch als relevant, die nun einige Jahre später den Faktor „Zeit“ operationalisierbar machen möchten.¹⁰ Zentrales Anliegen ist dabei vorwiegend die Modellierung von historischen Daten, um sie einerseits gemäß von informationellen Anforderungen formal adäquat, aber eben auch entsprechend der problembehafteten Quellen- und Datensituation passfähig erheben zu können. Fundamental treten dabei die Herausforderungen auf der technischen und informatischen Ebene zur „Modellierung des Zweifels“¹¹ hervor, die in direkter Liaison zur Ungenauigkeit und Relativität vieler Quelleninformationen stehen, die sich ja beileibe nicht nur auf Zeitdimensionen beziehen. Dabei konzentrieren sich viele Analysen – selbst bei einer Orientierung auf fachwissenschaftliche Voraussetzungen – dennoch vielmehr auf eine modellierungstechnische und informatisch möglichst präzise Be-

9 Malte Rehbein, „Digitalisierung“.

10 Vgl. Programm und Berichte zur Tagung: Data for History 2021: Modelling Time, Places, Agents, Berlin 2021, <https://d4h2020.sciencesconf.org/>, die bisher noch unpubliziert sind. Zugriff am 19.02.2022.

11 Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz, „Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Zur Einführung in diesen Band,“ in *Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten*, hg. v. Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz (Wolfenbüttel: Herzog August Bibliothek, 2019) 2019, doi: 10.17175/sb004_013.

schreibung und Sichtbarmachung von Unsicherheit, die vor allem an Lösungen zur Informationsweiterverarbeitung und des Informationsaustausches anknüpfen.¹² Dabei geht es vor allem um die semantische Verknüpfung (Linked Open Data, Ontologien) und Vernetzung von Wissenskonzepten in webbasierten Datenbanken und Angeboten, die heute auch als Wissensgraphen diskutiert werden. Im Ergebnis wird unter anderem moniert, dass die exakte Darstellung und Modellierung auch Erkenntnisprozesse und epistemologische Voraussetzungen der Analysen verändern, wie etwa Martin Raspe und Georg Schelbert auch anhand von aktuellen Modellierungs- und Visualisierungsprozessen von kulturhistorischen Objekten darlegen konnten und wie sie auch im Rahmen der ersten Digital History-Tagung ausgiebig problematisiert wurden. Weil solche Visualisierungen oder notwendigen Modellierungsvorgänge fehlende Informationen in analysierbare Details verwandeln, müssen wir eben methodische Ansätze diskutieren und zu Standards entwickeln, damit sie dann auch in den komplexen Datenstrukturen und im Gestrüpp des Internets noch sichtbar werden.¹³

Was sicherlich für die Weiterentwicklung von Modellierungstechniken, Recherche- und Kontextualisierungsoptionen von webbasierten Quellen innovative Impulse setzt und fraglos eine wichtige Rolle erfüllt, bleibt für die fragestellungsgeleitete aktuelle Forschung jedoch häufig noch zu rudimentär oder vielleicht besser gesagt, ein zu sehr auf seine technischen Dimensionen verengtes Problem, dass aber eben doch viele inhaltliche Herausforderungen adressiert. Allerdings werden mit der bisherigen Diskussion vor allem Dimensionen von fehlenden oder unsicheren Werten und grundsätzlich die entitätsspezifische Modellierung von Daten berührt, weil viele Projekte noch nicht mit der Analyse von Ergebnissen beschäftigt sind. Im Sinne der vorab geschilderten Anschlussfähigkeit zur Zeitforschung und zur Beschäftigung von Digital History mit Fragen des Wandels von Begrifflichkeiten müsste dagegen viel stärker die Operationalisierung und Modellierung von hermeneutischen Erkenntnisprozessen in den Vordergrund treten und die Überführung von subjektiven und objektiven Dimensionen in auswertbare Kategorien thematisiert werden.

¹² Dominik Kasper und Andreas Kuczera, „Modellierung von Zweifel – Vorbild TEI im Graphen,“ in *Modellierung des Zweifels*, hg. v. Kuczera, Wübbena und Kollatz, doi: 10.17175/sb004_003; Andreas Wagner, „Ambiguität und Unsicherheit: Drei Ebenen eines Datenmodells,“ in *Modellierung des Zweifels*, hg. v. Kuczera, Wübbena und Kollatz, doi: 10.17175/sb004_007.

¹³ Martin Raspe und Georg Schelbert, „Genau, wahrscheinlich, eher nicht: Beziehungsprobleme in einem kunsthistorischen Wissensgraph,“ in *Modellierung des Zweifels*, hg. v. Kuczera, Wübbena und Kollatz, doi: 10.17175/sb004_012.

Dieser Aspekt berührt sehr alte „Problemzonen“ der methodischen Herausforderungen von Natur- und Geisteswissenschaften und ihren jeweiligen Besonderheiten auch in der Geschichtswissenschaft. Sie nimmt damit auch grundsätzliche Herausforderungen und Diskussionspunkte wieder auf, die besonders in den 1990er Jahren zu einer langandauernden Trennung von Statistik und Geschichtswissenschaft führten.¹⁴ Dennoch knüpfen solche Forschungen der Digital History trotz aller fachlichen Detailfragen bisher weder direkt an eigene Themen der Geschichtswissenschaft¹⁵ noch an Themen der Historischen Hilfswissenschaften an. So schließen die Ansätze der Zeitmodellierung innerhalb von Digital History nur vermittelt an die Ziele der „Chronologie“ an. Vielleicht bildet sich die größte Schnittfläche noch dort aus, wo man mithilfe eines standardisierten Verfahrens und von Normdaten die verschiedenen Dimensionen von Zeit in eine objektive Messbarkeit¹⁶ und mit entsprechenden Standards der Gedächtnisinstitutionen in gemeinsame Notationen¹⁷ bringen möchte.

Auffällig kehrt sich bei diesen Bemühungen jedoch der eigentliche Umgang mit der Zeit um. Denn diese Ansätze verfolgen erst einmal nicht den Bruch in der Zeit, sondern schaffen vielmehr eine Basis zur Vergleichbarkeit von Prozessen. Ähnliche Ansätze lassen sich auch in den wirtschaftshistorischen, ökonomischen und sozialwissenschaftlichen Studien wiederfinden,¹⁸ die über quantitative Forschungen langfristige Vergleiche im Kontext von Zeitreihenanalysen erforschen und dabei auch die methodischen Schwierigkeiten der statistischen Forschung beschreiben. Ein wunderbares Beispiel für den hohen Wert und die Grenzen eines solchen Unterfangens bietet das von Thomas Rahlf herausgegebene Buch „Deutschland in Daten“. Wie alle reflektierten Autor:innen seiner Zunft fragt er natürlich auch nach der Sinnhaftigkeit und der Möglichkeit, über verschiedene politische Systeme und Brüche hinweg Jahr für Jahr Daten aneinanderzureihen, um eine Vergleichbarkeit der ermittelten Daten herzustellen. Vergleicht man also Äpfel mit Birnen, wenn man über lange Zeitreihen hinweg und aus aggregierten Statistiken heraus Informationen verschneidet? Oder

14 Thomas Rahlf, „Einleitung,“ in *Deutschland in Daten: Zeitreihen zur Historischen Statistik*, hg. v. Thomas Rahlf, 5–12 (Bonn: bpb Bundeszentrale für Politische Bildung, 2015), 6.

15 Thomas Rahlf, „Voraussetzungen für eine Historische Statistik von Deutschland (19./20. Jahrhundert),“ *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 101/3 (2014): 322–352.

16 Anna-Dorothee von den Brincken, *Historische Chronologie des Abendlandes: Kalenderreformen und Jahrtausendrechnungen; eine Einführung* (Stuttgart: Kohlhammer, 2000); Hermann Grotefend, *Taschenbuch der Zeitrechnung des deutschen Mittelalters und der Neuzeit* (Hannover: Hahn, 2007).

17 Zentral ist hier die ISO 8601: Date and Time Format; momentan in der Fassung ISO 8601-1:2019 und ISO 8601-2:2019.

18 Rahlf, „Voraussetzungen,“ 333–345.

müssten diese vielleicht nicht doch besser über perioden- und epochenspezifische zeitlich kurzatmigere Daten präsentiert werden? Trotz dieser Fragen entschied sich Rahlf nicht zufällig für die langfristigen Betrachtungen mit großen Entwicklungslinien und Strukturbrüchen, die heutige oft auf die Mikrogeschichte gerichtete Erkenntnisse wesentlich ergänzen können. Wie auch Rehbein oder die zahlreichen Autor:innen zur Modellierung des Zweifels in der Digital History kommt auch Rahlf letztlich zum Fazit: Alles ist eine Frage der wissenschaftlichen Redlichkeit, der Methodik und der Quellenkritik¹⁹ und schließlich natürlich auch einer passfähigen und realistischen Einordnung von Forschungsergebnissen. Dabei präsentiert auch sein Band jedoch für die Statistik des 19./20. Jahrhunderts nur einen Exkurs über die Besonderheiten der DDR-Statistik aufgrund ihrer Orientierung an den Berechnungsmodalitäten der Sowjetunion und der Funktion von Statistik in einer Diktatur,²⁰ während spezifische Herausforderungen der Interpretation beziehungsweise die ausführliche Quellenkritik für Statistiken vor der Gründung des Kaiserreiches und dem Beginn der offiziellen Statistik mit der Reichsgründung, in der Weimarer Republik und im Nationalsozialismus dann doch auffällig fehlen, für die natürlich auch strukturelle Besonderheiten prägend waren.

Letztlich liegt vielleicht genau hierin ein hohes Integrationspotential, dass nicht nur Geschichtswissenschaft, Historische Hilfswissenschaften, Wirtschaftsgeschichte und Digital History zusammenbringen kann, sondern eben auch einen wesentlichen Mehrwert für ein interdisziplinäres Setting von Methoden der Nachnutzung und Langzeitanalyse bietet.²¹ Eine ganz wesentliche Stärke und Fachkompetenz der Historiker:innen liegt ja genau darin, methodische Herangehensweisen für Längsschnittanalysen oder überhaupt zeitliche Vergleiche für den Wandel von Begrifflichkeiten zu entwickeln und den Umgang mit verschiedenen Herausforderungen zur Bestimmung von Repräsentativität in sich wandelnden und unsicheren Dimensionen, von Plausibilität und Modellierung und zur Imputation oder Visualisierung des Quellenverlusts zu markieren. Dabei können sich Digital History, klassische Historische Hilfswissenschaften und Fachwissenschaften äußerst positiv ergänzen.

¹⁹ Rahlf, „Einleitung“, 7.

²⁰ André Steiner, „Die DDR-Statistik: Probleme und Besonderheiten“, in *Deutschland in Daten*, hg. v. Rahlf, 13–17.

²¹ Katrin Moeller, „Grundwissenschaften als Masterdisziplin der Nachnutzung“, in: Eva Schlotheuber, Rüdiger Hohls und Claudia Prinz (Hg.), *Diskussionsforum: Grundwissenschaften als Masterdisziplin der Nachnutzung. H-Soz-Kult* 2015 (2015). <http://www.hsozkult.de/debate/id/diskussionen-2923>. Zugriff am 19.02.2022.

Dazu ist zunächst eine Systematisierung von methodischen Herausforderungen und ihren bestehenden Lösungsansätzen notwendig und die fachwissenschaftliche Klärung, wann und wie Zeitdimensionen nun überhaupt objektive oder subjektive Prozesse, Wandel, Bruch oder Ähnlichkeit messen beziehungsweise konstatieren können. Dabei ist zunächst kritisch anzumerken, dass der Vergleich erst einmal eine Methodik gewähren muss, die nicht Wandel/Persistenz vorwegnimmt, sondern beides als Ergebnis des Vergleichs analysieren kann, was nicht immer berücksichtigt wird.

Während diese Verfahren im Bereich der Statistik häufig schon allgemeiner, jedoch nicht unbedingt fachspezifisch beschrieben sind, bleiben sie im Bereich der qualitativen Forschung und Hermeneutik jedoch weniger konkret greifbar. Dabei ist zugleich unsere schroffe Grenzziehung zwischen quantitativen und qualitativen Methodiken zu hinterfragen. So räumt die Geschichtswissenschaft einem zeitübergreifenden Vergleich von metrischen Daten – etwa in Form einer Altersreihe – unwillkürlich mehr Plausibilität ein, als dies bei nominalen oder kategorialen Daten – wie etwa dem Beruf – der Fall ist. Es soll später noch hinterfragt werden, ob diese Annahmen so eigentlich plausibel sind. Vielmehr scheint es einer Diskussion wert, ob nicht die Altersangabe historisch viel früher einen Prozess der Systematisierung, Operationalisierung und Standardisierung erlebt hat, als viele andere Terminologien, hier als Beispiel eben der Beruf, und damit messbar geworden ist. Dabei handelt es sich letztlich darum, eine einzelne Eigenschaft eines komplexen Begriffes herauszugreifen und atomisiert messbar zu machen (eine Grundregel, die heute in der Informatik und in den Digital Humanities unter dem *Entity Relationship Modell* firmiert).²² Unter einer Altersangabe wird nun daher bevorzugt die metrisch messbare Altersangabe als Lebensspanne in Jahren vom Geburtsdatum an verstanden. Dennoch gibt es in den verschiedenen fachwissenschaftlichen Sparten wie Psychologie, Medizin oder Soziologie natürlich auch die verschiedensten Ansätze, um das biologische, gefühlte, geistige, generative oder körperliche Alter zu messen.²³ Gleiches gilt auch für den komplexen Begriff des Berufes, der natürlich auch in Tätigkeitsprofile, Anforderungsniveau, Identitätskonzepte, sozialhierarchische Strukturen, Karrieremodelle und weiteres zerfällt und für unterschiedliche Forschungen auch im Längsschnittkonzept messbar gemacht werden kann.

²² Ramez Elmasri und Sham Navathe, *Fundamentals of Database Systems*, Always Learning (Boston u. a.: Pearson, 2017).

²³ Pitirim A. Sorokin und Robert K. Merton, „Social Time: A Methodological and Functional Analysis.“ *The American Journal of Sociology* 42/5 (1937): 615–629. <https://www.jstor.org/stable/2767758>, 615–616. Zugriff am 19.02.2022.

Eine punktuelle Betrachtung von Herausforderungen in beiden Bereichen soll an dieser Stelle helfen, grundlegende Aufgaben zu skizzieren und zu systematisieren. Darüber hinaus sollen sie zeigen, dass gerade die Spezifik der Datenlandschaft des „vorstatistischen“ Zeitalters in Verknüpfung mit dem „statistischen“ Zeitalter ähnliche Herausforderungen erbringt, wie für die Verbindung von quantitativen und qualitativen Ansätzen. Dabei scheint es für die gesamte Geschichtswissenschaft sehr gewinnbringend zu sein, sich Gedanken über die Lösung solcher Problematiken zu machen, da diese im Jahrhundert des digitalen Zeitalters und der Daten wichtige interdisziplinäre Antworten auf methodische Herausforderungen geben können.²⁴

Quellengrundlage und Vergleichbarkeit – Mikrodaten und Makrodaten

Grundlegend braucht es auch in der digitalen Geschichtswissenschaft kreative Ideen und Verständigungsprozesse dazu, wie mit den Quellenverlusten und Quellenbrüchen beim zeitlichen Vergleich umgegangen werden kann, die immer mit veränderten Mess- und Erfassungsweisen verbunden sind. Statistisch gesehen ist es weitgehend möglich, über Imputationsverfahren fehlende Quellen mit Schätzungen oder Hochrechnungen auszugleichen.²⁵ Inhaltlich bleibt die Frage offen, welche Vorgehensweisen in der weiteren Fachwissenschaft Anwendung und Akzeptanz finden und wie diese sichtbar gemacht werden.²⁶ Zum Teil verzichten gerade Gedächtnisinstitutionen und historische Forschung explizit auf verlorengegangenes Wissen, um die Lücken demonstrativ sichtbar zu machen. Beide Richtungen existieren auch in der Statistik, was allerdings in der Analyse von Zeit dann wiederum zu fehlenden und damit verfälschenden Berechnungsgrundlagen führt. Während zudem in der quantitativen Methodik fehlende Werte und ihre Ergänzungen in den verschiedenen Qualitäten hervortreten, sucht man solche Betrachtungen in der hermeneutischen Forschung häufig vergeblich, weil hier das Existente unmittelbar in die Analyse eingeht und dann das „nicht Auffindbare“ quellenkritisch nicht immer unmittelbar in

²⁴ Moeller, „Grundwissenschaften“.

²⁵ Christian F. G. Schendera, *Datenqualität mit SPSS* (München: Oldenbourg, 2011), 119–161.

²⁶ In der Psychologie führte etwa ein Aufsatz zu Imputationsverfahren zu mehreren Repliken: Oliver Lüdtke et al., „Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen.“ *Psychologische Rundschau* 58 (2007): 103–117, doi: 10.1026/0033-3042.58.2.103.

der Analyse reflektiert wird. Zudem gibt es bekanntlich ganz unterschiedliche Ursachen für „fehlende Informationen“.

Welche weiteren Arten von Schwierigkeiten treten nun bei der Konzeption einer „langen Reihe“ auf? Worüber müssen wir also Verständigungsprozesse einleiten? Mit dieser Frage soll sich ein kleiner Exkurs zur Entstehung einer Zeitreihe des Sterbealters im Vergleich mit Berufstätigkeiten der gleichen Quelle beschäftigen, der zusätzlich den Mehrwert von Digital Humanities unterstreicht. Die hier verwendete Datenreihe stammt zum Teil aus einer Totalerhebung von Kirchbuchdaten der Gemeinde „Unser Lieben Frauen“ (heute Marktkirche) der Stadt Halle zwischen 1670 und 1820.²⁷ Ergänzt wurde dieser Datensatz im Rahmen eines Schülerforschungsprojekts²⁸ für das 19. und 20. Jahrhundert mit den aggregierten Sterbedaten. Die Zeitreihe erfasst damit jährliche Rohdaten über den Zeitraum von 450 Jahren (rohe Sterbeziffern ohne Alter) beziehungsweise 350 Jahren (altersspezifische Sterbeziffern) und repräsentiert damit einen ungewöhnlich langen Zeitraum, der in dieser Form nur sehr selten zur Analyse kommt. Für das Forschungsvorhaben waren die Totgeburten sowie die Sterblichkeit von Säuglingen und jeweils zehn Jahre umfassende Alterskohorten möglichst differenziert nach Geschlecht die Wunschvorstellung für das ideale Untersuchungsset. Dies diente zur Grundlegung einer Fragestellung, die im Anschluss an demografische Forschungen nach der prinzipiellen Begrenztheit oder Unbegrenztheit menschlicher Lebenszeit fragte.²⁹ Solche Analysen sind in Halle/Saale ab dem Ende des 16. Jahrhunderts möglich, weil erst ab diesem Zeitraum überhaupt serielle Quellen mit Altersangaben zur Verfügung stehen.

Für die Zeitreihe war einerseits die Ausdifferenzierung innerhalb der Quelle zu beobachten, wobei es auch immer wieder die bekannten Brüche in der Repräsentation der seriellen Quelle gab, wenn beispielsweise ein neuer Küster oder Pfarrer die Führung der Sterberegister übernahm. Zwar wurden von Beginn an mehr oder weniger individuelle Namen und das Geschlecht der verstorbenen Personen genannt, weitere Informationen gab die Quelle jedoch erst im Verlauf der Jahrhunderte. So traten ab ungefähr 1660 Informationen zu Berufen, ab den 1730er Jahren sogar die Todesursachen und ab den 1760er Jahren räumliche Informationen (Adressen) hinzu, die später aber auch zeitweise wie-

²⁷ Katrin Moeller und Petra Kühne, *Sterberegister der Mariengemeinde Halle (Saale) 1670-1820*, hg. v. Historischen Datenzentrum Sachsen-Anhalt, Halle 2014, doi: 10.25673/14112.

²⁸ Jakob Moeller und Katrin Moeller, „Pandemien und Sterblichkeitskrisen in der Geschichte Halles (1579 bis 1818).“ *Preprints and Working Paper*, hg. v. Historischen Datenzentrum Sachsen-Anhalt, Halle 2020, doi: 10.25673/32869.

²⁹ Jim Oeppen und James W. Vaupel, „Demography. Broken Limits to Life Expectancy,“ *Science* 296/5570 (2002): 1029–1031.

der verschwinden konnten. Grundsätzlich lieferte jedoch die Mikrodatenerhebung – wenn auch mit spezifischen Unschärfen (zum Beispiel beim Geschlecht von Totgeburten und Säuglingen) alle für die Forschungsfrage relevanten Daten nach dem Geschlecht, dem erreichten Lebensalter und dem Zeitpunkt des Todes.

Für die Zeit davor und danach musste aber aus forschungsökonomischen Gründen auf aggregierte Daten und Statistiken zurückgegriffen werden. Dies bedeutete zunächst eine Verkürzung der Analyse um knapp 100 Jahre (statt 1579 nun erst 1670 als Ausgangspunkt), weil die frühen Daten nur über eine eigene Datenaufnahme (Mikrodaten) erschließbar gewesen wären und weil statistische Auswertungen aus dieser Zeit vollständig fehlen. Man kann aber festhalten, dass sie prinzipiell erfassbar gewesen wären. Erst mit dem Jahr 1700 hätten dagegen erste Abhandlungen mit aggregierten Daten zu den Verstorbenen zur Verfügung gestanden,³⁰ die allerdings für die Forschungsanalyse auch nicht hinreichend genaue Angaben machten. Solche zusammenfassenden Angaben ließen sich zwar für die Zeit nach 1820 auf Grundlage der Kirchenbücher sowohl in Einzelstatistiken³¹ wie in zusammenfassenden Berichten³² finden, sie bleiben aber lückenhaft und sind eben für viele Fragestellungen dann nicht wirklich passfähig, weil sie andere Fragestellungen verfolgten.

Es ist diesen Jahrzehnten geschuldet, dass die Analyse nicht vergleichend für Frauen und Männer differenziert durchgeführt werden konnte, weil diese Geschlechtsspezifikation in den aggregierten Quellen fehlte. Überdies strukturierten diese Quellen die Altersgruppen erheblich vor, denn hier wurden zwar die Totgeburten, Säuglings- und Kindersterblichkeit angegeben, für die älteren Menschen erfolgte allerdings eine zusammenfassende Angabe, die lediglich die Gruppe der über 80-Jährigen heraushob, die mit der individuellen Altersangabe benannt wurden.

30 Heinrich Ernst Güte, *Angabe und Berechnung der Gebornen, Verstorbenen, Copulirten und Communicanten in den evangelisch=lutherischen Gemeinden der Stadt Halle und der damit verbundenen Amtsstädte vom Jahr 1701 bis 1800* (Halle, 1801).

31 Stadtarchiv Halle, Neujahrzettel: 1801–1802, 1808, 1810–1811, 1814–1815, 1817–1853, A 1.1.1, VA HA Kap. I, Abt. B, Nr. 11, Stadtarchiv Halle.

32 Stadt Halle, Bevölkerungsstatistik. Bericht über die Verwaltung und den Stand der Gemeinde=Angelegenheiten der Stadt Halle a./S., Marienbibliothek Halle/Saale.

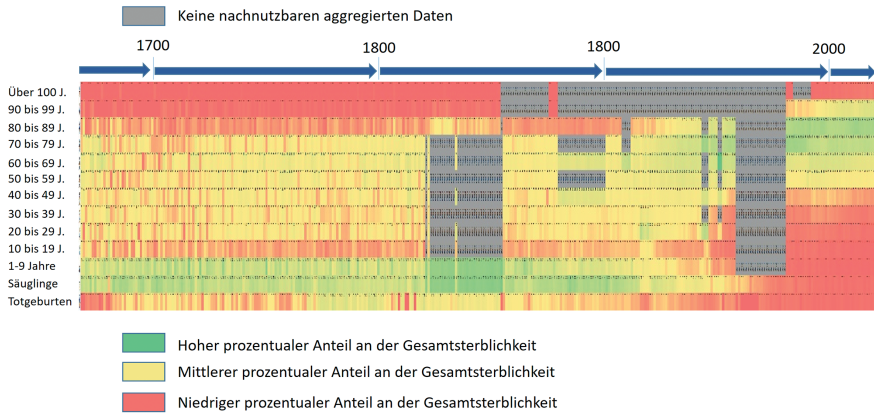


Abb. 1: Prozentuale Verteilung von Altersgruppen an der Sterblichkeit, 1670 bis 2018, Halle (Saale) (Grafik: Jakob und Katrin Moeller).

Festhalten kann man aus diesen Beobachtungen vier systematische Aspekte:

1. Nicht die in der Forschung immer wieder beschriebene Unterteilung in das „vorstatistische“ und „statistische“ Zeitalter machte den zeitlichen Vergleich schwierig, sondern der Bruch zwischen den genutzten Mikrodaten, die alle erforderlichen Informationen lieferten, und der Rückgriff auf die aggregierten Daten, die dies nicht in gleicher Weise oder auf einer direkt vergleichbaren Basis taten. Genau hier können Digital Humanities, breite Quelldigitalisierung (gerade von seriellen Quellen) oder digitales Forschungsdatenmanagement allein durch die Bereitstellung und Erschließung von bevorzugt Mikrodaten/Rohdaten (neben Analysedaten) wesentliche Beiträge zur Forschung leisten, selbst wenn sie nicht unmittelbar zu wissenschaftlichen Analysen dienen.
2. Der Bruch zwischen Mikro- und Makrodaten führt zu notwendig höheren Abstraktionsprozessen und ermöglicht nur eingeschränkte Kategorienbildungen. Dabei wurde nicht die Fragestellung verändert, wohl aber die Komplexität und Genauigkeit der Analyse (Verzicht der geschlechtsspezifischen Differenzierung, größere Alterskategorien). Es ist daher eine Aufgabe der digitalen Quellenkritik, Festlegungen oder Richtlinien zu entwickeln, in welchem Ausmaß solche Einschränkungen Akzeptanz finden, und so auch eine Verlässlichkeit für die Fachwissenschaft zu entwickeln. Dabei gilt natürlich, dass bei jeder Kategorienbildung unvermeidlich höhere Abstraktionsprozesse enthalten sind.

3. Zur digitalen Quellenkritik gehört auch immer eine reflexive Überprüfung der Aussagekraft, Repräsentativität und des Wandels der Quellengrundlage. Beispielsweise verloren die Kirchenbücher im Verlauf des 19. Jahrhunderts sukzessiv an Repräsentativität, weil Menschen zunehmend aufhörten, der Kirche anzugehören und hier eben auch nicht mehr erfasst wurden. Überdies ist bekannt, dass einzelne Bevölkerungsgruppen nie durch die kirchliche Registration erfasst wurden. Erst ab 1871 wurde dieser Schwund systematisch von der standesamtlichen Registrierung aufgefangen sowie durch weitere nunmehr sehr differenzierte statistische Dokumentationen ergänzt. Für die Stadt Halle/Saale wurde zwischen 1906/07 und 1950 ein statistisches Jahrbuch herausgegeben, dass vielfältige Einblicke in die Demografie und Altersentwicklung erlaubt und das seit 2004 eine Neuauflage erhielt. Insgesamt gibt es im Verlauf des Untersuchungszeitraumes eine erhebliche Veränderung des Einzugsbereiches. Wurde zu Beginn noch eine einzelne Kirchengemeinde erfasst, welche nur etwa die Hälfte der Bevölkerung repräsentierte, war es im 19. Jahrhundert dann der gesamte Stadt- raum, der durch fortwährende Eingemeindungen schnell wuchs. Natürlich können solche Probleme statistisch mit formalen Methoden eingefangen werden, sie müssen in der geschichtswissenschaftlichen Praxis jedoch auch Akzeptanz finden.
4. Wichtig erscheint noch ein Verweis auf die Zeit zwischen 1933 und 1945 sowie auf die DDR. Nicht im vorstatistischen Zeitalter traten die größten Schwierigkeiten der Quelleninformation auf, sondern die größten methodischen Probleme bereitete die DDR-Zeit, in der zwar viele Daten erfasst, aber aufgrund der Geheimniswahrung häufig nicht publiziert wurden. Herausgegeben wurden nur die statistischen Jahrbücher für die gesamte DDR,³³ bereits die Statistiken auf Bezirksebene³⁴ wurden nur für den Dienstgebrauch ausgewiesen, in den Zahlen auf Kreisebene³⁵ finden sich kaum differenzierte Angaben zur Bevölkerungsentwicklung. Quellen gerieten aufgrund der turbulenten Geschehnisse von Wende und Wiedervereinigung entweder gar nicht in Archive oder werden jetzt erst allmählich erschlos-

33 *Statistisches Jahrbuch der Deutschen Demokratischen Republik*, hg. v. Staatliche Zentralverwaltung für Statistik (Berlin: VEB Deutscher Zentralverlag, 1955 (Jg. 1) bis 1991 (Jg. 35)).

34 *Statistisches Jahrbuch des Bezirkes Halle*, hg. v. Staatliche Zentralverwaltung für Statistik, Bezirksstelle Halle (Halle: Statistisches Bezirksamt Halle, Bd. 1, Berichtsjahr 1963 (1964), letzter Bd. Berichtsjahr 1981 (1982)), vorhanden in der Sächsischen Staats- und Universitätsbibliothek, Bestand Y.8.997.

35 Die wenigen Exemplare befinden sich ebenfalls im Dresdner Bestand Y.8.997.

sen.³⁶ Im hier vorliegenden Datensatz wurden daher die Werte zum Durchschnittsalter auf Grundlage des erreichten Lebensalters im Verhältnis zur hallischen Bevölkerung imputiert.³⁷ Für die neuesten Daten seit der Wiedervereinigung bestand dann wieder das „Luxusproblem“, nicht über eine, sondern teilweise zwei oder mehr Datenreihen mit verschiedenen Zählweisen zu verfügen. Seit 1991 liegen sowohl die Daten des Landesamts für Statistik wie der städtischen Verwaltung „Statistik und Wahlen“ vor. Gleichzeitig unterliegen auch quantitative Daten dem Datenschutz. So dürfen Angaben von weniger als fünf Personen nicht publiziert werden, weil diese trotz der Anonymisierung einen Rückschluss auf einzelne Personen erlauben würden, statistisch aber dennoch relevant sind.

Auf Grund des Rückgriffs auf aggregierte Daten und von Geheimniswahrung oder heute Datenschutz sowie veränderten Erfassungsrichtlinien von Behörden in der heutigen Zeit führt dies fast zu mehr Analyseproblemen als die Zeit der Frühen Neuzeit, wo vorwiegend die Quellenerschließung geeigneter Ressourcen das eigentliche Hindernis bleibt, die heute mit Methoden der automatisierten Texterkennung aber aufgefangen werden könnten.

Fragen der Vergleichbarkeit: Wie und was wird gemessen?

Die grundlegendste Voraussetzung für die Verschneidung und Zusammenführung von Daten (und Big Data) ebenso wie für die Entwicklung von Wissensgraphen, Normdaten oder für wissenschaftliche Analysen ist und bleibt jedoch ihre inhaltliche Passfähigkeit, die auf Ebene jeder einzelnen Variablen geklärt werden muss. Dabei werden heute häufig statistische Verfahren verwendet, ohne die Grundprinzipien der Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit) in Massenquellen tatsächlich sicher abschätzen zu können. Dies gilt zum Teil auch, weil für qualitative und quantitative Datenanalysen variierende Prinzipien der Datenerhebung und Datenanalyse gelten, die in Mixed Methods-Ansät-

³⁶ André Steiner, „Die DDR-Statistik: Probleme und Besonderheiten,“ in *Deutschland in Daten*, hg. v. Rahlf, 15.

³⁷ Für den hier verwendeten Datensatz: LHASA, MER, Institut für Sozialhygiene und Organisation des Gesundheitsschutzes 1946–1991, M 527, Nr. 41, 43 und 45. Hier sind die Jahre 1976, 1978 bis 1989 vorhanden.

zen nun jedoch wechselseitig integriert werden.³⁸ Ausschlaggebender dürfte jedoch sein, dass grundlegende Fragen zur Reliabilität und Validität in Verfahren des *Distant Reading* trotz aller Bemühungen um Fehlerraten für Massendaten gerade in Digital History-Projekten schwer abschätzbar bleiben.³⁹ Das Auffinden einer bestimmten Information etwa über Text Mining, als Beispiel soll hier der Terminus „Bäcker“ verwendet werden, lässt sich noch nicht sicher mit dem Kontext einer Zuordnung dieser Informationen zu einer bestimmten Entität (statistisch: Variable, Kategorie) verorten. Gehört diese Angabe nun zu einem bestimmten Beruf, wird gerade eine Person benannt oder war im Text damit eigentlich die Zunft der Bäcker (Organisation), das Viertel der Bäcker (Ort) oder etwas völlig anderes gemeint? Überdies sind die Fragen nach der Erfassung einer fixierbaren Grundgesamtheit und der Klärung der Beobachtungseinheit, was bereits ganz basale Voraussetzungen für statistische Auswertungen bildet, häufig nicht sicher bestimmbar. Sowohl die statistische Auswertung noch mehr aber die fachliche Forschung im Rahmen von Digital History und Geschichtswissenschaft bedarf daher vor allem einer auf inhaltlichen Entitäten beruhenden Erkennung von Daten in ihrem Kontext, um eine differenzierte Quellenkritik überhaupt zu ermöglichen. Mittlerweile gibt es eine Vielzahl von Initiativen, die daran arbeiten, genau diese Herausforderung besser in den Griff zu bekommen.

Immerhin bieten gerade die in der historischen Forschung häufig genutzten seriellen Quellen, die etwa Angaben zu Alter und Beruf im langen Bogen thematisieren, den unschätzbaren Vorteil, eine Entitätserkennung wesentlich zu erleichtern, weil sie diese Angaben bereits in der jeweiligen Zeit separiert nennen und über OCR- und HTR-Erkennungen dann auch relativ schnell erfassbar sind. Die Problematik der Messbarkeit von Begrifflichkeiten im Wandel der Zeit bleibt jedoch auch hier bestehen.

Angesprochen wird damit die inhaltliche Konsistenz (was wird gemessen) ebenso wie die Frage der Methodik (wie wird gemessen), welche die grundlegende Voraussetzung für die zeitliche Vergleichbarkeit schaffen. Problemszenarien gibt es auch für die historische Forschung zahllose, einige davon illustriert auch Christopher Banditt in diesem Band mustergültig. Die Herausforderungen sind also hoch, die Möglichkeiten und Grenzen hierfür werden in Ausführungen

38 Rainer Diaz-Bone, Udo Kelle und Jo Reichertz, „Zur Frage der Integration von qualitativer und quantitativer Forschung,“ in *Qualitative Forschung, Analysen und Diskussionen – 10 Jahre Berliner Methodentreffen*, hg. v. Günter Mey und Katja Mruck, 227–232 (Wiesbaden: Springer VS, 2014), doi: 10.1007/978-3-658-05538-7_12.

39 Christian Fürber, *Data Quality Management with Semantic Technologies* (Wiesbaden: Springer Fachmedien Wiesbaden, 2016), doi: 10.1007/978-3-658-12225-6, 28–32; 74 f.

zur Methodik der Geschichtswissenschaft jedoch weniger direkt angesprochen⁴⁰ oder in der Ausführung tatsächlich konkretisiert, weil diese Herangehensweisen besonders hermeneutische Dimensionen reflektieren.⁴¹ Dabei wird für Kategorien wie das Alter oder andere mit naturwissenschaftlichen Prinzipien erfasste Entitäten häufig weit eher eine Möglichkeit zur Messbarkeit auch über eine lange Zeitreihe unterstellt, als dies für qualitative Variablen wie etwa den Beruf der Fall ist. Dies liegt wie gezeigt jedoch weniger an der Quellsituation selbst, sondern vielmehr an fachlichen Verständigungsprozessen über eine spezifische Entität.

Absolute Häufigkeit

	10 J.	20 J.	30 J.	40 J.	50 J.	60 J.	70 J.	80 J.	90 J.
- 1 z. B. 9 Jahre	225	251	208	230	253	269	286	191	34
Runde Zahl z. B. 10 Jahre	169	333	365	449	451	532	375	201	30
- 1 z. B. 11 Jahre	141	248	188	199	227	241	272	103	16

Abb. 2: Altersangabe in Sterberegistern Halle (Saale), 1680 bis 1820 (Abbildung von Katrin Moeller)

Denn natürlich bleibt auch die Information über das Alter einer Person in jeder historischen Quelle nur ein ungefährender Näherungswert. Sehr schön dokumentiert dies die in der Demografie bekannte Häufung von Altersnennungen: Für Gestorbene wurden in Halle/Saale im 18. Jahrhundert etwa doppelt so häufig die „runden“ Angaben gemacht, sie seien 20, 40 oder 60 Jahre alt geworden, statt „krumme“ Werte wie 21, 41 oder 61 Jahre zu nennen. Dabei war die Differenz in den höheren Altersgruppen größer, während bei den jüngeren Menschen aufgrund der zeitlichen Nähe offenbar noch ein spezifischeres Wissen bestand. Natürlich gibt es auch hier ganz allgemeine Angaben die beschreibend auf das Phänomen der Zeit eingingen wie „Wochenkind“, „Alt“, „im besten Alter“ oder „Uralt“, für die diskutiert werden muss, ob sie in konkrete Alterskategorien eingeordnet werden können. Andererseits wurden selbst bei sehr alten

⁴⁰ Stefan Jordan, *Theorien und Methoden der Geschichtswissenschaft*. Orientierung Geschichte 3104 (Paderborn: Ferdinand Schöningh, 2018).

⁴¹ Rüdiger Graf, *Zeit und Zeitkonzeptionen in der Zeitgeschichte*, Version: 2.0, in: Docupedia-Zeitgeschichte, 22.10.2012, doi: 10.14765/zzf.dok.2.266.v2.

Menschen von den Pastoren noch liebevoll die durchlebten Wochen, Stunden und manchmal gar Minuten notiert, was ebenfalls mehr über die sozialen Beziehungen zu dieser Person aussagt als gegebenenfalls zum tatsächlichen Alter. Diese Quellenprobleme und Ungenauigkeiten, für die heute so vielfältige Betrachtungen angestellt werden, sind letztlich für die eigentliche statistische Analyse allerdings gar nicht so erheblich, weil die Differenzierung der Kategorien gar nicht diese Tiefenschärfe erreicht.

Die Diskussion zur inhaltlichen Passfähigkeit des Alters könnte auch entlang der Chance/Differenz der erwartbaren Lebenszeit geführt werden, die über alle Altersgruppen in der Stadt Halle/Saale gerade mal bei 25 Jahren lag. Entspricht damit das heutige Lebensalter eines 80-Jährigen in der Frühen Neuzeit nicht eher dem Alter eines 25, 40, 50 oder 60-Jährigen? Die historische Demografie hat sich gegen solche Analysemethoden entschieden. Dies liegt auch darin begründet, dass der zeitliche Wandel der Sterblichkeit immens war: Während heute die 82-jährige Frau im Durchschnitt am häufigsten verstirbt, war es im 18. Jahrhundert der männliche Säugling. Letztlich würde es die Analyse des zeitlichen Wandels vielleicht verfälschen, wollte man also eigene Indikatoren entwickeln, welche den Wandel bereits zum Gegenstand der Kategorisierung machen.

Natürlich lassen sich dabei nur Aussagen zum zeitlichen Wandel des tatsächlich angegebenen Alters machen, nicht etwa zum gefühlten oder biologischen Alter, die in diesem Quellenbeispiel aufgrund der hier fehlenden Angaben nicht analysierbar sind. Insgesamt kann man nur immer wieder konstatieren, dass sich natürlich auch subjektive Wahrnehmungsformen und Prozesse in statistisch orientierte Kategorien bringen lassen, wenn denn Quellen dafür vorliegen. Interessanter Weise ordnete aber auch die Gesellschaft des frühen 19. Jahrhunderts nicht – wie man eigentlich erwarten könnte – die 40, 50 oder 60-jährigen Menschen der herausgehobenen Gruppe der „Alten“ zu, sondern diejenigen, welche die Altersgrenze der 80 Jahre überschritten. Dies war letztlich auch für das hier vorgestellte Projekt die ausschlaggebende Begründung, warum keine Methode zur Imputation differenzierter Altersgruppen entwickelt wurde, sondern sich die Analyse an den inhaltlichen Festsetzungen der Quelle selbst orientierte und schließlich nur nach der Sterblichkeit von ganzen vier Gruppen von Kindern, Jugendlichen, Erwachsenen und Alten verglich.

Was für das metrische Alter unberücksichtigt bleibt, wird für die qualitative Variable des Berufes mehr als genau unter die Lupe genommen, wie zahllose Diskussionen zeigen. Bereits weiter oben wurde die Berufsanalyse erwähnt, eine Variable, die sich neben dem Alter in der gleichen hier herangezogenen Quelle findet, wenn auch im 18. und 19. Jahrhundert noch unter der Bezeichnung des „Berufsstandes“ geführt. Diese Bezeichnung umfasste, wie der

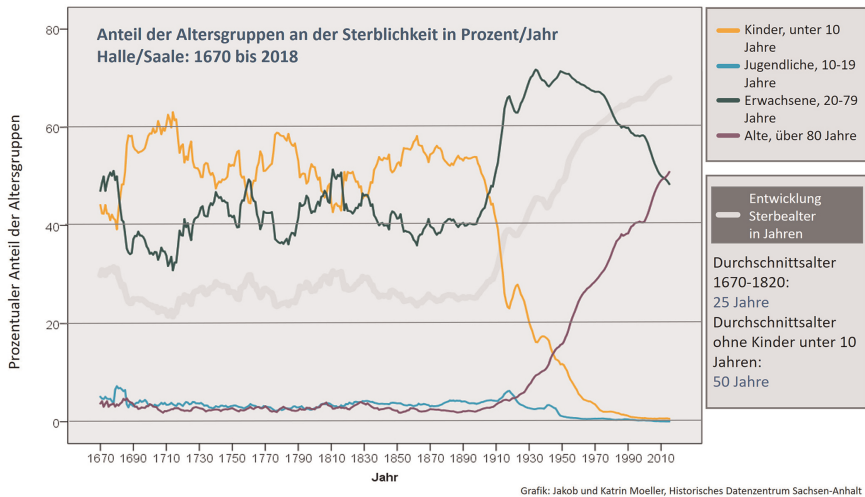


Abb. 3: Anteil der Altersgruppen an der Sterblichkeit in Prozent/Jahr, Halle/Saale: 1670 bis 2018 (Grafik: Jakob und Katrin Moeller)

heutige Berufsbegriff auch, neben den Aspekten der eigentlichen Erwerbstätigkeit auch soziokulturelle Aspekte wie etwa Elemente der (religiös geprägten) Lebensführung.⁴² Kann diese Entität des Berufsstandes nun für den Vergleich von Berufsangaben dienen, die heute mit spezifischen normativen Konzepten eines mit fachlicher Ausbildung und Zertifizierung verknüpften Erwerbsberufes versehen sind? Während moderne Soziologie und Wirtschaftswissenschaft international Berufe seit über 100 Jahren kategorisieren und vergleichen, und dabei natürlich auch den äthiopischen Injeraproduzenten⁴³ mit Berufen in einer hochmechanisierten, europäischen Großbäckerei vergleichen kann, lautete die Antwort der federführenden Kompetenz der DFG-Kommission Geschichtswissenschaft abschlägig: Der Berufsstand der Frühen Neuzeit und der moderne Erwerbsberuf im 20. Jahrhundert seien unvergleichbar, konstatiert dieses Gremium, das damit auch inhaltlich über Entitätsrekonstruktionen und Entwicklungen von Digital History entscheidet. Damit argumentierte die Kommission ganz analog der Ausgestaltung einer arbeitsorientierten geschichtswissenschaftlichen Forschungslandschaft, die mittlerweile zu relativ separierten Forschungen

⁴² Katrin Moeller, „Stand oder Beruf? Arbeitsspezifische und religiöse Konzepte von Beruf und Leistung im 18. Jahrhundert,“ in *Industrielle Arbeitsbeziehungen in Mitteldeutschland: Zwischen Klassenkampf, Sozialpartnerschaft und Betriebsgemeinschaft (1815–2015)*, hg. v. Michael Schäfer, Swen Steinberg und Veronique Töpel (Leipzig: Leipziger Universitätsverlag, 2018), 27–48.

⁴³ Injera ist ein Fladenbrot aus Teff.

einerseits zur ständischen Gesellschaft und andererseits zur industriellen Berufsgesellschaft geführt hat. Für letztere arbeitete die Zunft in der jüngeren Vergangenheit vor allem die normative Fixierung der Arbeit durch die reichsweite Sozialgesetzgebung sowie den Patent- und Urheberschutz als prägendes Merkmal der modernen Arbeit heraus, welche durch das Prinzip der sozialen Platzierung durch Arbeitsleistungen und die darauf aufbauende Ökonomisierung des gesamten Lebens sowie die Entwicklung der (spezifisch deutschen) Ausbildungsberufe im 20. Jahrhundert als Zäsur setzende Merkmale ergänzt werden.⁴⁴ Damit wird für das 19./20. Jahrhundert eine einzigartige Berufsgesellschaft konstatiert, die heute bereits im Zerfall begriffen ist. Die Frage ist nur, in welchem Verhältnis dazu die Erwerbsarbeit der Frühen Neuzeit steht, die letztlich durch ähnliche Merkmale von beruflichen Tätigkeitsfeldern und Anforderungsniveaus definiert ist, auch wenn die Rahmenbedingungen von Normen und sozialer Platzierung selbstverständlich andere waren. Gleiches könnte analog dazu allerdings auch für das Sterbealter in Anspruch genommen werden, wie ja allein die heute völlig andere Altersstruktur und das Sterbealter dokumentieren. Natürlich waren die Wohn- und Lebensbedingungen, die Hygiene, die Mittel der Gesundheitserhaltung durch die Medizin in der Frühen Neuzeit gänzlich andere als heute. Kann man daher das Lebensalter einst und jetzt nicht vergleichen? Oder wird hier nicht eher das Ergebnis des Vergleichs mit der Methode verwechselt?

Die Ablehnung einer Vergleichbarkeit von vormodernen und modernen Berufen speist sich letztlich auch aus den früheren Diskussionen der Überinterpretation des Berufsbegriffes in der früheren Forschung. Besonders die Schlussfolgerung von Sozialhierarchien aus reinen Berufsangaben führte zu falschen Schlussfolgerungen beziehungsweise Überinterpretationen, wie in den 1990er Jahren durchaus zu Recht kritisiert wurde.

Gleichzeitig führte noch eine zweite methodische Unschärfe zu dieser Ablehnung: Der Beruf wird nicht als spezifische Entität mit einer Eigenschaft gemessen oder interpretiert, sondern als komplexer analytischer Begriff im Kontext seiner vielfältigen Rahmenbedingungen, wie dies in der hermeneutischen Interpretation wahrgenommen wird. Wie bereits für die Zeit herausgearbeitet, folgt die informatische Modellierung und statistische Analyse strikt den Regeln einer Entitätsbildung, die sich auf einzelne Eigenschaften komplexer Begrifflichkeiten orientiert (ER-Modell). Im Fall des Berufes ist dies vor allem die Tätigkeit beziehungsweise das Tätigkeitskonzept, welche über die bloße Berufsbezeichnung wie sie uns in der Masse der historischen Quellen entgegentritt,

⁴⁴ Hier sei stellvertretend für eine sehr ausdifferenzierte Forschungslandschaft nur auf einen viele Aspekte bündelnden Tagungsband verwiesen: *Geschichte und Zukunft der Arbeit*, hg. v. Jürgen Kocka und Claus Offe (Frankfurt a. M.: Campus-Verlag, 2000).

überwiegend ermittelbar ist. In diesem Fall der Operationalisierung geht es also nicht, wie häufig unterstellt, um die soziale Hierarchie einer Gesellschaft oder die ständische Platzierung, nicht um bestimmte Ausbildungskonzepte oder die normative Verankerung eines Berufes. Diese Kategorien bedürften anderer Quellen, um sie zu messen. In einer Gesamtanalyse können sie aber natürlich wieder zu komplexeren Methoden zusammenfließen. Letztlich geht es bei der Berufsnennung, wie sie uns in vielen Quellen entgegentritt, eben um einen bestimmten Stand, der mit einer mehr oder weniger präzise beschreibbaren Tätigkeit zusammenfällt.

In Hinblick auf das Tätigkeitskonzept ist es daher doch weniger relevant, ob der Bäckerlehrling der Frühen Neuzeit seine Lossprechung vor einer örtlichen Zunft und unter Regelungen lokaler normativer Regelungen fand, während er dies heute im besten Fall unter dem Eindruck deutschlandweiter Regelung tut und morgen vielleicht unter EU-Normen. Dies wäre bevorzugt für bildungsbasierte Skalen und Messungen wichtig. Es hat für die Tätigkeitsbeschreibung auch weniger Relevanz, ob der Arbeitnehmer für seine Arbeit eine Altersabsicherung über individuelle, betriebs- oder gesellschaftsorganisierte Formen erhält oder ob er den Beruf frei wählen konnte. In der modernen Berufsforschung kennt man mehr als 25 verschiedene Klassifikationsansätze für ganz verschiedene Eigenschaften des Berufes.⁴⁵ Der am Historischen Datenzentrum Sachsen-Anhalt entwickelte Klassifikationsansatz der „Ontologie der historischen, deutschsprachigen Amts- und Berufsbezeichnungen“ schließt an internationale Standards zur Berufsklassifikation wie ISCO, HISCO in seiner deutschsprachigen Ausformung der Klassifikation der Berufe 2010 an und nimmt Tätigkeitsprofile und Kompetenzen in den Blick. Das heißt, ausgehend von den Überschneidungen oder Ähnlichkeiten einzelner Kompetenzen, Tätigkeiten und Anforderungsniveaus, hat man die einzelnen Berufsgruppen zusammengefasst.

Zentral sind dagegen Fragen, die sich tatsächlich auf die Tätigkeit und ihre Eigenschaften beziehen: Gehört der Bäcker der Frühen Neuzeit, der sein Brot mit der Hand knetete, in die gleiche Kategorie wie der heutige Bäcker, der die Maschinen dazu bedient? Moderne Klassifikationen gehen, wie oben geschildert davon aus, weil Tätigkeiten hier in einem Gesamtspektrum beruflicher Tätigkeit betrachtet werden.

Eine wichtige Frage für Digital History ist also: Wie viel Differenzierung ist für – vornehmlich standardisierte und für automatisierbare – Kategorienbildung von Massendaten notwendig, um zu einem guten Vergleich mit statisti-

⁴⁵ Forschungsdaten des BIBB-FDZ zu administrativen und nicht-administrativen Klassifikationen, in *FDZ Metadatenportal*, hg. v. Bundesinstitut für Berufsbildung (Bonn 2019), <https://metadaten.bibb.de/klassifikationenIndex>. Zugriff am 19.02.2022.

schen Mitteln und zur Akzeptanz im eigenen Fach zu kommen? In Hinblick auf die hier vorgestellten Beispiele wurde sichtbar, dass es sich kaum lohnt, hier überbordend hohe Barrieren mit einem unermesslichen Aufwand zu fixieren, was besonders für die Langzeitperspektive gilt. Die heute in Deutschland angewandte Klassifikation von Berufen im internationalen Standard (Klassifikation der Berufe 2010/20) kennt in ihrer minimalen Kategorienbildung zehn Berufshauptgruppen, und davon ausgehend differenzieren sie sich in ihrer maximalen Beschreibungsform in 1.286 Berufsuntergruppen.⁴⁶ Sie basiert auf einem rudimentären Konzept von verschiedenen Kompetenzen, Produkten und Sprachen, die eben selbst dem Unterschied zwischen manueller Tätigkeit des Backens und dem Bedienen von komplexen Backautomaten noch relativ wenig Bedeutung geben. Sie setzt aber sehr granulare Berufsamen zur Klassifikation an, um Einordnungen transparent und nach Bedürfnis auch änderbar auszugestalten. Solche Beispiele können für uns handlungsleitend sein, um gute Vokabulare oder gar Standards zu entwickeln, die zudem tatsächlich einer Analyse von Zeitdimensionen unmittelbar zugutekommen, gleichzeitig aber überhaupt effektiv einsetzbar und flexibel bleiben. Bei den Berufsbezeichnungen wird zudem schnell klar, dass sich verändernde Berufskonzepte häufig auch veränderte Begrifflichkeiten nach sich ziehen. Allein diese Analyse kann also Wandlungsprozesse in den Blick nehmen, wenn sie auch auf das Original und die Quelle zurückverweist und diese transparent sichtbar macht.

Zusammenfassung

Die Thematik des zeitlichen Vergleichs ist für Digital History und weit darüber hinaus für ein interdisziplinäres, datenbasiertes Forschungsfeld innovativ nutzbar. Über das Anwachsen von Datenbeständen werden Verfahren und Methoden der Langzeitanalyse möglich, für die Geschichtsforschung viele Kompetenzen mitbringt und diese systematisch ausbauen kann.

Wie argumentiert, wird das Ergebnis hermeneutischer Methoden des zeitlichen Vergleichs weitgehend in der Analyse des Wandels oder des Bruchs dargestellt, während die Gemeinsamkeiten fast unkommentiert „verschwinden“. Die grundsätzliche Verankerung des Vergleichs als Methode der Differenz und Übereinstimmung, wie sie bereits von John Stuart Mill 1843 skizziert wurde, hat sich

⁴⁶ *Klassifikation der Berufe 2010 (KldB 2010)* (Nürnberg, 2011). <https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/KldB2010-Fassung2020-Nav.html>. Zugriff am 19.02.2022.

in der hermeneutischen Praxis nicht gleichermaßen durchgesetzt.⁴⁷ Dies gelingt durch die formale Modellierung von zeitbezogenen Daten mithilfe definierter Eigenschaften von Entitäten. Die Modellierung zeitbezogener Daten verspricht also einen auf das eigene Fach bezogenen Mehrwert, wie Rehbein dies als wichtiges Kriterium für digitale Forschung gefordert hat. Während hermeneutische Forschung zudem eine komplexe Einbettung ermöglicht, ist dies bei der Modellierung von Daten der umgekehrte Fall.

Notwendig und wichtig sind jedoch Verständigungen über methodische Lösungskonzepte zur Abstraktion von Kategorien, zum Umgang mit Imputationsverfahren und ihre Weiterentwicklung sowie zu Forschungen zum genauen Begriffswandel und seiner Modellierung.

Bibliographie

- Brendecke, Arndt (Hg.). *Die Autorität der Zeit in der Frühen Neuzeit*. Berlin: LIT-Verl., 2007.
- Brincken, Anna-Dorothee von den. *Historische Chronologie des Abendlandes: Kalenderreformen und Jahrtausendrechnungen; eine Einführung*. Stuttgart: Kohlhammer, 2000.
- Chvojka, Erhard, Andreas Schwarcz und Klaus von Thienen (Hg.). *Zeit und Geschichte: Kulturgeschichtliche Perspektiven*. Wien, München: R. Oldenbourg Verlag, 2002.
- Deeg, Alexander. „Advent ist immer.“ In *Verweile doch ...: Über die Erforschung der Zeit*, hg. v. Elmar Schenkel und Kati Voigt. 99–106. Leipzig: Edition Hamouda, 2015.
- Diaz-Bone, Rainer, Udo Kelle und Jo Reichertz. „Zur Frage der Integration von qualitativer und quantitativer Forschung.“ In *Qualitative Forschung: Analysen und Diskussionen – 10 Jahre Berliner Methodentreffen*, hg. v. Günter Mey und Katja Mruck, 227–232. Wiesbaden: Springer VS, 2014. doi: 10.1007/978-3-658-05538-7_12.
- Elmasri, Ramez und Sham Navathe. *Fundamentals of Database Systems*. Always Learning. Boston u. a.: Pearson, 2017.
- Franz, Albrecht und Katja Patzel-Mattern. „Annäherungen an einen kulturwissenschaftlichen Zeitbegriff.“ In *Der Faktor Zeit: Perspektiven kulturwissenschaftlicher Zeitforschung*, hg. v. Katja Patzel-Mattern und Albrecht Franz, 7–14. Stuttgart: Franz Steiner Verlag, 2015.
- Fürber, Christian. *Data Quality Management with Semantic Technologies*. Wiesbaden: Springer Fachmedien Wiesbaden, 2016, doi: 10.1007/978-3-658-12225-6.
- Rüdiger Graf. *Zeit und Zeitkonzeptionen in der Zeitgeschichte*, Version: 2.0, in: Docupedia-Zeitgeschichte, 22.10.2012. doi: 10.14765/zzf.dok.2.266.v2
- Grotefend, Hermann. *Taschenbuch der Zeitrechnung des deutschen Mittelalters und der Neuzeit*. Hannover: Hahn, 2007.
- Güte, Heinrich Ernst. *Angabe und Berechnung der Gebornen, Verstorbenen, Copulirten und Communicanten in den evangelisch=lutherischen Gemeinden der Stadt Halle und der damit verbundenen Amtsstädte vom Jahr 1701 bis 1800*. Halle: Ruff, 1801.

⁴⁷ Hartmut Kaelble, „Historischer Vergleich,“ *Docupedia-Zeitgeschichte* 2012 (2012): 1, doi: 10.14765/zzf.dok.2.271.v1.

- Jochens, Birgit. *Deutsche Weihnacht: Ein Familienalbum 1900–1945*. Berlin: Nicolai, 2011.
- Jordan, Stefan. *Theorien und Methoden der Geschichtswissenschaft*. Paderborn: Ferdinand Schöningh, 2018.
- Kaelble, Hartmut. „Historischer Vergleich.“ *Docupedia-Zeitgeschichte* 2012 (2012): 1–12. doi: 10.14765/zzf.dok.2.271.v1.
- Kasper, Dominik und Andreas Kuczera. „Modellierung von Zweifel – Vorbild TEI im Graphen.“ In *Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten*. Hg. v. Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. Wolfenbüttel: Herzog August Bibliothek, 2019. doi: 10.17175/sb004_003.
- Kocka, Jürgen und Claus Offe (Hg.). *Geschichte und Zukunft der Arbeit*. Frankfurt a. M.: Campus-Verl., 2000.
- Koselleck, Reinhart. *Vergangene Zukunft: Zur Semantik geschichtlicher Zeiten*. Frankfurt a. M.: Suhrkamp, 2010.
- Kuczera, Andreas, Thorsten Wübbena und Thomas Kollatz (Hg.). *Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheit*. Wolfenbüttel: Herzog August Bibliothek, 2019, doi: 10.17175/sb004.
- Kuczera, Andreas, Thorsten Wübbena und Thomas Kollatz. „Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Zur Einführung in diesen Band.“ In *Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten*. Hg. v. Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. Wolfenbüttel: Herzog August Bibliothek, 2019. doi: 10.17175/sb004_013.
- Landwehr, Achim (Hg.). *Frühe Neue Zeiten: Zeitwissen zwischen Reformation und Revolution*. Bielefeld: transcript, 2012.
- Lüdtke, Oliver, Alexander Robitzsch, Ulrich Trautwein und Olaf Köller. „Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen.“ *Psychologische Rundschau* 58 (2007): 103–117. doi: 10.1026/0033-3042.58.2.103.
- Mey, Günter und Katja Mruck (Hg.). *Qualitative Forschung: Analysen und Diskussionen – 10 Jahre Berliner Methodentreffen*. Wiesbaden: Springer VS, 2014.
- Moeller, Jakob und Katrin Moeller. „Pandemien und Sterblichkeitskrisen in der Geschichte Halles (1579 bis 2018).“ In *Preprints und Working Paper*, hg. v. Historischen Datenzentrum Sachsen-Anhalt, Halle 2020, doi: 10.25673/32869.
- Moeller, Katrin. „Grundwissenschaften als Masterdisziplin der Nachnutzung“, in: Eva Schlotheuber, Rüdiger Hohls und Claudia Prinz (Hg.), *Diskussionsforum: Grundwissenschaften als Masterdisziplin der Nachnutzung. H-Soz-Kult* 2015 (2015). <http://www.hsozkult.de/debate/id/diskussionen-2923>. Zugriff am 19.02.2022.
- Moeller, Katrin. „Stand oder Beruf? Arbeitsspezifische und religiöse Konzepte von Beruf und Leistung im 18. Jahrhundert.“ In *Industrielle Arbeitsbeziehungen in Mitteldeutschland: Zwischen Klassenkampf, Sozialpartnerschaft und Betriebsgemeinschaft (1815–2015)*. Hg. v. Michael Schäfer, Swen Steinberg und Veronique Töpel, 27–48, Leipzig: Leipziger Universitätsverlag, 2018.
- Moeller, Katrin und Petra Kühne. *Sterberegister der Mariengemeinde Halle (Saale) 1670–1820*. Hg. v. Historischen Datenzentrum Sachsen-Anhalt, Halle 2014. doi: 10.25673/14112.
- Oeppen, Jim und James W. Vaupel. „Demography. Broken Limits to Life Expectancy.“ *Science* 296/5570 (2002): 1029–1031.
- Patzel-Mattern, Katja und Albrecht Franz (Hg.). *Der Faktor Zeit: Perspektiven kulturwissenschaftlicher Zeitforschung*. Stuttgart: Franz Steiner Verlag, 2015.

- Plattner, Ilse E., *Zeitbewußtsein und Lebensgeschichte: Theoretische und methodische Überlegungen zur Erfassung des Zeitbewußtseins*. Heidelberg: Asanger, 1990.
- Rahlf, Thomas. „Voraussetzungen für eine Historische Statistik von Deutschland (19./20. Jahrhundert).“ *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte* 101/3 (2014): 322–352.
- Rahlf, Thomas (Hg.). *Deutschland in Daten: Zeitreihen zur Historischen Statistik*. Bonn: bpb Bundeszentrale für Politische Bildung, 2015.
- Rahlf, Thomas. „Einleitung.“ In *Deutschland in Daten: Zeitreihen zur Historischen Statistik*, hg. v. Thomas Rahlf, 5–12. Bonn: bpb Bundeszentrale für Politische Bildung, 2015.
- Raspe, Martin und Georg Schelbert, „Genau, wahrscheinlich, eher nicht: Beziehungsprobleme in einem kunsthistorischen Wissensgraph.“ In *Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten*. Hg. v. Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. Wolfenbüttel: Herzog August Bibliothek, 2019. doi: 10.17175/sb004_012
- Rehbein, Malte. „Digitalisierung braucht Historiker/innen, die sie beherrschen, nicht beherrscht.“ <https://www.hsozkult.de/debate/id/diskussionen-2905>. Zugriff am 19.02.2022.
- Schäfer, Michael, Swen Steinberg und Veronique Töpel (Hg.). *Industrielle Arbeitsbeziehungen in Mitteldeutschland: Zwischen Klassenkampf, Sozialpartnerschaft und Betriebsgemeinschaft (1815–2015)*. Leipziger Universitätsverlag, 2018.
- Schendera, Christian F. G. *Datenqualität mit SPSS*. München: Oldenbourg, 2011.
- Schenkel, Elmar und Kati Voigt (Hg.). *Verweile doch ...: Über die Erforschung der Zeit*. Leipzig: Edition Hamouda, 2015.
- Schich, Maximilian, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási und Dirk Helbing. „Quantitative Social Science. A Network Framework of Cultural History.“ *Science* 345/6196 (2014): 558–562.
- Sorokin, Pitirim A. und Robert K. Merton. „Social Time: A Methodological and Functional Analysis.“ *The American Journal of Sociology* 42/5 (1937): 615–629. <https://www.jstor.org/stable/2767758>. Zugriff am 19.02.2022.
- Steiner, André. „Die DDR-Statistik: Probleme und Besonderheiten.“ In *Deutschland in Daten: Zeitreihen zur Historischen Statistik*, hg. v. Thomas Rahlf, 13–17. Bonn: bpb Bundeszentrale für Politische Bildung, 2015.
- Wagner, Andreas. „Ambiguität und Unsicherheit: Drei Ebenen eines Datenmodells.“ In *Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten*. Hg. v. Andreas Kuczera, Thorsten Wübbena und Thomas Kollatz. Wolfenbüttel: Herzog August Bibliothek, 2019. doi: 10.17175/sb004_007.
- Whitrow, Gerald James. *Time in History: Views of Time from Prehistory to the Present Day*. Oxford: Oxford Univ. Press, 1991.

Leif Scheuermann

Über die Rolle computerbasierter Modellrechnungen und Simulationen für eine digitale Geschichte

Abstract: The following article addresses the significance of simulations and model calculations for a digital history. In a first section, the term digital history is questioned in order to arrive at a general definition of digital history as an algorithmically composed approach to the past and algorithmic modeling of systems for the description and exploration of historical processes and states. A second section includes a definition of simulation and model computation as virtual attempts to recreate processes or states with the goal of modeling and testing complex systems. Both areas (digital history and simulation/model computation) are brought together in a concluding third section, with specific attention to the source value and limitations of simulations and model computations of historical processes and states.

Keywords: simulation, modelling, model calculation

Zusammenfassung: Der folgende Beitrag thematisiert die Bedeutung von Simulationen und Modellrechnungen für eine digitale Geschichte. Hierfür wird in einem ersten Abschnitt der Begriff Digital History hinterfragt, um letztlich zu einer allgemeinen Definition der digitalen Geschichte als algorithmisch verfasster Annäherung an die Vergangenheit und algorithmische Modellierung von Systemen zur Beschreibung und Exploration historischer Abläufe und Zustände zu gelangen. Ein zweiter Abschnitt beinhaltet eine Begriffsbestimmung von Simulation und Modellrechnung als virtuelle Versuche zur Nachbildung von Prozessen oder Zuständen mit der Zielsetzung der Modellierung und Erprobung komplexer Systeme. Beide Bereiche (Digital History und Simulation/Modellrechnung) werden in einem abschließenden dritten Abschnitt zusammengeführt, wobei ein spezifisches Augenmerk auf dem Quellenwert und den Grenzen von Simulationen und Modellrechnungen historischer Prozesse und Zustände gelegt wird.

Schlagwörter: Simulation, Modellierung, Modellrechnung

1 Digitale Geschichte – eine Begriffsbestimmung

Der Begriff Digital History erlangt nicht erst in Zeiten einer durch den grassierenden Corona-Virus bedingten räumlichen Einschränkung unserer Gesellschaft zunehmend an Bedeutung im internationalen Wissenschaftsbetrieb. Dabei ist die Beschäftigung der Historiker:innen mit den Neuen Medien ebenso wenig neu,¹ wie das Aufzeigen der Probleme und Risiken einer nicht hinterfragten Nutzung.²

Mit dem Aufkommen der ‚Digital Humanities‘ im deutschsprachigen Raum in den 2010er Jahren verschob sich allerdings der Fokus der Aufmerksamkeit zu Gunsten der Computerphilologie und Computerlinguistik, so dass die Übernahme der englischen Benennung und der Anfangsbuchstaben (welche regelmäßig zu Verwirrungen führt) als Gegenbewegung zum Anspruch einer generalisierten, die gesamten Geisteswissenschaften umfassenden Herangehensweise zu verstehen ist.³

Betrachtet man den Begriff der Digital History unabhängig von seiner historischen Genese, so stellt sich die Frage, was unter dem selbst im Alltagsleben omnipräsenten „digital“ zu verstehen ist. Bei einer näheren Betrachtung zeigt sich, dass es schwer fällt, den allgemein akzeptierten Begriff klar zu definieren.⁴

1 Hier sind an erster Stelle die Arbeiten von Manfred Thaller seit dem Ende der 1970er Jahre und damit eng verbunden die Forschungsgruppe am Max-Planck-Institut für Geschichte in Göttingen zu nennen. Zudem sei daran erinnert, dass bereits seit 1993 die Arbeitsgemeinschaft EDV und Geschichte (AGE) unter anderem mit einer Publikationsreihe sowie jährlichen Tagungen sich für die „Förderung des EDV-Einsatzes als Mittel der Forschung und Lehre in den Geschichtswissenschaften einsetzt“ (Homepage Arbeitsgemeinschaft Geschichte und EDV e.V., „Homepage“. <https://www.age-net.de/>. Zugriff am 26.04.2021. Siehe hierzu: Arbeitsgemeinschaft Geschichte und EDV, *20 Jahre Arbeitsgemeinschaft Geschichte und EDV*, hg. v. Jörn Kobes, Kai Ruffing und Wolfgang Spickermann (Gutenberg: Computus-Druck Satz & Verlag 2013); Patrick Reinard, „Zur Wirkungsgeschichte der AGE und der Reihe ‚Abhandlungen der Arbeitsgemeinschaft Geschichte und EDV‘ (AAGE)“, in *Funktion und Aufgabe digitaler Medien in Geschichtswissenschaft und Geschichtsunterricht*, hg. v. Krešimir Matijević, (Gutenberg: Computus Druck Satz & Verlag, 2020), 11–15.

2 In einem bis heute zentralen Aufsatz mit dem Titel „Die Geschichte: Wissenschaft und Fiktion“ zeigte Michel de Certeau (Reprint 2019, 61–95, bes. 72–82) bereits in den 1980er Jahren die zentralen Problematiken einer sich durch die vermeintliche Faktizität moderner Datenverarbeitung legitimierenden Historiographie.

3 Zur Frage einer Abgrenzung der Digitalen Geisteswissenschaften jenseits einer auf administrativen Grenzen begründeten Bestimmung des Faches siehe: Leif Scheuermann, „Die Abgrenzung der digitalen Geisteswissenschaften.“ *Digital Classics* 2/1 (2016). doi: 10.11588/dco.2016.1.22746.

4 So findet man im Duden beispielsweise unter Digital neben einem Verweis auf die Handchirurgie die selbstbezügliche Bestimmung „Digitalverfahren betreffend“. Dudenredaktion, „Digi-

Um dieses Phänomen der Vagheit des Begriffs zu fassen, hilft es, Ernesto Laclaus Überlegungen zur Rolle „Leerer Signifikanten“ heranzuziehen.⁵ Der argentinische Poststrukturalist sieht in dem Vakuum, das ein Bezeichner ohne bezeichneten Inhalt schafft, einen Leerraum, der für die Vereinigung unterschiedlichster oft auch divergenter Interessen mit hegemonialem Anspruch prädestiniert ist. Zudem, so Laclau, führt die Leere als Gegensatz zu allem Bestehenden dazu, dass das Bezeichnete etwas gänzlich Anderes, radikal Neues darstellen muss, für welches es sich lohnt, zu kämpfen. Der leere Signifikant wird so zum Politikum und zum Kampfbegriff gegen das Bestehende. Exakt dies zeigt sich in der inflationären Benutzung des Begriffs „digital“ in den öffentlichen Diskursen generell und speziell im Ringen der Historiker:innen um die begrenzten Drittmittelgelder.

Doch so erhellend und einleuchtend diese Erklärung des Phänomens des Digitalen erscheint, so wenig hilft sie für eine inhaltliche Bestimmung des Begriffs einer digitalen Geschichte. Es gilt also, einen anderen Weg einzuschlagen, weshalb im Folgenden eine Definition von „digital“ entgegengesetzt werden soll, die mit Max Eyth⁶ und in Nachfolge Georg Simmel⁷ folgendermaßen formuliert werden kann: Digital ist alles, was dem menschlichen Willen eine algorithmische Form gibt.⁸

Der Kern des Digitalen, so der hier vorgestellte Ansatz, ist der Algorithmus im Sinne Alan Turings als vollständige und eindeutige Handlungsanweisung zur Lösung eines Problems in endlich vielen Schritten. Aus dieser Definition ergeben sich nun weitere Bestimmungen, die gerade im Hinblick auf eine digitale Geschichte zentral sind. Ein Algorithmus beschreibt immer einen Prozess. Er ist ephemeral – eine Anleitung, ein Ablauf, aber kein Ergebnis oder Gegenstand. Er muss vollständig und eindeutig definiert sowie endlich sein. Zudem müssen die

tal auf Duden online“. <https://www.duden.de/node/134370/revision/134406>. Zugriff am 26.04.2021.

5 Ernesto Laclau, *Emanzipation und Differenz*, 2., unveränd. Aufl., Nachdr (Wien: Turia + Kant, 2007).

6 Max Eyth, „Poesie und Technik,“ in *Lebendige Kräfte*, hg. v. Max Eyth (Berlin, Heidelberg: Springer Berlin Heidelberg, 1908), 1.

7 Ernst Cassirer, „Form und Technik (1930),“ in *Aufsätze und kleine Schriften (1927–1931)*, hg. v. Tobias Berben, (Hamburg: Meiner, 2004), 139–183.

8 Eyth fährt fort: „Und da das menschliche Willen mit dem menschlichen Geist fast zusammenfällt, und dieser eine Unendlichkeit von Lebensäußerungen und Lebensmöglichkeiten einschließt, so hat auch die Technik trotz seines Gebundenseins an die stoffliche Welt, etwas von der Grenzenlosigkeit des reinen Geisteslebens überkommen“ (Eyth, *Poesie und Technik*, 1–2.). Überträgt man dies auf das Prozedurale und damit Unstoffliche des Algorithmus, so werden die überspitzten Träume und inhärenten Allmachtsphantasien, die mit einer vermeintlichen Digitalisierung der Welt einhergehen, umso verständlicher.

gleichen Bedingungen stets die gleichen Ergebnisse zeitigen und damit korrelierend, müssen zum Zeitpunkt ihrer Ausführung die einzelnen Handlungsschritte eindeutig definiert sein.⁹

Was bedeuten diese Konkretisierungen nun für die Möglichkeit einer digitalen Geschichte? Geht man nicht von einer eschatologischen Geschichtsphilosophie aus, ist eindeutig ersichtlich, dass der Ansatz, die Vergangenheit selbst als Algorithmus zu fassen oder zu verstehen, abzulehnen ist, da weder das System eindeutig bestimmt werden kann, noch Anforderungen der Determiniertheit, des Determinismus oder der Endlichkeit erfüllt sind. Dementsprechend ist es wenig verwunderlich, dass große Teile der bestehenden EDV-Nutzung in den Historischen Wissenschaften in der digitalen Zurverfügungstellung von Quellen oder Abhandlungen, der Erstellung digitaler Werkzeuge für die Analyse und nicht zuletzt in der digitalen Vermittlung für eine breite Öffentlichkeit bestehen. Dies ist zweifelsohne ein zentraler Bestandteil der Hilfswissenschaft „digitale Geschichte“ und wird sicher auch weiterhin ein wesentlicher Bestandteil der Teildisziplin bleiben. Doch ist es nicht zu verleugnen, dass es sich hier um eine „Digital aided History“ – also eine digital unterstützte Geschichte – handelt.¹⁰

Betrachtet man hingegen Geschichte als hermeneutische Annäherung an die Vergangenheit, als Modell und Übersetzung in die Gegenwart, so ist folgerichtig digitale Geschichte als algorithmische Annäherung an die Vergangenheit zu verstehen, als die algorithmische Modellierung von Systemen zur Beschreibung und Exploration historischer Abläufe und Zustände. Genau dies ist der Inhalt von Simulationen und Modellrechnungen.

2 Simulation und Modellrechnung

Spätestens mit der rasanten Zunahme der Leistungsfähigkeit von Prozessoren in den letzten 30 Jahren ist die computerbasierte Simulation komplexer Systeme in den Natur- und Technikwissenschaften zu einem zentralen Element der Forschung geworden.¹¹ Das Spektrum reicht hierbei von medizinischen Simulationen (zum Beispiel der Ausbreitung von Pandemien oder des Ablaufs komplizierter Operationen) über die Kerntechnik bis hin zur Meteorologie oder der Ver-

⁹ Einführend Thomas H. Cormen und Paul Molitor, *Algorithmen – eine Einführung*, 3., überarb. und erw. Aufl. (München: Oldenbourg, 2010).

¹⁰ Scheuermann, „Abgrenzung“.

¹¹ Siehe hierzu unter anderem Wissenschaftsrat, „Bedeutung und Weiterentwicklung von Simulation in der Wissenschaft.“ (2014), Positionspapier (Drs. 4032-14), Juli 2014 <https://www.wissenschaftsrat.de/download/archiv/4032-14.html>, Zugriff am 20.02.2022.

kehrstechnik. In den Geschichtswissenschaften hingegen sind Simulationen bis dato nur in Ansätzen vertreten.¹² Ein Grund mag sein, dass die Kenntnis vom Nutzen, aber auch von den Grenzen des Ansatzes in Historiker:innenkreisen noch unterrepräsentiert ist.¹³ Um hier Abhilfe zu schaffen, erscheint eine klare Definition der Thematik notwendig:

Eine Simulation ist ein im naturwissenschaftlichen Paradigma verfasstes virtuelles, reproduzierbares Experiment zur Nachbildung eines Prozesses oder Zustands. Zielsetzung ist die Modellierung und Erprobung komplexer Systeme für die Gewinnung eines besseren Verständnisses von Abläufen und Wertigkeiten einzelner Faktoren sowie möglichst realitätsnaher Vorhersagen bzw. starker Eingrenzungen der naturwissenschaftlich möglichen Zukünfte (Möglichkeitsraum).¹⁴

Bei Simulationen handelt es sich also um virtuelle naturwissenschaftliche Experimente, was wiederum eine klare Hypothese und eine wohldefinierte Versuchsanordnung impliziert. Simulationen kommen zum Einsatz, wenn ein realweltliches Experiment zu teuer oder gefährlich ist, der Untersuchungsgegenstand

12 Für die Alte Geschichte sind hier besonders die Arbeiten von Peter Scholz (Stuttgart) im Kontext des interdisziplinären Forscherverbunds „Reden ohne Mikrophon“, <https://www.hi.uni-stuttgart.de/ag/forschung/rom/> Zugriff am 11.01.2021 zu nennen sowie die Forschungen des Teams um Christoph Schäfer (Trier) zu Schifffahrt in der Antike. Einführend: Rudolf Aßkamp und Christoph Schäfer, „Projekt Römerschiff: Nachbau und Erprobung für die Ausstellung ‚Imperium, Konflikt, Mythos – 2000 Jahre Varusschlacht,‘“ in *2000 Jahre Varusschlacht. Imperium – Konflikt – Mythos*, hg. v. LWL Römermuseum, Landesverband Lippe und VARUSSCHLACHT im Osnabrücker Land GmbH – Museum und Park Kalkriese (Stuttgart: Theiss, 2009); Christoph Schäfer, „Experimentelle Archäologie trifft auf Schifffahrt: Ein römischer Prähm im Test.“ *Antike Welt* 5 (2017): 76–83. Besonders zu erwähnen ist hierbei die herausragende Dissertation von Pascal Warnking, in welcher er den Nutzen moderner Regatta-Navigationssoftware für die Alte Geschichte nachweist [Pascal Warnking, „Roman Trade Routes in the Mediterranean Sea: Modelling the routes and duration of ancient travel with modern offshore regatta software,“ in *Connecting the Ancient World: Mediterranean Shipping, Maritime Networks and their Impact*, hg. von Christoph Schäfer. 1. Auflage (Rahden/Westf.: Verl. Marie Leidorf, 2016), 40–90]. Für eine detailliertere Untersuchung bestehender Ansätze siehe: Leif Scheuermann (Hg.), *Simulation von Handel und Verkehr in Kriegs- und Friedenszeiten der Antike* (Heidelberg: Propyläeum, 2020).

13 So kritisiert Annette Vowinckel bereits 2009 die weit verbreitete, doch unzutreffende Ansicht, dass Geschichte und Simulation sich ausschließen würden: „History seems to be the only academic discipline which successfully avoids to discuss the impact of simulation on processes of gaining knowledge, even if at the same time non-scientific simulations of history – especially in computer games – have become very popular.“ [Annette Vowinckel, „Past Futures: From Reenactment to the Simulation of History in Computer Games.“ *Historical Research/Historische Sozialforschung* 34/2 (2009): 322–332]. Leider hat sich an dieser Aussage bis dato wenig geändert.

14 Scheuermann, *Simulation*, 9.

noch nicht oder nicht mehr existiert, der Untersuchungsgegenstand zu groß oder zu klein ist, Prozesse zu schnell oder zu langsam ablaufen oder die Strukturen des betrachteten Systems zu komplex oder noch nicht verstanden sind, wobei hier oft ein spezifisches Augenmerk auf der Exploration der Zusammenhänge verschiedener Parameter liegt. Ziel ist es dabei, den Einfluss einer begrenzten Menge an einzelnen das Gesamtsystem hinreichend abbildenden Faktoren zu isolieren und – nach Etablierung des Systems – Vorhersagen unter Annahme des naturwissenschaftlichen Paradigmas der Reproduzierbarkeit treffen zu können. Dies heißt letztlich nicht weniger, als dass ein Experiment mit gleichen Voraussetzungen stets gleiche Ergebnisse erzielt. Dabei bilden begleitende realweltliche Experimente die Basis, auf der das System validiert und schrittweise hin zu einer realweltlichen Evidenz adaptiert wird.

Zum Aufbau eines Simulationssystems (Abbildung 1) bedarf es einer Fragestellung, anhand welcher ein abstraktes Modell eines Ausschnitts der Realität erarbeitet wird. Dieses definiert die Eingangsdaten und die Algorithmen, welche in einem weiteren Schritt numerisch ausformuliert werden. Programm und Eingangsdaten werden in der Simulationsrechnung zusammengeführt und ergeben erste Simulationsergebnisse, welche (wie bereits gesagt) mit realweltlichen Benchmarkdaten abgeglichen werden. Diese Validierung führt zu einer Verfeinerung der Modellierung, bis eine hinreichende Deckungsgleichheit erreicht ist.

Am Schluss des Prozesses steht ein System, das den Anspruch erhebt, das Phänomen weitestgehend abbilden zu können, sodass es auf andere Szenarien angewandt werden kann.

Im Gegensatz zur Simulation sollen unter dem Begriff „Modellrechnung“ alle algorithmisch verfassten Modelle subsumiert werden, die nicht den strengen Richtlinien eines naturwissenschaftlichen Experiments unterliegen, da sie vom menschlichen Handeln abhängig sind oder dieses zu beschreiben versuchen. Der Ablauf einer Modellrechnung entspricht dabei weitestgehend dem der Simulation. Als zeitgenössisches Beispiel sind hier Wahlprognosen zu nennen. Diese basieren auf fundierten und statistisch abgesicherten Befragungen und bilden meist die Stimmungslage vor einer Wahl mit sehr hoher Genauigkeit ab. Als Benchmarkdaten zur Validierung der Rechnungen fungieren die erzielten Wahlergebnisse, wobei die Unschärfen in diesem Prozess weitaus größer sind als bei Simulationen. Grund hierfür ist, dass die tatsächlichen Entscheidungen der einzelnen Wähler:innen nicht durch Modellrechnungen ermittelt werden können, da diese von einer Vielzahl unterschiedlichster, oft nicht klar festzumachender Faktoren und schließlich der spontanen freien Entscheidung der Akteurin oder des Akteurs abhängen. Dennoch liegt der Wert einer solchen Prognose wie auch der algorithmischen Annäherung an den Entscheidungspro-

zess auf der Hand, ohne dass es sich im strengen Sinne um eine Simulation handelt.

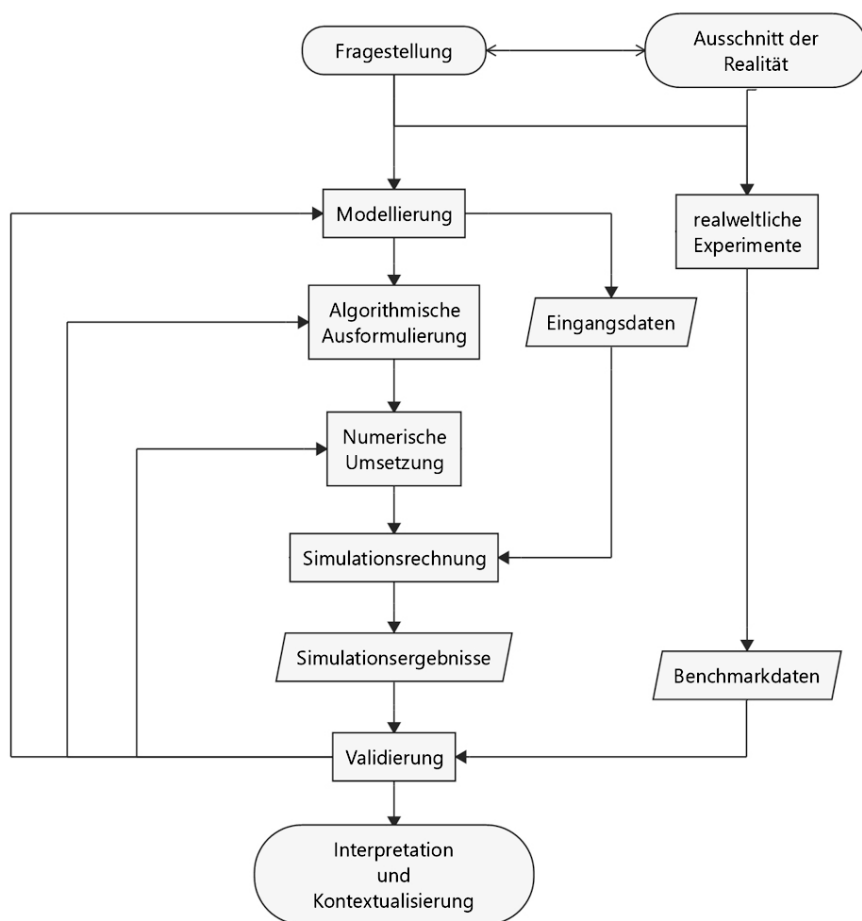


Abb. 1: Diagrammatische Darstellung des Aufbaus eines Simulationssystems. Abbildung von Leif Scheuermann.

3 Simulationen und Modellrechnungen als inhärenter Bestandteil einer digitalen Geschichte für die Historischen Wissenschaften?

Die ersten beiden Abschnitte dieses Beitrags beinhalteten Bestimmungen der Begriffe Digital History sowie Simulation und Modellrechnung. Es konnte bis hier gezeigt werden, dass Modellrechnungen und Simulationen einen inhärenten Bestandteil einer digitalen Geschichte ausmachen. In einem letzten Schritt werden im Folgenden nun beide Bereiche zusammengeführt, um Chancen und Grenzen eines solchen Vorgehens aufzuzeigen.

Die Sinnhaftigkeit von Modellrechnungen für die Historischen Wissenschaften erscheint bereits auf den ersten Blick als evident. Es können historische Modelle in algorithmischer Form formuliert, virtuell erprobt und anhand historischer Quellen validiert werden. So kann das Verständnis vom Einfluss der einzelnen Faktoren verfeinert und nicht haltbare Theorien verworfen werden. Idealtypische Einsatzbereiche sind hier die Wirtschaftsgeschichte, aber auch die historische Demographie. Weitaus problematischer ist hingegen der Versuch, nicht existente historische Quellen (zum Beispiel Geburtenraten oder die nicht dokumentierte Nutzung von Verkehrswegen) anhand von Modellrechnungen zu erfassen.

Dies wiederum ist im Kontext der Simulation ohne weiteres möglich. Ist beispielsweise einmal die Heizleistung eines antiken Ofens thermodynamisch durch eine Simulation ermittelt worden, so kann das Ergebnis problemlos in die Vergangenheit übertragen werden. Von daher erscheint das Anwendungsgebiet naturwissenschaftlicher Simulationen weitaus begrenzter. Doch gibt es historische Prozesse, die mit naturwissenschaftlichen Gesetzen zu fassen sind – zum Beispiel in der Technik- und Wissenschaftsgeschichte oder der Historischen Geographie. Das Ziel der Simulation liegt dabei anders als bei der Modellrechnung nicht in der Vorhersage menschlicher Handlungen, sondern vielmehr in der Rekonstruktion der naturwissenschaftlich erfassten Umwelt der historischen Akteur:innen. Die Reaktion auf oder der Umgang mit dieser Umwelt entzieht sich hingegen der Simulation. Hinzu kommt, dass Simulationen (wie alle naturwissenschaftlichen Verfahren) einen falsifizierenden Charakter besitzen.¹⁵ Es können also Zustände und Prozesse ausgeschlossen und so Möglichkeitsräu-

¹⁵ Karl R. Popper und Herbert Keuth, *Logik der Forschung*, 4., bearb. Aufl. (Berlin: Akad.-Verl., 2013).

me eröffnet werden. Positive Aussagen, wie sie sich aus traditionellen historischen Quellen erschließen lassen, sind jedoch nicht möglich. Um das gerade angeführte Beispiel wieder aufzugreifen: Die Heizleistung eines Ofens kann berechnet werden. Ob, wie häufig und auf welche Weise ein spezifischer Ofen dieser Bauart jedoch wirklich genutzt wurde, lässt sich ebenso wenig ermitteln, wie die Frage, ob einzelne historische Akteur:innen gefroren haben.

Es gilt also festzuhalten, dass Simulationen historisch-naturwissenschaftliche Möglichkeitsräume eröffnen, in denen Geschichte stattfinden kann, wohingegen Modellrechnungen erlauben, historische Modelle zu falsifizieren beziehungsweise die einzelnen Parameter des Systems zu validieren.

Für Simulationen wie Modellrechnungen gilt, dass sie prognostisch, also auf eine Zukunft ausgerichtet sind, wobei sich der Raum des Möglichen mit zunehmender Distanz von der Gegenwart weitet.¹⁶ Ob dieser Ausgangspunkt dabei in der Vergangenheit, der Gegenwart oder der Zukunft liegt, beeinflusst beide selbst nicht. Allein die Eingangsdaten sind von der zeitlichen Verortung der Simulation abhängig. Je exakter die historischen Umstände ermittelbar sind, desto besser sind die Eingangsdaten und damit auch die historisch-prognostischen Ergebnisse.

Im Gegensatz zu Simulation und Modellrechnung ist Historie immer von Heute auf die Vergangenheit hin ausgerichtet, wobei sich auch hier mit zunehmender Zeit der Raum des Möglichen weitet. Die historischen Quellen geben uns nun mit ihren individuellen Aussagen Wegmarken, aus denen wir im Rahmen des Möglichkeitsraums Narrative (re-)konstruieren können, die wiederum in eine vergangene Zukunft gerichtet sind. Ähnlich verhält es sich bei historischen Simulationen und Modellrechnungen. Diese benötigen einen Startpunkt, der von unseren Modellen abhängig ist. Von hier aus können sie in ihren jeweiligen Grenzen wiederum die Möglichkeitsräume bestimmen, wobei die historischen Quellen für eine Falsifikation der Algorithmen genutzt werden können. Man kann also von einer historisch prognostischen Vorgehensweise sprechen, welcher eine doppelte Unschärfe innewohnt, die des Ausgangspunktes und die der Simulation beziehungsweise im verstärkten Maße der Modellrechnung.

Doch wie ist nun das Verhältnis zu den traditionellen historischen Quellen? Betrachten wir dies beispielhaft anhand eines generischen Berichts über eine Flussschiffahrt und einer Simulation derselben. Im einfachsten Fall stimmen

16 Hierfür ein kurzes Beispiel. Meteorologische Simulationen erlauben uns für den heutigen Tag sehr klare und abgegrenzte Aussagen über das lokale Wetter. Betrachten wir den Bericht für eine Woche, so sind die Ergebnisse immer nicht mehr so exakt, wie bei der Tagesvorhersage. Eine Simulation des Wetters (nicht des Klimas) in 50 Jahren jedoch wird zumindest in absehbarer Zeit über eine generelle Aussage nicht hinausgehen.

die Daten beider Quellen überein. Dann ist die Simulation valide und der Reisebericht nennt zumindest realistische Zahlen, ohne dass jedoch von einem „Beweis“ oder gar der „Wahrheit“ gesprochen werden kann. Sind die im Reisebericht genannten Zeiten länger, so ist – vorausgesetzt, die Simulation ist nicht noch optimierungsbedürftig – anzunehmen, dass längere Pausen (mit Anlandungen) gemacht wurden, was wiederum historische Fragen zur Organisation der Schifffahrt eröffnet. Im schwierigeren Fall sind die im Reisebericht genannten Zahlen geringer als die errechneten. Dann gilt es zu hinterfragen, weshalb die/der Autor:in der Quelle (bewusst oder unbewusst) angibt, die Strecke jenseits des physikalisch Möglichen zurückgelegt zu haben. Hier ergeben sich eine ganze Reihe an weiteren Fragen, zum Beispiel nach dem Zweck der Quelle, dem Wissensstand der Autorin oder des Autors oder der Historizität der Reise.

Am Beispiel des Vergleichs eines Berichtes und einer Simulation einer geneesischen Flussreise zeigt sich der Nutzen von Simulationen in der Inbezugsetzung mit traditionellen Quellen. Nochmals hervorzuheben ist dabei, dass die Simulation lediglich einen naturwissenschaftlichen Kontext eröffnet, in welchem die individuell verifizierende historische Quelle verortet werden kann. Dabei ist das Aufzeigen von Widersprüchen und damit das Aufwerfen neuer Fragen die zentrale Aufgabe der Simulation. Die Beantwortung der Fragen hingegen liegt im Zuständigkeitsbereich der Historiker:innen. Ähnlich verhält es sich im größeren Kontext. Simulationen sind prädestiniert dazu, größere und längerfristige Entwicklungen wie beispielsweise den Klimawandel nachzuvollziehen und auf einer Makroebene Möglichkeiten und Unmöglichkeiten menschlichen Handelns zu skizzieren. Hier besitzen Simulationen auch den Vorteil, dass mehrere potentielle Entwicklungen parallel zueinander berechnet werden können, was in begrenztem Rahmen „Was-wäre-wenn“-Fragen erlaubt.¹⁷

4 Fazit

Zwei Begriffsdefinitionen standen im Fokus dieses Beitrages – zum einen die der Digital History als algorithmische Annäherung an die Vergangenheit und zum anderen die von historischer Simulation und Modellrechnung als Nachbildung eines Prozesses oder Zustands mit der Zielsetzung der Modellierung und Erprobung historischer Systeme für die Gewinnung eines besseren Verständnisses von Abläufen und Wertigkeiten einzelner Faktoren sowie der Eingrenzung

¹⁷ Zum Beispiel: Wie hätte sich die historische Landschaft entwickelt, wenn es fünf Grad kälter gewesen wäre?

gen vergangener Zukünfte. Als solche sind historische Simulationen und Modellrechnungen ein inhärenter Bestandteil einer im Entstehen begriffenen historischen Teildisziplin Digital History. Für eine breite Akzeptanz in der Fachcommunity wird dabei zukünftig entscheidend sein, dass nicht der Anspruch erhoben wird, zu berechnen ‚wie es gewesen ist‘ und sich nicht in romantizistischen Wunschträumen einer virtuellen Zeitmaschine zu verstricken, sondern klar die Grenzen und Möglichkeiten solcher Ansätze für die Geschichtswissenschaften und die in der Interpretation der Ergebnisse benötigten hermeneutischen Arbeitsweise aufzuzeigen.

Bibliographie

- Aßkamp, Rudolf und Christoph Schäfer, „Projekt Römerschiff: Nachbau und Erprobung für die Ausstellung ‚Imperium – Konflikt – Mythos – 2000 Jahre Varusschlacht‘.“ In *2000 Jahre Varusschlacht. Imperium – Konflikt – Mythos*, hg. v. LWL Rörmuseum, Landesverband Lippe und VARUSSCHLACHT im Osnabrücker Land GmbH – Museum und Park Kalkriese. Stuttgart: Theiss, 2009.
- Certeau, Michel de, Andreas Mayer und Luce Giard, *Theoretische Fiktionen: Geschichte und Psychoanalyse*, 2., erweiterte Auflage. Turia Reprint. Berlin: Turia + Kant, 2019.
- Geschichte und EDV, Arbeitsgemeinschaft, *20 Jahre Arbeitsgemeinschaft Geschichte und EDV*, hg. v. Jörn Kobes, Kai Ruffing und Wolfgang Spickermann. Gutenberg: Computus-Druck, Satz & Verlag 2013.
- LWL Rörmuseum, Landesverband Lippe und VARUSSCHLACHT im Osnabrücker Land GmbH – Museum und Park Kalkriese (Hg.). *2000 Jahre Varusschlacht. Imperium – Konflikt – Mythos*. Stuttgart: Theiss, 2009.
- Matijević, Krešimir (Hg.). *Funktion und Aufgabe digitaler Medien in Geschichtswissenschaft und Geschichtsunterricht*. Gutenberg: Computus Druck-Satz & Verlag, 2020.
- Popper, Karl R. und Herbert Keuth, *Logik der Forschung*. Unter Mitarbeit von Herbert Keuth, 4., bearb. Aufl. Berlin: Akad.-Verl., 2013.
- Reinard, Patrick, „Zur Wirkungsgeschichte der AGE und der Reihe ‚Abhandlungen der Arbeitsgemeinschaft Geschichte und EDV‘ (AAGE).“ In *Funktion und Aufgabe digitaler Medien in Geschichtswissenschaft und Geschichtsunterricht*, hg. v. Krešimir Matijević, 11–15. Gutenberg: Computus Druck Satz & Verlag, 2020.
- Schäfer, Christoph (Hg.), *Connecting the Ancient World: Mediterranean Shipping, Maritime Networks and their Impact*, 1. Auflage. Rahden/Westf.: Verl. Marie Leidorf, 2016.
- Schäfer, Christoph. „Experimentelle Archäologie trifft auf Schifffahrt: Ein römischer Prahm im Test.“ *Antike Welt* 5 (2017): 76–83.
- Scheuermann, Leif, „Die Abgrenzung der digitalen Geisteswissenschaften.“ *Digital Classics* 2/1 (2016). doi: 10.11588/dco.2016.1.22746
- Scheuermann, Leif (Hg.), *Simulation von Handel und Verkehr in Kriegs- und Friedenszeiten der Antike*. Heidelberg: Propylaeum, 2020.
- Vowinckel, Annette, „Past Futures: From Reenactment to the Simulation of History in Computer Games.“ *Historical Research / Historische Sozialforschung* 34/2 (2009): 322–332.

Warnking, Pascal, „Roman Trade Routes in the Mediterranean Sea: Modelling the routes and duration of ancient travel with modern offshore regatta software.“ In *Connecting the Ancient World: Mediterranean Shipping, Maritime Networks and their Impact*, hg. v. Christoph Schäfer. 1. Auflage, 40–90. Rahden/Westf.: Verl. Marie Leidorf, 2016.

Wissenschaftsrat, „Bedeutung und Weiterentwicklung von *Simulation in der Wissenschaft*.“ Positionspapier (Drs. 4032-14), Juli 2014, <https://www.wissenschaftsrat.de/download/archiv/4032-14.html>. Zugriff am 20.02.2022.

Christopher Banditt

Die Rekonstruktion der materiellen Ausstattung von Haushalten in Ostdeutschland 1980–2000

Ein Beispiel für quantitative und digitale Methoden in den Geschichtswissenschaften

Abstract: In the field of study of recent contemporary history, the application possibilities of quantification, understood as the translation of historical and social phenomena into numerical expression, have expanded immensely. On the one hand, this goes along with the easier availability of application-oriented statistical software. On the other hand, there are the broad quantitative legacies of contemporary empirical social research and official statistics, whose data collections can be used today for historical reconstruction. This paper traces the development of income disparities – as a core dimension of social inequality – in East German employee households on the basis of household surveys between 1980 and 2000. It illustrates the methodological and epistemological possibilities and limitations of working with historical survey data.

Keywords: quantification, income, social inequality, East German transformation, survey-data

Zusammenfassung: Bei der Erforschung der jüngsten Zeitgeschichte haben sich die Anwendungsmöglichkeiten für Quantifizierung, verstanden als Übersetzung von historischen und sozialen Phänomenen in zahlenmäßigen Ausdruck, immens erweitert. Auf der einen Seite geht dies mit der leichteren Verfügbarkeit anwendungsorientierter Statistiksoftware einher. Auf der anderen Seite stehen die breiten quantitativen Hinterlassenschaften der jeweils zeitgenössischen empirischen Sozialforschung und amtlichen Statistik, deren seinerzeitige Datenerhebungen heute zur historischen Rekonstruktion nutzbar sind. Im vorliegenden Beitrag wird die Entwicklung von Einkommensdisparitäten – als Kerndimension sozialer Ungleichheit – in ostdeutschen Arbeitnehmerhaushalten auf Basis von Haushaltsbefragungen zwischen 1980 und 2000 nachgezeichnet. Hieran werden exemplarisch die methodischen und erkenntnismäßigen Möglichkeiten und Grenzen bei der Arbeit mit historischen Surveydaten aufgezeigt.

Schlagwörter: Quantifizierung, Einkommen, soziale Ungleichheit, ostdeutsche Transformation, Surveydaten

1 Quantifizierung in der Geschichtswissenschaft

Quantitative Methoden sind nichts Neues in den Geschichtswissenschaften. Mitte der 1970er Jahre hielt Konrad Jarausch fest, dass es sich bei quantitativen Ansätzen „nicht um einen radikalen Bruch mit der herkömmlichen Geschichtsschreibung, sondern eher um eine Anknüpfung an eine eigene Minoritätstradition handelt, deren Anfänge im neunzehnten Jahrhundert von der politischen Ereignis- und Persönlichkeitsgeschichte überschattet wurden.“¹ Bereits 1976 stellte also quantifizierende historische Forschung kein neues Phänomen dar, wenn Jarausch in seinem Beitrag etwa auf die Historische Schule der Nationalökonomie im 19. Jahrhundert anspielt, die allerdings nie eine hervorgehobene oder gar bestimmende Stellung wie der Historismus in der Geschichtswissenschaft einnehmen sollte.² Zwar hat es mit der Konjunktur der Sozialgeschichte in den 1970er Jahren durchaus eine zwischenzeitliche Zunahme quantitativ ausgerichteter historischer Arbeiten gegeben.³ Jedoch gehören heute die „historischen Quantifizierer:innen“ wieder beziehungsweise weiterhin nicht gerade zum Mainstream in der Geschichtswissenschaft, wenngleich mit der Cliometrie/*New Economic History* eine Schule existiert, die wirtschafts- und sozialhistorische Fragestellungen mit dezidiert quantitativ-statistischen Methoden (und in der Regel theoriegeleitet) bearbeitet.

Abseits der Fragen, wie alt oder neu quantitative Methoden sind und welchen Stellenwert sie in der historischen Wissenschaft nun genau einnehmen, ist hier vor allem der semantische Gehalt des Begriffs „Quantifizierung“ von Relevanz. Dem Soziologen Steffen Mau zufolge bedeutet Quantifizierung, dass „Phänomene, Eigenschaften oder Beschaffenheiten eines Sachverhalts in einer

1 Konrad H. Jarausch, „Möglichkeiten und Probleme der Quantifizierung in der Geschichtswissenschaft,“ in *Quantifizierung in der Geschichtswissenschaft. Probleme und Möglichkeiten*, hg. v. Konrad H. Jarausch (Düsseldorf: Droste Verlag, 1976), 13.

2 Zu den unterschiedlichen Auffassungen der Schulen über den Wert historischer Statistiken siehe die prägnante und anschauliche Schilderung bei: Thomas Rahlf, „Einleitung,“ in *Deutschland in Daten. Zeitreihen zur Historischen Statistik*, hg. v. Thomas Rahlf (Bonn: Bundeszentrale für politische Bildung, 2015), 5–7.

3 Vgl. Michael Buchner et al., „Zur Konjunktur des Zählens – oder wie man Quantifizierung quantifiziert. Eine empirische Analyse der Anwendung quantitativer Methoden in der deutschen Geschichtswissenschaft,“ *Historische Zeitschrift* 310 (2020): 580–621.

allgemeinen, abstrakten und universell anschlussfähigen Sprache repräsentiert werden, der der Mathematik.“⁴ Demnach bringt Quantifizierung „eine unübersichtliche und komplexe Welt in die standardisierte Sprache der Zahlen, in welcher eindeutige Ordnungsverhältnisse von größer oder kleiner (oder von mehr oder weniger) herrschen.“⁵ Indem komplexe Realitäten in Zahlenform übersetzt werden, werden sie mess- und skalierbar und lassen sich folglich rechnerisch-statistisch beschreiben und analysieren. Auf die Handhabbarmachung von sozialer Komplexität mithilfe mathematischen Ausdrucks wird im Weiteren noch zurückzukommen sein, wenn es um die Veranschaulichung der eingesetzten Methodik in meinem Dissertationsprojekt geht.

2 Historische Daten zu ostdeutschen Haushalten

Das näher vorzustellende Projekt „Soziale Ungleichheit in Ostdeutschland 1980–2000. Die materiellen Lagen von Arbeitnehmerhaushalten im Wandel“ widmet sich folgender Themenstellung: Welche Ausprägungen hatte die materielle Ausstattung ostdeutscher Arbeitnehmerhaushalte in der letzten sozialistischen Dekade und im ersten „Nachwendejahrzehnt“? Die materielle Lage und ökonomische Situation von Haushalten wird generell als eine der wichtigsten Dimensionen von sozialer Ungleichheit verstanden.⁶ Primärindikatoren sind das verfügbare Einkommen der Haushalte, Vermögenswerte und die damit verbundenen Konsummöglichkeiten, die sich insbesondere in der Ausstattung mit langlebigen, technischen Konsumgütern niederschlagen. Allerdings geht es nicht allein um die deskriptive Abbildung von Kenngrößen und Verteilungsmaßzahlen. Da soziale Ungleichheit immer ein Produkt der sozialen Strukturen und Beziehungsgefüge ist,⁷ sind auch die Bestimmungsfaktoren im gesellschaftlichen Rahmen für das Zustandekommen und die Entwicklungen bestimmter Ungleichheitsrelationen zu analysieren.

Im Sozialismus vorherrschende Disparitäten sind insofern von besonderem Interesse, als doch dort das Marxsche Diktum galt, wonach „mit der Abschaf-

⁴ Steffen Mau, *Das metrische Wir. Über die Quantifizierung des Sozialen* (Berlin: Suhrkamp Verlag, 2017), 27.

⁵ Ibid.

⁶ Vgl. Stefan Hradil, „Soziale Ungleichheit, soziale Schichtung und Mobilität,“ in *Einführung in Hauptbegriffe der Soziologie*, hg. v. Hermann Korte und Bernhard Schäfers (Wiesbaden: Springer VS, 2016), 9. Auflage, 250.

⁷ Vgl. Stefan Hradil, *Soziale Ungleichheit in Deutschland* (Wiesbaden: Springer VS, 2005), 8. Auflage, 30.

fung der Klassenunterschiede von selbst alle aus ihnen entspringende soziale und politische Ungleichheit verschwindet.“⁸ Für den Kapitalismus ist soziale Ungleichheit als systemerhaltendes Wesensmerkmal identifiziert worden.⁹ Wird nun mit einem Untersuchungszeitraum von 1980 bis 2000 eine integrale, zäsurübergreifende Perspektive eingenommen, lässt sich das Zutreffen dieser Systemzuschreibungen ebenso untersuchen, wie die Frage, ob neben den Umbrüchen auch mögliche Kontinuitäten über 1989/90 vorherrschten. Inwieweit waren sozioökonomische Entwicklungen im Ostdeutschland der 1990er Jahre bereits durch soziale Anordnungen in der DDR historisch vorgründet? Die Beschäftigung mit vertikaler Ungleichheit wirft naturgemäß Fragen nach einem „Mehr“ und „Weniger“ zwischen sozialen Gruppen auf. Die Rekonstruktion vergangener materieller Verteilungen erfordert geradezu deren Übersetzung in mathematische Relationen, sie legt folglich quantitative Quellen und Auswertungsmethoden nahe.

Um zu historischen Statistiken beziehungsweise Datenreihen zu gelangen, gibt es grundsätzlich zwei Wege. Zum einen lassen sich qualitative Quellen auswerten und durch Klassifizierung, Kodierung und Kategorisierung in numerische Werte übertragen.¹⁰ Hierbei wären beispielsweise Oral-History-Interviews mit ostdeutschen Arbeitnehmerhaushalten zu ihren Einkommen in den 1980er und 1990er Jahren denkbar. Die so erfragten Angaben könnten dann das Basismaterial einer Zeitreihenkonstruktion bilden. Jedoch ist es nicht nur sehr unwahrscheinlich, dass bei den retrospektiv Interviewten ausreichend exakte Erinnerungen zu ihrer materiellen Ausstattung vor 20, 30 oder gar 40 Jahren vorliegen. Zugleich wäre diese Methode mit einem unmäßigen – und letztlich unrealistischen – Aufwand verbunden, um ein aussagekräftiges Sample zu generieren.

Zum anderen lassen sich aus Quellen, die bereits in quantitativer Form vorliegen, die gesuchten Variablenwerte extrahieren. Ist das Basismaterial als serielle Quelle angelegt, können die gewonnenen Angaben sogar zu Datenreihen kompiliert werden, die intertemporale Vergleiche und eine Beschreibung der Entwicklung im Zeitverlauf erlauben.¹¹ In diesem Sinne würden sich etwaige

8 Karl Marx, „Kritik des Gothaer Programms,“ in *Werke*, Bd. 19, Karl Marx und Friedrich Engels (Berlin (Ost): Dietz Verlag, 1962 [1875]), 26.

9 Vgl. Werner Plumpe, *Das kalte Herz. Kapitalismus: Die Geschichte einer andauernden Revolution* (Berlin: Rowohlt Verlag, 2019), 23.

10 Vgl. Tobias A. Jopp und Mark Spoerer, *Historische Statistik lehren. Quellenkritische Vermittlung von Zielen und Methoden statistischen Arbeitens in der Geschichtswissenschaft* (Schwalbach/Ts.: Wochenschau Verlag, 2017), 13.

11 Vgl. *ibid.*, 12–13.

jährliche Daten aus der steuerlichen Veranlagung von Einkommen und Vermögen ostdeutscher Arbeitnehmer:innen heranziehen lassen. Zu bedenken ist jedoch, dass Steuerdaten lediglich Aufschluss über steuerpflichtige Einkommen, wie das Arbeitseinkommen, geben und dass sie zumeist auf Individualebene angelegt sind, womit der Haushaltszusammenhang vernachlässigt würde.

Um Informationen über die sämtlichen Einnahmen eines Haushalts – wozu neben Arbeitseinkommen auch Sozialtransfers oder Vermögenseinkommen zählen – zu sammeln, eignen sich letztlich nur zeitgenössisch durchgeführte Haushaltsbefragungen.¹² Idealerweise haben die Haushalte dort nicht nur Auskunft über ihr Haushaltseinkommen, ihre Vermögenswerte und ihre Ausstattung mit Konsumgütern zum Befragungszeitpunkt gegeben. Mitabgefragte soziodemografische Items, wie Personenzahl, regionale Verortung sowie Alter, Bildungsgrad oder soziale Stellung der jeweiligen Haushaltsmitglieder, sind für weiterführende Analysen ebenfalls von großer Wichtigkeit.

Zugangsmöglichkeiten zu historischen Surveys bestehen zum Beispiel bei den Forschungsdatenzentren des Statistischen Bundesamts und der Statistischen Landesämter. So bietet sich die größte deutsche Haushaltsbefragung an, der jährliche amtliche Mikrozensus, der 1991 in Ostdeutschland eingeführt worden ist.¹³ Das *Statistische Bundesamt* stellt die faktisch anonymisierten Datensätze als *Scientific Use Files* in einer 70-Prozent-Sub-Stichprobe zur Verfügung; in diesen beträgt die Fallzahl ostdeutscher Arbeitnehmerhaushalte für die 1990er Jahre mehr als 20 000. In der DDR war die *Staatliche Zentralverwaltung für Statistik* eine der wenigen Institutionen, die bevölkerungsrepräsentative Befragungen durchführen durfte. Die Hinterlassenschaften ihrer regelmäßigen Einkommensstichprobe in rund 30 000 Arbeiter- und Angestelltenhaushalten aus den 1980er Jahren werden heute vom Bundesarchiv verwaltet. Allerdings liegen diese Archivalien sehr häufig nicht in maschinenlesbarer Form vor, sondern als analoge Ergebnisberichte¹⁴, was zunächst deren Übertragung in ein digitales Format erfordert(e).

¹² Vgl. Manfred Euler, „Genauigkeit von Einkommensangaben in Abhängigkeit von der Art der statistischen Erfassung. Dargestellt am Beispiel der Einkommens- und Verbrauchsstichprobe 1978,“ *Wirtschaft und Statistik* (1983): 813.

¹³ Vgl. Hans-Joachim Heidenreich, „Einführung des Mikrozensus in den neuen Bundesländern: Probleme und Erfahrungen,“ in *Amtliche Daten der DDR und der neuen Bundesländer: Informationsquelle für die Sozialwissenschaften. Tagungsdokumentation*, hg. v. Paul Lüttinger und Heike Wirth (Mannheim: ZUMA, 1993), 11–26.

¹⁴ Vgl. Bernhard Schimpl-Neimanns und Heike Wirth, *Bestandsaufnahme und Nutzungsmöglichkeiten amtlicher Mikrodaten der DDR für Sekundäranalysen zur Bildungs- und Einkommensungleichheit* (Mannheim: ZUMA, 1994), 7–9.

Während solche amtlichen Erhebungen, die vornehmlich zur Informationsbereitstellung für Politik und Öffentlichkeit eingeführt worden sind, mit relativ hohen Fallzahlen aufwarten, sind die Surveys sozialwissenschaftlicher Produzent:innen in dieser Hinsicht meistens kleiner dimensioniert. Letztere warten dafür oftmals mit einer größeren Bandbreite an abgefragten Merkmalen auf, die dann über die „harten“ sozioökonomischen Items der amtlichen Befragungen hinausgehen, indem sie etwa auch Fragen zu mentalen Settings und Perzeptionsmustern einbeziehen. Zu nennen ist in diesem Zusammenhang das Datenangebot des *GESIS – Leibniz-Instituts für Sozialwissenschaften*. Über dessen Internetpräsenz ist eine Vielzahl an Befragungen auffindbar, die hauptsächlich von sozialwissenschaftlichen Instituten durchgeführt worden sind – so auch aus der DDR-Zeit.¹⁵ Für die 1990er-Jahre kann zudem auf den Datensatz des *Sozio-oekonomischen Panels (SOEP)* zurückgegriffen werden, das 1990 – sechs Jahre nach der ersten Erhebungswelle in der alten Bundesrepublik – mit einer Stichprobengröße von mehr als 2000 Haushalten auch in Ostdeutschland etabliert worden ist.¹⁶ Im SOEP werden (mit zwischenzeitlichen Auffrischungsstichproben) immer dieselben Haushalte befragt, was Zeitverlaufsanalysen, wie zu sozialer Mobilität, sehr begünstigt.

3 Bevölkerungsbefragungen als geschichtliche Quelle

Mit den höheren Fallzahlen, die Surveys in der Regel aufweisen, geht eine größere Aussagekraft beziehungsweise Verallgemeinerbarkeit einher – je nach methodischem Design, wenn es sich beispielsweise um eine Zufallsstichprobe handelt, ließe sich auch von Repräsentativität sprechen –, als sie mit qualitativen Erhebungen, wie Interviews, gemeinhin erreicht wird. Allerdings lassen Surveydaten kein so tiefes Eindringen in die Umstände und Motivlagen der Untersuchungsobjekte zu. Ja/Nein-Abfragen oder die Erfassung von Mentalitäten und Einstellungen über graduelle Antwortskalen von „starker“ bis „geringer Zustimmung“ sind hierbei etwa gegenüber Ego-Dokumenten, wie Tagebücher oder Briefe, klar im Hintertreffen. Auf eher „schweigende Gruppen“, die selten für

¹⁵ Vgl. Michael Wettengel, „Sozialwissenschaftliche Daten aus der DDR – ein Katalog archivierter Studien“, *Historische Sozialforschung* 20 (1995): 127–134.

¹⁶ Vgl. Gert G. Wagner, „Die Längsschnittstudie Sozio-oekonomisches Panel (SOEP) – Die Jahre von der Wende zur Jahrtausendwende“, *Vierteljahrshefte zur Wirtschaftsforschung* 77 (2008): 43–62.

(Zeitzeugen-)Interviews zur Verfügung stehen oder kaum persönliche Dokumente hinterlassen, bieten Befragungsdaten wiederum potenziell größere Zugriffsmöglichkeiten. Dies umso mehr, je geringer der Antwortausfall und *Non Response Bias* einer Erhebung ist; beim Mikrozensus besteht sogar eine gesetzliche Auskunftspflicht (die wiederum freilich nicht vor bewussten Falschangaben schützen kann).

Regelmäßig durchgeführte Wiederholungsbefragungen, die durch ein gleichbleibendes Fragenprogramm charakterisiert sind, und speziell Panelerhebungen, in die darüber hinaus immer dieselben Personen und Haushalte einbezogen sind, ermöglichen konzise Zeitverlaufsanalysen. Quantitative Quellen sind für die Zeitgeschichte insbesondere aus Phasen schnellen Umbruchs überliefert. Wenn rasche Ergebnisse gefragt sind, stellen leichter aufzusetzende und durchzuführende quantitative Befragungen für Sozialwissenschaftler:innen zu meist das Instrument der Wahl vor qualitativen Erhebungen dar.¹⁷ Beobachten ließ sich das beispielsweise während der Corona-Krise 2020/21, als eine Vielzahl von Surveys den Umgang mit der Pandemie zu beschreiben suchte. Auch während des ostdeutschen Umbruchs 1989/90 kam es zu einer „explosionsartige[n] Zunahme von Wahlforschung, Meinungsumfragen, Mediaanalysen und Umfragen auf Demonstrationen“¹⁸, um den sich verändernden Lebensverhältnissen der Ostdeutschen wie auch ihren Zukunftserwartungen nachzuspüren.¹⁹

Die Überlieferungsform eines Surveys ist in aller Regel ein numerisch codierter digitaler Datensatz. Wenn eine Erhebung direkt als Online-Befragung oder als sogenanntes *Computer Assisted Personal Interview* (CAPI) durchgeführt worden ist, liegt eine genuin digitale Quelle vor. Auch aus Papierbefragungen (PAPI: *Paper and Pencil Interview*) entstehen digitale Datensätze, wenn die Rohdaten in Form ausgefüllter Fragebögen eingescannt und analoge in digitale Informationen umgewandelt werden. Dies erleichtert ihre Auswertung mittels softwarebasierter Analyseprozeduren, faktisch wird die Nutzbarmachung so überhaupt erst möglich. Bei Befragungen mit mehreren hundert oder tausend Fallzahlen ist eine händische Auswertung der zugrundeliegenden Fragebögen kaum realisierbar.

Wie der Großteil geschichtlicher Quellen sind auch Surveydatensätze originär nicht zum Zwecke der späteren historischen Rekonstruktion vergangener Realitäten entstanden. Die jeweiligen Sozial- und Wirtschaftswissenschaftler:in-

¹⁷ Vgl. Kerstin Brückweh, „Wissen über die Transformation. Wohnraum und Eigentum in der langen Geschichte der ‚Wende‘“, *Zeithistorische Forschungen* 16 (2019): 36.

¹⁸ Wettengel, „Sozialwissenschaftliche Daten“, 132.

¹⁹ Vgl. Christopher Banditt, „Einheitserwartungen. Präferenzen, Hoffnungen und Befürchtungen im ‚Wendejahr‘ 1989/90“, *Berliner Debatte Initial* 30 (2019): 67–84.

nen oder amtlichen Statistiker:innen haben mit ihnen ganz eigene und hauptsächlich gegenwartsbezogene Absichten verfolgt. Folglich sind auch an diese Daten klassische „W-Fragen“ einer traditionellen Quellenkritik zu stellen: Wer hat was, wann, wie und warum erhoben? Dies ist notwendig, um zu prüfen, ob überlieferte Datensätze aus bestimmten Haushalts- und Personensurveys grundsätzlich tauglich sind, die eigene historische Fragestellung zu beantworten.²⁰ Für die äußere Quellenkritik sind Metadaten nutzbar, die ein gutes Forschungsdatenzentrum als Host von Datensätzen zur Verfügung stellt. Zu solchen Metadaten zählen Informationen über jeweilige Erhebungsmethoden, Codebücher, damals eingesetzte Fragebögen oder Leitfäden für Interviewer:innen. Oftmals werden in Codebüchern bereits Häufigkeitsverteilungen der Variablen mit ausgegeben, die sich zu rein deskriptiven Zwecken bereits nutzen lassen.²¹ Jedoch bleibt ohne Auswertung möglicher Kontextvariablen im Datensatz ein bestimmtes Antwortverhalten in der Tat „[s]chwierig zu interpretieren.“²²

Für das hier anzuzeigende Projekt wurden über die Metadaten zunächst Erhebungen in Ostdeutschland ausfindig gemacht, die die Variable „Haushaltseinkommen“ beinhalten. Dabei wurde für die DDR-Einkommensstichprobe anhand des überlieferten Fragebogens deutlich, dass die Nettogeldeinnahmen der Haushaltsmitglieder zum einen als exakte Werte und zum anderen differenziert nach Einkommensarten (Nettolohn/-gehalt, Kindergeld, sonstige Einnahmen und so weiter) erhoben wurden. Dahingegen zeigten die Metadaten zum bundesdeutschen Mikrozensus, dass das Haushaltseinkommen ganz pauschal und klassiert in einem „Von-bis-Bereich“ abgefragt wurde. Zur Einschätzung der Verwertbarkeit von Datensätzen und der Adäquanz der zugrunde liegenden Erhebungsmethodik sind Grundkenntnisse der empirischen Sozialforschung notwendig, um nicht etwa unbeabsichtigt einem unsauber arbeitenden Befragungsinstitut aufzusitzen.²³

Für eine innere Quellenkritik sind darüber hinaus digitale Methoden im engeren Sinne notwendig, wenn es darum geht, die tatsächlichen Daten auf ihre Verwertbarkeit zu überprüfen. Bei welchen Items fallen *Missings* an, wie hoch liegt der Antwortausfall? Gibt es Gruppen, die bei bestimmten Fragen mit Ant-

²⁰ Vgl. Jopp und Spoerer, *Historische Statistik*, 14; Eva Pfanzer, „Die historische Quellenkritik und das Digitale“, *Archiv und Wirtschaft* 48 (2015): 11.

²¹ Vgl. z. B. Brückweh, „Wissen über die Transformation,“ *passim*.

²² *Ibid.*, 39.

²³ Dies ist beispielsweise der Fall bei Manfred Görtemaker, „Gestaltung der Wiedervereinigung,“ *Informationen zur politischen Bildung* 250 (2015): 68, wo auf unplausible Zahlen des verrufenen Wickert-Instituts rekuriert wird. Vgl. „Falsche und gefälschte Zahlen. Demoskopen in Deutschland: Wickert intern,“ *Der Spiegel*, 9. Oktober 1978.

wortverweigerung reagieren (zum Beispiel Arbeitslose oder Rentner:innen beim Thema Arbeitseinkommen)? Wie plausibel sind die Daten insgesamt?

Tab. 1: Persönliches monatliches Bruttoeinkommen vorzeitig exmatrikulierter Student:innen 1986 (in Mark)

Wert	Häufigkeit	Prozent	Kumulierte Prozente
150	1	0,4	0,4
190	1	0,4	0,8
200	24	9,2	9,8
215	4	1,5	11,5
250	1	0,4	11,8
260	3	1,1	13,0
...	...	...	...
1 300	4	1,5	98,5
1 350	1	0,4	98,9
1 500	1	0,4	99,2
1 800	1	0,4	99,6
7 000	1	0,4	100,0
Gesamt	262	100,0	

Ausgewertet nach: Zentralinstitut für Jugendforschung, „Studenten-Intervallstudie Leistung SIL-EX II 1986,“ *GESIS Datenarchiv*, Köln. ZA6137, Datenfile Version 1.0.0. doi: 10.4232/1.6137

Vom DDR-Zentralinstitut für Jugendforschung in Leipzig ist eine Befragung aus dem Jahr 1986 überliefert, in der ohne Abschluss exmatrikulierte Student:innen um die Angabe ihres monatlichen Bruttoeinkommens gebeten worden sind. Die (hier mit Aussparung versehene) Häufigkeitsverteilung zeigt einen bei 150 Mark beginnenden Wertebereich, der bei 1800 Mark enden würde, wenn nicht noch ein:e Respondent:in mit 7 000 Mark an Einkommen in der Stichprobe wäre (siehe Tabelle 1). Angesichts der Tatsache, dass Erich Honecker 1980 über ein monatliches Gehalt von 6 500 Mark verfügte²⁴, könnte man einerseits von gelebtem Sozialismus sprechen, in dem manch ehemalige:r Student:in ebenso viel verdiente wie der Staatsratsvorsitzende. Andererseits sollte man wohl eher anneh-

²⁴ Vgl. Jens Gieseke, „Soziale Ungleichheit im Staatssozialismus. Eine Skizze,“ *Zeithistorische Forschungen* 10 (2013): 182.

men, dass der Wert nicht korrekt ist: dass beim Ausfüllen des Fragebogens oder bei der Transformation der analogen Information in digitales Format schlichtweg eine Null hinzugekommen ist.

4 Methoden der Auswertung

Um in Datensätze „hinzuschauen“ und Merkmalsausprägungen gegebenenfalls zu plausibilisieren – indem man solch einen extremen Ausreißerwert auf 700 korrigiert oder ihn von der Verteilung abschneidet –, sind Kenntnisse entsprechender Statistiksoftware nötig. Daneben können Filterungen oder Umformungen bestimmter Variablen notwendig werden. Mit klassiert vorliegenden Einkommensangaben (wie im Mikrozensus) können keine Berechnungen vorgenommen werden, sodass hier den Einkommensgruppen Repräsentanzwerte zuzuweisen sind.²⁵ Vorbereitende statistische Verfahren sind mitunter auch relevant, um Daten unterschiedlicher Provenienz miteinander passgerecht zu machen. Beispielsweise wurde in den Datensätzen der amtlichen DDR-Einkommensstichprobe als Haushaltsvorstand die:der Haupteinkommensbezieher:in definiert. Im Mikrozensus der amtlichen Statistik der Bundesrepublik wurde hingegen der Haushaltsvorstand vom Haushalt selbst benannt (wobei die Wahl im Großteil der Fälle auf den „Mann im Haus“ fiel). Anhand der sozialen Stellung des Haushaltsvorstands wird zugleich der gesamte Haushalt klassifiziert, zum Beispiel als Arbeiter- oder Angestelltenhaushalt. Hieran wird deutlich, wie wichtig die Vergleichbarkeit der statistischen Quellen bei der Generierung von Datenreihen ist.²⁶ Entsprechende Datensatzpräparationen – wie das behutsame Anpassen der Methodik des einen an die Methodik des anderen Datensatzes –, die der eigentlichen Auswertung vorgelagert sind, erfordern bereits statistische Kenntnisse und Softwarekompetenzen.

Verbreitete Statistikprogramme sind STATA, R und insbesondere das einsteigerfreundliche SPSS. Standardauswertungen sind mit SPSS über die Dialogfelder der grafischen Benutzeroberfläche möglich. Dies beinhaltet einfache Häufigkeitsauszählungen (wie oft kommt eine Merkmalsausprägung innerhalb der Stichprobe vor) oder die Erstellung von Kreuztabellen (wie sind bestimmte Va-

²⁵ Vgl. Johannes Stauder und Wolfgang Hüning, „Die Messung von Äquivalenzeinkommen und Armutsquoten auf der Basis des Mikrozensus,“ *Statistische Analysen und Studien Nordrhein-Westfalen* 13 (2004): 9–31.

²⁶ Vgl. Anthony B. Atkinson, *Ungleichheit. Was wir dagegen tun können* (Stuttgart: Klett-Cotta Verlag, 2016), 66.

riablen in verschiedenen Gruppen ausgeprägt). Und auch die Durchführung komplexerer statistischer Verfahren, wie lineare multiple Regressionen (welchen Einfluss haben mehrere Variablen jeweils auf die Ausprägung einer zu untersuchenden Variable), gestattet SPSS über eine Schaltflächenauswahl und ohne dass Programmieraufwand zwingend notwendig würde. Daneben lassen sich spezielle Analysen und Berechnungen, die im SPSS-Menü nicht vorgesehen sind, mithilfe von Syntax-Befehlen programmieren.

Zum Beispiel bietet SPSS originär keine Berechnung des sogenannten Gini-Koeffizienten an, sodass hierfür eigens eine SPSS-Syntaxroutine geschrieben werden musste (siehe nachfolgende Programmieranwendung). Der Gini-Koeffizient zählt zu den am häufigsten verwendeten Ungleichheitsmaßen und nimmt einen Wert von 0 bis 1 an. Wenn alle Untersuchungseinheiten (etwa Haushalte oder Personen) über dasselbe Einkommen verfügen, beträgt der Gini einer Einkommensverteilung 0. Erhält im Extremfall ein Haushalt oder eine Person das gesamte verfügbare Einkommen, während alle anderen gar kein Einkommen haben, liegt der Gini bei 1.²⁷ Je größer die Ungleichheit, umso höher also der Gini.

SPSS-Syntaxroutine zur Berechnung des Gini-Koeffizienten (eigener Programmcode für Variable „EINKOMMEN“):

```
COMPUTE N=1.
EXECUTE.
AGGREGATE
/OUTFILE=* MODE=ADDVARIABLES
/BREAK=N
/FAELLE=N
/EINKOMMEN_SUM=SUM(EINKOMMEN).
EXECUTE.

SORT CASES BY EINKOMMEN(A).
EXECUTE.

COMPUTE ANTEIL_EINK = EINKOMMEN / EINKOMMEN_SUM.
EXECUTE.
```

²⁷ Rechnerisch kann der Wert 1 nur vom normierten und nicht vom einfachen Gini erreicht werden; vgl. Henriette Engelhardt, „Modelle zur Messung und Erklärung personeller Einkommensverteilungen,“ in *Handbuch der Demographie, Bd. 2: Anwendungen*, hg. v. Ulrich Mueller, Bernhard Nauck und Andreas Dieckmann (Berlin: Springer, 2000), 1075–1076. Mit dem Syntaxbeispiel wird der normierte Gini berechnet.

```

COMPUTE ANTEIL_KUM = ANTEIL_EINK.
IF ($CASENUM > 1) ANTEIL_KUM = ANTEIL_EINK + LAG(ANTEIL_KUM).
EXECUTE.

COMPUTE TRAPEZ = ANTEIL_KUM / (2 * FAELLE).
IF ($CASENUM > 1) TRAPEZ = (ANTEIL_KUM + LAG(ANTEIL_KUM)) / (2 * FAELLE).
EXECUTE.

AGGREGATE
/OUTFILE=* MODE=ADDVARIABLES
/BREAK=N
/TRAPEZ_SUM=SUM(TRAPEZ).
EXECUTE.

COMPUTE GINI = (1/2 - TRAPEZ_SUM) * 2 * FAELLE / (FAELLE - 1).
EXECUTE.

ALTER TYPE GINI (f8.5).

FREQUENCIES GINI.

DELETE VARIABLES N FAELLE EINKOMMEN_SUM
ANTEIL_EINK ANTEIL_KUM TRAPEZ TRAPEZ_SUM.
EXECUTE.

```

Wenn eine komplexe soziale Realität wie Einkommensungleichheit in einer einzigen mathematischen Kennzahl verdichtet wird, mag das bisweilen seltsam erscheinen. Und zweifellos geht mit einer solchen Reduktion ein großer Informationsverlust einher. Jedoch wird somit die Beobachtung der Entwicklung von Einkommensungleichheit im Zeitverlauf erst ermöglicht oder zumindest erleichtert. Für die ostdeutschen Arbeitnehmerhaushalte zwischen 1980 und 2000 zeigt der Gini-Koeffizient eine augenfällige Zunahme der Ungleichheit bei den Pro-Kopf-Haushaltseinkommen mit dem Systemumbruch (siehe Abbildung 1). Neben einer stärkeren Differenzierung der Arbeitseinkommen in der Marktwirtschaft zeichnete hierfür vor allem die Zunahme der Arbeitslosigkeit verantwortlich (Haushalte von Arbeitslosen sind als „potentielle Arbeitnehmerhaushalte“ miteinbezogen).

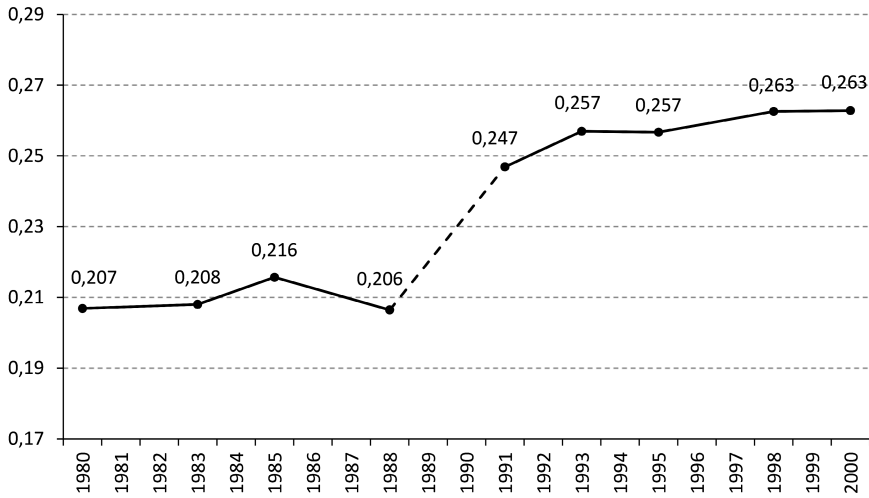


Abb. 1: Gini-Koeffizient für die Haushaltseinkommen pro Kopf in ostdeutschen Arbeitnehmerhaushalten. Eigene Berechnung nach: *Einkommensstichprobe 1980–1988*, BArch DE2/60151–60156; FDZ der Statistischen Ämter des Bundes und der Länder, *Mikrozensus 1991–2000*.

Insbesondere, wenn sie auf Grundlage subjektiver Selbstauskünfte von Personen berechnet werden, ist zu bedenken, dass mathematische Resultate weniger mit der objektiven Exaktheit, die ihnen innezuwohnen scheint, sondern eher als Größenordnung zu interpretieren sind. In diesem Sinne weist die Entwicklung des Gini-Koeffizienten darauf hin, dass es zwischen 1985 und 1988 auch eine Phase der Abmilderung von Ungleichheit gab. Ursächlich hierfür war primär eine Erhöhung des Kindergeldes, von der in erster Linie die in der Einkommenshierarchie häufiger unten stehenden „kinderreichen“ Haushalte profitierten. Denn auch im Sozialismus war, wie später in der Marktwirtschaft, die Anzahl der Kinder ein bedeutsamer Bestimmungsfaktor der materiellen Lage – dies ist zugleich ein Beispiel für vorherrschende Kontinuitäten zwischen den 1980er und 1990er Jahren.²⁸

Methodisch geht es folglich nicht darum, Datenreihen um ihrer selbst willen zu produzieren, sondern auf ihnen aufbauend historische Erklärungsfaktoren für soziale Ungleichheit ausfindig zu machen. Die mit digitalen quantifizie-

²⁸ Vgl. Christopher Banditt, „Sozioökonomische Lagen ostdeutscher Arbeitnehmerhaushalte in der Systemtransformation,“ in *Der Zusammenbruch der alten Ordnung? Die Krise der Sozialen Marktwirtschaft und der neue Kapitalismus in Deutschland und Europa*, hg. v. Christoph Lörke und Rüdiger Schmidt (Göttingen: Wallstein Verlag, 2020), 224–226.

renden Methoden generierten Statistiken und Zeitreihen sind immer auch zu historisieren, indem sie im gesellschaftlichen, politischen und wirtschaftlichen Rahmen kontextualisiert werden. Wenn zum Beispiel überlieferte quantitative Erhebungen der DDR-Soziologie offenbaren, dass Haushalte mit SED-Mitgliedern häufiger über eine moderne Wohnung (mit Fernheizung, Dusche/Bad, Innen-WC) verfügten als „Nicht-SED-Haushalte“, und sich für diesen Zusammenhang aus den Datensätzen eine hohe statistische Signifikanz ergibt, lassen sich daraus noch keine eindeutigen historischen Wirkkräfte ableiten. Pointiert gefragt: Erhielten SED-Kader aufgrund ihrer SED-Mitgliedschaft öfter eine moderne Wohnung? Oder wurde ihnen und ihrem Haushalt zuerst die moderne Wohnung zugewiesen und sie sind dann – etwa aus Dankbarkeit dafür – in die SED eingetreten? Die statistischen Zusammenhangsmaße, die sich aus den Mikrodaten als Quellenmaterial berechnen lassen, müssen in den jeweiligen geschichtlichen Kontext eingeordnet werden, um Wirkungsmechanismen zu identifizieren. Erst daraus ergibt sich das historische Narrativ.

Um aber überhaupt statistische Zusammenhänge als Anhaltspunkte zu gewinnen, sind entsprechende Kompetenzen bei der Verarbeitung und Auswertung digitaler Datensätze aus Bevölkerungsbefragungen notwendig. Doch warum sollten Historiker:innen überhaupt Surveydaten auswerten und nicht einfach auf die Ergebnisse derer Produzenten, der Sozial- und Wirtschaftswissenschaften, rekurrieren? Zunächst einmal ergeben sich aus historischer Perspektive mitunter andere Fragen ans Material, als sie zeitgenössisch von den Soziolog:innen und Ökonom:innen gestellt worden sind. Ferner lassen sich bestimmte leitende theoretische Annahmen bei der ursprünglichen Befragungskonzeption mit ihren Implikationen für die Forschungsergebnisse zumeist erst identifizieren, wenn das Datenmaterial konkret bewertet wird. Dies gilt auch für bestimmte Besonderheiten von Erhebungen und methodische Unterschiede zwischen verschiedenen Befragungen (wie es hier anhand der Art der Einkommensabfrage knapp skizziert worden ist). Zudem liegen sozialwissenschaftlichen Ergebnissen oftmals noch Hochrechnungs- und Gewichtungszugriffe zugrunde, die nur mit einer eigenen Beschäftigung mit den Primärdatenfiles nachvollzogen werden können. Ähnliches gilt für den Umgang mit *Missings*, statistischen Ausreißern und unplausiblen Werten. Es ist letztlich unerlässlich, dass die historische Quellenkritik solche Punkte adressieren kann.

5 Schlussbemerkungen

Wie eingangs angedeutet sind digitale quantifizierende Methoden an sich nichts Neues. Sozial- und Wirtschaftswissenschaftler:innen oder auch Psycholog:innen nutzen schon lange digitale Prozeduren zur Auswertung von Befragungsdaten. Für Wirtschaftshistoriker:innen gilt dasselbe hinsichtlich Einkommens- und Verbrauchsstichproben. Neuartig sind meines Erachtens allerdings Umfang und Menge an überlieferten Surveys, die heute potenziell für historische Re-Analysen zur Verfügung stehen. Je näher die zeitgeschichtlich untersuchte Vergangenheit an der Gegenwart liegt, umso größer ist der Fundus an solchen quantitativen und digitalen Quellen. Mit dem „Aufstieg der Sozialwissenschaften“ in der zweiten Hälfte des 20. Jahrhunderts ging gleichsam die „statistische Vermessung der Gesellschaft“²⁹ mittels Datenerhebungen einher. Das daraus hervorgegangene Material beinhaltet freilich nicht nur „harte Items“, wie zur materiellen Ausstattung von Haushalten. Es finden sich in Erhebungsdatensätzen – je nach disziplinärer Provenienz – auch „weichere Merkmale“, wie zum Beispiel die persönliche politische Einstellung oder auch der Bücherbestand eines Haushalts, womit etwa milieu- oder kulturgeschichtlichen Forschungen Anknüpfungspunkte gegeben werden.

Gelangt nun die Zeitgeschichtsforschung mit einem stärkeren Einbezug quantitativer Quellen – ganz gleich, welche Themenstellung mit ihnen bearbeitet wird – ins „Zeitalter der Quantifizierung“, das George Stigler bereits in den 1960er Jahren für die Wirtschaftswissenschaft ausgerufen hat?³⁰ „[U]nzweifelhaft gehört die Zukunft den Zahlen,“ ließ der französische Schriftsteller Romain Gary eine seiner Romanfiguren feststellen: „Seit dem Ende des Rittertums hat die Welt gelernt zu rechnen, und das wird immer schlimmer. Wir werden noch das Verschwinden all dessen erleben, was nicht beziffert werden kann.“³¹ Droht in der Tat eine Marginalisierung nichtquantifizierbarer Sujets oder der Geschichtswissenschaft gar die Mathematifizierung, wie sie die Ökonomie erlebt hat? Ganz sicher nicht. Vor allem dann nicht, wenn der Tatsache Rechnung getragen wird, dass die Mathematik „auch nur eine Sprache [ist], sie kann – wie alle Sprachen – nicht alles ausdrücken.“³² Und wie wichtig Mehrsprachigkeit in

29 Eva Schlotheuber und Frank Bösch, „Quellenkritik im digitalen Zeitalter: Die Historischen Grundwissenschaften als zentrale Kompetenz der Geschichtswissenschaft und benachbarter Fächer,“ *H-Soz-Kult*, 16. November 2015, aufgerufen am 15. Juli 2021, www.hsozkult.de/debate/id/diskussionen-2866.

30 George Stigler, „The Economist and the State,“ *The American Economic Review* 55 (1965): 16.

31 Romain Gary, *Gedächtnis mit Flügeln* (Berlin (Ost): Aufbau-Verlag, 1989 [1980]), 74.

32 Tomáš Sedláček, *Die Ökonomie von Gut und Böse* (München: Carl Hanser Verlag, 2012), 370.

der Geschichtswissenschaft ist, verdeutlicht auch das hier vorgestellte und sich eigentlich stark auf zahlenmäßige Ausdrucksformen stützende Beispiel historischer materieller Disparitäten.

So wurden Ungleichheitsverhältnisse in der späten DDR zu einem guten Teil durch das Verfügen über D-Mark mitbestimmt. Dies lag durchaus quer zu anderen Ungleichheitsdimensionen, wie beruflicher Status, Arbeitseinkommen (in DDR-Mark) oder gesellschaftliches Prestige. So bot das als „kapitalistisches Rudiment“ nur leidlich geduldete private Handwerk seine gefragten Leistungen bevorzugt gegen D-Mark an. Beschäftigte in der Gastronomie, die ansonsten kaum zu den Besserverdiener:innen in der DDR gehörten, profitierten von Trinkgeldern westdeutscher Tourist:innen. Und als „asozial“ kategorisierte Prostituierte boten etwa ihre Dienste westlichen Messebesuchern oder Unternehmensreisenden ebenfalls gegen die begehrte Zweitwährung an.³³ Dagegen hatten es SED-Kader mangels aktiver Westkontakte deutlich schwerer, West-Mark und damit Zugang zu Intershop-Waren zu erhalten. Jedoch liegen über die jeweiligen D-Mark-Bestände in Händen einzelner privater Haushalte keine quantitativen Erhebungen vor, sodass sich diese Ausprägung materieller Ungleichheit kaum in verallgemeinerbarer Weise beziffern respektive „mathematisieren“ lässt. Gleichwohl ist der Aspekt zu wichtig, um ihn vernachlässigen zu können – sodass sich ihm auf Basis des qualitativen Quellenmaterials zu nähern ist.

Wenn quantitative Hinterlassenschaften die Primärquelle für eine Forschungsfrage darstellen, eignen sich qualitative Überlieferungen oft genug für komplementäre Tiefensonden. Bei qualitativ zu untersuchenden Themenstellungen können wiederum entsprechend vorhandene quantitative Erhebungen mitunter einen Rahmen setzen, indem sie zunächst den (repräsentativen) Umfang eines Phänomens beschreiben. Grundsätzlich gilt es also, wo möglich, quantitatives und qualitatives Material gemeinsam zu analysieren. Während qualitative Auswertungsmethoden in den Geschichtswissenschaften traditionell fest verankert sind, bilden sozialwissenschaftlich-statistisch inspirierte Ansätze eine Ergänzung. Entsprechende digitale Kompetenzen zur Auswertung von Surveydaten sind unabdingbar, wenn die (zeit-)historische Forschung die Quellenbestände aus quantitativen Bevölkerungserhebungen stärker einbeziehen möchte. Dass sie es sollte, ist mein Plädoyer.

33 Vgl. Steffi Brüning, *Prostitution in der DDR. Eine Untersuchung am Beispiel von Rostock, Berlin und Leipzig, 1968 bis 1989* (Berlin: be.bra Verlag, 2020), passim.

Bibliographie

- Atkinson, Anthony B. *Ungleichheit. Was wir dagegen tun können*. Stuttgart: Klett-Cotta Verlag, 2016.
- Banditt, Christopher. „Einheitserwartungen. Präferenzen, Hoffnungen und Befürchtungen im ‚Wendejahr‘ 1989/90.“ *Berliner Debatte Initial* 30 (2019): 67–84.
- Banditt, Christopher. „Sozioökonomische Lagen ostdeutscher Arbeitnehmerhaushalte in der Systemtransformation.“ In *Der Zusammenbruch der alten Ordnung? Die Krise der Sozialen Marktwirtschaft und der neue Kapitalismus in Deutschland und Europa*, hg. v. Christoph Lorke und Rüdiger Schmidt, 213–234. Göttingen: Wallstein Verlag, 2020.
- Brückweh, Kerstin. „Wissen über die Transformation. Wohnraum und Eigentum in der langen Geschichte der ‚Wende‘.“ *Zeithistorische Forschungen* 16 (2019): 19–45.
- Brüning, Steffi. *Prostitution in der DDR. Eine Untersuchung am Beispiel von Rostock, Berlin und Leipzig, 1968 bis 1989*. Berlin: be.bra Verlag, 2020.
- Buchner, Michael, Tobias A. Jopp, Mark Spoerer und Lino Wehrheim. „Zur Konjunktur des Zählens – oder wie man Quantifizierung quantifiziert. Eine empirische Analyse der Anwendung quantitativer Methoden in der deutschen Geschichtswissenschaft.“ *Historische Zeitschrift* 310 (2020): 580–621.
- Engelhardt, Henriette. „Modelle zur Messung und Erklärung personeller Einkommensverteilungen.“ In *Handbuch der Demographie, Bd. 2: Anwendungen*, hg. v. Ulrich Mueller, Bernhard Nauck und Andreas Dieckmann, 1066–1091. Berlin: Springer Verlag, 2000.
- Euler, Manfred. „Genauigkeit von Einkommensangaben in Abhängigkeit von der Art der statistischen Erfassung. Dargestellt am Beispiel der Einkommens- und Verbrauchsstichprobe 1978.“ *Wirtschaft und Statistik* (1983): 813–818.
- „Falsche und gefälschte Zahlen. Demoskopien in Deutschland: Wickert intern.“ *Der Spiegel*, 9. Oktober 1978.
- Gary, Romain. *Gedächtnis mit Flügeln*. Berlin (Ost): Aufbau-Verlag, 1989 [1980].
- Gieseke, Jens. „Soziale Ungleichheit im Staatssozialismus. Eine Skizze.“ *Zeithistorische Forschungen* 10 (2013): 171–198.
- Görtemaker, Manfred. „Gestaltung der Wiedervereinigung.“ *Informationen zur politischen Bildung* 250 (2015): 62–74.
- Heidenreich, Hans-Joachim. „Einführung des Mikrozensus in den neuen Bundesländern: Probleme und Erfahrungen.“ In *Amtliche Daten der DDR und der neuen Bundesländer: Informationsquelle für die Sozialwissenschaften. Tagungsdokumentation*, hg. v. Paul Lüttlinger und Heike Wirth, 11–26. Mannheim: ZUMA, 1993.
- Hradil, Stefan. *Soziale Ungleichheit in Deutschland*. Wiesbaden: Springer VS, 2005. 8. Auflage.
- Hradil, Stefan. „Soziale Ungleichheit, soziale Schichtung und Mobilität.“ In *Einführung in Hauptbegriffe der Soziologie*, hg. v. Hermann Korte und Bernhard Schäfers, 247–275. Wiesbaden: Springer VS, 2016.
- Jarausch, Konrad H. „Möglichkeiten und Probleme der Quantifizierung in der Geschichtswissenschaft.“ In *Quantifizierung in der Geschichtswissenschaft. Probleme und Möglichkeiten*, hg. v. Konrad H. Jarausch, 11–30. Düsseldorf: Droste Verlag, 1976.
- Jopp, Tobias A. und Mark Spoerer. *Historische Statistik lehren. Quellenkritische Vermittlung von Zielen und Methoden statistischen Arbeitens in der Geschichtswissenschaft*. Schwalbach/Ts.: Wochenschau Verlag, 2017.

- Marx, Karl. „Kritik des Gothaer Programms.“ In *Werke, Bd. 19*, Karl Marx und Friedrich Engels, 11–32. Berlin (Ost): Dietz Verlag, 1962 [1875].
- Mau, Steffen. *Das metrische Wir. Über die Quantifizierung des Sozialen*. Berlin: Suhrkamp Verlag, 2017.
- Pfanzelter, Eva. „Die historische Quellenkritik und das Digitale.“ *Archiv und Wirtschaft* 48 (2015): 5–19.
- Plumpe, Werner. *Das kalte Herz. Kapitalismus: Die Geschichte einer andauernden Revolution*. Berlin: Rowohlt Verlag, 2019.
- Rahlf, Thomas. „Einleitung.“ In *Deutschland in Daten. Zeitreihen zur Historischen Statistik*, hg. v. Thomas Rahlf, 5–12. Bonn: Bundeszentrale für politische Bildung, 2015.
- Schimpl-Neimanns, Bernhard und Heike Wirth. *Bestandsaufnahme und Nutzungsmöglichkeiten amtlicher Mikrodaten der DDR für Sekundäranalysen zur Bildungs- und Einkommensungleichheit*. Mannheim: ZUMA, 1994.
- Schlotheuber, Eva und Frank Bösch. „Quellenkritik im digitalen Zeitalter: Die Historischen Grundwissenschaften als zentrale Kompetenz der Geschichtswissenschaft und benachbarter Fächer.“ *H-Soz-Kult*, 16. November 2015. Aufgerufen am 15. Juli 2021. www.hsozkult.de/debate/id/diskussionen-2866.
- Sedláček, Tomáš. *Die Ökonomie von Gut und Böse*. München: Carl Hanser Verlag, 2012.
- Stauder, Johannes und Wolfgang Hüning. „Die Messung von Äquivalenzeinkommen und Armutsquoten auf der Basis des Mikrozensus.“ *Statistische Analysen und Studien Nordrhein-Westfalen* 13 (2004): 9–31.
- Stigler, George. „The Economist and the State.“ *The American Economic Review* 55 (1965): 1–18.
- Wagner, Gert G. „Die Längsschnittstudie Sozio-oekonomisches Panel (SOEP) – Die Jahre von der Wende zur Jahrtausendwende.“ *Vierteljahrshefte zur Wirtschaftsforschung* 77 (2008): 43–62.
- Wettengel, Michael. „Sozialwissenschaftliche Daten aus der DDR – ein Katalog archivierter Studien.“ *Historische Sozialforschung* 20 (1995): 127–134.

Florian Windhager, Eva Mayr, Matthias Schlögl und
Maximilian Kaiser

Visuelle Analyse und Kuratierung von Biographiedaten

Abstract: Large volumes of biographical text collections have been digitized and developed into structured research databases in recent years. In order to strengthen the historiographical relevance and acceptance of these data collections, this paper discusses two interdependent technological objectives. First, it summarizes advanced visual analysis techniques that can support formal and distant reading approaches to individual and aggregated biography data from multiple perspectives. On the other hand, it suggests to continuously improve structured data collections through an interplay of visual interfaces, manual procedures, and semi-automated annotation techniques. This aims at closing the operational gap between data analysis and data curation, and at developing new tools for the creation and maintenance of biographical data even for historians without computer-linguistic skills.

Keywords: biographical data, visualization, distant reading, data curation, digital humanities

Zusammenfassung: Große Mengen von biographischen Textsammlungen wurden in den letzten Jahren digitalisiert und zu strukturierten Forschungsdatenbanken weiterentwickelt. Um die effektive Nutzung dieser Daten strategisch weiterzudenken diskutiert der Beitrag zwei interdependente technologische Zielsetzungen. Zum einen werden avancierte Techniken der visuellen Analyse resümiert, welche formale und distante Lektüren von individuellen und aggregierten Biographiedaten aus multiplen Perspektiven unterstützen können. Zum anderen wird angeregt, strukturierte Daten durch ein balanciertes Zusammenspiel von visueller Analyse, manuellen Verfahren und semi-automatisierten Annotationstechniken laufend zu verbessern. Dies zielt auf die Schließung der operativen Lücke zwischen Datenanalyse und Datenkuratierung, und auf die Schaffung neuer Werkzeuge zur Erstellung und Pflege biographischer Daten auch für Historiker:Innen ohne computer-linguistische Methodenkompetenz.

Danksagung: Das Projekt In/Tangible European Heritage: Visual Analysis, Communication and Curation wird von der Europäischen Kommission im Rahmen des H2020 Forschungs- & Innovationsprogramms gefördert (Projekt ID: 101004825).

Schlagwörter: Biographiedaten, Visualisierung, Distant Reading, Datenkuratierung, Digitale Geisteswissenschaften

1 Einleitung

In den letzten Jahren wurden große Mengen von biographischen Textsammlungen und Lexika digitalisiert und zu strukturierten Forschungsdatenbanken weiterentwickelt. Die resultierenden prosopographischen Datenbanken eröffnen neue Möglichkeiten für die historische Forschung und die Vermittlung von relevanten Themen an eine interessierte Öffentlichkeit – beispielsweise durch deren Verknüpfung mit kulturellen Objekten aus Museen oder Archiven. Bevor Historiker:innen jedoch die bereits vorhandene Fülle an Informationen ausschöpfen und in Forschung und Lehre ebenso wie in Wissenschaftskommunikation und Kulturvermittlung produktiv anwenden können, sind zahlreiche Herausforderungen zu bewältigen.

Um die effektive Nutzung digitaler Biographiedaten strategisch weiterzudenken, diskutiert dieser Beitrag zwei interdependente technologische Zielsetzungen, die die Produktivität von hermeneutischen und kuratorischen Zugängen zu prosopographischen Informationssystemen auf eine neue Ebene führen können:

- Zum einen wird resümiert, wie prosopographische Informationssysteme mittels avancierter Techniken der visuellen Analyse besser erschlossen werden können. In diesem Kontext sehen wir primär Bedarf für die fachspezifische Adaption und Kombination von Techniken der *visuellen Analyse* für die Biographieforschung, die gleichzeitig die Kommunikation der Ergebnisse an die wissenschaftliche Community wie auch an die interessierte Öffentlichkeit erleichtern können.
- Zum anderen ist zu diskutieren, wie strukturierte biographische Daten durch ein balanciertes Zusammenspiel von visuellen Interfaces, manuellen Verfahren und semi-automatisierten Extraktionstechniken laufend verbessert, differenziert und neu kreiert werden können. Dies würde eine Schließung der operativen Lücke zwischen Datenanalyse und Datenkuratierung erlauben, und neue Ansätze zur Erstellung und Pflege von Daten auch für Historiker:innen ohne computerlinguistische Expertise erschließen.

Ausgehend von diesen Betrachtungen geben wir einen Ausblick darauf, welche Möglichkeiten sich durch die Optionen der visuellen Analyse und Kuratierung von Biographiedaten für die prosopographische Forschung einerseits und die Kommunikation kultureller Themen andererseits eröffnen.

2 Visuelle Analyse von Biographiedaten

Prosopographische Datenbanken nutzen je nach Entstehungszeit, Kontext und Forschungsfrage unterschiedliche Datenmodelle, die aber zumeist mehrere Entitäten (zum Beispiel Personen, Institutionen und Orte) enthalten, denen verschiedene Attribute zugeordnet (zum Beispiel Berufe) sind und die zueinander durch vielfältige Relationen in Beziehung stehen (zum Beispiel Ausbildung, Verwandtschaft). Damit bieten sie Informationen über Akteure (einzelne oder Gruppen von Personen), deren Aktivitäten (Ereignisse) und Beziehungen zu anderen Entitäten (Relationen) zu bestimmten Zeitpunkten (Zeitstempel). Vor dem Hintergrund solch facettenreicher Daten bedarf es aus unserer Sicht neuer – insbesondere auch visualisierungsgestützter – Interfaces, die dieser Datenkomplexität gerecht werden und die damit neue Möglichkeiten für die historische Forschung und Wissensvermittlung eröffnen.

Dabei sollten solche visuellen Interfaces den gesamten Forschungsprozess – von der Suche nach relevanten Daten bis zu deren Kuratierung, Analyse und Kommunikation – unterstützen. Um der Komplexität solcher multidimensionaler, prosopographischer Datenbanken gerecht zu werden, bedarf es dazu mehrerer, miteinander verknüpfter visueller Analysetechniken (zum Beispiel einer geographischen, einer netzwerkanalytischen und einer kategorialen Perspektive) unter besonderer Berücksichtigung der zeitlichen Datendimension, der oft heterogenen Datenqualität und der Datenprovenienz (siehe auch 3.). Darüber hinaus sollten entsprechende Visualisierungssysteme Techniken des „close readings“ und des „distant readings“ von individuellen und aggregierten Biographiedaten aus den verschiedenen Perspektiven unterstützen.

Ein Blick auf den Stand der Visualisierungsforschung zeigt, dass bereits zahlreiche Techniken im Kontext biographischer Daten angewandt oder entwi-

ckelt wurden:¹ Zeitvisualisierungen², kartographische Ansichten³, und netzwerkanalytische Ansätze⁴. Die Attribute von einzelnen oder Gruppen von Personen wurden mittels hierarchischer Visualisierungstechniken dargestellt.⁵ Zahlreiche Forscher nutzen auch multiple Visualisierungstechniken, um komplementäre analytische Perspektiven zu bündeln.⁶

1 Vgl. auch, Florian Windhager et al., „Beyond One-Dimensional Portraits: A Synoptic Approach to the Visual Analysis of Biography Data,“ in *BD 2017*, CEUR Workshop Proceedings, 2017, hg. v. Antske Fokkens et al., 67–75.

2 Stephen Boyd Davis und Florian Kräutli, „Scholarly chronographics: can a timeline be useful in historiography?“ In *Proceedings of 10th European Social Science History Conference*, no. Web, Web-Web. 2014; Matthew Brehmer et al., „Timelines revisited: A design space and considerations for expressive storytelling,“ *IEEE transactions on visualization and computer graphics* 23/9 (2016): 2151–2164. doi: 10.1109/TVCG.2016.2614803; Marc Champagne, „Diagrams of the past: How timelines can aid the growth of historical knowledge,“ *Cognitive Semiotics* 9/1 (2016): 11–44. doi: 10.1515/cogsem-2016-0002.

3 Ryan Eccles et al., „Stories in geotime,“ *Information Visualization* 7/1 (2008): 3–17. doi: 10.1109/VAST.2007.4388992; Tiago Gonçalves, Ana Paula Afonso und Bruno Martins, „Cartographic visualization of human trajectory data: Overview and analysis,“ *Journal of Location Based Services* 9/2 (2015): 138–166. doi: 10.1080/17489725.2015.1074736; Shakespear, Daniel, „Interactive Genealogy Explorer: Visualization of Migration of Ancestors and Relatives,“ in *BD 2015*, CEUR Workshop Proceedings, 2015, hg. v. Serge ter Brake et al., 94–100.

4 Elisa Beshero-Bondar und Elizabeth Raisanen, „Recovering from Collective Memory Loss: The Digital Mitford’s Feminist Project,“ *Women’s History Review* 26/5 (2017): 738–750. doi: 10.1080/09612025.2016.1166882; Theresia Biehl, Anne Lorenz, Dirk Osierenski und Exilnetz33, „Exilnetz33. Ein Forschungsportal als Such- und Visualisierungsinstrument, in *Grenzen und Möglichkeiten der Digital Humanities* (Zeitschrift für digitale Geisteswissenschaften, Sonderband 1) (2015); Lonneke Geerlings, „A Visual Analysis of Rosey E. Pool’s Correspondence Archives. Biographical Data, Intersectionality, and Social Network Analysis,“ in *BD2015*, CEUR Workshop Proceedings, 2015, hg. v. Serge ter Braake et al., 61–67; Maximilian Schich et al., „A network framework of cultural history,“ *science* 345/6196 (2014): 558–562. Patrick Schmitz und Laurie Pearce, „Humanist-centric tools for big data: berkeley prosopography services,“ in *Proceedings of the 2014 ACM symposium on Document engineering*, 179–188. ACM: 2014.

5 Cristian Jara-Figueroa, Amy Z. Yu und César A. Hidalgo, „Estimating technological breaks in the size and composition of human collective memory from biographical data,“ Preprint auf <https://arxiv.org/abs/1512.05020> (2015). Amy Zhao Yu et al., „Pantheon: visualizing historical cultural production,“ in *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 289–290. IEEE, 2014.

6 Neil Armitage, „The biographical network method.“ *Sociological Research Online* 21/2 (2016): 165–179. doi: 10.5153/sro.3827; Olivier Gergaud, Morgane Laouenan und Etienne Wasmer, „A Brief History of Human Time. Exploring a database of ‘notable people’,“ (2017). Working paper, HAL Id: hal-01440325/; Eero Hyvönen et al., „BIOGRAPHYSAMPO—A Paradigm Shift for Publishing and Using Biography Collections on the Semantic Web,“ in *ESWC 2019: The Semantic Web*, 574–589. Cham: Springer; Richard Khulusi et al., „An interactive chart of biography,“ in *2019 IEEE Pacific Visualization Symposium (PacificVis)*, 257–266. IEEE, 2019; Christofer Mein-

Während diese Forschungsarbeiten den Möglichkeitsraum für die Visualisierung von biographischen Daten aufzeigen, so wurden diese bislang kaum mit prosopographischen Datensammlungen verknüpft und zu deren Erschließung und Vermittlung genutzt. Das möchten wir durch die laufende Methoden- und Technologieentwicklungen im H2020-Projekt InTaVia (In/Tangible European Heritage – Visual Analysis, Curation and Communication, <https://inta-via.eu>) ändern: Das Projekt hat eine Verbesserung der Datenbasis und die Entwicklung eines visuellen Analysestudios für Daten und Themen des materiellen und immateriellen Kulturerbes zum Ziel (Abbildung 1). Die Basis des Projekts bildet die Schaffung einer europäischen und synoptischen Datenbasis durch (i) die Zusammenführung und Integration verschiedener europäischer Nationalbiographien,⁷ (ii) deren Optimierung durch den Einsatz fortgeschrittener *Natural Language Processing*-Verfahren, und (iii) deren Verknüpfung mit kulturellen Objektdatenbanken (zum Beispiel Europeana). Diese erweiterte Datenbasis, welche Informationen zu Leben und Werken zahlreicher europäischer Künstler:innen, aber auch anderer bedeutender Akteure enthalten wird, liefert das Ausgangsmaterial für multiple Perspektiven der visuellen Analyse. Ausgehend von statischen Blicken auf die Sammlungen von individuellen Lebenswerken (inklusive Karten, Netzwerke und Mengendiagramme, Abbildung 1, Sektion 1a) sollen verschiedene Visualisierungstechniken mit einem Fokus auf die Zeitdimension (Abbildung 1, Sektion 1b) genutzt werden, um historische Lebenswege und damit verknüpfte kulturelle Objekte synoptisch zu analysieren (Abbildung 1, Sektion 2). Daneben wird auch die prosopographische Forschung visuell unterstützt werden, indem aggregierte Gruppen von Personen und Objekten, die miteinander in Relation stehen, gemeinsam analysiert werden können (Abbildung 1, Sektion 3).

ecke und Stefan Jänicke, „Visual Analysis of Engineers’ Biographies and Engineering Branches,“ in *LEVIA18: Leipzig Symposium on Visualization in Applications*. 2018; Irene Russo, Tommaso Caselli und Monica Monachini, „Extracting and Visualising Biographical Events from Wikipedia,“ in *BD 2015*, hg. v. Serge ter Braake et al., 111–115. CEUR Workshop Proceedings, 2015; Schmitz und Pearce, „Humanist-centric tools for big data“.

⁷ Die Kollaboration inkludiert die Nationalbiographie von Finnland mit dem digitalen Portal *BiographySampo* (<https://seco.cs.aalto.fi/projects/biografiasampo/en/>), das holländische *BiographyNet* (<http://www.biografischportaal.nl/en>), das österreichische *Austrian Prosopographical Information System* (<https://apis.acdh.oeaw.ac.at/>) und die slowenische *Slovenska biografija* (<http://www.slovenska-biografija.si/>), Zugriff am 16.03.2021.

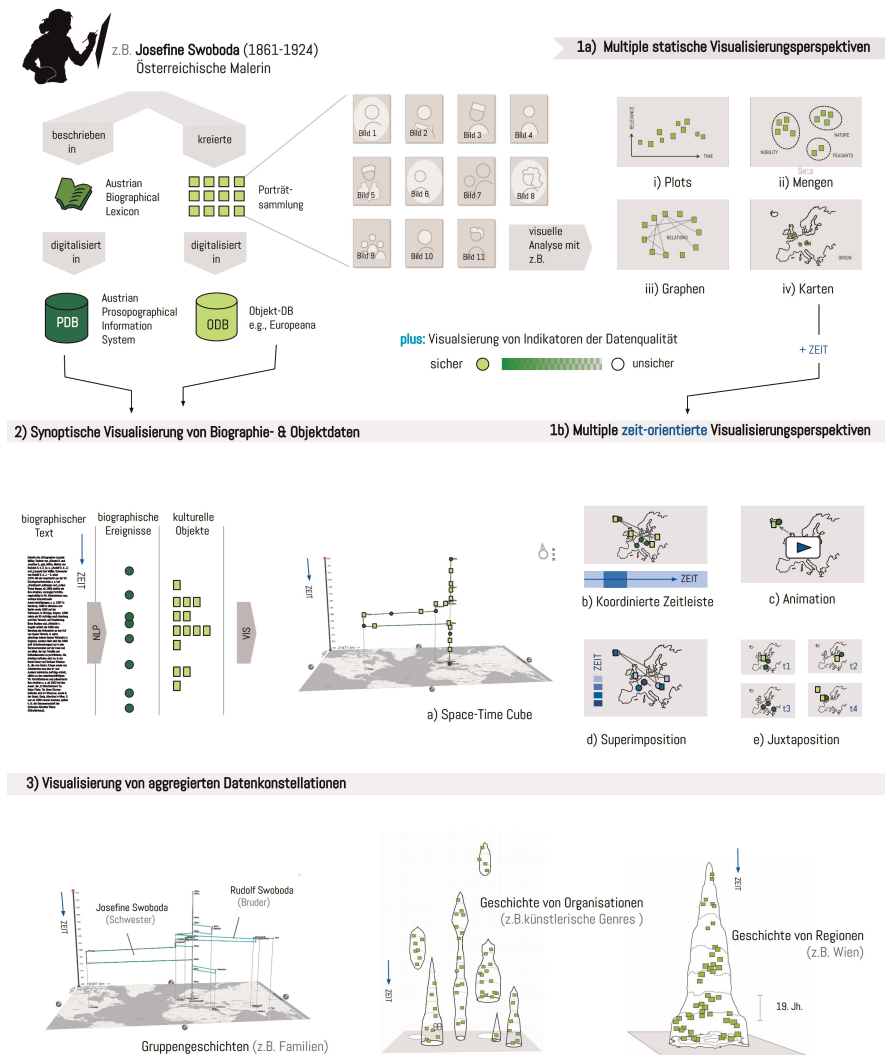


Abb. 1: Synoptische Visualisierung von materiellem Kulturerbe (kulturelle Objekte) und immateriellem Kulturerbe (Biographien) in InTaVia. Abbildung von Florian Windhager.

Visualisierungen erleichtern es zudem, die Ergebnisse dieser Analysen an diverse Zielgruppen zu kommunizieren. Neben Interfaces für Expert:innen, ermöglichen es Storytelling-Methoden biographische Narrationen visuell auch für eine breite Öffentlichkeit aufzubereiten und leichter verständlich zu kommunizie-

ren.⁸ Biographien als zeitliche Abfolge von Ereignissen eines Protagonisten sind in ihrem Kern bereits narrativ und legen daher eine solche Aufbereitung von kulturellen Informationen nahe.

3 Visuelle Kuratierung von Biographiedaten

Für eine effizientere Aufbereitung und strukturierte Annotation von biographischen Texten kommen in den letzten Jahren vermehrt Methoden des Natural Language Processing (NLP) zum Einsatz⁹ – teilweise auch durch Verknüpfung mit weiteren offen zugänglichen Datenquellen¹⁰. So wurde exemplarisch das Österreichische Biographische Lexikon (ÖBL) digitalisiert und mithilfe von NLP-Verfahren aufbereitet, sodass es nun als prosopographische Datenbank (Austrian Prosopographical Information System, APIS)¹¹ zugänglich ist, die zu Personen strukturierte, maschinenlesbare Informationen in der Form von Relationen mit Orten, Institutionen oder anderen Personen bietet.

Diese Art der Datenaufbereitung durch Prozessierung von (manchmal bereits im Original fehlerhaften)¹² historischen Textquellen durch stochastische NLP-Verfahren¹³ erzeugt jedoch auch Datensätze, die zahlreiche Lücken, Fehler und Unsicherheiten aufweisen und die damit der qualitativen Korrektur, Anrei-

8 Eva Mayr und Florian Windhager, „Once upon a spacetime: Visual storytelling in cognitive and geotemporal information spaces,“ *ISPRS International Journal of Geo-Information* 7/3 (2018): 96. doi: 10.3390/ijgi7030096.

9 John Bradley und Harold Short, „Texts into databases: The evolving field of new-style prosopography,“ *Literary and linguistic computing* 20 (2005): 3–24. doi: 10.1093/lilc/fqi022; Antske Fokkens et al., „BiographyNet: Methodological Issues when NLP supports historical research,“ in *LREC* (2014), 3728–3735.

10 Judith Brouwer und Harm Nijboer, „Golden Agents. A Web of Linked Biographical Data for the Dutch Golden Age,“ in *Proceedings of the Second Conference on Biographical Data in a Digital World 2017*, hg. v. Antske Fokkens et al., 33–38. CEUR Workshop Proceedings, 2017; Petri Leskinen, Eero Antero Hyvönen und Jouni Antero Tuominen, „Analyzing and visualizing prosopographical linked data based on biographies,“ in *Proceedings of the Second Conference on Biographical Data in a Digital World 2017* (BD2017), hg. v. Antske Fokkens et al., 39–44. CEUR Workshop Proceedings, 2018.

11 Vgl. Matthias Schlögl und Katalin Lejtovicz, „A Prosopographical Information System (APIS),“ in *BD 2017*, hg. v. Antske Fokkens et al., 53–58. CEUR Workshop Proceedings, 2017.

12 Vgl. Ágoston Zénó Bernád und Maximilian Kaiser, „The Biographical Formula: Types and Dimensions of Biographical Networks,“ in *BD 2017*, hg. v. Antske Fokkens et al., 45–52. CEUR Workshop Proceedings, 2017.

13 Zu deren Reliabilität vgl. auch Fokkens et. al., „BiographyNet, 3728–3735; Schlögl und Lejtovicz, „Prosopographical Information System“.

cherung und Kuratierung durch Historiker:innen bedürfen. Gleichzeitig sind die Expert:innen, die eine fachlich fundierte Kuratierung der Daten vornehmen könnten, nicht zwingendermaßen auch diejenigen, die die Funktionsweise von NLP-Verfahren verstehen und so gezielt jene Datensätze identifizieren könnten, die besondere Aufmerksamkeit und Nachbesserungen benötigen. Auch in diesen Fällen können visuelle Interfaces¹⁴ Unterstützung bieten, um die Ergebnisse von NLP-Verfahren weniger technisch zu kommunizieren, die Aufmerksamkeit auf unsichere Ergebnisse dieser Verfahren zu lenken und so eine benutzerfreundliche Umgebung für die Kuratierung prosopographischer Daten bereitzustellen. Gleichzeitig können durch die Korrektur fehlerhafter Ergebnisse NLP-Verfahren in einem aktiven Lernprozess trainiert und damit optimiert werden.

Im Kontext von InTaVia streben wir danach, die Korrektur, Anreicherung und Aufbereitung von Objekt- und Biographiedaten in einem eigenen Kuratierungs-Modul zu ermöglichen, welches eng mit dem visuellen Analysestudio verknüpft ist, um Kuratierungstätigkeiten auf textlicher und visueller Ebene zu ermöglichen (Abbildung 2): So können biographische Daten die durch NLP-basierte, probabilistische Verfahren gewonnen werden – ebenso wie Metadaten von historischen Objektsammlungen – durch Visualisierung einer multimodalen Qualitätskontrolle unterzogen werden und entsprechenden manuellen Korrekturen zugeführt werden – zum Beispiel durch die Hervorhebung von probabilistisch unsicheren Extraktionen, die besonderer Aufmerksamkeit durch Expert:innen bedürfen. Die Visualisierung von Datenqualität, Unvollständigkeit und Unsicherheit bietet hierzu ein wachsendes Spektrum von Optionen.¹⁵ Durch Markierung solch unsicherer Datenpunkte können neben Ansatzpunkten für manuelle Revision auch alternative NLP-Ergebnisse sichtbar gemacht werden – und damit Vorschläge für die fortgesetzte Kuratierung der Datenbasis unterbreitet werden.

14 Steffen Koch et al., „VarifocalReader – in-depth visual analysis of large text documents,“ *IEEE transactions on visualization and computer graphics* 20/12 (2014): 1723–1732. doi: 10.1109/TVCG.2014.2346677.

15 Florian Windhager, Saminu Salisu und Eva Mayr, „Exhibiting uncertainty: Visualizing data quality indicators for cultural collections,“ *Informatics* 6/3 (2019): 29, doi: 10.3390/informatics6030029.

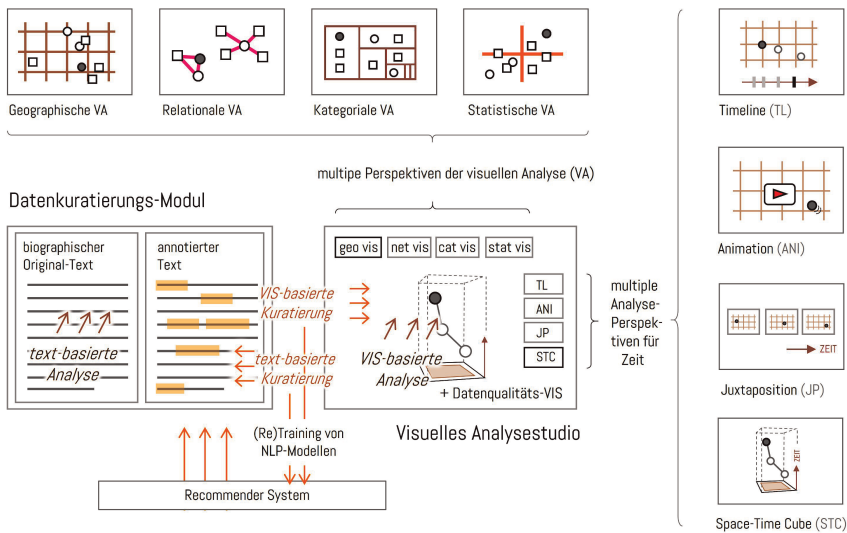


Abb. 2: Konzept der visuell gestützten Kuratierung von Biographiedaten. Abbildung von Florian Windhager.

4 Diskussion

Mit Blick auf die Geschichte verschiedener historischer und geisteswissenschaftlicher Disziplinen wirft die Entwicklung von digitalen prosopographischen Forschungsumgebungen auch Fragen der Interoperabilität mit traditionellen, hermeneutischen und interpretativen Ansätzen der Biographie- und Geschichtsforschung auf. Vor dem Hintergrund des etablierten historiographischen Methodenspektrums erscheinen prosopographische Datenbanken als zweierlei: als fachdienliche Werkzeuge auf dem neuesten Stand der Informationsmodellierung, wie auch als epistemisch riskante und methodisch experimentelle Blackbox-Systeme. Mit ihnen werden historische Informationen und Entitäten aus vormals rigiden syntaktischen Vernetzungen überlieferter Texte befreit und neu verfügbar für ein vielgestaltiges Spektrum von Abfragen, Analysen, Verknüpfungen und (visuellen) Transformationen. Laut zahlreicher Beobachter:innen bereichern sie damit aber nicht nur das Methodenrepertoire der historischen Forschung, sondern remedieren und deformieren deren Prozesse in ebensolchem Ausmaß. Mit exemplarischem Fokus auf das kunstgeschichtli-

che Feld haben zum Beispiel Drucker¹⁶, Schelbert¹⁷, und Bishop¹⁸ diverse Aspekte und Dimensionen aufgezeigt, in denen Digital Humanities-Technologien traditionelle historiographische Standards und ihre komplexen interpretative und diskursiven Zielsetzungen verfehlen – oder ihnen noch nicht entsprechen. In diesem Kontext wird auch explizit auf die Bedeutung von benutzerzentriertem Design verwiesen, der angestrebte Nutzer:innen in den gesamten Entwicklungszyklus einbindet.¹⁹

Vor diesem Hintergrund ergibt sich die zentrale Validierungsstrategie für die skizzierten Methoden und Technologien des InTaVia-Projekts aus seinem partizipativen Design-Ansatz und aus dem geplanten Aufbau eines ganzen Repositoriums von kultur- und kunsthistorischen Fallbeispielen. Schon zu Beginn des Projekts wurden Expert:innen aus kunsthistorischer Forschung und praktischer Kulturvermittlung zu Workshops eingeladen, zu ihren Forschungspraktiken befragt und um Feedback zu den Konzepten und Komponenten der geplanten Plattform gebeten. Dies erlaubte einen unmittelbaren Abgleich von forschungsgetriebenen Entwicklungszielen mit den Bedürfnissen von etablierten Praktiker:innen, sowie den Aufbau von multifokalen und differenzierten Zielgruppen-Definitionen. So konnten Nutzer:innen-Profile für diverse Praxisfelder abgeleitet werden, inklusive den Bereichen der Ausstellungskuratierung, des Sammlungsmanagements, der schulischen und musealen Vermittlung von Kultur und (Kunst-)Geschichte, der (digitalen) geisteswissenschaftlichen und historischen Forschung, der Mediengestaltung im Bereich von Gedächtnisinstitutionen, der Fremdenführung, des Kulturtourismus und des Kulturjournalismus. In diesen unterschiedlichen Feldern sollen bereits im Laufe des Projekts gemeinsam mit praktischen und thematischen Expert:innen ein ganzes Repertoire von Fallstudien aufgebaut werden, um von den verschiedenen Fragestellungen, Praktiken, Ausgangsdaten und Zielsetzungen für die Anforderungen an die

16 Johanna Drucker, „Is there a ‚digital‘ art history?“ *Visual Resources* 29/1–2 (2013): 5–13. doi: 10.1080/01973762.2013.761106.

17 Georg Schelbert, „Art history in the world of digital humanities. Aspects of a difficult relationship,“ *Kunsttexte.de* 4 (2017): 1–10, <https://edoc.hu-berlin.de/bitstream/handle/18452/19404/Schelbert%20-%20final.pdf>, Zugriff am 25.02.2022.

18 Claire Bishop, „Against digital art history,“ *International Journal for Digital Art History* 3 (2018): 121–131.

19 Houda Lamqaddam et al., „When the tech kids are running too fast: Data visualisation through the lens of art history research,“ in *Proceedings of the 3rd IEEE Workshop for Visualization for the Digital Humanities (vis4dh)*, 2018, Berlin; Eva Mayr, Günther Schreder und Florian Windhager, „Digital HUMANities-Eine benutzerzentrierte Perspektive,“ in *Digital Humanities im deutschsprachigen Raum* (DHd 2018), Book of Abstracts, 2018, Köln.

Technologien zu lernen – und um unmittelbares Nutzer:innen-Feedback für die iterativen Zyklen der Technologieentwicklung zu generieren.

Eine weitere Strategie zur Harmonisierung von digitalen und traditionellen geisteswissenschaftlichen Perspektiven wird sich aus der Sicherstellung von konsistenten Zugriffsmöglichkeiten auf originale Texte und Objekte ergeben. So soll gerade im Kontext einer multiperspektivischen „Distant Reading & Viewing“-Umgebung die Verfügbarkeit von „Close Reading“-Konstellationen Historiker:innen eine direkte Einsicht in den biographischen Quelltext ermöglichen oder Kunsthistoriker:innen den Zugang zu einzelnen Werken und Objekten sichern. Entsprechende Benutzeroberflächen führen idealerweise die abstrakten Optionen der digitalen Makroanalyse mit dem Potential der hermeneutischen Mikroanalyse zusammen und erlauben variable Zyklen von elastischer Exploration und skalierbarer Lektüre.²⁰

Zu den wertvollsten Einsichten aus der ersten Feedback-Phase im Projekt gehörte auch die Erkenntnis der Bedeutung von individuellen Datenbeständen. Obwohl InTaVia durch die Zusammenführung von nationalen Datenbeständen (das heißt, den Nationalbiographien) und internationalen Aggregatoren (zum Beispiel Europeana) eine verhältnismäßig umfangreiche Datenbasis schaffen wird, argumentierten Forscher:innen und Praktiker:innen stringent für die Öffnung der Analyseumgebung in Richtung lokaler Biographie- und Objektbestände. Vor diesem Hintergrund wurden Konzepte und Optionen für den Import von lokalen Daten gestärkt. Dies wird es zum einen möglich machen, zahllose regionale Biographien und Sammlungen visuell zu analysieren, zu kommunizieren, und mit einem größeren europäischen Kontext kulturgeschichtlicher Daten in Beziehung zu setzen. Zum anderen kann so auch lokal und effektiv den diversen Formen von Bias und von (kultur-)historischer Selektivität entgegengewirkt werden, welche sich über die historische Kanonisierung der „ausgezeichnetsten Maler, Bildhauer und Baumeister“²¹ auch unübersehbar in digitale Datenbanken eingeschrieben haben.²²

20 Vgl. Mark-Jan Bludau et al., „Reading Traces: Scalable Exploration in Elastic Visualizations of Cultural Heritage Data,“ *Computer Graphics Forum* 39/3 (2020): 77–87, doi: 10.1111/cgf.13964.

21 Giorgio Vasari, *The lives of the most excellent painters, sculptors, and architects* (Überarbeitete Neuauflage, Original veröffentlicht um 1550) (Oxford, UK: Oxford University Press, 2008).

22 Katrin Glinka, Sebastian Meier und Marian Dörk, „Visualising the ‚Un-seen‘: Towards Critical Approaches and Strategies of Inclusion in Digital Cultural Heritage Interfaces,“ *Kultur und Informatik XIII* (2015): 105–118; Roopika Risam, „Decolonizing the digital humanities in theory and practice,“ in *The Routledge companion to media studies and digital humanities* (London: Routledge, 2018), 78–86.

Visuelle Analyseumgebungen, wie sie im Forschungsprojekt InTaVia entwickelt werden, können einen niederschweligen Zugang zu prosopographischen Datenbanken bieten ohne deren Funktionsumfang einzuschränken. Dadurch werden neue Dimensionen (zum Beispiel Anzahl der Personen, die gemeinsam betrachtet werden können) und Möglichkeiten (zum Beispiel Verknüpfung mit weiteren Datenquellen, wie beispielsweise dem materiellen kulturellen Erbe) auch für nicht- oder post-digitale Historiker:innen erschlossen. Ein zentraler Erfolgsfaktor für die Akzeptanz solcher Technologien ist dabei die konstante Einbeziehung der Zielgruppe(n) in den Entwicklungsprozess, da so nicht nur die Passung der Technologien mit nicht-digitalen Forschungsprozessen sichergestellt werden kann, sondern auch Einsichten in technologische Möglichkeiten und Grenzen zu erwarten sind, die für die wechselseitige Vermittlung von (digitalen) Humanities-Praktiken produktiv gemacht werden können.

Bibliographie

- Armitage, Neil. „The biographical network method.“ *Sociological Research Online* 21/2 (2016): 165–179. doi: 10.5153/sro.3827.
- Bernád, Ágoston Zénó und Maximilian Kaiser. „The Biographical Formula: Types and Dimensions of Biographical Networks.“ In *BD 2017*, hg. v. Antske Fokkens et al., 45–52. CEUR Workshop Proceedings, 2017.
- Beshero-Bondar, Elisa und Elizabeth Raisanen. „Recovering from Collective Memory Loss: The Digital Mitford’s Feminist Project.“ *Women’s History Review* 2/5 (2017): 738–750. doi: 10.1080/09612025.2016.1166882.
- Biehl, Theresia, Anne Lorenz und Dirk Osierenski. „Exilnetz33. Ein Forschungsportal als Such- und Visualisierungsinstrument.“ *Grenzen und Möglichkeiten der Digital Humanities*, hg. v. Constanze Baum und Thomas Stäcker (Wolfenbüttel: Herzog August Bibliothek, 2015). doi: 10.17175/sb001_011.
- Bishop, Claire. „Against digital art history.“ *International Journal for Digital Art History* 3 (2018): 121–131.
- Blessing, André, Andrea Glaser und Jonas Kuhn. „Biographical Data Exploration as a Test-bed for a Multi-view, Multi-method Approach in the Digital Humanities.“ In *BD2015*, hg. v. Serge ter Braake et al., 53–60, CEUR Workshop Proceedings, 2015.
- Bludau, Mark-Jan, Viktoria Brüggemann, Anna Busch und Marian Dörk. „Reading Traces: Scalable Exploration in Elastic Visualizations of Cultural Heritage Data.“ In *Computer Graphics Forum* 39/3 (2020): 77–87, doi: 10.1111/cgf.13964.
- Boyd Davis, Stephen und Florian Kräutli. „Scholarly chronographics: can a timeline be useful in historiography?“ In *Proceedings of 10th European Social Science History Conference* 2014.
- Bradley, John und Harold Short. „Texts into databases: The evolving field of new-style prosopography.“ *Literary and linguistic computing* 20 (2005): 3–24. doi: 10.1093/lilc/fqi022.

- Brehmer, Matthew, Bongshin Lee, Benjamin Bach, Nathalie Henry Riche und Tamara Munzner. „Timelines revisited: A design space and considerations for expressive storytelling.“ *IEEE transactions on visualization and computer graphics* 23/9 (2016): 2151–2164. doi: 10.1109/TVCG.2016.2614803.
- Brouwer, Judith und Harm Nijboer. „Golden Agents. A Web of Linked Biographical Data for the Dutch Golden Age.“ In *Proceedings of the Second Conference on Biographical Data in a Digital World 2017*, hg. v. Antske Fokkens et al., 33–38. CEUR Workshop Proceedings, 2017.
- Champagne, Marc. „Diagrams of the past: How timelines can aid the growth of historical knowledge.“ *Cognitive Semiotics* 9/1 (2016): 11–44. doi: 10.1515/cogsem-2016-0002.
- Drucker, Johanna. „Is there a ‚digital‘ art history?“ *Visual Resources* 29/1–2 (2013): 5–13. doi: 10.1080/01973762.2013.761106.
- Eccles, Ryan, Thomas Kapler, Robert Harper und William Wright. „Stories in geotime.“ *Information Visualization* 7/1 (2008): 3–17. doi: 10.1109/VAST.2007.4388992.
- Fokkens, Antske, Serge Ter Braake, Niels Ockeloen, Piek Vossen, Susan Legêne und Guus Schreiber. „BiographyNet: Methodological Issues when NLP supports historical research.“ In *LREC – Proceedings* (2014): 3728–3735.
- Fokkens, Antske, Serge Ter Braake, Niels Ockeloen, Piek Vossen, Susan Legêne, Guus Schreiber und Victor de Boer. „Biographynet: Extracting relations between people and events.“ In *Europa baut auf Biographien*. Aspekte, Bausteine, Normen und Standards für eine europäische Biographik, hg. v. Ágoston Zénó Bernád, Christine Gruber, Maximilian Kaiser unter Mitarbeit von Matthias Schlögl und Katalin Lejtovicz. 193–224. Wien: new academic press, 2017.
- Geerlings, Lonneke. „A Visual Analysis of Rosey E. Pool’s Correspondence Archives. Biographical Data, Intersectionality, and Social Network Analysis.“ In *BD2015*, hg. v. Serge ter Braake et al., 61–67. CEUR Workshop Proceedings, 2015.
- Gergaud, Olivier, Morgane Laouenan und Etienne Wasmer. „A Brief History of Human Time. Exploring a database of ‚notable people‘.“ (2017). Working paper, HAL Id: hal-01440325/
- Glinka, Katrin, Sebastian Meier und Marian Dörk. „Visualising the ‚Un-seen‘: Towards Critical Approaches and Strategies of Inclusion in Digital Cultural Heritage Interfaces.“ *Kultur und Informatik XIII* (2015): 105–118.
- Gonçalves, Tiago, Ana Paula Afonso und Bruno Martins. „Cartographic visualization of human trajectory data: Overview and analysis.“ *Journal of Location Based Services* 9/2 (2015): 138–166. doi: 10.1080/17489725.2015.1074736.
- Hyvönen, Eero, Petri Leskinen, Minna Tamper, Heikki Rantala, Esko Ikkala, Jouni Tuominen und Kirsi Keravuori. „BIOGRAPHYSAMPO—A Paradigm Shift for Publishing and Using Biography Collections on the Semantic Web.“ In *ESWC 2019: The Semantic Web*, 574–589. Cham: Springer.
- Jara-Figueroa, Cristian, Amy Z. Yu und César A. Hidalgo. „Estimating technological breaks in the size and composition of human collective memory from biographical data.“ *Preprint auf* <https://arxiv.org/abs/1512.05020> (2015).
- Khulusi, Richard, Jakob Kusnick, Josef Focht und Stefan Jänicke. „An interactive chart of biography.“ In *2019 IEEE Pacific Visualization Symposium (PacificVis)*, 257–266. IEEE, 2019.
- Koch, Steffen, Markus John, Michael Wörner, Andreas Müller und Thomas Ertl. „VarifocalReader – in-depth visual analysis of large text documents.“ *IEEE transactions on visualization and computer graphics* 20/12 (2014): 1723–1732. doi: 10.1109/TVCG.2014.2346677.

- Lamqaddam, Houda, Koenraad Brosens, Frederik Truyen, Rudy Jos Beerens, Inez De Prekel und Katrien Verbert. „When the tech kids are running too fast: Data visualisation through the lens of art history research.“ In *Proceedings of the 3rd IEEE Workshop for Visualization for the Digital Humanities (vis4dh)*, 2018, Berlin.
- Leskinen, Petri, Eero Antero Hyvönen und Jouni Antero Tuominen. „Analyzing and visualizing prosopographical linked data based on biographies.“ In *Proceedings of the Second Conference on Biographical Data in a Digital World 2017 (BD2017)*, hg. v. Antske Fokkens et al., 39–44. CEUR Workshop Proceedings, 2018.
- Mayr, Eva, Günther Schreder und Florian Windhager. „Digital HUMANities-Eine benutzerzentrierte Perspektive.“ In *Digital Humanities im deutschsprachigen Raum (DHD 2018), Book of Abstracts*, 2018, Köln.
- Mayr, Eva und Florian Windhager. „Once upon a spacetime: Visual storytelling in cognitive and geotemporal information spaces.“ *ISPRS International Journal of Geo-Information* 7/3 (2018): 96. doi: 10.3390/ijgi7030096.
- Meinecke, Christofer und Stefan Jänicke. „Visual Analysis of Engineers’ Biographies and Engineering Branches.“ In *LEVIA18: Leipzig Symposium on Visualization in Applications*. 2018.
- Risam, Roopika. „Decolonizing the digital humanities in theory and practice.“ In *The Routledge companion to media studies and digital humanities*, 78–86. London: Routledge, 2018.
- Russo, Irene, Tommaso Caselli und Monica Monachini. „Extracting and Visualising Biographical Events from Wikipedia.“ In *BD 2015*, hg. v. Serge ter Braake et al., 111–115. CEUR Workshop Proceedings, 2015.
- Schelbert, Georg. „Art history in the world of digital humanities. Aspects of a difficult relationship.“ (2017). *Kunsttexte.de*, 4, 1–10.
- Schich, Maximilian, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási und Dirk Helbing. „A network framework of cultural history.“ *science* 345/6196 (2014): 558–562.
- Schlögl, Matthias und Katalin Lejtovicz. „A Prosopographical Information System (APIS).“ In *BD 2017*, hg. v. Antske Fokkens et al., 53–58. CEUR Workshop Proceedings, 2017.
- Schmitz, Patrick und Laurie Pearce. „Humanist-centric tools for big data: berkeley prosopography services.“ In *Proceedings of the 2014 ACM symposium on Document engineering*, 179–188. ACM: 2014.
- Shakespeare, Daniel. „Interactive Genealogy Explorer: Visualization of Migration of Ancestors and Relatives.“ In *BD 2015*, hg. v. Serge ter Brake et al., 94–100. CEUR Workshop Proceedings, 2015.
- Vasari, Giorgio. „The lives of the most excellent painters, sculptors, and architects“ (Überarbeitete Neuauflage, Original veröffentlicht um 1550). Oxford, UK: Oxford University Press, 2008.
- Windhager, Florian, Matthias Schlögl, Maximilian Kaiser, Ágoston Zénó Bernád, Saminu Salisu und Eva Mayr. „Beyond One-Dimensional Portraits: A Synoptic Approach to the Visual Analysis of Biography Data.“ In *BD 2017*, hg. v. Antske Fokkens et al., 67–75. CEUR Workshop Proceedings, 2017.
- Windhager, Florian, Saminu Salisu und Eva Mayr. „Exhibiting uncertainty: Visualizing data quality indicators for cultural collections.“ In *Informatics*, 6/3 (2020): 29. doi: 10.3390/informatics6030029.
- Yu, Amy Zhao, Kevin Zeng Hu, Deepak Jagdish und Cesar A. Hidalgo. „Pantheon: visualizing historical cultural production.“ In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 289–290. IEEE, 2014.

Clemens Beckstein, Robert Gramsch-Stehfest, Clemens Beck,
Jan Engelhardt, Christian Knüpfer, Georg Zwillling

Digitale Prosopographie

Die automatisierte Auswertung des *Repertorium Germanicum*,
eines Quellenkorpus zur Geschichte geistlicher Eliten des
15. Jahrhunderts

Abstract: This paper presents a new approach to the automatic evaluation of the Repertorium Germanicum (RG), a collection of registers on late medieval German ecclesiastical history, based on modern compiler construction technology. The RG, which is very large by medieval standards, uses an elaborate system of abbreviations, which makes the contents of the records accessible in a summarised form. For the automatic evaluation of these registers, it is necessary to record the semantics of the compressed text. With ANTLR, a parser generator for structured text from compiler construction, the individual components in the text can be identified by using a grammar. The creation of the grammar is explained with an example and its individual components are described. Subsequently, it will be considered how the results of such a parsing process can be used for the development of a general ontology for the prosopography of clerical elites, which will form the starting point for comprehensive data collection and its further analysis.

Keywords: Repertorium Germanicum, Vatican sources, historical network analysis, language processing, prosopography, Church history, Middle Ages

Zusammenfassung: In diesem Beitrag wird ein neuer Ansatz zur automatischen Auswertung des Repertorium Germanicum (RG), eines Regestenwerks zur spätmittelalterlichen deutschen Kirchengeschichte, auf der Basis moderner Compilerbau-Technologie vorgestellt. Um das für mittelalterliche Verhältnisse sehr umfangreiche Quellenmaterial zu erschließen, verwendet das RG ein elaboriertes System von Abkürzungen, auf dessen Basis die Inhalte der Akten komprimiert dargestellt werden. Zur automatischen Auswertung dieser Regesten ist es notwendig, die Semantik des komprimierten Textes zu erfassen. Dafür wird ANTLR, ein Parsergenerator für strukturierten Text aus dem Compilerbau, verwendet, mit dem die einzelnen Bestandteile im Text mittels einer von uns entwickelten Grammatik identifiziert werden können. An einem Beispiel werden die Erstellung der Grammatik erläutert und ihre einzelnen Bestandteile erklärt. Dar-

an schließen sich Überlegungen an, wie die Ergebnisse eines solchen Parserprozesses für die Entwicklung einer allgemeinen Ontologie zur Prosopographie geistlicher Eliten genutzt werden können, die die Ausgangsbasis für eine umfassende Datenerhebung und deren weitere Analyse bilden.

Schlagwörter: Repertorium Germanicum, Vatikanische Quellen, Historische Netzwerkanalyse, Language Processing, Prosopographie, Kirchengeschichte, Mittelalter

Denn auf diesem Ozean von Papier, wo keine Sterne den Weg zeigen, würde der suchende Forscher lange umherfahren können, ehe er die Inseln findet, von denen er nicht einmal mit Sicherheit voraussagen kann, ob sie überhaupt existieren.

Johannes Haller (1903)

Als der bekannte Mittelalter- und Papsthistoriker Johannes Haller 1892 frisch promoviert nach Rom kam, bewarb er sich am dortigen Königlich Preußischen Historischen Institut als „Hilfsarbeiter“ bei einem Projekt, das seine wissenschaftliche Laufbahn mitbestimmen und dem er selbst seinen Stempel aufdrücken sollte, dem *Repertorium Germanicum* (RG).¹ In dieses „Verzeichnis der in den päpstlichen Registern und Kameralakten vorkommenden Personen, Kirchen und Orte des deutschen Reiches“ sollten alle Aktenstücke aus dem kurz zuvor für die wissenschaftliche Forschung geöffneten Vatikanischen Geheimarchiv Aufnahme finden, welche sich auf die deutsche Geschichte zwischen 1378 und 1521 bezogen. Angesichts von tausenden dickleibigen Folianten, die sorgsam auf deutsche Betreffe durchzusehen, zu exzerpieren und in Regestenform zu bringen waren, handelte es sich um eine wahre Mammutaufgabe. Und so veranlassten die Erfahrungen aus diesem Projekt den mittlerweile zum Marburger Professor aufgestiegenen Haller elf Jahre später zu einer denkwürdigen Stellungnahme:

¹ Zur Person vgl. Benjamin Hasselhorn, *Johannes Haller: Eine politische Gelehrtenbiographie: Mit einer Edition des unveröffentlichten Teils der Lebenserinnerungen Johannes Hallers* (Göttingen: Vandenhoeck & Rupprecht, 2015), darin 47–57 zu seiner römischen Zeit. Die Geschichte des Editionsprojekts *Repertorium Germanicum* wird geschildert in: Dieter Brosius, „Das Repertorium Germanicum“, in *Das Deutsche Historische Institut in Rom 1888–1988*, hg. v. Reinhard Elze und Arnold Esch (Tübingen: Niemeyer, 1990), 123–165; dazu auch Brigide Schwarz, „Das Repertorium Germanicum: Eine Einführung“, *Vierteljahrsschrift für Sozial- und Wirtschaftsgeschichte* 90 (2003): 429–440. Zur Benutzung des RG immer noch unentbehrlich: Walter Deeters, „Über das Repertorium Germanicum als Geschichtsquelle. Versuch einer methodischen Anleitung“, *Blätter für deutsche Landesgeschichte* 105 (1969): 27–43.

Mit der Quelle, aus der hier geschöpft wird, kann sich für diese Zeit keine andere in der Welt messen, sie übertrifft an Umfang und Mannigfaltigkeit schlechterdings alles, was aus dem ausgehenden Mittelalter überhaupt bekannt ist. [...]. So reich der Inhalt, so schwierig ist der Zugang. Vor diesen rund 4000 Bänden steht der einzelne Forscher ratlos und hilflos da [...]. Hier kann man wohl finden, aber nicht suchen.²

Scharfsinnig analysierte Haller, wie diese unübersehbare Aktenfülle effizient für die Zwecke mediävistischer Forschung erschlossen werden könne. Sein visionärer Vorschlag war, den Stoff über ein höchst platzsparendes Indizierungs-system zu erschließen, das an eine moderne Datenbank erinnert: „[...] was wir vor allem brauchen, ist ein alphabetischer Index. [...] Wird zu den Namen (Personen und Orten) außer dem Datum auch noch mit kurzem Schlagwort die Natur der Erwähnung hinzugesetzt [...] so ersetzt für den weiteren Kreis der Benutzer der Index die Regestensammlung.“³

Extrem knappe, ja geradezu kryptische Regesten in der Quellsprache Latein, die auf engstem Raum alle wichtigsten Informationen beinhalten, dazu umfangreiche Indices: Dieser Aufbau wurde zum „Markenzeichen“ des *Repertorium Germanicum*, das mittlerweile auch in die digitale Sphäre umgezogen ist (RG Online).⁴ Für mediävistische Verhältnisse handelt es sich regelrecht um „Big Data“ und so stellt sich die naheliegende Frage, ob dieses Material auch mit fortgeschrittenen Technologien automatisiert erschlossen werden könnte. Denn solange das Material nur von Einzelforschenden lesend im Buch oder am Bildschirm rezipiert wird, gilt noch immer das Wort von Johannes Haller, dass „auf diesem Ozean von Papier [...] der suchende Forscher lange umherfahren“ muss.⁵ An dieser Stelle setzt unser Jenaer Forschungsprojekt Core-H („Collaborative Open Research Environment for the Humanities“) an, in dem sich die Erschließung der RG-Regesten als Fallstudie (*case study*) in den breiteren Kontext der Entwicklung computergestützter Workflows für die historische Forschung einordnet.⁶

2 Johannes Haller, „Über das Repertorium Germanicum, Gutachten (Marburg, 5.4.1903)“, Archiv des DHI in Rom, Dienstakten A/8 (Acta generalia, Bd. VIII), 19 Bl., hier fol. 1. Zu dem Gutachten siehe auch Hasselhorn, *Johannes Haller*, 84–86.

3 Haller, „Repertorium Germanicum“, fol. 10.

4 Zugang über die Website des „Romana Repertoria“-Online-Portals des DHI in Rom: <http://www.romana-repertoria.net/993.html>, Zugriff am 06.10.2021. Vgl. dazu Jörg Hörnschemeyer, „Repertorium Germanicum Online“, in *Friedensnobelpreis und historische Grundlagenforschung. Ludwig Quidde und die Erschließung der kurialen Registerüberlieferung*, hg. v. Michael Matheus (Berlin: De Gruyter, 2012), 605–615.

5 Haller, „Repertorium Germanicum“, fol. 10.

6 Vgl. die Projektwebsite unter <https://www.mephisto.uni-jena.de/projekte/core-h> sowie die Blogbeiträge unter <https://www.mephisto.uni-jena.de/blog>, Zugriff am 06.10.2021.

Schauen wir uns einige typische Regesten aus dem RG Online (RG V) an – einen Ausschnitt aus der Pfründenvita⁷ des Klerikers Johannes (Alberti) von Echte⁸ aus den Jahren 1431 und 1437 (Abbildung 1).



RG V 04399

zur Merkliste hinzufügen

Johannes de Echte Johannes (Alberti) de Echte Johannes de Eckt: l.b. iul. 1431 T 3 25v.

cler. Magunt. dioc.: de vicar. ad alt. ss. Laurentii, Cosme et Damiani atque Liberii in eccl. s. Crucis Northusen. Magunt. dioc. (4 m. arg.) vac. p. resign. Johannis Wittenborn; n.o. disp. sup. def. nat. (subdiac., c.) 24. nov. 1431 S 272 81v.

24. nov. 1431

(...)

pres. in R. cur.: de vicar. ad alt. ss. Andree et Nicolai in colleg. eccl. s. Crucis Northusen. Magunt. dioc. (5 m. arg.) vac. p. resign. Hermannii Rulen presb. vel p.o. Johannis de Bendeleyben; n.o. vicar. in capel. s. Spiritus e.m. op. Duderstat d. dioc. (2 m. arg.), disp. sup. def. nat. (subdiac., c.) et quod pater ipsius J. cantor. c. can. et maior. preb. d. eccl. s. Crucis Northusen. obtin. 2. nov. 1437 S 341 224vs.

2. nov. 1437

Abb. 1: Ausschnitt aus der Pfründenvita des Klerikers Johannes (Alberti) von Echte im RG V, Nr. 4399 (Screenshot nach RG Online, RG IV 00133, URL: <<http://rg-online.dhi-roma.it/RG/4/133>> (Datum 14.10.2021)).

Im ersten Sublemma finden wir den vollständigen Namen des Geistlichen. Es folgt ein auf den 24. November 1431 datiertes Schreiben, in dem sich Johannes auf eine Vikarie am Altar St. Laurentius und weiteren in der Stiftskirche Heiligkreuz Nordhausen bewirbt, welche nach seiner Angabe ein jährliches Einkommen von vier Silbermark hatte. Weiter wird angegeben, dass die Pfründe durch den Verzicht (Resignation) eines gewissen Johannes Wittenborn frei (vakant)

⁷ Pfründen sind mit konkreten Einkünften versehene kirchliche Ämter.

⁸ Zu dessen Biographie vgl. Robert Gramsch, „Kurientätigkeit als ‚Berufsbild‘ gelehrter Juristen. Der Beitrag Roms zur Akademisierung Deutschlands im Spätmittelalter. Eine personengeschichtliche Betrachtung,“ *Quellen und Forschung aus italienischen Archiven und Bibliotheken* 80 (2000): 138–143. Eine Pfründenvita fasst alle Quellenbelege aus einem Pontifikat zusammen, die einem bestimmten Kleriker als Bittsteller (sogenannter Petent) zuzuordnen sind. Dieselbe Person kann darüber hinaus auch in den Pfründenviten anderer Petenten genannt sein, etwa als Konkurrent in einem Prozess. Im RG Online können Personen, Orte und so weiter über eine Stringsuche gefunden werden, die allerdings wegen variierender Schreibweisen nicht treffgenau ist.

geworden war, und dass Johannes von Echte eine Geburtsdispens besaß, eine Ausnahmegenehmigung, Pfründen erwerben zu dürfen, obwohl er aus der Verbindung eines Subdiakons und einer verheirateten Frau hervorgegangen war.⁹ Das schließende Kürzel „S 272 81v“ gibt die Fundstelle im Vatikanischen Archiv an.

Im Regest zum 2. November 1437 erfahren wir, dass sich Johannes von Echte damals am Papsthof (*in Romana curia*) aufhielt und dass er sich um eine weitere Altarvikarie in Nordhausen bewarb. Diese war durch den Pfründenverzicht eines gewissen Hermann Rulen frei geworden, sowie (zuvor) durch den Tod des Johannes von Bendeleyben. Nach einem Semikolon folgt das Kürzel „n.o.“, welches den Beginn der sogenannten Nonobstanzenformel markiert, der Angabe, welche Pfründen Johannes von Echte bereits im Besitz hatte und deren Besitz dem Erwerb des erbetenen Altars nicht im Wege stehen sollte. Auch hier wird nochmals die Geburtsdispens genannt, diesmal konkretisiert durch die Angabe, dass Johannes' Vater Kantor am Heiligkreuzstift war.

Regesten dieser Art automatisiert beziehungsweise mit gewisser menschlicher Hilfe semiautomatisiert auszuwerten, wird durch mehrere Umstände entschieden vereinfacht: Zum einen sind es immer wieder dieselben Gegenstände, die den Inhalt der Regesten ausmachen: Bittschriften, päpstliche Gratialschreiben und weitere Aktenstücke, die sich um den Erwerb von kirchlichen Ämtern und Einkünften drehen.¹⁰ Schon im spätmittelalterlichen Verwaltungsbetrieb der kurialen Behörden herrschten hier bürokratische Effizienz und Routine, wenn mit den immer gleichen Rechtsformeln ähnliche Sachverhalte, bei denen nur die Namen der beteiligten Personen und zur Disposition stehenden Pfründen variieren, dokumentiert wurden. Der Inhalt dieser Schreiben lässt sich somit durch den modernen Bearbeiter recht einfach auf den wesentlichen Gehalt reduzieren, hinreichend charakterisiert durch einige hundert *termini technici* (wobei freilich so manches individualisierende Detail wegfällt, das erst der Blick auf das Originaldokument enthüllt). Mit dem entsprechenden kirchenrechtlichen Hintergrundwissen lässt sich mithin aus vergleichsweise wenigen Worten, die zudem noch zu platzsparenden Abkürzungen ohne grammatikalische Finessen verkürzt worden sind, schon das Wesentliche erkennen. Das RG Online macht dieses Material digital abfragbar – allerdings durch den Flaschen-

⁹ Zum Phänomen der Klerikerkinder, die wegen dieses „Geburtsdefekts“ (*defectus natalium*) einer päpstlichen Dispens bedurften, um in der Kirche Karriere machen zu können, vgl. Ludwig Schmugge, *Kirche, Kinder, Karrieren: Päpstliche Dispense von der unehelichen Geburt im Spätmittelalter* (Zürich: Artemis, 1995).

¹⁰ Zu diesem Phänomen vgl. etwa Brigide Schwarz, „Römische Kurie und Pfründenmarkt im Spätmittelalter,“ *Zeitschrift für Historische Forschung* 20 (1993): 129–152.

hals einer Suchmaske, die nur Einzelabfragen und ein relativ mühseliges Datensammeln per *copy & paste* erlaubt. Hat man hingegen die Regestentexte digital zur Verfügung, so rückt die Möglichkeit eines automatisierten *Information Retrieval* in greifbare Nähe.

Gestützt auf den digitalen 10. Band (zum Pontifikat Sixtus' IV. 1471–1484) haben wir 2019 eine erste „Probeforschung“ unternommen. Ihr Ziel war, jene Belege aufzufinden und automatisch auszuwerten, die den Geldtransfer zwischen Deutschland und Rom in Form der sogenannten Annaten zum Gegenstand haben. Diese wurden fällig, wenn ein Kleriker auf päpstliche Verleihung hin eine geistliche Pfründe erhielt, deren ersten halben Jahresertrag er als Gebühr an den Papst zu entrichten hatte. Uns interessierte vor allem, welche individuellen und institutionellen Akteure (Bankhäuser, aber auch Kurienpersonal) an diesen Transfers beteiligt waren. Eine solche Forschungsaufgabe wäre früher nur mit hohem Aufwand zu lösen gewesen, denn es hätte endlosen Nachschlagens in der Edition und einer peniblen Listenführung bedurft, um allein die Daten zu erheben. Unser Workflow ermöglichte es hingegen, die einschlägigen Regesten automatisch aus der Stoffmasse des RG herauszufiltern. Dafür wurden die Einträge in der TUSTEP-Datei des RG-Bandes mit einem Python-Skript nach der Zeichenkette „*solv. pro annat.*“ durchsucht. Zu einer so gefundenen Annatenzahlung wurden dann das Datum des Vorgangs, die Regestennummer sowie der Auftraggeber der Zahlung aus dem Lemma extrahiert. Weiterhin wurden der Übermittler der Zahlung, die zur Disposition stehende Pfründe und die Gebührenhöhe ermittelt. Diese Daten wurden zur weiteren Verarbeitung in eine CSV-Datei geschrieben. Mit der Programmiersprache R wurde anschließend das Personennetzwerk der Annatenzahlungen analysiert und visualisiert. Die Knoten in einem solchen Netzwerk sind die historischen Personen, eine gerichtete Kante verbindet den Gebührenzahler mit dem involvierten Finanzakteur (ein Bankhaus oder auch eine konkrete Person). Über Zentralitäts- und Clusteranalysen konnten so die wichtigsten Agenten in diesem Netzwerk identifiziert und Strukturmuster des damaligen kurialen Finanzsystems herausgearbeitet werden.¹¹

Freilich hat sich der hier eingeschlagene Weg, über *regular expressions* den Inhalt der Lemmata zu erschließen, als problematisch für komplexere Auswertungen erwiesen. Benötigt wurde ein digitales Werkzeug, das in der Lage ist, die charakteristische Struktur eines jeden Regestentyps eindeutig zu identifizieren und dafür in einem ersten Schritt der späteren digitalen Auswertung einen kodierenden Parsebaum zu generieren. Eine Grammatik im informatischen Sin-

¹¹ Vgl. hierzu Robert Gramsch-Stehfest, „Von der Metapher zur Methode: Netzwerkanalyse als Instrument zur Erforschung vormoderner Gesellschaften,“ *Zeitschrift für Historische Forschung* 47 (2020): 20–28.

ne bildet dabei alle Optionen ab, in der ein bestimmter Regestentyp ausgeformt sein kann. Der Prozess der Erstellung einer solchen Grammatik setzt einen engen Feedbackloop zwischen den daran zwingend beteiligten Historiker:innen und Informatiker:innen voraus. Für die Erstellung unserer RG-spezifischen Grammatik und die sich darauf stützende Extrahierung der Informationen aus dem Text nutzen wir seit dem Frühjahr 2021 mit bemerkenswertem Erfolg ANTLR (ANother Tool for Language Recognition)¹², ein ursprünglich für den Compilerbau entwickeltes, sehr mächtiges Werkzeug zur schnellen und komfortablen Implementierung domänenspezifischer Sprachen (DSLs), die sich über eine formale Grammatik charakterisieren lassen, also zum Beispiel von Programmiersprachen, Datenbeschreibungssprachen, Interpretern oder Textanalyse- und Texttransformationswerkzeugen jedweder Art.

ANTLR ist ein Programmgenerator, der mit Hilfe des rekursiven Abstieges, passend zu einer vorgegebenen Grammatik einen top-down Parser, also das eigentliche, später für die Textanalyse eingesetzte digitale Werkzeug erzeugt.¹³ Bei diesem Verfahren kommt jedem Nichtterminalsymbol der Grammatik eine (von ANTLR automatisch generierte) Prozedur zu, die die Grammatikregel zu diesem Symbol für die (gegebenenfalls rekursive) Konkretisierung des zu diesem Symbol gehörenden Unterbaums im generierten Parsebaum benutzt. Die digitale Analyse und Transformation von strukturierten Texten (DSLs im oben genannten Sinne) erfolgt mit und durch ANTLR grob vereinfacht in drei aufeinanderfolgenden Schritten:

1. Es wird eine lexikalische Analyse des Textes vorgenommen, das heißt im Eingabetext werden die von der Grammatik zugelassenen elementaren Textbausteine (die Tokens, das heißt die „Wörter“ der von der Grammatik implizit definierten formalen Sprache) identifiziert, die dann im weiteren Analyseprozess die Grundlage für die Konstruktion von Parsebäumen bilden.
2. Der von ANTLR automatisch generierte Parser startet den Prozess der schrittweisen Parsebaumgenerierung, indem er die Prozedur zu derjenigen Grammatikregel aktiviert, die nach Einschätzung von ANTLR den größtmöglichen Teil des Eingabetextes bestmöglich abdeckt. Die von ANTLR er-

¹² Siehe <https://www.antlr.org/>.

¹³ Vgl. Alfred V. Aho, Ravi Sethi und Jeffrey D. Ullman, *Compilers. Principles, Techniques, and Tools* (Reading, Massachusetts et al.: Addison-Wesley, 1986).

zeugten Parser verwenden dafür die ALL(*) Parsertechnik, die Terence Parr, der Entwickler des Werkzeuges, zusammen mit Sam Harwell entwickelte.¹⁴

3. ANTLR baut nach von Nutzenden vorgebbaren Transformationsregeln (unter anderem über sogenannte *tree-walker*, siehe zum Beispiel Abbildung 3 unten) den im zweiten Verarbeitungsschritt berechneten „besten“ Parsebaum im Rahmen einer automatischen Traversierung in eine Form um, die für die weitere domänenspezifische Nutzung, Analyse und Visualisierung der in dem Baum kodierten Informationen optimal ist.

Diese für ANTLR typische Untergliederung des Verarbeitungsprozesses in drei jeweils für sich getrennt spezifizierbare Verarbeitungsschritte gestattet es, die eigentliche Analyse der Quelle weitgehend unabhängig von den sich daran anschließenden, domänenspezifischen digitalen Nutzungen des Analyseergebnisses zu halten: Ein und dieselbe ANTLR-Grammatik lässt sich so für die unterschiedlichsten Auswertungen von Quelltexten nutzen, sofern diese Texte nur korrekt durch diese Grammatik beschrieben werden. Und man ist bei diesem Ansatz nicht nur frei darin, welche Art von weitergehender Nutzung der von ANTLR generierten, sprachspezifischen Analyseergebnisse man realisieren möchte, sondern auch darin, wie man diese Nutzung unter Verwendung der dafür gegebenenfalls vorhandenen Werkzeuge und Infrastrukturelemente (Zielprogrammiersprachen, Visualisierer, Nonstandard-Datenbanken, Plattformen und vieles mehr), konkret realisiert. ANTLR ermöglicht so auf eine unkomplizierte und gerade auch für Nicht-Informatiker:innen nachvollziehbare Weise eine komfortable Entwicklung vergleichsweise komplexer sprachtechnologischer Werkzeuge.

Der für Anwender:innen herausforderndste Schritt in diesem Entwicklungsprozess ist natürlich die Identifikation der für die beiden ersten Verarbeitungsphasen benötigten Grammatik. Denn diese muss für die ersten beiden, primär quellenabhängigen Verarbeitungsschritte die Struktur der zu analysierenden Quelltexte angemessen charakterisieren. In unserem konkreten Anwendungsfall erfordert das insbesondere die Ausdifferenzierung der einzelnen Regestentypen, eine Aufgabe, die hauptverantwortlich durch Historiker:innen vorgenommen wird. Eine notwendige Voraussetzung dafür ist die genaue Kenntnis des *Repertorium Germanicum* und seiner standardisierten Abkürzungen sowie der Syntax und Semantik der Regesten selbst. Historiker:innen müssen dafür wissen, welche Abkürzungen zu welchem Rechtsakt gehören und welche von uns so genannten „Markerworte“ den entsprechenden Vorgang kennzeichnen.

¹⁴ Terence Parr, *Language Implementation Patterns*, (Dallas, Texas et al.: Pragmatic Bookshelf, 2010).

Beispielsweise steht im dritten Band des RG ein *solv.* immer am Anfang einer Annatenzahlung oder ein *prov. super* immer am Anfang einer Provision (das heißt der Verleihung einer kirchlichen Pfründe durch den Papst).

Über ein systematisches Studium der Viten müssen zu Beginn des Prozesses zur Erstellung von Grammatikregeln nicht nur die Regelmäßigkeiten im Aufbau eines bestimmten Regestentyps herausgearbeitet werden, sondern auch, welche Anteile bei Vorliegen dieses Typs optional sind. Mittels einer Suche nach Markern im RG Online lässt sich dafür zunächst eine Auswahl von Regesten zusammenstellen, die zur gleichen Kategorie gehören. Dann lassen sich Regelmäßigkeiten innerhalb dieser Regesten identifizieren und mögliche variierende Ausprägungen ermitteln.

Erst durch das Zerlegen vieler gleichartiger Einträge nach dem gleichen Schema wird offensichtlich, welche Bestandteile dieses Schemas in den Regesten optional und welche obligatorisch sind. Auch hier ist historisches Wissen notwendig, um konkrete Abkürzungen, die an bestimmten Stellen eines Eintrages auftreten können, korrekt zu interpretieren. Beispielsweise können bei einer Provision nach dem Markerwort *prov. de* eine Pfarrkirche (*par. eccl.*) oder ein Kanonikat mit Pfründe (*can. et preb.*) stehen. Dies bedeutet, dass an jener Stelle wahrscheinlich alle Arten von kirchlichen Pfründen auftreten können. Anstatt nun in einer Grammatikregel alle möglichen Pfründen an jeder Stelle ihres Auftretens in der Grammatik aufzuzählen, wird für die Grammatik besser eine eigene Kategorie „Pfründe“ definiert, unter die sich unter Heranziehung des Abkürzungsverzeichnisses des RG Online alle Abkürzungen subsumieren lassen, die in diese Kategorie fallen.

Im nächsten Schritt wird die Grammatik anhand von weiteren Sublemmata getestet, die bei der eigentlichen Entwicklung der Regeln noch nicht in Betracht gezogen wurden, jedoch mit dem gleichen Markerwort beginnen. Immer wenn die Grammatik nicht in der Lage ist, das entsprechende Sublemma in einem Teilbaum korrekt abzubilden, müssen Anpassungen (ein *debugging*) der Grammatik vorgenommen werden. Im einfachsten Fall muss dazu nur eine bestimmte Ausprägung in einer schon bestehenden Kategorie ergänzt werden, in komplexeren Fällen muss die entsprechende Regel für einen Vorgangstyp grundsätzlich alteriert werden.

Im Folgenden wollen wir dies an einem Beispiel, nämlich dem oben bereits vorgestellten ersten Sublemma der Vita des Johannes von Echte (Abbildung 1) demonstrieren.

In einer mit XML (Extensible Markup Language) ausgezeichneten Version wird dieses Regest wie folgt dargestellt:

```
<lemma id="10504399" spStart="" spEnd=""><reg>
```

```
<head> Johannes de Echte Johannes (Alberti) de Echte </head>
```

```
<sublemma vol="5"> Johannes de Eck: <abk ref="415">l.b.</abk> <date  
norm="14310700" year="1431">iu1. 1431</date> <fund norm="T:3:25v" l1="T"  
l2="3" l3="25v">T 3 25v</fund>.</sublemma>
```

```
<sublemma vol="5"><abk ref="202">cler.</abk> <abk ref="33">Magunt.  
</abk> <abk ref="296">dioc.</abk>: de <abk ref="792">vicar.</abk> ad <abk  
ref="109">alt.</abk> <abk ref="726">ss.</abk> Laurentii, Cosme et Da-  
miani atque Liberii in <abk ref="312">eccl.</abk> <abk ref="671">s.</abk>  
Crucis Northusen. <abk ref="33">Magunt.</abk> <abk ref="296">dioc.</abk>  
(4 <abk ref="438">m. arg.</abk>) <abk ref="787">vac.</abk> <abk  
ref="521">p.</abk> <abk ref="658">resign.</abk> Johannis Wittenborn;  
<abk ref="468">n.o.</abk> <abk ref="297">disp.</abk> <abk ref="743">sup.  
</abk> <abk ref="267">def.</abk> <abk ref="469">nat.</abk> (<abk  
ref="732">subdiac.</abk>, <abk ref="166">c.</abk>) <date  
norm="14311124" year="1431">24. nov. 1431</date> <fund norm="S:272:81v"  
l1="S" l2="272" l3="81v">S 272 81v</fund>.</sublemma>
```

Deutlich erkennbar ist eine Differenzierung nach Head und mehreren Sublemmata. Im Head stehen Angaben zu dem Antragsteller. Handelt es sich um eine natürliche Person (und nicht beispielsweise um eine juristische Person wie ein Kloster oder ein Domkapitel), dann werden regelmäßig ihr Name und gegebenenfalls weitere individualisierende Angaben wie beispielsweise der akademische Grad und die absolvierten Studienfächer, ihr Weihegrad, eine Ordenszugehörigkeit oder weltliche und geistliche Ämter, die die Person bekleidet, genannt. Diese persönlichen Angaben fassen wir in der Grammatik unter einer Regel mit dem (noch vorläufigen) Namen „Vita“ zusammen. In der ANTLR-Grammatik kann unsere Vita-Regel wie folgt Ausprägungen festlegen, die ihrerseits wieder über speziellere Unterregeln zu charakterisieren sind:

```
vita : ehestand? (funktion | inkardination | akadGrad | kirchAmt | herkunft  
| stand | orden | familienstand | pfruende | kardinal | titel )+
```

Bei unserem Beispiel Johannes de Echte sind diese einleitenden biographischen Informationen sehr überschaubar. Angesichts der fehlenden biographischen Angaben im Head konzentrieren wir uns im Folgenden auf eine genauere Erläuterung des Sublemmas. Ein solches Sublemma wird in ANTLR zum Beispiel wie folgt dargestellt:

sublemma : SUBLEMMATAG vita? person? abks?

(provision | annatenVerkehr | dispense | gunstGP | schutzbrieff | nachfolge
| eid | expektanz | verlaengerung | nonObs | ausstellung | erlaubnis |
gruendung)+

(date | fund)+ CLOSESUBLEMMATAG ;

Hier ist zunächst einmal allgemein festgehalten, dass in einem Sublemma biographische Informationen zum Petenten (im obigen Sinne), eine (weitere) Person oder eine Abkürzung auftreten können. Anschließend folgen in ihm regelmäßig einer oder auch mehrere der im RG vorkommenden Rechtsvorgänge. Dies wird im Sublemma durch das „+“ hinter dem geklammerten Ausdruck signalisiert, der die jeweils möglichen Ausprägungen des Sublemmas durch ein „|“ (es steht logisch für ein „oder“) voneinander getrennt aufzählt. Abgeschlossen wird jedes solche Sublemma durch Tokens des Typs Datum und Fundstelle. Da es sich um eine Arbeitsversion des Parsers handelt, ist diese Liste bei Weitem noch nicht vollständig. Endziel ist es, alle möglichen (respektive die häufigen) Ausprägungen von Sublemmata in jeweils eigenen Kategorien fassen zu können.

Das erste Sublemma der Vita Johannes von Echtes beinhaltet eine Provision. Bei dem Beispiel handelt es sich um eine „normale“ Provision, daher kürzen wir sie mit „nProvision“ ab. Ihr Sublemma ist wie folgt aufgebaut:

nProvision: intro (pfruende acquType? Schätzwert?)+;

Eine normale Provision erwähnt also nach dem typischen Intro, das im 5. Band des RG nur aus „de“ besteht, eine Pfründe, einen optionalen Erwerbstyp („AcquType“, das nachgestellte „?“ signalisiert, dass er auch fehlen kann) und den optionalen Schätzwert der Pfründe. Das „+“ nach dem geklammerten Ausdruck zeigt wieder an, dass sich der eingeklammerte Teil mehrmals wiederholen und dadurch beispielsweise auch der Schätzwert vor dem Erwerbstypen stehen kann. Die Erwerbstypen sind selbst wieder ausdifferenziert, je nachdem um welche Arten von Pfründen es sich handelt.

Da sich Johannes von Echtes um Altäre bewarb, leitet ANTLR den entsprechenden Teil der Regesten korrekt mit der Regel für Altar-Pfründen ab. Diese haben wir wie folgt definiert:

pfruendenAlt : altarIntro (patroName | 'et' | 'atque')* #'situm'? instType
patroName? location;

Das Altar-Intro kennzeichnet dabei die verschiedenen möglichen Stellen wie zum Beispiel die Altarvikarie (oder das Fehlen einer konkreten Angabe, wenn sich direkt um den Altar beworben wurde):

```
('<abk ref="791">vic.</abk>' | '<abk ref="792">vicar.</abk>')? 'ad'?  
('<abk ref="109">alt.</abk>' | 'altare')
```

Deutlich zu sehen ist, dass im RG nach einem Altar-Intro ein Patrozinium (oder mehrere durch *et* oder *atque* verbundene Patrozinien) des Altars sowie die Institution (in der Regel eine Kirche), in der der Altar gelegen ist, stehen kann. Diese Institution kann auch selbst über ein Patrozinium verfügen, wobei die Optionalität wieder durch das nachgestellte „?“ signalisiert wird. Der Institution folgen dann der Standort und die Diözese, was wir in unserer ANTLR-Grammatik in einer sogenannten Location-Regel zusammengefasst haben. Zuletzt muss der Schätzwert der Pfründe charakterisiert werden, um die Definition zu vervollständigen. Unserer Regel zufolge besteht er nur aus einem Integer-Wert und einer durch XML-Tags ausgezeichneten Abkürzung für eine Währung:

```
schätzwert : INT?  
( '<abk ref="354">fl.</abk>' |  
  | '<abk ref="716">sol.</abk>' |  
  | '<abk ref="277">den.</abk>' |  
  | '<abk ref="354">fl.</abk> Renen.' |  
  | '<abk ref="368">gr.</abk>' |  
  | '<abk ref="438">m. arg.</abk>' |  
  | 'fl' )
```

Nach dem Schätzwert folgt entsprechend unserer Grammatikregeln der Erwerbstyp. Die im Beispiel des Johannes von Echte genannte Vakanz durch Resignation des Vorbesitzers, Johannis Wittenborn, ist nur eine mögliche Erwerbs-Ausprägung. Weitere Vakanzgründe umfassen beispielsweise den Tod oder die Absetzung des Vorbesitzers; die Möglichkeit, dass der Vorbesitzer die Pfründe mit dem Antragsteller getauscht hat; die Heirat des Vorbesitzers, womit er aus dem geistlichen Stand ausgeschieden war; oder sogar, dass die Gründe der Vakanz unbekannt sind.

Nach dem Namen des Vorbesitzers können in einem Regest zusätzliche persönliche Informationen zu dieser Person stehen, beispielsweise deren Ordenszugehörigkeit oder weitere Pfründen und Ämter, die der Vorbesitzer bekleidet. Eine weitere wichtige (kirchenrechtlich und damit historisch relevante) Information ist der sogenannte Kurien-Status der Pfründe, da das Besetzungsrecht

auf Pfründen, die während eines Aufenthaltes des Vorbesitzers in Rom frei geworden waren, exklusiv beim Papst lag.

```
acquType: (tausch | tod | verzicht | klostereintritt | hochzeit |
translation | absetzung | uebertritt | befoerderung | widerruf | unbekannt
) natPerson (('de' | 'ad') (orden | pfruende))? kurienStatus? ('seu'
acquType)? /*imAngesicht?*/;
```

Ebenfalls zu berücksichtigen sind die unterschiedlichen Tempora für das Freiwerden der Pfründe. Im Beispiel des Johannes von Echte handelt es sich aufgrund der Formel „vac. p. resign.“ um einen Verzicht in der Gegenwart („verzichtistfrei“).

```
verzichtistfrei: '<abk ref="787">vac.</abk>'? ('per' | 'per.' | p. |
'post')? '<abk ref="658">resign.</abk>' ; //
```

Nach dem Erwerbstyp folgt in dem ersten Beispiel von Johannes von Echte das Kürzel „n.o.“, was den Beginn der Nonobstanzenformel markiert. Als Nonobstantien werden in der Regel vor allem Pfründen genannt, die der Petent bereits im Besitz hat. Hierfür enthält unsere Grammatik aus den vorab genannten Gründen bereits die entsprechende Kategorie „pfruende“. Als weitere mögliche Kategorie treten die Dispense hinzu – im vorliegenden Fall die Geburtsdispens. Somit lautet eine naheliegende Grammatikregel für die Nonobstanzenformel:

```
nonObs : '<abk ref="468">n. o.</abk>' ('et')? (pfruende | dispense))+';
```

Der auf der Basis dieser Grammatik durch ANTLR nach den ersten beiden Verarbeitungsschritten generierte Baum sieht wie folgt aus:

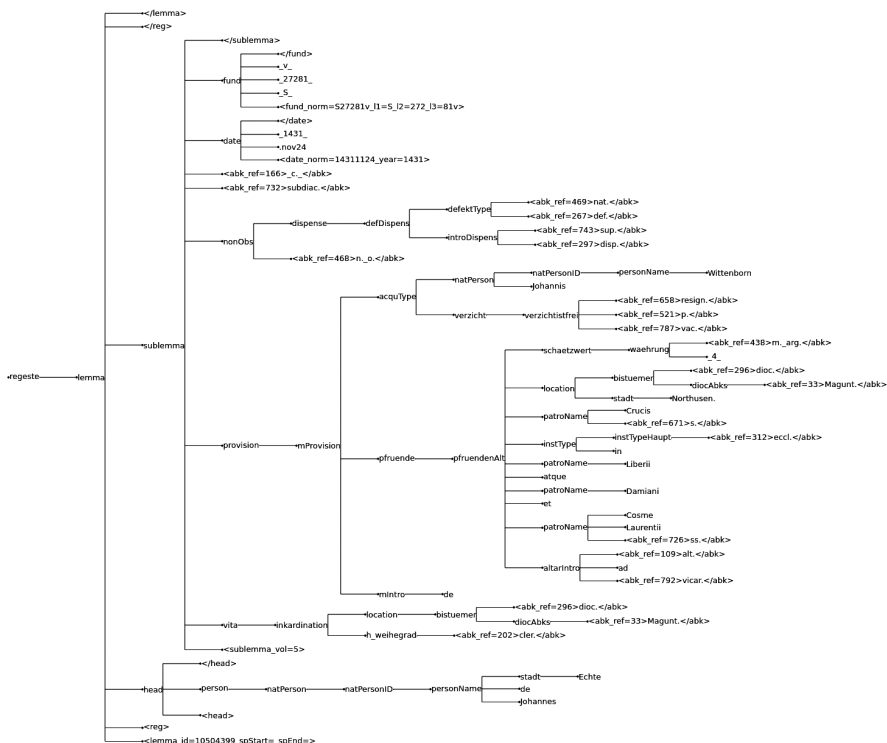


Abb. 2: ANTLR-Baum des ersten Sublemmas der Pfründenvita des Klerikers Johannes (Alberti) von Echte (RG V 4399)

Abgesehen von einem kleinen, noch unaufgelösten Bestandteil (erst das RG V erhält erläuternde Angaben zur Geburtsdispens, dass nämlich der Vater Subdiakon, die Mutter eine verheiratete Frau – *c[oniugata]* – war), ist es uns also gelungen, mit Hilfe unserer anhand vom 3. Band entwickelten Grammatik alle Bestandteile des Sublemmas aus dem RG V inhaltlich sinnvoll mit dem von ANTLR dafür generierten Textanalysetool zu parsen. Den dabei entstandenen Parsebaum ließen wir dann noch im Rahmen des (hier nicht dargelegten) dritten Verarbeitungsschrittes von ANTLR visualisieren und nach XML transformieren, wodurch die in ihm kodierten Informationen sowohl für vertiefte algorithmische Auswertungen als auch generell für eine funktionale Erweiterung der RG Online-Version zur Verfügung stehen.

Die Entwicklung einer ANTLR-Grammatik für die RG-Regesten stellt bei einem geschickten Design der dafür gewählten Regelverfeinerungsstruktur quasi als „Nebenprodukt“ ein strukturiertes Begriffsgebäude zur Verfügung, um die

im *Repertorium Germanicum* enthaltenen biographischen und anderen Informationen in inhaltlich angemessener Weise kategorial zu ordnen und zu repräsentieren. Das Spezialwissen von Historiker:innen, wie bestimmte Quellenbefunde zu lesen und auszudeuten sind, wird somit explizit in das Datenmaterial inkorporiert, was dessen weitergehende Analyse nach unterschiedlichsten Fragestellungen gestattet. Ein Sublemma wie zum Beispiel das in der Vita des Johannes von Echte zum 2. November 1437 überlieferte, kann dann in ganz verschiedener Hinsicht personen- und institutionengeschichtlich oder auch netzwerkanalytisch ausgewertet werden (Abbildung 3).

pres. in R. cur.: de vicar. ad alt. ss. Andree et Nicolai in colleg. eccl. s. Crucis Northusen. Magunt. dioc. (5 m. arg.) vac. p. resign. Hermann Rulen presb. vel p.o. Johannis de Bendeleyben; n.o. vicar. in capel. s. Spiritus e.m. op. Duderstat d. dioc. (2 m. arg.), disp. sup. def. nat. (subdiac., c.) et quod pater ipsius J. cantor. c. can. et maior. preb. d. eccl. s. Crucis Northusen. obtin. 2. nov. 1437 S 341 224vs.

2. nov. 1437

Aus der Vita des Johannes von Echte im RG V

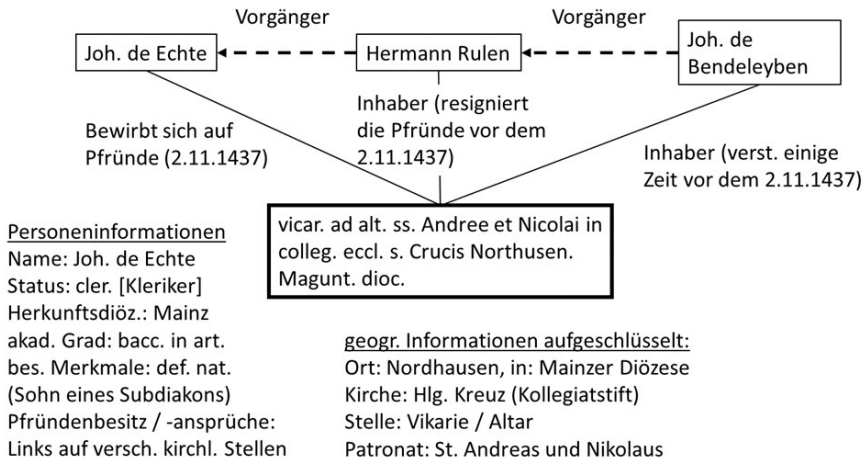


Abb. 3: Unterschiedliche Informationen aus einem Sublemma der Vita des Johannes von Echte (RG V 4399)

Die Abbildung zeigt, wie umfangreich und komplex die Sachverhalte sind, die schon in einem einzigen von zigtausenden RG-Einträgen dokumentiert sind. Diese Informationen möglichst erschöpfend automatisiert aus der Quelle herauszulesen, ist die Herausforderung, vor der wir stehen. Ermöglicht werden wird dies aber erst durch eine Ontologie, die nicht nur die entsprechenden *termini technici* enthält, sondern auch Abhängigkeiten zwischen diesen Begriffen beschreibt; damit kann sie einerseits als ein standardisiertes Vokabular dienen und andererseits interessante Arten von Schlussfolgerungen auf den Quelltext-

ten ermöglichen.¹⁵ So ist es letztlich ein Fernziel unseres Core-H-Projektes, exemplarisch eine auf die Themenstellung der Prosopographie spätmittelalterlicher gelehrt-klerikaler Eliten spezialisierte Ontologie zu entwickeln, die aus Gründen der Generalisierbarkeit und Verknüpfbarkeit mit anderen Datenbeständen als Spezialisierung entsprechender Konzepte des bekannten CIDOC-CRM-Standards formalisiert ist. Diesem Ziel sind wir durch die Arbeit an der ANTLR-Grammatik bereits ein gutes Stück nähergekommen.

All dies ist beileibe kein einfacher Stoff. Es gibt eine Reihe von Spezialaufsätzen, die man studieren muss, wenn man die Regesten des *Repertorium Germanicum* richtig lesen und verstehen will. Die Herausforderung, dieses Wissen in ein digitales Begriffs- und Regelsystem zu überführen, ist entsprechend groß. Dennoch erscheint die Aufgabe, diese konkreten Verstehensvorgänge der Historiker:innen zu einem wesentlichen Teil zu algorithmisieren, grundsätzlich lösbar. Die Entwicklung einer entsprechenden Ontologie und die darauf aufbauende Erschließung von digitalen Quellenkorpora in einer mehr und mehr umfassenden Wissensbasis, würden historisches Wissen in einer völlig neuen Form verfügbar machen. Am Ende winkt schließlich jene Verheißung, die Johannes Haller schon vor über hundert Jahre mit Blick auf die Akten des Vatikanischen Archivs aussprach: „So bietet, bei aller Unscheinbarkeit im einzelnen, dieses Material, im großen genommen, ein Hilfsmittel zur Belebung des geschichtlichen Bildes dar, wie es reicher und zuverlässiger nicht leicht gedacht werden kann.“¹⁶

Bibliographie

Aho, Alfred V., Sethi, Ravi und Jeffrey D. Ullman. *Compilers. Principles, Techniques, and Tools*. Reading, Massachusetts et al.: Addison-Wesley, 1986.

Brosius, Dieter. „Das Repertorium Germanicum.“ In *Das Deutsche Historische Institut in Rom 1888–1988*, hg. v. Reinhard Elze und Arnold Esch, 123–165. Tübingen: Niemeyer, 1990.

Deeters, Walter. „Über das Repertorium Germanicum als Geschichtsquelle. Versuch einer methodischen Anleitung.“ *Blätter für deutsche Landesgeschichte* 105 (1969): 27–43.

Gramsch, Robert. „Kuriertätigkeit als ‚Berufsbild‘ gelehrter Juristen. Der Beitrag Roms zur Akademisierung Deutschlands im Spätmittelalter. Eine personengeschichtliche Betrachtung.“ *Quellen und Forschung aus italienischen Archiven und Bibliotheken* 80 (2000): 117–163.

¹⁵ Vgl. hierzu etwa Wolfgang Hesse, „Ontologie(n),“ in: *Informatiklexikon der Gesellschaft für Informatik*, 2005. <https://gi.de/informatiklexikon/ontologien> (Zugriff am 13.10.2021) sowie allgemein John F. Sowa, *Knowledge Representation. Logical, Philosophical, and Computational Foundations* (Pacific Grove, California: Brooks/Cole Publishing, 2000).

¹⁶ Haller, „Repertorium Germanicum,“ fol. 9.

- Gramsch-Stehfest, Robert. „Von der Metapher zur Methode: Netzwerkanalyse als Instrument zur Erforschung vormoderner Gesellschaften.“ *Zeitschrift für Historische Forschung* 47 (2020): 1–39.
- Hasselhorn, Benjamin. *Johannes Haller: Eine politische Gelehrtenbiographie: Mit einer Edition des unveröffentlichten Teils der Lebenserinnerungen Johannes Hallers*. Göttingen: Vandenhoeck & Rupprecht, 2015.
- Hesse, Wolfgang. „Ontologie(n).“ In: *Informatiklexikon der Gesellschaft für Informatik*, 2005. <https://gi.de/informatiklexikon/ontologien>. Zugriff am 13.10.2021.
- Hörschemeyer, Jörg. „Repertorium Germanicum Online.“ In *Friedensnobelpreis und historische Grundlagenforschung. Ludwig Quidde und die Erschließung der kurialen Registerüberlieferung*, hg. v. Michael Matheus, 605–615. Berlin: De Gruyter, 2012.
- Parr, Terence. *Language Implementation Patterns*. Dallas, Texas et al.: Pragmatic Bookshelf, 2010.
- Schmugge, Ludwig. *Kirche, Kinder, Karrieren: Päpstliche Dispense von der unehelichen Geburt im Spätmittelalter*. Zürich: Artemis, 1995.
- Schwarz, Brigide. „Römische Kurie und Pfründenmarkt im Spätmittelalter.“ *Zeitschrift für Historische Forschung* 20 (1993): 129–152.
- Schwarz, Brigide. „Das Repertorium Germanicum: Eine Einführung.“ *Vierteljahrsschrift für Sozial- und Wirtschaftsgeschichte* 90 (2003): 429–440.
- Sowa, John F. *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Pacific Grove, California: Brooks/Cole Publishing, 2000.

Digitale Praktiken in der Reflexion

Georg Vogeler, Christopher Pollin, Roman Bleier
„Ich glaube, Fakt ist...“:

Der geschichtswissenschaftliche Zugang zum digitalen Edieren

Abstract: Editions of primary sources are of major interest to historical research. They serve as a means to falsify statements about “historical facts”. Digital editions attempt to capture the “facts” in formalised form. Therefore, content-oriented digital editing for historians does not always require to reproduce the text in its entirety, but rather to capture and formally describe the semantic structures of a source and its content. In this chapter, the content-oriented approach to digital editing will be discussed using concrete examples. In these editions, the attempt is made to represent subject-specific questions about the historical texts in formalised form as data. Such databases, in which the historical data and its source are made available together with a so-called “knowledge domain” – conceived as an ontology – defines the framework of formal processing of the edition’s data.

Keywords: digital edition, assertive edition, historical facts, historical information, semantic web, ontology

Zusammenfassung: Quelleneditionen sind von zentralem Interesse für die historische Forschung. Essenziell für geschichtswissenschaftliche Editionen ist es, Aussagen von Historiker:innen über „historische Fakte“ falsifizierbar zu machen. In ihrer digitalen Version geht es darum, sie in formalisierter Form zu erfassen. Zu diesem Zweck ist es für eine inhaltsorientierte digitale Edition von Quellen nicht immer nötig, den Text in seiner Gesamtheit wiederzugeben, sondern die semantischen Strukturen einer Quelle zu erfassen und formal zu beschreiben. Anhand konkreter Beispiele aus dem Umfeld der digitalen Edition wird der inhaltsorientierte Zugang zum Edieren erörtert. In diesen Editionen wird versucht, fachspezifische Fragestellungen an die historischen Texte in formalisierter Form als Daten abzubilden. Solche quellenbasierten Datenbanken, in denen die historischen Daten und ihre Quelle zur Verfügung gestellt werden, zusammen mit einer sogenannten „knowledge domain“ – als Ontologie gedacht – definieren den Rahmen der formalen Verarbeitung der Editionsdaten.

Schlagwörter: Digitale Edition, Assertive Edition, historische Fakten, historische Information, Semantic Web, Ontologie

Historiker:innen argumentieren mit Fakten, die in Quellen belegt sind. Sind Quelleneditionen für Historiker:innen also „Faktengeneratoren“? Wir möchten im Folgenden die epistemische Funktion digitaler Editionen für die Geschichtswissenschaft reflektieren. Wir behaupten, dass digitale Editionen zwei methodische Grundlagen besonders gut abbilden: den *Perspektivismus* und die *Intersubjektivität* historischer Erkenntnis. Der Beitrag versucht deshalb zunächst, die geschichtswissenschaftliche Vorstellung von Fakten zu beschreiben, um das Verhältnis zwischen Quellen und Fakten zu bestimmen. Daraus können Aufgaben geschichtswissenschaftlicher Edition allgemein abgeleitet werden. Wie digitale Editionen diese Aufgaben erfüllen, wird schließlich theoretisch und an Beispielen diskutiert.¹

1 Fakten, Quellen und historische Interpretation

Die Phrase „Fakt ist...“ ist geläufig und ein gern verwendetes rhetorisches Mittel. Häufig ist diese Phrase bei Politiker:innen zu finden, die sie noch mit dem Zusatz „evidenzbasiert“ aufzuwerten versuchen, oder die eben mal „die Fakten auf den Tisch legen“. Die politische Rhetorik bedient sich dabei des umgangssprachlichen Verständnisses von „Fakt“ als einem „wirklichen, nachweisbaren, bestehenden, wahren oder anerkannten Sachverhalt.“² So zumindest definiert die Wikipedia „Tatsachen“, die sie als Synonym für „Fakten“ setzt.³ Historische Fakten sind besonders konfiguriert. In der Geschichtswissenschaft sind historische Fakten Grundlage für Aussagen über vergangene Wirklichkeit, mit denen die Historiker:innen ihre historischen Deutungen begründen.⁴ Da die menschlichen Handlungen der Vergangenheit nicht mehr existieren, sind sie im Sinne

¹ Dieser Text ist unter maßgeblichen Beiträgen von Hans Clausen entstanden. Wir möchten uns besonders bei Patrick Sahle, unserem Respondenten auf der Digital History Konferenz, sowie Gabriele Haug-Moritz und Gunter Vasold für ihr konstruktives Feedback zu einer frühen Fassung des Aufsatzes bedanken. Die Arbeiten an den vorgestellten Projekten werden von der Mellon-Foundation (Projekt 1907-06980) und dem FWF (Projekt I 3446) gefördert.

² Wikipedia, „Tatsache“, Zugriff am 20.07.2021, <https://de.wikipedia.org/wiki/Tatsache>. Wir werden im Folgenden ebenso „Fakten“ und „Tatsachen“ gleichsetzen.

³ Eine Übersicht über die philosophischen Diskussionen über „Fakten“ geben Kevin Mulligan und Fabrice Correia, „Facts“, in *The Stanford Encyclopedia of Philosophy* (Stanford, Winter Edition, 2020), hg. v. Edward N. Zalta, Zugriff am 20.07.2021, <https://plato.stanford.edu/archives/win2020/entries/facts/>.

⁴ Richard J. Evans, *Fakten und Fiktionen* (Frankfurt/New York: Campus Verlag, 1998), 78–103.

von Searle⁵ „beobachterabhängig“. Sie existieren nur in der Perspektive der Historiker:innen auf die vergangene Wirklichkeit (Perspektivismus).⁶ Historische Fakten sind abhängig von den Theorien, von denen sich die Historiker:innen bei der Interpretation von Quellen leiten lassen. „Die Entscheidung darüber, welche Faktoren zählen sollen, oder nicht, fällt zunächst auf der Ebene der Theorie, welche die Bedingungen möglicher Geschichte setzt.“⁷

Dass Geschichte vor dem Hintergrund variabler theoretischer Vorannahmen und Metaerzählungen entsteht, bedeutet jedoch nicht, dass es sich bei ihr um ein Sammelsurium individueller Interpretationen handelt. Im klassischen Verständnis ist Geschichte kollektiv geteilte Auffassung über die Vergangenheit (Intersubjektivität). Die Möglichkeit der Übereinkunft über den Status von historischen Fakten ist an geteilte Forschungsmethoden gebunden, unter denen die „historische Methode“ besonders prominent ist. Dazu gehört die kritische Interpretation von Quellen für eine Darstellung der vielen möglichen Vorstellungen von der Vergangenheit, unter Offenlegung der Darstellungsabsichten.⁸ Die Auswertung von Quellen ist der Weg, gemeinsame Erkenntnisse über die Vergangenheit zu gewinnen.⁹ In Jörn Rüsen's Definition von historischer Forschung wird das Wissen „über die menschliche Vergangenheit aus den empirischen Gegebenheiten vergangenen menschlichen Lebens“ ermittelt.¹⁰ Auch in modernen medientheoretisch fundierten Geschichtstheorien, wie sie zum Beispiel Tschiggerl, Walach und Zahlmann vertreten, sind die Quellen Kern des überindividuellen Diskurses in der Geschichtswissenschaft, der ihnen ihre historische Bedeutung einschreibt.¹¹

5 John R. Searle, „Social ontology: Some basic principles,“ *Anthropological theory* 6/1 (2006): 12–29, doi: 10.1177/1463499606061731.

6 Zum Perspektivismus der Geschichtswissenschaft vgl. Jörn Rüsen, *Grundzüge einer Historik*, Bd. II, (Göttingen: Vandenhoeck & Ruprecht, 1986).

7 Reinhart Koselleck, „Standortbindung und Zeitlichkeit: Ein Beitrag zur historiographischen Erschließung der geschichtlichen Welt,“ in *Objektivität und Parteilichkeit in der Geschichtswissenschaft*, hg. v. Reinhart Koselleck et al. (München: dtv, 1977), 45.

8 Johann Gustav Droysen, *Historik: Vorlesungen über Enzyklopädie und Methodologie der Geschichte* (1882), hg. v. Rudolf Hübner (München: R. Oldenbourg Verlag, ⁸1977); Ernst Bernheim, *Lehrbuch der historischen Methode und der Geschichtsphilosophie: mit Nachweis der wichtigsten Quellen und Hilfsmittel zum Studium der Geschichte* (Leipzig: Duncker & Humblot, 1908); Evans, *Fakten und Fiktionen*, 78–103.

9 Bernheim, *Lehrbuch der historischen Methode*, 185; Jörn Rüsen, *Geschichte denken: Erläuterungen zur Historik* (Wiesbaden: Springer VS, 2020), 8.

10 Rüsen, *Geschichte denken*, 7.

11 Martin Tschiggerl, Thomas Walach und Stefan Zahlmann, *Geschichtstheorie* (Wiesbaden: Springer Fachmedien Wiesbaden, 2019), doi: 10.1007/978-3-658-22882-8, 92–93.

Dabei sind die Quellen auch Teil des grundsätzlichen Perspektivismus von Geschichtswissenschaft. An sich können alle Relikte früheren menschlichen Handelns¹² von Historiker:innen als solche empirischen Gegebenheiten verwendet werden und so zum Ankerpunkt der Erkenntnis von der Vergangenheit werden. Mit Gustav Droysens methodischer Trias von Heuristik, Kritik und Interpretation der Quellen wird aber deutlich, dass nicht jedes Relikt von sich aus Quelle ist. Erst im Zuge der Heuristik werden beliebige Relikte zu Quellen, indem sie von Forschenden für die eigene Fragestellung für relevant gehalten werden. So unterliegt schon die Quellenauswahl dem Perspektivismus der Geschichtswissenschaft.¹³ Die Quellen sind nicht Ursprung der *einen* historischen Wirklichkeit, vielmehr sind sie der Ausgangspunkt eines Diskurses, der durch den Hintergrund theoretischer Vorentscheidungen und impliziter Annahmen der sie interpretierenden Historiker:innen geprägt ist.¹⁴

Auch der Prozess der hermeneutischen Quelleninterpretation ist davon bestimmt, dass die Quellen nicht einfach historische Fakten liefern, sondern sie Teil des geschichtswissenschaftlichen Diskurses sind, in dem ein:e Historiker:in die eigenen Auffassungen von der Vergangenheit intersubjektiv etablieren will. Die Quelle ist nicht ein absolut gültiger Beweis, sondern sie macht die eigene Sichtweise plausibel oder widerlegt bestehende historische Fakten. Epistemologisch kommt den Quellen primär ein „Vetorecht“ zu: die Quellen „verbieten uns, Deutungen zu wagen oder zuzulassen, die aufgrund eines Quellenbefundes schlichtweg als falsch oder als nicht zulässig durchschaut werden können“.¹⁵ In ihrer Funktion, bestimmte Aussagen über die Geschichte zu stützen oder zu falsifizieren, „überträgt [die Rede vom Vetorecht der Quellen] gewissermaßen das Falsifizierungsprinzip von Karl R. Popper (1902–1994) in den Bereich historischer Erkenntnistheorie.“¹⁶

12 Wir folgen hier Hüttenbergers Überlegungen, dass der Unterschied zwischen „Tradition“ und „Überrest“ erst ein sekundärer ist, weil beide quellenkritisch zunächst einmal ein Überrest der Handlungen der Autor:innen sind und beide von Handlungen der Menschen in der Vergangenheit berichten. Peter Hüttenberger, „Überlegungen zur Theorie der Quelle,“ in *Einführung in die Interpretation historischer Quellen. Schwerpunkt: Neuzeit*, hg. v. Bernd A. Rusinek et al. (Paderborn/München/Wien/Zürich: Ferdinand Schöningh, 1992), 253–265.

13 Gabriele Haug-Moritz hat uns vorgeschlagen, diese Funktion der Quellenedition als „Schaufenster“ des spezifischen Forschungsinteresses zu bezeichnen, was schön in das Bild des Perspektivismus passt.

14 Koselleck, „Standortbindung und Zeitlichkeit,“ 45.

15 Koselleck, „Standortbindung und Zeitlichkeit,“ 45–46.

16 Stefan Jordan, „Vetorecht der Quellen,“ *Docupedia-Zeitgeschichte*, (2011): 1–7, doi: 10.14765/ZZF.DOK.2.570.V1.

2 Edieren für historische Fakten

Historiker:innen ermitteln also mit einer intersubjektiv akzeptierten Methode aus Quellen historische Fakten in ihrer jeweiligen Forschungsperspektive. Historische Forschung profitiert davon, wenn Quellen für historische Aussagen intersubjektiv leichter nachvollziehbar sind. In der vordigitalen Zeit waren die öffentliche Zugänglichkeit von Archiven und die Verbreitung der Quellentexte durch Druckeditionen dafür die wichtigsten Mittel. In den letzten Jahrzehnten haben sich ihnen digitale fotografische Reproduktionen der Quellen hinzugesellt, die ebenso die Nachvollziehbarkeit von Argumentationen erleichtern. Wissenschaftliches Edieren geht aber über die reine Reproduktion hinaus.¹⁷ Historiker:innen müssen die Quelle in der Edition nämlich einer Kritik unterziehen, die mehr ist als die korrekte Reproduktion des Textes: Sie müssen überprüfen, ob die edierte Quelle sich als Argument für historische Fakten eignet, so wie das zum Beispiel in der Echtheitskritik von Urkundeneditionen der Fall ist.

Eine Edition macht die Quellen auch als Argument nutzbar, indem sie sie in den historischen Diskurs einbindet. Auch damit trägt die Edition zur Glaubwürdigkeit, der aus den Texten extrahierten Fakten bei. Umstrittene oder ohne Kontext missverständliche Aussagen in Quellen werden kommentiert – und die kritische Edition von Hitlers *Mein Kampf* durch das Institut für Zeitgeschichte zeigt gut,¹⁸ wie weit das gehen kann. Editionen von mittelalterlichen Chroniken weisen beispielsweise nach, ob Textpassagen nur eine Kopie von anderen Texten sind, oder, ob es keine Vorlage dazu gibt. So bekommen die in den kopierten und den nicht kopierten Passagen formulierten Aussagen unterschiedliche epistemische Qualitäten: von Selbsterlebtem bis zu Selbsterfundem, von unreflektierter Aufnahme vorhandenen Wissens bis zum Hinweis auf eine Geisteshaltung, weil der Text Teil einer identifizierbaren intellektuellen Tradition ist.

Die Anschlussfähigkeit von Editionen an den geschichtswissenschaftlichen Diskurs führt so zu einer eigenen geschichtswissenschaftlichen Art und Weise des Edierens: Nicht nur der Text, auch die in der Quelle ermittelbaren historischen Fakten sollen der geschichtswissenschaftlichen Argumentation leicht zugänglich sein. Die Historiker:innen interessieren sich für einen Zugang zu Text,

17 Manfred Thaller, „Reproduktion, Erschließung, Edition, Interpretation. Ihre Beziehungen in einer digitalen Welt,“ in *Vom Nutzen des Edierens. Akten des internationalen Kongresses zum 150-jährigen Bestehen des Instituts für Österreichische Geschichtsforschung*, Wien, 3.–5. Juni 2004, hg. v. Brigitte Merta et al. (Wien et al., 2005), 205–227.

18 Adolf Hitler, *Hitler, Mein Kampf: eine kritische Edition*, hg. v. Christian Hartmann et al. (München: Institut für Zeitgeschichte, 2016).

die Patrick Sahle in seinem „Textrad“ mit „Text als Inhalt“ bezeichnet.¹⁹ Die oben angeführten Kommentare gehören ebenso dazu, wie die Identifikation von Handlungen, Ereignissen und den damit verbundenen Daten über Personen, Orte, Datumsangaben und ähnlichem. Je nach Forschungsschwerpunkt und Quellentyp kann das Verständnis darüber variieren, was zentrale Inhalte eines Textes sind.

Ziel von Editionen, die den Inhalt in den Mittelpunkt rücken, ist es die historischen Dokumente so aufzubereiten, dass Historiker:innen zielgerichtet zu den aus ihrer Perspektive ermittelbaren Fakten kommen. Hierfür haben sich in der Druckkultur Verfahren wie etwa Regesten etabliert.²⁰ Die historische Quelle wird dabei nicht vollständig wiedergegeben, sondern es reicht, dass die zentralen Aussagen paraphrasiert werden, und sei es nur im sogenannten Kopfregeß oder Betreff als Kurzinformation über die Inhalte. Ebenso wie in einem Regest sind auch in Registern nur die zentralen Entitäten²¹ erschlossen.²² Bedeutende historische Editionsreihen im deutschsprachigen Raum wie zum Beispiel die *Monumenta Germaniae Historica*, die „Deutschen Reichstagsakten“, oder die Akten der Reichskanzlei des 20. Jahrhunderts demonstrieren diesen Ansatz, indem sie solche inhaltlichen Zusammenfassungen in die Edition integrieren. In der Edition der „Deutschen Reichstagsakten“ geht das Verfahren so weit, dass Textpassagen ausschließlich regestiert und manche Quellengruppen nur durch Zusammenfassungen („Aktenreferate“) erschlossen werden.²³ Aus philologi-

19 Patrick Sahle, *Digitale Editionsformen: zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, Teil 3: *Textbegriffe und Recodierung* (Norderstedt: Books on Demand, 2013), 37–41.

20 Vgl. zum Beispiel Arbeitskreis für Editionsgrundsätze 1965 oder die Regelwerke von Archiven: Joachim Kemper, „Neue Richtlinien der staatlichen Archive Bayerns für die Erstellung von Urkundenregesten“, *Archivalische Zeitschrift* 91 (2009): 209–219, doi: 10.7788/az.2009.91.1.209; oder Überlegungen aus dem Umfeld der Regesta Imperii: Regesta Imperii, „Richtlinien für die Registrierung von Urkunden (1978)“, in *Die Regesta Imperii im Fortschreiten und Fortschritt*, hg. v. Harald Zimmermann (Köln: Vandenhoeck & Ruprecht, 2000), 87–94; Johannes Mötsch, „Vorteile und Grenzen der Regestentechnik“, in Zimmermann (Hrsg.), *Fortschreiten und Fortschritt*, 115–127.

21 Unter Entitäten verstehen wir Ereignisse, Themen, Personen, Orte und andere Daten, die sich als Instanz aus einer Klasse ableiten lassen.

22 Zu den Eigenschaften von Registern über die Funktion als Pfad zu Textstellen hinaus vgl. Reinhard Härtel, „Mehr als ein Anhang. Das computererstellte Register“, in *Historische Edition und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung. Berichte von der Internationalen Tagung „Historische Edition und Computer“ Karl-Franzens-Universität Graz*, 26.–30. Oktober 1988, hg. v. Anton Schwob et al. (Graz: Leykam 1989), 67–84; Horst Kunze, *Über das Registermachen* (München/London/New York/Paris: K G Saur, 1992).

23 Gabriele Haug-Moritz et al., „Der Regensburger Reichstag des Jahres 1576 – ein Pilotprojekt zum digitalen Edieren frühneuzeitlicher Quellen“ (gekürzter DFG-Antrag), 2018, Zugriff am

scher Sicht ist ein solches Verfahren problematisch, da eine fragmentarische Wiedergabe eines historischen Textes einen gravierenden Eingriff in einen historischen Text selbst bedeutet und dieser nicht in seiner Gesamtheit wiedergegeben wird. Gleichzeitig erscheint in einer die inhaltliche Dimension bevorzughenden Perspektive eine detaillierte Erfassung der Textgestalt und der Textüberlieferung für „inhaltsarme“ Textstellen nicht notwendig, und wird dann, auch im Blick auf die zur Verfügung stehenden Ressourcen, reduziert.

In der Gutenberg-Galaxis hat es die Edition aber schwer, den Perspektivismus der Quelleninterpretation abzubilden. Die Unveränderbarkeit der gedruckten Edition macht es unmöglich, nachträglich andere Perspektiven einzuschließen. Editionstexte sind nur über den umständlichen Weg einer Neuedition mit Überarbeitung in einer neuen Perspektive erschließbar, während digitale Editionen offen für Veränderungen sind. Gedruckte Editionen arbeiten platzsparend, was mögliche Perspektiven auf den Text bewusst ausschließt – Regestierung und ähnliche Verfahren wurden bereits angesprochen – da der Umfang der Publikation einerseits eine Kostenfrage ist, andererseits verteilte Information (Register am Ende des Bandes, alternative Darstellungsformen in eigenen Bänden) das Arbeiten mit den edierten Quellen erschwert. Ein Register ist eine stark reduzierte Präsentation der Wissensumgebung, die sich auf das Alphabet oder eine andere vorgegebene Ordnung (zum Beispiel Begriffshierarchie) und Verweise als Suchmechanismen reduziert. Im Digitalen können in Suchmasken oder facettierten Suchen mehrere Erschließungskriterien gleichzeitig angewendet werden, wodurch die Historiker:innen mehrere für ihr jeweiliges Forschungsinteresse nützliche Eigenschaften in die Auswahl von Quellenstellen einbeziehen können. Trefferlisten solcher Suchen, Visualisierungen oder Text-Bild Synopsen der Editionstexte ordnen die einzelnen Quellenstellen anders als in der fixierten Abfolge der Buchseiten und können so neue Interpretationskontexte für die Historiker:innen erzeugen.

3 Änderung im Digitalen

Mit der Digitalisierung prägen Algorithmen und Informationssysteme immer mehr, „wie Historikerinnen Geschichte denken und folglich auch, wie sie histo-

rischen Sinn produzieren.“²⁴ Die digitale Edition ist eine Ausprägung dieser Entwicklung. Digitale Methoden erlauben es, Texte mehrdimensional wiederzugeben, also historische Inhalte der Texte zugänglich zu machen und dabei gleichzeitig Ansprüchen philologischer Disziplinen an kritisch edierte Texte gerecht zu werden.²⁵

Georg Vogeler hat für einen Typ von digitaler Edition,²⁶ der den Anforderungen der Historiker:innen entspricht, den englischen Namen *assertive edition* vorgeschlagen, den man vielleicht mit *Aussagenedition* übersetzen könnte. Wir werden im Folgenden bei einer Eindeutschung als assertive Edition bleiben. Eine *Assertion* ist dabei eine Tatsachenbehauptung auf Grundlage von Inhalten historischer Quellen, die in einem digitalen Informationssystem abgebildet ist, das aus also *Behauptungsdatenbanken* besteht. Damit sind nicht die Expertensysteme der 1980er gemeint, die aus einer *closed world* Schlüsse ziehen sollten, sondern die Behauptungsdatenbanken dokumentieren nur, was die Editor:innen als historische Fakten in der von ihnen gewählten Perspektive auf die Quelle für ermittelbar halten. Charakteristisch für die assertive Edition ist, dass formale Aussagen über historische Sachverhalte in eine Edition eingebettet und mit den Repräsentationen der Quelle und editorischen Kommentaren verknüpft sind, so dass im Sinne der Quellenkritik der Interpretationsprozess der Editor:innen transparent wird. Die Idee der assertiven Edition ist also, dass der historische Gehalt einer Quelle in einer Menge von Assertionen steckt, die jeweils mit bestimmten Elementen des Textes korrespondieren oder auf diesen basieren.

„Historische Fakten“ in formalisierter Form, also die Assertionen, sind nämlich auch nur Ergebnisse einer Perspektive auf die Quelle. Diese Perspektivität kann sich in der digitalen Repräsentation einer bestimmten Wissensdomäne (*Knowledge Domain*) abbilden. Die digitale Repräsentation der Wissensdomäne bildet das *Knowledge Environment* oder die *Wissensumgebung* einer Edition.²⁷ In dieser Umgebung wird das zur formalen Repräsentation und darauf aufbauenden Verarbeitung historischer Quellen notwendige Wissen zusammengeführt. Dies beinhaltet Wissen über in Quellen angewandte Orthographie, Fachbegriffe, Währungen, soziale Verhältnisse oder viele andere von der Wissensdomäne ab-

24 Mareike König, „Geschichte digital Zehn Herausforderungen,“ in *Geschichtswissenschaft im 21. Jahrhundert* (Berlin/Boston: De Gruyter Oldenbourg, 2020), 68, doi: 10.1515/9783110689143-008.

25 Sahle, *Digitale Editionsformen*, 37–41.

26 Georg Vogeler, „The ‚Assertive Edition‘,“ *International Journal of Digital Humanities* 1/2 (2019): 309–322, doi: 10.1007/s42803-019-00025-5.

27 Manfred Thaller, „Historical Information Science: Is there such a Thing? New Comments on an old Idea [1993],“ *Historical Social Research/Historische Sozialforschung, Supplement* 29 (2017): 260–286, doi: 10.12759/hsr.suppl.29.2017.260-286.

hängige Aspekte. Unschärfen, Unsicherheiten oder Kontextabhängigkeiten wie zum Beispiel Währung im Verhältnis zu Zeit und Raum, oder zeitgenössische Terminologien, die historischen Quellen immanent sind, werden nur innerhalb einer gewählten Forschungsperspektive normalisiert. Die Wissensumgebung definiert den terminologischen Rahmen und den Geltungsbereich der formalisierten Quellenaussagen. Jede weiterführende Verarbeitung, wie zum Beispiel Aggregation von Datenpunkten für eine Visualisierung, ist ausschließlich in dieser Wissensumgebung zu verstehen. Sie fasst kontrollierte Vokabularen, strukturierte Daten und formale Modelle zusammen.

Die Wissensumgebung reicht über eine einzelne Quellenedition hinaus: Zum Beispiel lässt sich der Fluss von ökonomischen Gütern und Geldbeträgen, also von Transaktionen, nicht nur in Rechnungsbüchern finden, die für sich schon in tabellarischen oder listenähnlichen Formaten existieren. Auch in anderen historischen Quellen sind Transaktionen dokumentiert und können für Forschungsvorhaben relevant sein: Friedrich Schleiermacher hat beispielsweise in seinen Tageskalendern nicht nur tagebuchartige Notizen vermerkt, sondern auch Einkäufe und Verkäufe dokumentiert.²⁸ Ein Modell, das versucht, Transaktionen auf abstrakter Ebene zu beschreiben, kann auf unterschiedliche Quellen und sogar Quellengattungen angewendet werden. Gleichzeitig kann eine Quelle in unterschiedliche Wissensumgebungen integriert werden, kann *schleiermacher digital* also auch für prosopographische Forschung Personen im Tageskalender identifizieren oder für ideengeschichtliche Fragestellungen die Werke ermitteln, die er rezipiert. Die assertive Edition ermöglicht, historische Quellen aus unterschiedlichen Perspektiven zu bearbeiten, und diese zugleich als Wissensumgebung formalisiert mitzuliefern.

Die formalisierte Wissensdomäne in der assertiven Edition erlaubt auch die Verknüpfung der Quelle mit Datensätzen bestehender Ressourcen, wie etwa Normdateien. Die Wissensumgebung trägt damit zur Intersubjektivität der ermittelten Fakten bei, indem diese Verknüpfung die Systematizität des von der Edition repräsentierten Wissens erhöht.

Aus der Funktion der digitalen Edition in der Geschichtswissenschaft, nicht nur einen Text, sondern auch eine Behauptungsdatenbank darzustellen, folgen weitere Anforderungen:

- Die digitalen Mittel, die bei der Erarbeitung der Edition eingesetzt werden, müssen kritisch reflektiert werden. Die Diskussion über die Zuverlässigkeit

²⁸ Wolfgang Virmond (unter Mitarbeit von Holden Kelm), „Schleiermachers Tageskalender 1808–1834,“ in *schleiermacher digital*, hg. v. Elisabeth Blumrich et al. (Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, seit 2012). Zugriff am 23.07.2021, <https://schleiermacher-digital.de/tageskalender/index.xql>.

automatischer Transkription von handgeschriebenen Texten ist ein Beispiel dafür. Unter dem Begriff *Digitale Hermeneutik* wird über die kritische Reflexion der geschichtswissenschaftlichen Methode unter diesen Bedingungen und über den Wandel der Arbeitsweisen von Historiker:innen von der Quelle zum Dokument zu Daten diskutiert. Formale Methoden in der Verarbeitung von historischer Information suggerieren eine Objektivität, die es zu dekonstruieren gilt.²⁹ Die Aufarbeitung und die Interpretation historischer Quellen mit digitalen Methoden bedarf auch einer Kritik an Daten und Werkzeugen.³⁰ Die von der assertiven Edition gebildeten Datenbanken müssen die Herkunft der in ihnen gespeicherten Fakten offenlegen.

- Eine digitale Edition erzeugt Informationen, die als Daten in weiterer historischer Forschung mit algorithmischen Mitteln verarbeitet werden sollen. Das gilt auch für die von Editor:innen ermittelten Tatsachendarstellungen der Quellen. Textuelle Beschreibungen reichen deshalb nicht aus. Mit den Ergebnissen assertiver Editionen soll man Statistiken erstellen können, Filter darauf anwenden, sie mit anderen Datensätzen als Linked Data verknüpfen oder, wenn die Wissensdomäne in Sprachen zur Abbildung formaler Logik wie OWL³¹ oder SWI-Prolog³² ausgedrückt werden, formale Schlüsse ziehen.

4 Wie erzeugt man Editionen als Behauptungsdatenbanken?

Die technische Umsetzung von assertiven Editionen erfolgt gemeinhin über die Auszeichnung textlicher Repräsentanten semantischer Strukturen. Der Web-of-Data-Stack kann für diese Zwecke herangezogen werden. Dabei lassen sich die

²⁹ Joris J. van Zundert, „Screwmenetics and Hermenumericals: The Computationality of Hermeneutics,“ in *A New Companion to Digital Humanities*, hg. v. Susan Schreibman et al. (Hoboken, New Jersey: John Wiley & Sons, 2015), doi: 10.1002/9781118680605.ch23.

³⁰ Andreas Fickers, „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?“ *Zeithistorische Forschungen* 17/1 (2020): 157–168, 10.14765/zzf.dok-1765 sowie den Beitrag in diesem Band.

³¹ OWL Web Ontology Language, <https://www.w3.org/TR/owl-features>.

³² Ein Beispiel für diesen Ansatz: Jana Kittelmann und Christoph Wernhard, „KBSET – Knowledge-Based Support for Scholarly Editing and Text Processing with Declarative LaTeX Markup and a Core Written in SWI-Prolog,“ *Lecture Notes in Computer Science*, arXiv:2002.10329 [cs] 12057 (2020): 178–196, doi: 10.1007/978-3-030-46714-2_12. SWI-Prolog, <https://www.swi-prolog.org>.

einzelnen Assertions in RDF³³ und notwendige Modelle als Ontologien beschreiben. Für die digitale Edition können semantische Strukturen in XML/TEI³⁴ über das ana-Attribut³⁵ auf Ontologien referenziert werden. Daraus lassen sich Triple-Aussagen ableiten, die Tatsachenbehauptung im Sinne einer Ontologie darstellen. Ein Personennamen in einer Quelle kann dadurch zu einer Akteurin in einer Transaktion (Rechnungsbücher) oder zu einem Kommunikationspartner (Besprechungsprotokolle, Briefe) werden.

Die Normalisierung und Identifikation von Entitäten wird im Web-of-Data-Stack über unique identifiers (URI) realisiert. Dadurch werden internetweit eindeutige Namen für Konzepte erzeugt, auf die dann aus den Editionen heraus verwiesen werden kann. URIs ermöglichen die Repräsentation von Entitäten ohne die möglichen Mehrdeutigkeiten natürlicher Sprache, wie etwa orthographische Variationen. Diese Methode kann zum Beispiel für die Erweiterung von Registern verwendet werden, um Personen und Orte über URIs zu identifizieren und damit mit externen Ressourcen zu vernetzen.

Für menschliche Nutzer:innen werden die Assertions im User Interface der Edition in der Form von erweiterten Registern, Kalendern und Timelines und Karten nutzbar gemacht, in Editionen von Wirtschaftsdokumenten vor allem als Diagramme, aber auch als Netzwerkvisualisierungen, um zum Beispiel Korrespondenznetzwerke zu veranschaulichen.³⁶ Der Web-of-Data-Stack und RESTful APIs verbessern die Nachnutzung der Daten durch Maschinen. Die Einbindung und Vernetzung der Daten mit bereits existierenden Ressourcen wie etwa Normdaten und die Anbindung an existierende Top-Level Ontologien wie zum Beispiel das CIDOC CRM³⁷ machen die assertive Edition zu einer Linked Open Data Ressource und zu einem Baustein im historischen *Knowledge Graph*.

5 Beispiele

Assertive Editionen müssen nicht unter dieser Bezeichnung erscheinen. So sind die Editionen der Notizbücher von Paolo Bufalini³⁸ ebenso assertive Editionen

³³ Resource Description Framework, <https://www.w3.org/TR/rdf11-concepts>.

³⁴ Text Encoding Initiative, <https://tei-c.org>.

³⁵ <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.global.analytic.html>.

³⁶ Für Beispielditionen siehe Vogeler, „The ‚Assertive Edition‘“, 312–313.

³⁷ CIDOC Conceptual Reference Model, <http://www.cidoc-crm.org/>.

³⁸ Francesca Tomasi, „L’informazione digitale e il Web semantico. Il caso delle scholarly digital editions,“ in *Informatica umanistica: risorse e strumenti per lo studio del lessico dei beni culturali*, hg. v. Valeria Zotti und Ana Pano Alamán (Florenz: Firenze University Press, 2017),

wie Sandrart.net,³⁹ die Edition der Werke von Johannes Friedrich Blumenbach⁴⁰ oder verschiedene Editionsprojekte im Symogih-System.⁴¹ An der Universität Graz werden verschiedene assertive Editionen realisiert, die illustrieren können, wie sich das Verhältnis zwischen Quelle und Fakten mit digitalen Mitteln ausdrücken lässt.

5.1 Rechnungsbücher

Güter- und Geldflüsse zu dokumentieren ist der Zweck von Rechnungsbüchern. Historisches Forschungsinteresse daran liegt nicht nur in den Einzeleinträgen, sondern auch in der Aggregation von Daten aus größeren Quellenbeständen. In einer assertiven Edition werden die Einträge in Rechnungsbüchern zu formalisierten Aussagen über Transaktionen. Diese Perspektive auf die Rechnungsbücher wird in einer Domänenontologie beschrieben.

Die *Bookkeeping Ontology*⁴² ist ein Modell zur formalen Beschreibung von Transaktionen in historischen Rechnungsunterlagen. Darin wird eine Transaktion als eine Menge von mindestens einem Transfer definiert. Jeder Transfer umfasst den Austausch von wirtschaftlichen Gütern oder Geldbeträgen, die von Akteur:in zu anderen fließen. Geldbeträge sind dadurch gekennzeichnet, dass sie aus einer Zahl und einer Währung bestehen und es können damit auch Spezialfälle wie Steuern oder Preise abgebildet werden. Die wirtschaftlichen Güter

157–74. Marilena Daquino, Francesca Giovannetti und Francesca Tomasi, „Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell’edizione semantica del quaderno di appunti di Paolo Bufalini,“ *Umanistica Digitale* 3/7 (2019): 49–75, doi: 10.6092/issn.2532-8816/9091.

³⁹ Sandrart.net: Eine netzbasierte Forschungsplattform zur Kunst- und Kulturgeschichte des 17. Jahrhunderts, hg. v. Anna Schreurs, Carsten Blüm und Thorsten Wübbena (Wolfenbüttel: Herzog-August-Bibliothek, [2012]), Zugriff am 21.07.2021, <http://www.sandrart.net/de>.

⁴⁰ Jörg Wettlaufer, Christopher Johnson, Martin Scholz, Mark Fichtner und Sree Ganesh Thotempudi, „Semantic Blumenbach: Exploration of Text-Object Relationships with Semantic Web Technology in the History of Science,“ *Digital Scholarship in the Humanities* 30 (supl_1) (2015): i187–i198, doi: 10.1093/llc/fqv047.

⁴¹ The symogih.org project: a modular system for managing historical information. <http://symogih.org/>. Francesco Beretta, „Pour une annotation sémantique des textes: le projet symogih.org et la Text encoding initiative,“ *Bruniana e Campanelliana, Ricerche filosofiche e materiali storico – testuali* 22/2 (2016): 453–465, doi: 10.19272/201604102005.

⁴² Christopher Pollin, „Digital Edition Publishing Cooperative for Historical Accounts and the Bookkeeping Ontology,“ in *Proceedings of the Doctoral Symposium on Research on Online Databases in History (RODBH 2019)*, hg. v. Thomas Riechert et al. 7–14. Leipzig: CEUR-WS 2020. Zugriff am 20.07.2021, <http://ceur-ws.org/Vol-2532/paper1.pdf>.

wiedermum werden aufgeteilt in Waren – bestehend aus Menge, Einheit und Art – und Dienstleistungen, die über eine zeitliche Komponente verfügen. Akteur:innen in Transfers können Individuen, Gruppen oder Organisationen oder Kategorien innerhalb von Konten (zum Beispiel Ausgaben für Bautätigkeiten) sein. Der Eintrag in der Quelle fungiert als Beleg der Assertion vom Typ „Transaktion“ und wird als solcher abgebildet. Die verwendeten Kernbegriffe lassen sich in das konzeptuelle Referenzmodell des CIDOC CRM integrieren.

So ergibt sich die Möglichkeit, jede Transaktion in ihrer zeitlichen, räumlichen und inhaltlichen Dimension, also der Zuordnung zu einem bestimmten, durch die Forschungsfrage geprägten Zusammenhang zu beschreiben. Die semantische Struktur „Transaktion“ kann unabhängig von einer spezifischen historischen Fragestellung beschrieben werden. Aber erst in der konkreten Kontextualisierung einer Transaktion bekommen Währungen, Maßeinheiten, Personen, Preise einen Wert. Die Zuordnung eines Wertes zu diesen Entitäten im Kontext einer wissenschaftlichen Edition sind Annahmen in der Form von historischen Fakten.

Auf Basis dieser Annahmen und ihrer Nachvollziehbarkeit durch die Wissensdomäne können weitere formale Methoden angewandt werden. Dann werden die RDF-Daten, modelliert nach der *Bookkeeping Ontology*, die Primärdaten (XML/TEI, CSV), das Digitalisat der Quelle und die Verknüpfungen zu anderen Linked Open Data-Vokabularen oder Authority-Files zu einer Wissensumgebung für eine historische Interpretation, wie zum Beispiel eine Finanzgeschichte der Stadt Basel im 16. Jahrhundert.

5.2 Deutsche Reichstagsakten – der Reichstag von 1576

Die digitale Edition der Akten des Reichstags von 1576 ist geprägt von einer langen, im Falle der Abteilung „Reichsversammlungen“ in die ausgehenden 1980er Jahre zurückreichenden Tradition und einer modernen historischen Fragestellung. Die digitale Methode vermittelt zwischen beiden und kann so eine assertive Edition erzeugen, die über die reine Übersetzung etablierter Erschließungsmethoden (zum Beispiel einen Index) in geschichtswissenschaftlichen Editionen hinausgeht. In der neueren Forschung zu vormodernen parlamentarischen Versammlungen begegnet international in den vergangenen Jahren nämlich immer mehr ein Erkenntnisinteresse, das nicht mehr „nur“ an der inhaltlichen Seite dieser Versammlungsereignisse interessiert ist, sondern Ständeversammlungen als Kommunikations- und Interaktionszusammenhänge in den

Blick nimmt.⁴³ Die von der Edition zu dokumentierenden Fakten sind damit nicht nur die verhandelten Themen und getroffenen Entscheidungen, nicht nur die inhaltlichen Positionen, die von den politischen Akteur:innen geäußert wurden, sondern auch Personen, die miteinander sprechen, die Gespräche protokollieren oder darüber berichten, oder sich schriftliche Anweisungen und Berichte schicken.⁴⁴ Die digitale Edition erzeugt deshalb eine Datenbank von Kommunikationssituationen. In diese fließen nicht edierte, aber detailliert erschlossene Archivalien ebenso ein wie Teile von umfangreichen Protokollen oder Berichte, die das Versammlungsgeschehen dokumentieren.

Realisiert wird die digitale Edition mithilfe von an EAD⁴⁵ angelehnte Archivalienbeschreibungen, TEI-Transkriptionen und RDF-Extrakten. Dem RDF liegt eine im Projekt entwickelte Domänenontologie von Kommunikation in vormodernen parlamentarischen Versammlungen zugrunde. Sie verwendet Kernbegriffe, die sich auf die Klassen des CIDOC CRM abbilden lassen. Die Grundannahme ist, dass sich die politische Interaktion auf dem Reichstag in den erfassten Kommunikationsakten abgespielt hat. Landeshistorische Forschung kann also zum Beispiel die Sitzungen filtern, in denen Repräsentanten des jeweiligen Territorialfürsten anwesend waren. Die These über die zentrale – und bislang editorisch „unterbelichtete“ – Rolle des Kaisers auf dem Reichstag⁴⁶ kann durch Befunde über die Beteiligung des Kaiserhofs an den vielfältigen Beratungen untersucht werden. Eine solche Fragestellung wird durch die digitale Edition auf neue Art und Weise beantwortbar, da die Behauptungsdatenbank der Edition auch Kommunikationsakte enthält, die von nicht im Volltext edierten Dokumenten belegt sind: Ladungsschreiben (durch den Kaiser) oder Ereignisse aus den Hofratsprotokollen (mit den zum Anlass des Reichstags eingereichten Suppliken). Die Wissensressource integriert sich auch in eine nach RiC modellierte Archiverschließung,⁴⁷ auch wenn die Wissensdomäne derzeit den in Entwicklung befindlichen Standard noch nicht berücksichtigt. Dass die Verknüpfung der erwähnten Personen mit einschlägigen prosopographischen Ressourcen⁴⁸

⁴³ Michel Hébert, *Parlementer. Assemblées représentatives et échange politique en Europe occidentale à la fin du Moyen Age* (Paris: Editions de Boccard, 2014).

⁴⁴ Gabriele Haug-Moritz, „Deliberieren. Zur ständisch-parlamentarischen Beratungskultur im Lateineuropa des 16. Jahrhunderts,“ *Historisches Jahrbuch* 141 (2021): 115–155.

⁴⁵ *Encoded Archival Description*, <https://www.loc.gov/ead/>.

⁴⁶ Gabriele Haug-Moritz et al., „Pilotprojekt zum digitalen Edieren frühneuzeitlicher Quellen“.

⁴⁷ International Council on Archives – Experts Group on Archival Description: *Records in Contexts. A Conceptual Model for Archival Description* (Consultation Draft), 2016, bzw. der aktuelle Stand der Ontologie in https://www.ica.org/standards/RiC/RiC-O_v0-2.html.

⁴⁸ Zum Beispiel Kaiser und Höfe. Personendatenbank der Höflinge der österreichischen Habsburger des 16. und 17. Jahrhunderts, <https://kaiserhof.geschichte.lmu.de>.

neue Fakten zum historischen Knowledge Graph als Linked Open Data beitragen wird, ist eine Selbstverständlichkeit.

6 Fazit

Im methodischen Diskurs der Geschichtswissenschaft spielen Editionen bislang keine besondere Rolle. Sie werden in der Praxis hingenommen, eingefordert, in ihrer Qualität gelobt oder kritisiert, aber nicht in Beziehung gesetzt zur Epistemologie der Geschichtswissenschaft. Sie sind gewissermaßen transparent zu den Quellen selbst. Medienwandel kann auch hier zu einer neuen Reflexion über den methodischen Status der Edition führen: Mit der digitalen Transformation wird die historische Aussage der Quellen Teil des elektronisch auswertbaren Datenbestandes, den Historiker:innen als Grundlage ihrer digital gestützten Forschung verwenden können. Kritik an der Erschließungstiefe einer Edition kann sich damit in Kritik zum Beispiel an der Verwendbarkeit einer digitalen Edition für Netzwerkanalysen oder statistische Auswertungen übersetzen. Die Geschichtswissenschaft kann in der assertiven Edition eine eigene Editionsform entwickeln.

Die assertive Edition baut auf zwei geschichtstheoretischen Grundlagen auf: Historische Fakten entstehen erstens erst durch die Akzeptanz im Diskurs der Historiker:innen, sind also intersubjektiv. Historische Fakten sind zweitens perspektivenabhängig und theoriegebunden. Quellen sind das empirische Werkzeug der Argumentation im Diskurs über historische Fakten. Sie unterliegen ebenso in Heuristik und Kritik immer der Perspektivität und der Intersubjektivität von Geschichtswissenschaft. Editor:innen als „Anwälte des Veto-rechts“ der Quellen müssen deshalb sowohl die Aussagen der Quellen über historische Fakten zu ermitteln versuchen („Text als Inhalt“) als auch offenlegen, dass sie nur eine Perspektive auf die Quelle wiedergeben, die aus einer bestimmten Wissensdomäne gespeist ist. Die assertive digitale Edition erreicht das, indem sie erstens neben der Textwiedergabe auch RDF-Aussagen anbietet und sie eng an die Quelle bindet. Um die Perspektivität der Aussagen deutlich zu machen, bettet sie zweitens die RDF-Aussagen in eine formalisierte Wissensdomäne ein, als Linked Open Data, als kontrollierte Vokabularien und als TBox im Sinne der semantischen Technologien.

Die assertive Edition steht natürlich selbst unter einem historischen Vorbehalt: Fragestellungen können sich ändern und mit ihnen der Bedarf neuer Fakten für die wissenschaftliche Argumentation. Welche Zeremonien auf einem Reichstag stattfanden, war vor 30 Jahren im Mainstream der Geschichtswissen-

schaft vielleicht noch keine relevante Information. Das Interesse an geschichtswissenschaftlichen Editionen als zuverlässige Lieferanten von Fakten ändert sich jedoch nicht. Neue Fragestellungen erfordern dann vielleicht Neueditionen, die sich mit den editorischen Entscheidungen über die zu dokumentierenden Aussagen einer älteren assertiven Edition auseinandersetzen. Das Semantic Web ermöglicht aber auch, einen existierenden Editionstext mit einer neuen Ebene an formalisierten Aussagen zu versehen. Wenn eine assertive Edition der Reichstagsakten jetzt die historischen Kommunikationssituationen in den Mittelpunkt rückt, ist damit nicht ausgeschlossen, die Aussagen über die politische Wirklichkeit in den Berichten und Protokollen oder die von den Akteur:innen verwendeten Argumente als RDF-Aussagen zu repräsentieren, und die gewählte Perspektive in einer Ontologie der verwendeten Wissensdomäne auszudrücken.

Das gilt auch für die Methode, mit der die Aussagen aus den Quellen ermittelt werden. Die assertive Edition steht in einem hoffentlich fruchtbaren Austausch zu den aktuellen Bemühungen, Informationen automatisch aus digitalisierten Originalquellen zu extrahieren,⁴⁹ die in der eher visionären Konzeption einer vollständigen Transformation historischer Datenspeicherformen und den Vorstellungen einer europäischen Zeitmaschine kulminieren,⁵⁰ aber sehr wohl auch praktische Verfahren in kleinen Anwendungsszenarien testen. Die automatischen Verfahren unterliegen aber den gleichen Überlegungen, wie wir sie für die assertive Edition angestellt haben: Wenn die wissenschaftliche Edition als Treuhänder der Historiker:innen über die intersubjektive Verwendbarkeit der Quellenaussagen als beobachterabhängige und im gemeinsamen Diskurs legitimierte Fakten wacht, dann gilt das ebenso für die Algorithmen „künstlicher Intelligenz“, die mit maschinellern Lernen typische menschliche Interpretationsvorgänge nachbilden und auf unbekanntes Material übertragen. Die dabei erzeugten Daten sind als Interpretationen zu kennzeichnen, die Verfahren ihrer Erzeugung zu dokumentieren und zu kontextualisieren, die Beziehung zur Originalquelle ist nachvollziehbar zu machen – alles Verfahren, die aus der Tradition kritischen Edierens wohlbekannt sind.

⁴⁹ Zum Beispiel Emanuela Boroş et al., „A comparison of sequential and combined approaches for named entity recognition in a corpus of handwritten medieval charters,“ in *17th International Conference on Frontiers in Handwriting Recognition (ICFHR), Sep 2020, Dortmund, Germany*, 2020, 79–84, doi: 10.1109/ICFHR2020.2020.00025; Juri Opitz und Anette Frank, „Deriving Players & Themes in the Regesta Imperii Using SVMs and Neural Networks,“ in *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*. Berlin, Germany, hg. v. Nils Reiter et al., 74–83 (Association for Computational Linguistics: Berlin, 2016), doi: 10.18653/v1/W16-2108.

⁵⁰ Frédéric Kaplan und Isabella di Lenardo, „Big Data of the Past,“ *Frontiers in Digital Humanities* 4 (2017): 1–12, doi: 10.3389/fdigh.2017.00012.

Bibliographie

- Arbeitskreis für Editionsgrundsätze. „Richtlinien für die Regestierung von Urkunden.“ *Blätter für Deutsche Landesgeschichte* 101 (1965): 1–7.
- Bernheim, Ernst. *Lehrbuch der historischen Methode und der Geschichtsphilosophie: mit Nachweis der wichtigsten Quellen und Hilfsmittel zum Studium der Geschichte*. Leipzig: Duncker & Humblot, 1908.
- Beretta, Francesco. „Pour une annotation sémantique des textes: le projet symogh.org et la Text encoding initiative.“ *Bruniana e Campanelliana, Ricerche filosofiche e materiali storico – testuali* 22/2. (2016): 453–465, doi: 10.19272/201604102005.
- Boroš, Emanuela, Verónica Romero, Martin Maarand, Kateřina Zenklová, Jitka Křečková, Enrique Vidal, Dominique Stutzmann und Christopher Kermorvant. „A comparison of sequential and combined approaches for named entity recognition in a corpus of handwritten medieval charters.“ In *17th International Conference on Frontiers in Handwriting Recognition (ICFHR), Sep 2020, Dortmund, Germany*. 2020. 79–84. doi: 10.1109/ICFHR2020.2020.00025.
- Burger, Daniel, Joachim Kemper und Christian Kruse. *Richtlinien der staatlichen Archive Bayerns für die Erstellung von Regesten*. München, 2010. Zugriff am 16.07.2021. http://www.gda.bayern.de/download/erschliessung/richtlinien_regesten.pdf.
- Clavaud, Florence und International Council on Archives Expert Group on Archival Description (ICA EGAD). „International Council on Archives Records in Contexts Ontology (ICA RIC-O) version 0.2.“ 2021. Zugriff am 20.07.2021. https://www.ica.org/standards/RiC/RiC-O_v0-2.html.
- Daquino, Marilena, Francesca Giovannetti und Francesca Tomasi. „Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell’edizione semantica del quaderno di appunti di Paolo Bufalini.“ *Umanistica Digitale* 3/7 (2019): 49–75. doi: 10.6092/issn.2532-8816/9091.
- Droysen, Johann Gustav. *Historik. Vorlesungen über Enzyklopädie und Methodologie der Geschichte* (1882). 8. Aufl., hg. v. Rudolf Hübner, München: R. Oldenbourg Verlag, 1977. *Encoded Archival Description*. Zugriff am 15.10.2021, <https://www.loc.gov/ead/>.
- Evans, Richard J. *Fakten und Fiktionen*. Frankfurt/New York: Campus Verlag, 1998.
- Fickers, Andreas. „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?“ *Zeithistorische Forschungen* 17/1 (2020): 157–168. Zugriff am 13.02.2021. <https://zeithistorische-forschungen.de/1-2020/5823>.
- Gabriel, Gottfried. „Fakten oder Fiktion? Zum Erkenntniswert der Geschichte.“ *Historische Zeitschrift* 297/1 (2013): 1–26. doi: 10.1524/hzhz.2013.0305.
- Haug-Moritz, Gabriele et al. „Der Regensburger Reichstag des Jahres 1576 – ein Pilotprojekt zum digitalen Edieren frühneuzeitlicher Quellen“ (gekürzter DFG-Antrag). 2018. Zugriff am 20.07.2021. https://static.uni-graz.at/fileadmin/projekte/reichstagsakten-1576/Projektbeschreibung/DFG-Antrag_gekuerzt_dt.pdf.
- Haug-Moritz, Gabriele. „Deliberieren. Zur ständisch-parlamentarischen Beratungskultur im Lateineuropa des 16. Jahrhunderts.“ *Historisches Jahrbuch* 141 (2021): 115–155.
- Härtel, Reinhard. „Mehr als ein Anhang. Das computererstellte Register.“ In *Historische Edition und Computer. Möglichkeiten und Probleme interdisziplinärer Textverarbeitung und Textbearbeitung. Berichte von der Internationalen Tagung „Historische Edition und Computer“*

- Karl-Franzens-Universität Graz, 26.-30. Oktober 1988. Hg. v. Anton Schwob, Karin Kranich-Hofbauer und Diethard Suntinger, 67–84. Graz, 1989.
- Hébert, Michel. *Parlementer. Assemblées représentatives et échange politique en Europe occidentale à la fin du Moyen Age*. Paris, 2014.
- Hitler, Adolf. *Hitler, Mein Kampf: eine kritische Edition*. Hg. v. Christian Hartmann, Thomas Vordermayer, Othmar Plöckinger und Roman Töppel. München: Institut für Zeitgeschichte, 2016.
- Hüttenberger, Peter. „Überlegungen zur Theorie der Quelle.“ In *Einführung in die Interpretation historischer Quellen. Schwerpunkt: Neuzeit* (UTB Bd. 1674). Hg. v. Bernd A. Rusinek, Volker Ackermann und Jörg Engelbrecht, 253–265. Paderborn/München/Wien/Zürich: Ferdinand Schöningh, 1992.
- Jordan, Stefan. „Vetorecht der Quellen.“ *Docupedia-Zeitgeschichte*, (2010): 1–7. doi: 10.14765/ZZF.DOK.2.570.V1.
- Kaplan, Frédéric und Isabella di Lenardo. „Big Data of the Past.“ *Frontiers in Digital Humanities* 4 (2017): 1–12. doi: 10.3389/fdigh.2017.00012.
- Kemper, Joachim. „Neue Richtlinien der staatlichen Archive Bayerns für die Erstellung von Urkundenregesten.“ *Archivalische Zeitschrift* 91 (2009): 209–219, doi: 10.7788/az.2009.91.1.209.
- Kittlmann, Jana und Christoph Wernhard. „KBSET – Knowledge-Based Support for Scholarly Editing and Text Processing with Declarative LaTeX Markup and a Core Written in SWI-Prolog.“ *Lecture Notes in Computer Science*, arXiv:2002.10329 [cs] 12057 (2020): 178–196. doi: 10.1007/978-3-030-46714-2_12.
- König, Mareike. „Geschichte digital Zehn Herausforderungen.“ In *Geschichtswissenschaft im 21. Jahrhundert*. 67–76. Berlin/Boston: De Gruyter Oldenbourg, 2020. doi: 10.1515/9783110689143-008.
- Koselleck, Reinhart. „Standortbindung und Zeitlichkeit: Ein Beitrag zur historiographischen Erschließung der geschichtlichen Welt.“ In *Objektivität und Parteilichkeit in der Geschichtswissenschaft*. Hg. v. Reinhart Koselleck, Wolfgang J. Mommsen und Jörg Rüsen, 17–46. München: dtv, 1977.
- Kunze, Horst. *Über das Registernmachen*. 4., erw. und verb. Aufl. München/London/New York/Paris: K G Saur, 1992.
- Mötsch, Johannes. „Vorteile und Grenzen der Regestentechnik.“ In *Die Regesta Imperii im Fortschreiten und Fortschritt* (Forschungen zur Kaiser- und Papstgeschichte des Mittelalters 20). Hg. v. Harald Zimmermann. 115–127. Köln, 2000.
- Mulligan, Kevin und Fabrice Correia. „Facts.“ In *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition). Hg. v. Edward N. Zalta. Zugriff am 20.07.2021. <https://plato.stanford.edu/archives/win2020/entries/facts/>.
- Opitz, Juri und Anette Frank. „Deriving Players & Themes in the Regesta Imperii Using SVMs and Neural Networks.“ In *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*. Berlin, Germany. Hg. v. Nils Reiter, Beatrice Alex and Kalliopi A. Zervanou. 74–83. Berlin: Association for Computational Linguistics, 2016. doi: 10.18653/v1/W16-2108.
- Pollin, Christopher. „Digital Edition Publishing Cooperative for Historical Accounts and the Bookkeeping Ontology.“ In *Proceedings of the Doctoral Symposium on Research on Online Databases in History (RODBH 2019)*. Hg. v. Thomas Riechert, Francesco Beretta und Georg Bruseke. 7–14. Leipzig, 2020. Zugriff am 20.07.2021. <http://ceur-ws.org/Vol-2532/paper1.pdf>.

- Regesta Imperii. „Richtlinien für die Registrierung von Urkunden (1978).“ In *Die Regesta Imperii im Fortschreiten und Fortschritt* (Forschungen zur Kaiser- und Papstgeschichte des Mittelalters 20). Hg. v. Harald Zimmermann. 87–94. Köln, 2000. Rüsen, Jörn. *Grundzüge einer Historik*. II. Göttingen: Vandenhoeck & Ruprecht, 1986.
- Rüsen, Jörn. *Geschichte denken: Erläuterungen zur Historik*. Wiesbaden: Springer VS, 2020.
- Sahle, Patrick. *Digitale Editionsformen: zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 3: Textbegriffe und Recodierung*. Schriften des Instituts für Dokumentologie und Editorik 9. Norderstedt: Books on Demand, 2013.
- Schreurs, Anna, Carsten Blüm und Thorsten Wübbena (eds.). *Sandart.net: Eine netzbasierte Forschungsplattform zur Kunst- und Kulturgeschichte des 17. Jahrhunderts*. Wolfenbüttel: Herzog-August-Bibliothek, [2012]. Zugriff am 20.07.2021. <http://www.sandart.net/de>.
- Searle, John R. „Social ontology: Some basic principles.“ *Anthropological theory* 6/1 (2006): 12–29. doi: 10.1177/1463499606061731.
- Stollberg-Rilinger, Barbara. *Des Kaisers alte Kleider: Verfassungsgeschichte und Symbolsprache des Alten Reiches*. 2., durchgesehene und aktualisierte Edition. München: C.H.Beck, 2013.
- Thaller, Manfred. „Reproduktion, Erschließung, Edition, Interpretation: Ihre Beziehungen in einer digitalen Welt.“ In *Vom Nutzen des Edierens: Akten des internationalen Kongresses zum 150-jährigen Bestehen des Instituts für Österreichische Geschichtsforschung, Wien, 3.–5. Juni 2004*. Hg. v. Brigitte Merta, Andrea Sommerlechner und Herwig Weigl. (Wien: Oldenbourg, 2005), 205–227.
- Thaller, Manfred. „Historical Information Science: Is there such a Thing? New Comments on an old Idea [1993].“ *Historical Social Research/Historische Sozialforschung, Supplement* 29 (2017): 260–286. doi: 10.12759/hsr.suppl.29.2017.260-286.
- Tomasi, Francesca. „L’informazione digitale e il Web semantico. Il caso delle scholarly digital editions.“ In *Informatica umanistica: risorse e strumenti per lo studio del lessico dei beni culturali*. (A cura di) Valeria Zotti und Ana Pano Alamán, 157–174. Firenze: Firenze University Press, 2017.
- Torstendahl, Rolf. „Fact, Truth, and Text: The Quest for a Firm Basis for Historical Knowledge Around 1900.“ *History and Theory* 42/3 (2003): 305–331. doi: 10.1111/1468-2303.00246.
- Tschiggerl, Martin, Thomas Walach und Stefan Zahlmann. *Geschichtstheorie*. Wiesbaden: Springer Fachmedien Wiesbaden, 2019. doi: 10.1007/978-3-658-22882-8.
- van Zundert, Joris J. „Screwmenetics and Hermeneumericals: The Computationality of Hermeneutics.“ In *A New Companion to Digital Humanities*. Hg. v. S. Schreibman, R. Siemens, und J. Unsworth, 331–347. John Wiley & Sons, Ltd., 2015. doi: 10.1002/9781118680605.ch23.
- Virmond, Wolfgang (unter Mitarbeit von Holden Kelm). „Schleiermachers Tageskalender 1808–1834.“ Hg. v. Elisabeth Blumrich, Christiane Hackel und Wolfgang Virmond. Berlin-Brandenburgische Akademie der Wissenschaften, Berlin. Zugriff am 23.07.2021. <https://schleiermacher-digital.de/tageskalender/index.xql>.
- Vogeler, Georg. „The ‚Assertive Edition‘.“ *International Journal of Digital Humanities* 1/2 (2019): 309–322. doi: 10.1007/s42803-019-00025-5.
- Wettlaufer, Jörg, Christopher Johnson, Martin Scholz, Mark Fichtner und Sree Ganesh Thotempudi. „Semantic Blumenbach: Exploration of Text-Object Relationships with Semantic Web Technology in the History of Science.“ *Digital Scholarship in the Humanities* 30 (supl_1) (2015): i187–i198. doi: 10.1093/llc/fqv047.
- Wikipedia. „Tatsache.“ Zugriff am 20.07.2021. <https://de.wikipedia.org/wiki/Tatsache>.

Silke Schwandt

Geschichte visualisieren: Digitale Praktiken in der Geschichtswissenschaft als Praktiken der Wissenschaftsreflexion

Abstract: Digitization and digital data play a major role in our society today and also have a strong impact on our scholarly practices. This paper will take a closer look at how practices in historical scholarship are changing and thus how we are currently able to conceptualize and analyze historical change as a research object in historical scholarship. Practices of visualization in the humanities serve as an example to demonstrate the transformation of our scholarly practices. It will be shown how conceptions of change associated with digitization and new digital methods can be brought together in historical research: How are our practices of theorizing change evolving in the digital age?

Keywords: visualization, history, time, change, practice

Zusammenfassung: Digitalisierung und digitale Daten spielen eine große Rolle in unserer heutigen Gesellschaft und haben auch einen starken Einfluss auf unsere wissenschaftliche Praxis. Der Beitrag wird sich näher damit befassen, wie sich geschichtswissenschaftliche Praktiken und damit die Art und Weise verändern, wie wir gegenwärtig in der Lage sind, historischen Wandel als Forschungsgegenstand in der Geschichtswissenschaft zu konzeptualisieren und zu analysieren. Praktiken der Visualisierung in den Geisteswissenschaften dienen als Beispiel, um die Transformation unserer Wissenschaftspraxis zu demonstrieren. Es wird gezeigt, wie sich die mit der Digitalisierung verbundenen Konzeptionen des Wandels und die neuen digitalen Methoden in der historischen Forschung zusammenführen lassen: Wie entwickeln sich unsere Praktiken der Theoretisierung von Wandel im digitalen Zeitalter?

Schlagwörter: Visualisierung, Geschichte, Zeit, Wandel, Praktiken

1 Doing History Digitally: „Geschichte“ und „Wandel“ im digitalen Zeitalter

Die Frage, was Geschichtswissenschaftler:innen eigentlich tun, ist oft beantwortet worden und entzieht sich fast genauso oft trotzdem einer klaren Definition. Jürgen Kocka hat 2008 formuliert: „Geschichte gehört in sehr unterschiedlichen Formen zum Leben. [...] Wer sich und seine Welt verstehen und sich in ihr verhalten will, kommt ohne Geschichte nicht aus.“¹ Das bezieht sich vor allem auf das, was Jürgen Sarnowsky 2016 als „Wissensarchive“ der Menschheit“ beschrieben hat: die Funktion der Geisteswissenschaften im Allgemeinen und der Geschichtswissenschaft im Besonderen als ein Archiv, das als ein „kollektives, sich selbst reflektierendes, kulturelles Gedächtnis der Menschheit“ diene.² Der Gegenstand dieser Geschichtswissenschaft ist die Auseinandersetzung mit der Vergangenheit und dem, was man historische Fakten nennt. Schon Edward Carr hat sich in einem inzwischen als Klassiker geltenden Werk zur Frage „Was ist Geschichte?“ mit der Frage nach dem Wesen dieser Fakten auseinandergesetzt und Geschichte als einen Prozess der Wechselwirkung zwischen Tatsachen und Interpretationen, „zwischen dem Historiker und seinen Fakten“ beschrieben, der „ein unendlicher Dialog zwischen Gegenwart und Vergangenheit“ sei.³ Zentral für diesen Dialog sei stets die Standortgebundenheit der Historikerin oder des Historikers: „Only if the empirical evidence of the past is shaped according to the standpoints of those who communicate in the realm of historical consciousness and according to the value system which is derived from this standpoint does it acquire the quality of a plausible history; neutrality is the end of history.“⁴ Diese Standortgebundenheit betrifft aber nicht allein die Zuordnung von Ereignissen und historischen Fakten, sondern auch das Bild der Geschichte selbst – und eng damit verbunden die Frage nach „Zeit“ und „Wandel“. Geschichtswissenschaft beschäftigt sich zentral mit der Beschreibung der Wahrnehmung von Wandlungsprozessen sowie mit deren Erklärung. „Doing History“ – *Geschichte machen* – ist also immer auch die Konzeptualisierung von Zeit und Wandel. Gerade diese Konzepte sind es, die im Zeitalter der Digitalisierung neu diskutiert werden müssen. Welchen Einfluss hat die Digitalisierung

1 Jürgen Kocka, „Geschichte als Wissenschaft,“ in *Geschichte: Studium, Wissenschaft, Beruf*, hg. v. Gunilla Budde (Berlin: Akad.-Verl., 2010), 12.

2 Jürgen Sarnowsky, „Was sind und zu welchem Ende studiert man Geisteswissenschaften?“, <http://mittelalter.hypotheses.org/7972>. Zugriff am 02.08.2021.

3 Edward H. Carr, *Was ist Geschichte?*, 6. Aufl. (Stuttgart u.a.: Kohlhammer, 1981), 30.

4 Jörn Rüsen, „Historical Objectivity as a Matter of Social Values,“ in *Historians and Social Values*, hg. v. Joep Leerssen (Amsterdam: Amsterdam Univ. Press, 2000), 63.

auf unsere Wahrnehmung von Zeit? Welche neuen Praktiken zur Konzeptualisierung von Zeit und Wandel ergeben sich unter digitalen Bedingungen und mithilfe von digitalen Methoden? Wie verändert sich der Standort der Historiker:innen unter dem Paradigma der Digital History?

Moderne Gesellschafts- und Geschichtstheorien sprechen oft von Beschleunigungswahrnehmung, wenn es um Geschichte und Wandel im digitalen Zeitalter geht. Hartmut Rosa sieht die Moderne dem Prinzip der Dynamisierung unterworfen, die ihrerseits von Beschleunigung geprägt ist,⁵ und schreibt den „neueren Beschleunigungsdiskurs“ seit 1989 auch der von ihm so genannten digitalen Revolution zu.⁶

Reinhart Koselleck korreliert in seinem Entwurf der *Zeitschichten* technologische und soziale Beschleunigung in folgender Weise: Er behauptet, dass die Erfahrung beschleunigten Wandels nicht notwendigerweise mit technologischem Fortschritt verbunden sei, wie man vermuten könnte: Die Erwartung von Beschleunigung im Sinne einer antizipierten, ja erhofften Verkürzung der Zeit gebe es seit der jüdisch-christlichen apokalyptischen Literatur, aber tatsächliche Beschleunigung, die das Potenzial hat, die Realität zu verändern, gebe es erst in der technologisch geprägten Neuzeit.⁷ Insofern müssten technologische Beschleunigung und gesellschaftliche Beschleunigung getrennt voneinander betrachtet werden, auch wenn das Gefühl der Beschleunigung, das unsere gegenwärtige Zeitwahrnehmung präge, mit der Technologisierung unserer Welt verbunden sein mag, damit aber nicht gleichzusetzen sei.⁸

Paul Virilio beschäftigt sich in seinem im Jahr 1999 auf Englisch erschienenen Essay „Polar Inertia“ mit der Rolle der Technologie in unserem Verständnis von Zeit. Er behauptet: „[...] in our ordinary everyday life, we are passing from the extensive time of history to the intensive time of an instantaneity without history made possible by the technologies of the hour.“⁹ In gewisser Weise lässt sich dieser Gedanke mit dem verbinden, was andere Theoretiker das „Ende der Geschichte“ genannt haben. Während Francis Fukuyama das Ende der Geschichte darin sah, dass wir mit der Demokratie die ultimative Regierungsform

⁵ Vgl. Hartmut Rosa, *Beschleunigung: Die Veränderung der Zeitstrukturen in der Moderne*, 11. Aufl., (Frankfurt a. M.: Suhrkamp, 2016), 430.

⁶ Rosa, *Beschleunigung*, 40.

⁷ Reinhart Koselleck, „Einleitung,“ in *Zeitschichten: Studien zur Historik*, hg. v. Reinhart Koselleck (Frankfurt a. M.: Suhrkamp, 2003), 15.

⁸ Koselleck, „Einleitung,“ 15.

⁹ Paul Virilio, *Polar Inertia*, Published in association with Theory, Culture & Society (London: Sage Publications, 1999), 9.

erreicht haben,¹⁰ sieht Virilio das Ende der Geschichte im Verständnis der Zeit als „instantaneity without history“ sowie in der Herrschaft der modernen Technologie.¹¹ Leben wir also in einer Zeit des „ewigen Jetzt“?

François Hartog bezeichnet das gegenwärtige zeitliche Regime als das eines Präsentismus, in dem wir unsere Beziehung zur Vergangenheit und zur Zukunft verloren haben.¹² Hartog betont, dass es die soziale Position oder der Standpunkt sei, der das Empfinden von und die Perspektive auf Zeit präge. Was für den einen ein Gefühl von Beschleunigung und Globalisierung darstelle, könne sich für den anderen wie Stagnation und der Verlust jeglicher Zukunftsperspektive anfühlen. Statt das digitale Zeitalter als das Ende der Zeit oder unsere Gegenwart als das ‚ewige Jetzt‘ zu sehen, plädiert Helge Jordheim für ein Verständnis von Zeit als „a new multiplicity of time“, der wir „in the new immediacies and accelerating rhythms of digital technology“ gegenüberstehen.¹³ Es seien Praktiken der Synchronisation, der Vergleichzeitung, die der Instabilität einer Vielzahl von Zeiten entgegenwirke.¹⁴ Für Jordheim sind die Vielheit der Zeit und die Praktiken der Synchronisation zwei Teile derselben Gleichung, die Koselleck als „Gleichzeitigkeit des Ungleichzeitigen“ gerahmt hat.¹⁵ Dabei ist Synchronisation eine soziale und kulturelle Praxis, etwas sozial und kulturell Konstruiertes, das Orientierung in der Vielfalt der Zeiten bietet. Sie soll dem Gefühl der Zeitlosigkeit oder der Unmittelbarkeit der Gegenwart etwas entgegensetzen.

Als soziale Praxis betrifft Synchronisation die Gesellschaft als Ganzes. Zeit und Raum sind die materiellen Grundlagen der Gesellschaft und werden durch neue Technologien und Praktiken transformiert. Manuel Castells beschreibt diese Transformation als den Beginn der „Netzwerkgesellschaft“.¹⁶ Der Handel und insbesondere die stark technologisierten Praktiken der Börse, die auch von Karin Knorr Cetina als Motoren des (gesellschaftlichen) Wandels beschrieben wur-

10 Francis Fukuyama, *The End of History and the Last Man: With a New Afterword*, 1. Free Press trade paperback ed., [Nachdr.] (New York, NY: Free Press, 2006).

11 Virilio, *Polar Inertia*, 78.

12 François Hartog, *Regimes of Historicity: Presentism and Experiences of Time*, (New York: Columbia University Press, 2017).

13 Helge Jordheim, Introduction: Multiple Times and the Work of Synchronization, *History and Theory* 53 (2014): 499.

14 Jordheim, „Introduction: Multiple Times and the Work of Synchronization,“ 502.

15 Reinhart Koselleck, „„Neuzeit“: Zur Semantik moderner Bewegungsbegriffe,“ in *Vergangene Zukunft: Zur Semantik geschichtlicher Zeiten*, hg. v. Reinhart Koselleck, 1. Aufl., Theorie (1979), 325.

16 Manuel Castells, *The Rise of the Network Society*, 2. Aufl., with a new pref., [Nachdr.], 3 vols., *The Information Age Economy, Society and Culture* / Manuel Castells; Vol. 1 (Chichester: Wiley-Blackwell, 2010), doi: 10.1002/9781444319514.

den,¹⁷ beeinflussen Konzepte von Zeit und Raum grundlegend. Raum wird in Netzwerken aufgelöst, Zeit wird zeitlos in Netzwerken von Strömen, die Vergangenheit und Zukunft ignorieren. Castells analysiert: „infinite social distance is created between this meta-network and most individuals, activities, and locales around the world.“¹⁸ Er geht so weit, zu behaupten, dass es nicht mehr Menschen oder Praktiken seien, sondern „[...] unseen logic of the meta-network where value is produced, cultural codes are created, and power is decided.“¹⁹ Zeit und Raum scheinen sich unserer Kontrolle zu entziehen und werden gleichzeitig durch soziale Interaktion geschaffen, wobei sich diese soziale Interaktion im digitalen Zeitalter grundlegend verändert hat. Die virtuelle Kommunikation über das Internet und in sozialen Medien macht Zeit und Raum fluide wie nie zuvor. Wir können zu jeder Zeit überall sein und trotzdem an dem teilnehmen, was Castells „a purely cultural pattern of social interaction and social organization“ nennt.²⁰ Das scheint der Moment der ultimativen Gleichzeitigkeit zu sein.

Aber (nicht nur) als Historiker:in muss man fragen: Wie stellen sich, unter der Prämisse dieser Diagnosen, Konzepte des Wandels dar? Leben wir in einer präsentistischen Gesellschaft, die die Vergangenheit ignoriert und Angst vor der Zukunft hat? Hat die Beschleunigung unserer Gesellschaft zu einem Moment der Unmittelbarkeit ohne Geschichte geführt? Folgen wir nur noch den Strömen des Netzes?

2 Geschichte visualisieren: Vergleichzeitigung im Bild?

Geschichte zu visualisieren ist keine neue Idee. Helge Jordheim erinnert uns daran, dass schon im Laufe des 18. Jahrhunderts synchronisierende Tabellen zu den Werken der Universalgeschichte gehörten.²¹ Es gibt aber auch neuere Beispiele. Im Jahr 1952 veröffentlichte Arno Peters, ein deutscher Historiker und Kartograph, seine *Synchroptische Weltgeschichte*, die hauptsächlich aus verschiedenen Tabellen besteht (so genannte *Zeittafeln*), die verschiedene Ereignisse in

¹⁷ Karin K. Cetina und Alex Preda, „The Temporalization of Financial Markets: From Network to Flow,“ *Theory, Culture & Society* 24/7–8 (2007), 10.1177/0263276407084700.

¹⁸ Castells, *The Rise of the Network Society*, 507.

¹⁹ Ibid., 508.

²⁰ Ibid.

²¹ Jordheim, „Introduction: Multiple Times and the Work of Synchronization,“ 515.

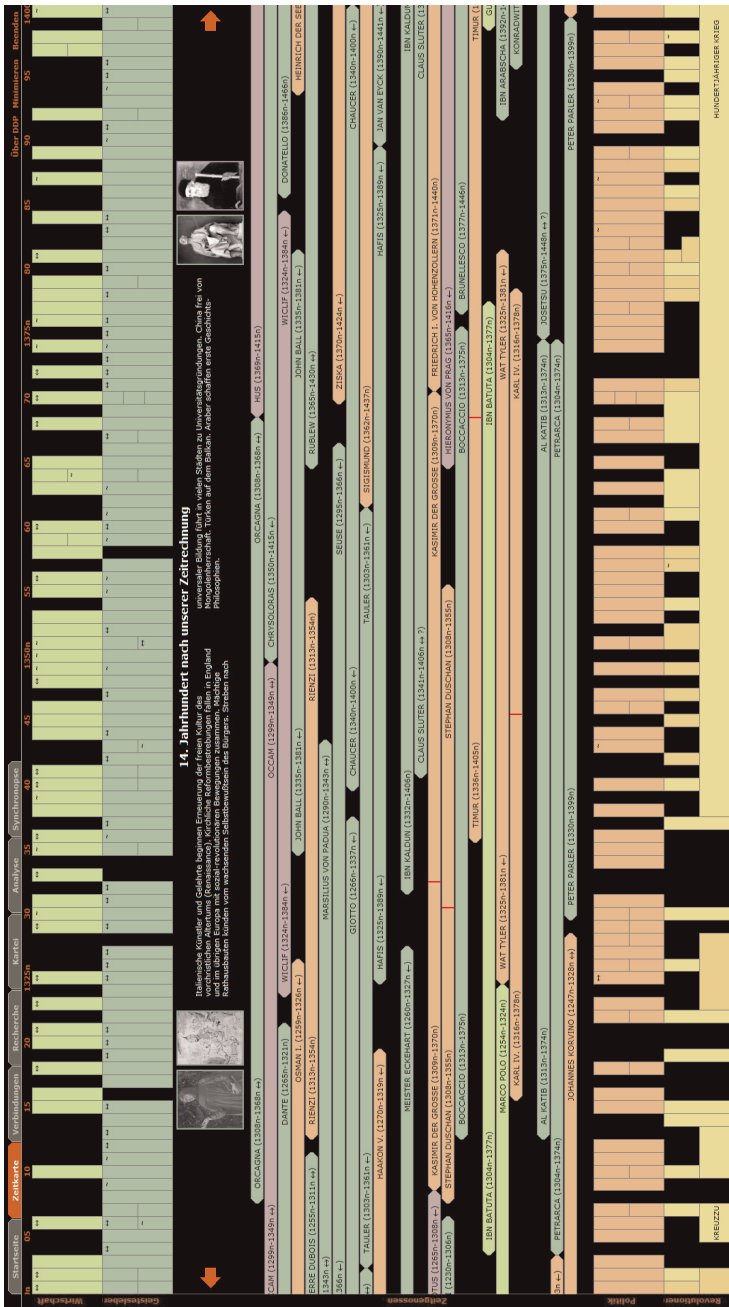


Abb. 1: Screenshot von ddp (Der Digitale Peters) mit Daten für das vierzehnte Jahrhundert CE.

verschiedenen Regionen der Welt zeigen, die zur gleichen Zeit stattfanden.²² Diese umfasst den Zeitraum von 3000 v. Chr. bis 2000 n. Chr. und zeigt auf jeder Doppelseite jeweils ein Jahrhundert. Peters kategorisiert die Ereignisse und Prozesse, die er katalogisiert, in vier Hauptkategorien: Wirtschaft, Geistesleben, Politik und Krieg beziehungsweise Revolutionen. Der Mittelteil jeder Seite ist den *Lebenslinien* von zeitgenössischen Persönlichkeiten aus diesen Bereichen gewidmet.

Die in Abbildung 1 dargestellte synchrone Ansicht²³ der Geschichte des vierzehnten Jahrhunderts ermöglicht es uns, Ereignisse zu sehen, die in verschiedenen Teilen der Welt zur gleichen Zeit stattfanden sowie die *Lebenslinien* von Menschen, die Zeitgenossen waren, sich aber vielleicht nie getroffen haben. Sie suggeriert eindeutig eine Synchronizität von Ereignissen und vielleicht sogar eine Vielzahl von Zeiten [multiplicity of times], die durch die Visualisierung synchronisiert werden. In diesem Sinne ist die Abbildung mehr als nur eine Illustration. Sie bietet durch die Art der Darstellung zusätzliche Informationen über die dargestellten Daten und kann auf explorative Weise verwendet werden, um weitere Einsichten hervorzurufen.²⁴ Um die Recherche mit diesem Werkzeug weiter zu erleichtern, enthält die digitalisierte Version der „Synchronoptischen Weltgeschichte“, die 2010 veröffentlicht wurde, Notizen und Karteikarten, die von Peters und seinem Team während des Schreibprozesses geschrieben wurden.²⁵ So können Nutzer:innen nicht nur von einer synchronen Sicht auf die Geschichte profitieren, sondern auch einen Einblick in die Entscheidungen und Ideen gewinnen, die der Konzeption der *Weltgeschichte* zugrunde liegen.

3 Visualisierungen in den Geisteswissenschaften: Versuch einer Einordnung

Visualisierungen in den Geisteswissenschaften fordern unsere Wissenschaftspraktiken auf mehreren Ebenen heraus. Ausgehend von der geschichtswissen-

²² Arno Peters, *Synchronoptische Weltgeschichte* (Frankfurt a. M.: Universum-Verlag, 1952).

²³ Arno Peters et al., *Der digitale Peters: Arno Peters synchronoptische Weltgeschichte*, Programmversion DDP-2010-03 (Wiener Neudorf, Wiesbaden: ÖBZ; Büro W, 2010).

²⁴ Vgl. Martyn Jessop, „Digital Visualization as a Scholarly Activity“, *Literary and Linguistic Computing* 23/3 (2008), doi: 10.1093/lc/fqn016 über die Unterscheidung zwischen „Illustration“ und „Visualisierung“ und deren explorativen Wert für die Informationsgenerierung.

²⁵ Peters et al., *Der digitale Peters*.

schaftlichen Aufgabe der Beschreibung, Interpretation und auch Konzeption von Wandel und (historischer) Zeit stellt sich die Frage, welchen Mehrwert das Visualisieren haben kann.

Abbildung 1 oben konstituiert eine synchrone Betrachtung durch die gleichzeitige Darstellung verschiedener Aspekte oder Ereignisse, die im Wesentlichen durch den Vergleich im Hinblick auf das Datum ihres Auftretens miteinander in Verbindung gebracht werden. Auch wenn dies alles in einem Bild geschieht und nicht der Vergleich mehrerer Bilder im Mittelpunkt steht, kann die Praxis der Synchronisation als „vergleichendes Sehen“ konzipiert werden. Johannes Grave hat dieses Konzept für die Kunstgeschichte entwickelt und warnt davor, sich auf das Herausgreifen und Isolieren bestimmter Elemente eines Bildes zu konzentrieren, obwohl er vermutet, dass Praktiken der vergleichenden Betrachtung die Tendenz haben, ein Bild in verschiedene, kleinere Elemente zu zerlegen, die sich leichter vergleichen lassen.²⁶ Für die Kunstgeschichte würde dies der Annahme widersprechen, dass ein Bild eine Botschaft als Ganzes vermittele. Gilt das auch für (geistes-)wissenschaftliche Visualisierungen?

Es scheint notwendig zu sein, zwischen Bildern und Visualisierungen, zwischen ästhetischer und wissenschaftlicher Bildgebung zu unterscheiden – auch wenn die Unterscheidung am Ende irreführend sein könnte. Bettina Heintz und Jörg Huber haben fünf Kriterien für eine solche Unterscheidung vorgeschlagen, die dabei helfen, die produktive Kapazität wissenschaftlicher Visualisierungen als Werkzeuge für geisteswissenschaftliche Forscher:innen herauszuarbeiten: Visualisierungen sind *Werkzeuge*. Sie unterscheiden sich von (künstlerischen) Bildern in Fragen der *Indexikalität*, der *Interikonizität*, der *Performativität* sowie als Produkte von *Mensch-Maschine-Interaktion*.²⁷ Jörg Huber betont ihre „Vollzugsorientiertheit“ im Gegensatz zur Selbstreferenzialität eines künstlerischen Bildes.²⁸ Sie hätten zwar ästhetische Qualitäten, diese würden aber in der Regel nicht als Aufwertung der Aussage einer Visualisierung interpretiert. Visualisierungen schafften dagegen wissenschaftliche Legitimität durch visuelle Evidenz, eine Funktionalität, die ihr eigenes soziokulturelles Gewicht hat.²⁹ Während Bil-

²⁶ Johannes Grave, „Vergleichen als Praxis: Vorüberlegungen zu einer praxistheoretisch orientierten Untersuchung von Vergleichen,“ in *Die Welt beobachten: Praktiken des Vergleichens*, hg. v. Angelika Eppele und Walter Erhart (Frankfurt a. M. u.a.: Campus Verlag, 2015), 151.

²⁷ Vgl. Bettina Heintz und Jörg Huber, „Der verführerische Blick: Formen und Folgen wissenschaftlicher Visualisierungsstrategien,“ in *Mit dem Auge denken: Strategien der Sichtbarmachung in wissenschaftlichen und virtuellen Welten. Theorie – Gestaltung 1*, hg. v. Bettina Heintz und Arnold Benz (Zürich: Ed. Voldemeer, 2001).

²⁸ Heintz und Huber, „Der verführerische Blick,“ 28.

²⁹ Ibid., 29.

der auf etwas Äußeres verwiesen, seien Visualisierungen Modelle, die nicht abschließend durch die Objekte bestimmt sind, die sie darstellen sollen.³⁰

Da wissenschaftliche Visualisierungen zwangsläufig nicht eine gegebene Realität abbilden, sondern auf Messdaten und Algorithmen der Bilderzeugung verweisen, sind ihre Referenzen Daten und Software, nicht die Realität.³¹ Dennoch bleibt der Eindruck, dass Visualisierungen die Wahrheit sichtbar machen – erst recht, wenn sie automatisch generiert werden. Heintz spricht von einer „Illusion der [...] Objektivität“, die diesen Bildern innewohne,³² die aber unbedingt hinterfragt werden müsse. Auch Algorithmen sind menschengemacht und unterliegen kulturellen Konventionen, auch wenn ihnen oft ein hohes Maß an Evidenz zugeschrieben und Vertrauen entgegengebracht wird. Auch die Visualisierungen selbst sind an Konventionen gebunden.³³ Das hat mit der *Mensch-Maschine-Interaktion* zu tun, die für Prozesse der computerbasierten Bilderzeugung grundlegend ist. Wer entwirft und produziert eine Visualisierung? „Die Macht der Apparatur liegt in der Visualisierung, die Kompetenz des Forschers bleibt bei der gestalteten Darstellung und damit bei der Interpretation.“³⁴ Bettina Heintz nennt dieses Zusammenspiel von Bilderzeugung und visueller Interpretation einen „genuin experimentellen Prozess“.³⁵ Während sie diesen Prozess als einen problematisiert, bei dem die Bildinformation „verändert, gefiltert, geglättet und angepasst wird, bis eine Relation zwischen dem Erwarteten und dem Dargestellten besteht“ – kurz: manipuliert wird, bis es zur Geschichte passt –, sieht Martyn Jessop weitere Vorteile und neue Potenziale im Einsatz von Visualisierungswerkzeugen für die geisteswissenschaftliche Forschung. Er behauptet, dass diese Werkzeuge die Fähigkeiten hätten „to allow visual perception to be used in the creation or discovery of new knowledge.“³⁶ Die Beschäftigung mit Visualisierungen ist seiner Ansicht nach ein dynamischer Prozess, der auch erkenntnistheoretische Fragen zur Visualisierung aufwirft.³⁷ Ausschlaggebend ist hier die *Performativität* der Visualisierung. Das Visualisieren von Informationen erzeugt neue Informationen – es ist keine bloße Illustration, sondern demonstriert die Verschiebung in der Praxis der Bilderzeugung „von der Ontologie des Bildes zur Performativität des Bildes, von der formalen Quali-

30 Ibid., 12.

31 Ibid., 30.

32 Ibid., 19.

33 Vgl. auch Katharina Zweig, *Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können* (München: Heyne, 2019).

34 Heintz und Huber, „Der verführerische Blick“, 31.

35 Ibid., 23.

36 Jessop, „Digital Visualisation as a Scholarly Activity“, 282.

37 Ibid.

tät zur kommunikativen Funktion.“³⁸ Martyn Jessop fügt hinzu, „that an image can only be a true visualization when it is the primary carrier of information.“³⁹

Aber warum Geschichte visualisieren? Das Argument war bisher, dass synchronisierende Visualisierungen uns erlauben, Zeit und Wandel anders zu sehen, dass die Visualisierung eine Praxis der Vergleichzeitigung ist. Visualisierungen sind Datenmodelle. Sie umfassen und komprimieren große Datenmengen zu einem Bild, das diese Daten gleichzeitig in einer räumlichen Konfiguration präsentieren kann.⁴⁰ Es ist dieses Zusammentreffen von Ereignissen, diese Verdichtung von Informationen, die Visualisierungen für die geisteswissenschaftliche Forschung produktiv macht. Wenn wir Abbildung 1 betrachten, sehen wir Dinge, die wir vorher vielleicht nicht in Relation gesehen haben, und das löst neue Forschungsfragen und damit neue historische Erkenntnisse aus. Wenn es stimmt, dass Visualisierungen mit uns kommunizieren, indem sie unsere Erwartungen konfrontieren, dann ist es das Brechen dieser Erwartung durch Irritation, das Wissen schafft.⁴¹ Genau in dieser kreativen Chance liegt auch die größte Herausforderung für den Einsatz von Visualisierungen in der geisteswissenschaftlichen Forschung.

Noah Bubenhofer formuliert drei kritische Thesen zu Visualisierungspraktiken in den Digital Humanities: Erstens entsprächen Visualisierungen einem „utilitaristischen und positivistischen Denkstil“, der mit den Geisteswissenschaften nicht vereinbar sei. Sie kämen aus einem ganz anderen wissenschaftlichen Paradigma, nämlich dem der Naturwissenschaften. Zweitens unterschätzten Geisteswissenschaftler:innen die Bedeutung der technischen und algorithmischen Grundlagen von Visualisierungsmethoden. Die Frage, ob ein Verständnis von Programmierung und anderen Praktiken der Informatik notwendig sei, um Visualisierungen zu nutzen und zu interpretieren, sei obendrein sehr umstritten.⁴² Jenna Burrell weist darauf hin, dass „[...] the workings of machine learning algorithms can escape full understanding and interpretation by humans, even for those with specialized training, even for computer scientists.“⁴³ Was bedeutet das aber für den wissenschaftlichen Forschungsprozess

³⁸ Ibid., 35.

³⁹ Ibid., 283.

⁴⁰ Heintz und Huber, „Der verführerische Blick“, 13.

⁴¹ Vgl. zur Idee der produktiven Irritation durch computergestützte Methoden in der Geschichtswissenschaft: Silke Schwandt, „Methoden für die Historische Semantik“, *Geschichte und Gesellschaft* 44/1 (2018), doi: 10.13109/gege.2018.44.1.107.

⁴² Noah Bubenhofer, „Drei Thesen zu Visualisierungspraktiken in den Digital Humanities“, *Rechtsgeschichte – Legal History* 24 (2016), doi: 10.12946/rg24/351-355, 351.

⁴³ Jenna Burrell, „How the Machine ‚Thinks‘: Understanding Opacity in Machine Learning Algorithms“, *Big Data & Society* 3/1 (2016), doi: 10.1177/2053951715622512, 10.

und unsere Praktiken? Die Anwendung von algorithmischen Verfahren in der Geschichtswissenschaft verlangt nach einer präzisen Beschreibung dessen, was diese überhaupt leisten sollen – es bedarf also einer Analyse der eigenen Forschungspraxis, um von den neuen Methoden zu profitieren, so hat es schon John Unsworth gefordert.⁴⁴ Wenn wir uns dann aber nicht um ein Verständnis der algorithmischen Verfahren bemühen, geben wir die Kontrolle über einen Teil unseres Forschungsprozesses ab. Hier ist die Reflexion der eigenen Wissenschaftspraxis besonders wichtig. Was machen wir eigentlich, wenn wir als Geisteswissenschaftler:innen forschen?

Bubenhofers konstatiert drittens, dass die Programmierung von Visualisierungswerkzeugen dazu neigt, traditionelle Interpretationsmethoden nachzuahmen, anstatt die Chancen eines neuen datengetriebenen Paradigmas zu nutzen.⁴⁵ Auch Franco Moretti, einer der wichtigsten Pioniere auf dem Gebiet der digitalen Literaturwissenschaft, warnt vor dem Auseinanderfallen von neuen Methoden und Daten auf der einen Seite und alten Methoden auf der anderen: „Algorithms generate new facts, whose interpretation continues however to rely on a different hermeneutic tradition.“⁴⁶

Abgesehen von der offenen Frage, ob Visualisierungen wie Bilder behandelt werden können oder ob wir Algorithmen verstehen müssen, um Visualisierungen zu nutzen, bleibt die Tatsache bestehen, dass Visualisierungspraktiken in den Geisteswissenschaften neue Herausforderungen darstellen und neue Perspektiven bieten.

4 Geschichte auf einen Blick: Methoden und Interpretationen

Visualisierungen sind Komprimierungen von großen Datenmengen in einem einzigen Bild. Auf diese Weise sind sie Datenmodelle, „a corrected, rectified, regiminted, and in many instances idealized version of the data we gain from im-

⁴⁴ John Unsworth, „Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?“, <https://johnunsworth.name/Kings.5-00/primitives.html>. Zugriff am 19.02.2022.

⁴⁵ Vgl. Bubenhofers, „Drei Thesen zu Visualisierungspraktiken in den Digital Humanities“, 351.

⁴⁶ Franco Moretti, „Patterns and Interpretations“, *Pamphlets of the Stanford Literary Lab* 15 (2017): 6.

mediate observation, the so-called raw data.“⁴⁷ Wie passt das zu unserem vorherigen Beispiel der visualisierten Geschichte? Welche Rohdaten liegen dem „Digitalen Peters“ zugrunde?

Historische Ereignisse oder Entwicklungen sowie die Informationen zu historischen Persönlichkeiten, die Peters und sein Team recherchierten, wurden auf Karteikarten dokumentiert. Abbildung 2 zeigt eine Karteikarte für das Jahr 1331, dem Jahr des ersten dokumentierten Einsatzes von Kanonen in Europa, auf italienisch-deutschem Grenzgebiet im Friaul. Neben dem Ereignis sind auf der Karte auch die Informationsquellen, die weiteren wirtschaftlichen und technologischen Entwicklungen im Zusammenhang mit der Verwendung von Schießpulver im vierzehnten Jahrhundert in Europa und sogar das Entstehungsdatum der Karteikarte selbst vermerkt (10.5.51).

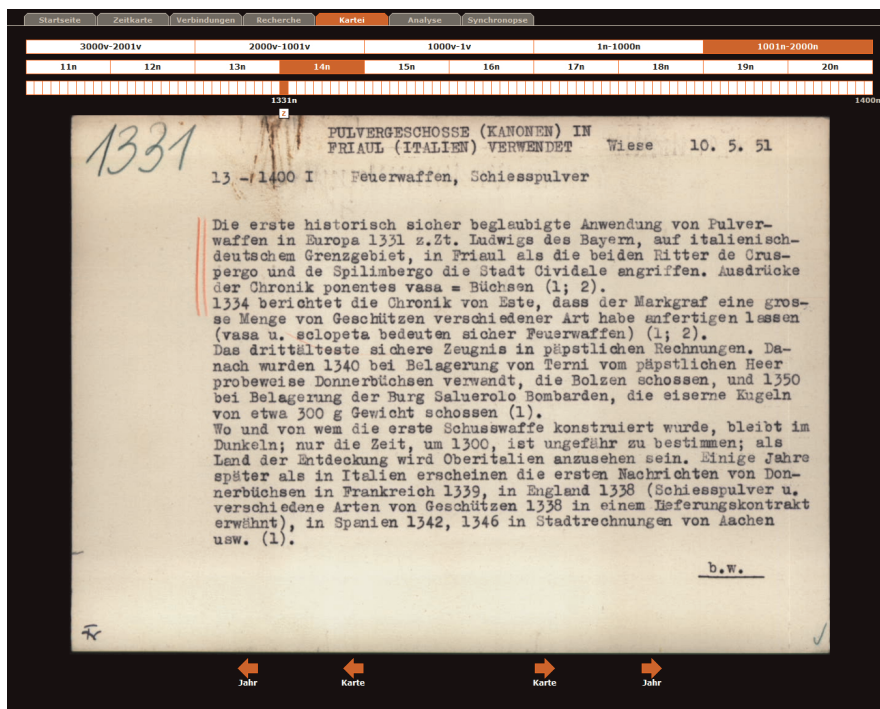


Abb. 2: Screenshot einer Karteikarte mit der Überschrift „Pulvergeschosse (Kanonen) in Friaul (Italien) verwendet“ aus ddp.

⁴⁷ Roman Frigg und Stephan Hartmann, „Models in Science,“ in *The Stanford Encyclopedia of Philosophy*, hg. v. Edward N. Zalta, Spring 2020 (Metaphysics Research Lab, Stanford University, 2020).

Die Daten auf den Karteikarten sind zwar den Visualisierungen der Zeittafeln vorangestellt, aber sie sind bereits manipuliert und verändert worden; Fakten wurden ausgewählt und nach den bereits vorhandenen Kategorien Wirtschaft, Geistesleben, Politik, Krieg und Revolution sortiert. Die Veröffentlichung der Karten im Rahmen des „Digitalen Peters“ legt zumindest die der Visualisierung zugrundeliegenden Informationen offen und erfüllt die Forderung einiger Forscher:innen, bei Visualisierungen nicht nur das transparent zu machen, was man sehen kann, sondern auch „what is omitted and the levels of uncertainty of that which is present.“⁴⁸

Was die Zeitkarte in Abbildung 1 visualisiert, ist eher ein Zeitverlauf als ein Konzept historischen Wandels. Es gibt kein Konzept des Dazwischen, des Übergangs oder der Transformation. Der dominante Darstellungsmodus ist der der linearen Zeit – ein Modell, das Praktiken der Synchronisation widerspiegelt und selbst eine lange Geschichte hat. Wie die Historiker des achtzehnten Jahrhunderts fallen wir immer noch unserem Bedürfnis nach einer stabilen Ordnung der Ereignisse zum Opfer. Die meisten Visualisierungen folgen diesem Modell, es sind Datenmodelle, die die Daten geordnet darstellen, indem sie zum Beispiel eine regelmäßige Kurve durch eine Menge von Datenpunkten zeichnen.⁴⁹ Dies entspricht dem vermeintlich affirmativen Charakter von wissenschaftlichen Visualisierungen im Gegensatz zum offeneren und selbstreflexiven Charakter künstlerischer Bilder. Aber auch Visualisierungen können explorativ und kreativ sein. Abbildung 3 zeigt eine andere Art der Visualisierung, die vom „Digitalen Peters“ angeboten wird. Unter dem Stichwort Verbindungen werden die Daten als Zeitscheibe visualisiert. Die Herausgeber der digitalen Version beschreiben die Scheibe als ein Modell eines ‚früheren Jetzt‘ mit der ‚gegenwärtigen Vergangenheit‘ des Ereignisses im Zentrum – in diesem Fall der erste dokumentierte Einsatz von Kanonen in Friaul im Jahr 1331. Andere Ereignisse oder Entwicklungen sowie historische Persönlichkeiten sind auf konzentrischen Kreisen um das ausgewählte Ereignis herum angeordnet, wobei die Größe des Punktes (genannt Zeitplanet) die Relevanz oder Stärke der Verbindung repräsentiert, während der Abstand zum Zentrum die zeitliche Distanz anzeigt. Im Gegensatz zum Zeitdiagramm gibt es keine Richtung in der Zeit. Eine große Entfernung kann gleichbedeutend sein mit einer langen Zeit in der Vergangenheit oder in der Zukunft. Die Anordnung der Zeit in konzentrischen Kreisen erinnert wieder an die Zeitschichten von Koselleck. Die Farben der Keile sind die gleichen Farben wie im Zeitdiagramm und symbolisieren die Kategorien.

⁴⁸ Jessop, „Digital Visualization as a Scholarly Activity,“ 287.

⁴⁹ Frigg und Hartmann, „Models in Science.“

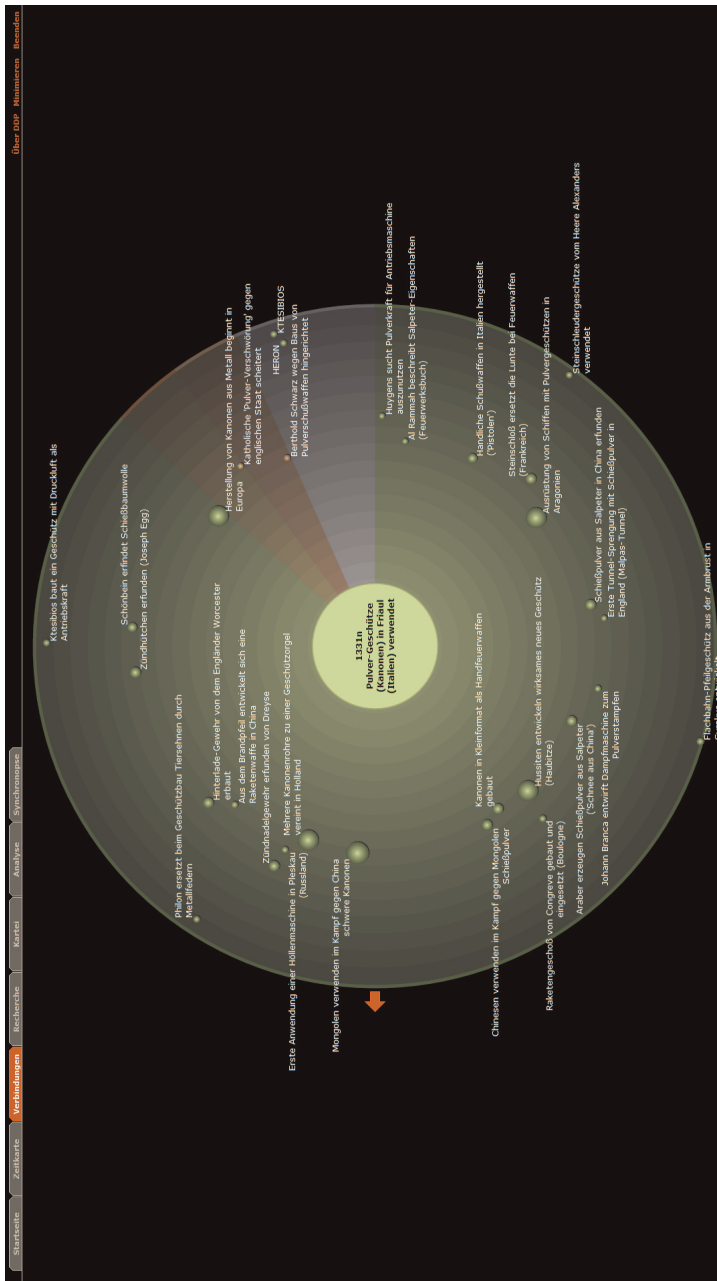


Abb. 3: Screenshot der Visualisierung von „Verbindungen“ aus ddp, die verschiedene Entwicklungen, Ereignisse und Persönlichkeiten im Zusammenhang mit „Schießpulvereinsatz“ darstellt.

Offensichtlich bestehen die meisten Verbindungen zwischen Einträgen innerhalb der Kategorie „Wirtschaft“, die in Grün dargestellt ist.

Die Visualisierung der Zeitscheibe erlaubt es der oder dem Betrachter:in, in der interaktiv gestalteten digitalen Version seine Perspektive auf zeitliche Zusammenhänge zu ändern, indem er ein anderes Ereignis (einen anderen Punkt) für die Position in der Mitte des Kreises wählt, um damit die Datensätze und ihre Verbindungen zu explorieren. Auf diese Weise verändert sich das Bild und damit auch die Perspektive auf die Zeit. Es ist nicht mehr das Ereignis von 1331, das die historische Gegenwart repräsentiert. In Abbildung 4 hat sich die ‚Gegenwart‘ auf das Jahr 1420 verlagert, in dem die Hussiten ein neues Geschütz, die Haubitze, entwickelten. Einige der umgebenden Zeitplaneten sind gleichgeblieben, andere haben sich verändert. Das Festlegen eines bestimmten Ereignisses als Gegenwart verändert die historischen Schichten von Vergangenheit und Zukunft.

Und mit diesem Perspektivwechsel kommen zugleich mehr politische Entwicklungen (orange dargestellt) in den Blick. Im selben Jahr, 1420, ruft Papst Martin V. zum Kreuzzug gegen die Hussiten auf. Mit diesem im Zentrum der Scheibe (Abbildung 5) gewinnt der oder die Betrachter:in eine weitere Perspektive auf die zeitlichen Zusammenhänge des Jahres 1420. Nun dominieren die politischen Entwicklungen (orange dargestellt) den Blick, während die wirtschaftlichen und technischen Entwicklungen auf einen Nebenschauplatz reduziert werden. Die Namen von Zeitgenossen, militärische Auseinandersetzungen (wie der Kreuzzug gegen die Hussiten) und Ereignisse aus dem Bereich des Geisteslebens prägen nun das Bild. Vergleicht man die verschiedenen Zeitscheiben, verändert sich die Geschichte vor unseren Augen.

Die vorgestellten Visualisierungen aus dem „Digitalen Peters“ funktionieren auf zwei Arten: Sie visualisieren Geschichte entweder als lineares, synchronisiertes, gerichtetes Vergehen der Zeit oder als vielschichtiges Modell, das Zeit als ungerichtet darstellt, als etwas, das Verbindungen zwischen singulären Ereignissen herstellt, ohne sie chronologisch ausrichten zu müssen – dies bleibt der oder dem Betrachter:in und der eigenen Wahrnehmung und Interpretation überlassen. Letzteres, die Zeitscheibe, könnte auch als Visualisierung einer *Chronoferenz* dienen, wie sie von Achim Landwehr beschrieben wurde.⁵⁰ Ereignisse werden auf ihre eigenen (historischen) Versionen der Vergangenheit, der Gegenwart und der Zukunft bezogen.

⁵⁰ Vgl. Achim Landwehr, *Die anwesende Abwesenheit der Vergangenheit: Essay zur Geschichtstheorie* (Frankfurt a. M.: S. Fischer, 2016).

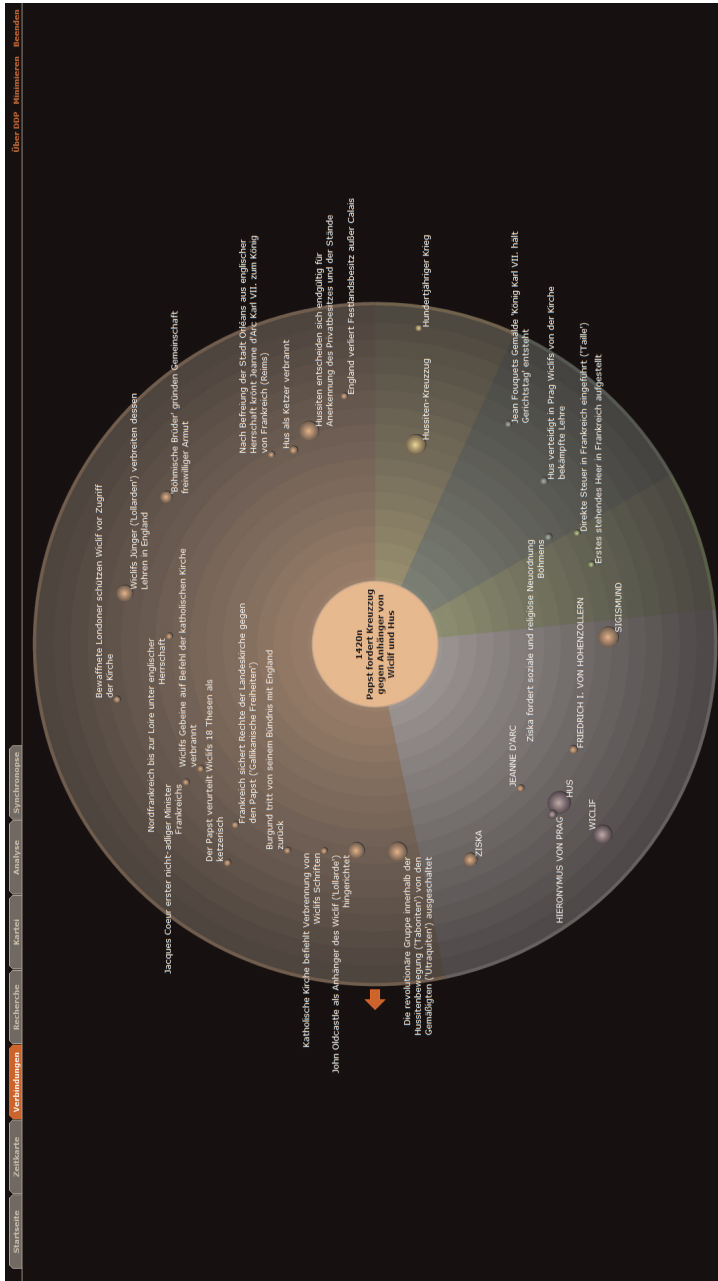


Abb. 5: Screenshot einer Zeitscheibe aus ddp mit dem Aufruf zum Kreuzzug gegen die Hussiten im Jahr 1420 in der Mitte.

5 Digitale Wissenschaftspraktiken: Modellieren, Visualisieren und Vergleichen

Konzepte von Zeit und Wandel sind nie unschuldig. Sie repräsentieren immer die Ansichten und Überzeugungen ihrer Autor:innen und ihrer eigenen Zeit, sie werden von einem bestimmten politischen, kulturellen oder sozialen Standpunkt aus entworfen; diese Perspektive gilt auch im digitalen Zeitalter und für die neuen Möglichkeiten der Visualisierung von Geschichte. Anstatt eine neue Sichtweise auf Zeit und Wandel selbst zu präsentieren, hat sich dieser Artikel vor allem mit den Praktiken beschäftigt, die mit der Konzeptualisierung und Visualisierung von Zeit- und Wandelmodellen verbunden sind. Er will die Bedeutung der Analyse von wissenschaftlichen Praktiken, von Visualisierungen und ihren Voraussetzungen und Ergebnissen hervorheben; es geht um die Forscher:innen selbst und ihre Modelle von Geschichte, Zeit und Wandel.

In diesem Sinne sind auch Visualisierungen nicht unschuldig. Obwohl Bilder oft selbsterklärend, objektiv und evident erscheinen mögen, sind sie Produkte von Praktiken. Die dargestellten Daten (in den Geisteswissenschaften sind dies meist Textdaten) sind dekontextualisiert, ihre Sequenzialität ist aufgelöst. Ebenso werden sie in der visuellen Darstellung durch weitere Dimensionen wie Distanz, Gleichzeitigkeit oder Räumlichkeit angereichert – Qualitäten, die nicht vorhanden sind, wenn wir die Textdaten selbst betrachten. Die Visualisierung von Daten ermöglicht es uns, abstrakte Phänomene oder Zusammenhänge zu sehen, die sich aus der Betrachtung einzelner Ereignisse oder Datenpunkte nicht hätten erschließen lassen.⁵¹ Sie bietet Möglichkeiten der wissenschaftlichen Exploration. Die digitale Wende hat zahlreiche Diskussionen hervorgebracht über künstliche Intelligenz und die Art und Weise, wie Algorithmen unsere Denkweise verändern oder sogar überflüssig machen könnten. In diesem Artikel wird argumentiert, dass die menschlichen Fähigkeiten der Interpretation und des „Sinnmachens dessen, was wir sehen“ umso notwendiger sind, wenn wir mit computergenerierten Daten und Visualisierungen interagieren.⁵² Nur das Zusammenspiel zwischen menschlichen Fähigkeiten der Interpretation und der digitalen Mustererkennung produziert Wissen und bietet neue Perspektiven. Das vergleichende Betrachten ist das Herzstück dieses Unternehmens. Für die Geschichte hat dies einen zentralen Vorteil: Das vergleichende Betrachten von

51 Noah Bubenhofer, „Visual Linguistics: Plädoyer für ein neues Forschungsfeld,“ in *Visualisierung sprachlicher Daten: Visual Linguistics – Praxis – Tools*, hg. v. Noah Bubenhofer und Marc Kupietz (Heidelberg: Heidelberg University Publishing, 2018), 54–56.

52 Heintz und Huber, „Der verführerische Blick,“ 21.

Gemeinsamkeiten und Unterschieden in den Daten provoziert neue Erzählungen des Wandels, die die Notwendigkeit von Chronologien überwinden und die Multidimensionalität der Zeit betonen.

Interaktive Praktiken der Visualisierung erlauben es uns, mit verschiedenen Ansichten von Zeit und Geschichte zu spielen, wobei wir uns immer wieder auf das beziehen, was wir sehen und was wir wissen. William L. Benzon hat diese Praxis als die Aktivierung eines kognitiven Netzwerks beschrieben. Das Lesen einer Visualisierung aktiviere verschiedene Teile eines Teilnetzes von Konzepten und Wissen, die mit den Elementen des Bildes verbunden sind.⁵³ Karin Knorr Cetina hat einen anderen Weg vorgeschlagen, wie Praktiken der Visualisierung Praktiken der Repräsentation und Generierung von Information beeinflussen. Sie entwickelt die Idee des „Viskurses“ in Analogie zum Diskurs, womit die Präsentation von visuellem Material anstelle von verbalen Diskussionen gemeint ist. Die Visualisierungen laufen durch viele Hände; sie werden zur Kenntnis genommen, untersucht und dienen als Anregung für weitere Experimente und als Quelle für Imagination in den Naturwissenschaften wie der Physik.⁵⁴ In ihrer Analyse werden „Viskurse“ schnell den Diskursen überlegen, da letztere bloßes Gerede seien, während Visualisierungen Ergebnisse zeigten.⁵⁵ Obwohl dieses letzte Beispiel wiederum die klassische Verwendung von Visualisierungen in den Wissenschaften demonstriert, betont es die Analyse von Modellierungspraktiken als zentral für die Entschlüsselung der Arbeit von Forscher:innen. In den Geisteswissenschaften spielen Praktiken der Visualisierung und der vergleichenden Betrachtung zum Beispiel bei historischen Analysen erst ansatzweise eine Rolle. Aber es ist sicherlich vielversprechend, visuelle, interaktive Exploration in unsere Forschungspraktiken zu implementieren. Das gilt nicht zuletzt auch, um das eigene Arbeiten besser verstehen zu lernen. Das Bemühen um die Modelle macht die eigenen Wissenschaftspraktiken bewusst und ermöglicht, die Vorannahmen, Einschränkungen und Chancen der historischen Methoden sowie unsere Vorstellungen von Geschichte zu reflektieren.

53 William L. Benzon, „Toward a Computational Historicism: From Literary Networks to the Autonomous Aesthetic,“ *SSRN Electronic Journal*, 2014, doi: 10.2139/ssrn.2432653, 13.

54 Karin K. Cetina, „„Viskurse“ in der Physik: Konsensbildung und visuelle Darstellung,“ in *Mit dem Auge denken: Strategien der Sichtbarmachung in wissenschaftlichen und virtuellen Welten. Theorie – Gestaltung 1*, hg. v. Bettina Heintz und Arnold Benz (Zürich: Ed. Voldemeer, 2001), 308.

55 Cetina, „„Viskurse“ in der Physik,“ 309.

Bibliographie

- Benzon, William L. „Toward a Computational Historicism: From Literary Networks to the Autonomous Aesthetic.“ *SSRN Electronic Journal*, 2014. doi: 10.2139/ssrn.2432653.
- Bubenhofer, Noah. „Drei Thesen zu Visualisierungspraktiken in den Digital Humanities.“ *Rechtsgeschichte – Legal History* 2016/24 (2016): 351–55. doi: 10.12946/rg24/351-355.
- Bubenhofer, Noah. „Visual Linguistics: Plädoyer für ein neues Forschungsfeld.“ In *Visualisierung sprachlicher Daten: Visual Linguistics – Praxis – Tools*. hg. v. Noah Bubenhofer und Marc Kupietz, 25–62. Heidelberg: Heidelberg University Publishing, 2018.
- Burrell, Jenna. „How the Machine ‚Thinks‘: Understanding Opacity in Machine Learning Algorithms.“ *Big Data & Society* 3/1 (2016): 205395171562251. doi: 10.1177/2053951715622512.
- Carr, Edward H. *Was ist Geschichte?* 6. Aufl. Stuttgart, Berlin, Köln, Mainz: Kohlhammer, 1981.
- Castells, Manuel. *The Rise of the Network Society*. 2. ed., with a new pref., [Nachdr.]. 3 vols. The Information Age: Economy, Society and Culture. Vol. 1. Chichester: Wiley-Blackwell, 2010. doi: 10.1002/9781444319514.
- Cetina, Karin Knorr. „Viskurse‘ in der Physik: Konsensbildung und visuelle Darstellung.“ In *Mit dem Auge denken: Strategien der Sichtbarmachung in wissenschaftlichen und virtuellen Welten. Theorie – Gestaltung 1*, hg. v. Bettina Heintz und Arnold Benz, 305–320. Zürich: Ed. Voldemeer, 2001.
- Cetina, Karin Knorr und Alex Preda. „The Temporalization of Financial Markets: From Network to Flow.“ *Theory, Culture & Society* 24/7–8 (2007): 116–38. doi: 10.1177/0263276407084700.
- Frigg, Roman und Stephan Hartmann. „Models in Science.“ In *The Stanford Encyclopedia of Philosophy*, hg. v. Edward N. Zalta. Spring 2020. Metaphysics Research Lab, Stanford University, 2020.
- Fukuyama, Francis. *The End of History and the Last Man: With a New Afterword*. 1. Free Press trade paperback ed., [Nachdr.]. New York, NY: Free Press, 2006.
- Grave, Johannes. „Vergleichen als Praxis: Vorüberlegungen zu einer praxistheoretisch orientierten Untersuchung von Vergleichen.“ In *Die Welt beobachten: Praktiken des Vergleichens*, hg. v. Angelika Eppe und Walter Erhart, 135–59. Frankfurt, New York: Campus Verlag, 2015.
- Hartog, François. *Regimes of Historicity: Presentism and Experiences of Time*. European Perspectives: A Series in Social Thought and Cultural Criticism. Paperback edition. New York: Columbia University Press, 2017.
- Heintz, Bettina und Arnold Benz (Hrsg.). *Mit dem Auge denken: Strategien der Sichtbarmachung in wissenschaftlichen und virtuellen Welten. Theorie – Gestaltung 1*. Zürich: Ed. Voldemeer, 2001.
- Heintz, Bettina und Jörg Huber. „Der verführerische Blick: Formen und Folgen wissenschaftlicher Visualisierungsstrategien.“ In *Mit dem Auge denken: Strategien der Sichtbarmachung in wissenschaftlichen und virtuellen Welten. Theorie – Gestaltung 1*, hg. v. Bettina Heintz und Arnold Benz, 9–40. Zürich: Ed. Voldemeer, 2001, 9–40.
- Jessop, Martyn. „Digital Visualization as a Scholarly Activity.“ *Literary and Linguistic Computing* 23/3 (2008): 281–93. doi: 10.1093/llc/fqn016.
- Jordheim, Helge. „Introduction: Multiple Times and the Work of Synchronization.“ *History and Theory* 53 (2014): 498–518.

- Kocka, Jürgen. „Geschichte als Wissenschaft.“ In *Geschichte: Studium, Wissenschaft, Beruf*, hg. v. Gunilla Budde, 12–30. Berlin: Akad.-Verl., 2010.
- Koselleck, Reinhart. „Neuzeit. Zur Semantik moderner Bewegungsbegriffe.“ In *Vergangene Zukunft: Zur Semantik geschichtlicher Zeiten*, hg. v. Reinhart Koselleck. 1. Aufl., 300–348. Theorie, 1979.
- Koselleck, Reinhart. „Einleitung.“ In *Zeitschichten: Studien zur Historik*, hg. v. Reinhart Koselleck, 9–16. Frankfurt a. M.: Suhrkamp, 2003.
- Landwehr, Achim. *Die anwesende Abwesenheit der Vergangenheit: Essay zur Geschichtstheorie*. Frankfurt a. M.: S. Fischer, 2016.
- Moretti, Franco. „Patterns and Interpretations.“ *Pamphlets of the Stanford Literary Lab* 15 (2017).
- Peters, Arno. *Synchronoptische Weltgeschichte*. Frankfurt a. M.: Universum-Verlag, 1952.
- Peters, Arno, Hans Rudolf Behrendt, Thomas Burch, Martin Weinmann und Andreas Kaiser. *Der digitale Peters: Arno Peters synchronoptische Weltgeschichte*. Programmversion DDP-2010-03. Wiener Neudorf, Wiesbaden: ÖBZ; Büro W, 2010.
- Rosa, Hartmut. *Beschleunigung: Die Veränderung der Zeitstrukturen in der Moderne*. 11. Auflage, Originalausgabe. Suhrkamp TB Wissenschaft 1760. Frankfurt a. M.: Suhrkamp, 2016.
- Rüsen, Jörn. „Historical Objectivity as a Matter of Social Values.“ In *Historians and Social Values*, hg. v. Joep Leerssen, 57–66. Amsterdam: Amsterdam Univ. Press, 2000.
- Sarnowsky, Jürgen. „Was sind und zu welchem Ende studiert man Geisteswissenschaften?“ <http://mittelalter.hypotheses.org/7972>. Zugriff am 02.08.2021./
- Schwandt, Silke. „Digitale Methoden für die historische Semantik.“ *Geschichte und Gesellschaft* 44./1 (2018): 107–34. doi: 10.13109/gege.2018.44.1.107.
- Unsworth, John. „Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?“ <https://johnunsworth.name/Kings.5-00/primitives.html>. Zugriff am 02.08.2021.
- Virilio, Paul. *Polar Inertia*. Published in association with Theory, Culture & Society. London: Sage Publications, 1999. <http://gbv.eblib.com/patron/FullRecord.aspx?p=1024103>. Zugriff am 02.08.2021.
- Zweig, Katharina. *Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können*. München: Heyne, 2019.

Visual Heuristics

Abstract: Research highly depends on third-party funds, raising the importance of preliminary results and research prospects in grant applications. At the same time, historians increasingly disseminate interim findings or data. This development contrasts an older tradition when historians mainly publicized finished works. Therefore, historians face the challenge of thoroughly making the “unfinished” plausible more than ever. Our article focuses on heuristics as an early stage of research and visualizations as suitable heuristics tools to address this challenge. Network graphs, in particular, enable historians (1) to grasp complex historical constellations and patterns better and (2) to lay those configurations open transparently. This helps historians specifically to grapple with complex temporal constellations beyond linear chronology. Unfortunately, the qualities of visualizations responsible for this potential still are underexposed. Moreover, when utilized, those qualities seldomly are being considered. For this reason, our article outlines the heuristic potential of visualizations borrowing from multimodality research.

Keywords: visualization, heuristics, visual analytics, temporality, theory

Zusammenfassung: Im Zeitalter drittmittelabhängiger Wissenschaft gewinnt die Formulierung vorläufiger Ergebnisse und Ergebniserwartungen in Forschungsanträgen an Bedeutung. Gleichzeitig steht eine wachsende Kultur der Veröffentlichung von Zwischenergebnissen oder Forschungsdaten einer geschichtswissenschaftlichen Tradition entgegen, der es hauptsächlich um die Präsentation von Endergebnissen geht. Diese Entwicklung verlangt nach einer stärkeren Plausibilisierung des noch „Unfertigen“. Unser Beitrag nimmt deshalb die Forschungsphase der Heuristik in den Fokus und präsentiert Visualisierungen als heuristische Mittel zur Plausibilisierung. Besonders Netzwerkgraphen dienen als Werkzeuge, um komplexe Zusammenhänge zu erschließen und transparent offenzulegen. Dies gilt vor allem für zeitliche Zusammenhänge, die sich einer linearen Chronologie entziehen. Die bei solchen Visualisierungen zum Tragen kommenden medialen Qualitäten sind bisher jedoch unterreflektiert und werden eher unbeschwert angewendet. Unser Beitrag will daher anknüpfend an die Multimodalitätsforschung das heuristische Potenzial von Visualisierungen kennzeichnen.

Schlagwörter: Visualisierung, Heuristik, Visual Analytics, Zeitlichkeit, Theorie

Im folgenden Text wollen wir für eine stärkere Rationalisierung und Modernisierung der Heuristik in der Geschichtswissenschaft plädieren. Als Heuristik wird jene Phase des wissenschaftlichen Forschungsprozesses angesehen, in dem mit verhältnismäßig geringem Arbeitsaufwand ein bereits hoch wahrscheinliches Ergebnis des Vorhabens als Hypothese formuliert wird. Dass diese Phase zunehmend an Bedeutung gewinnt, liegt an einigen grundlegenden Entwicklungen im Wissenschaftssystem, in denen die Drittmittelabhängigkeit nach überzeugenden Ergebnisprospektionen verlangt. Doch auch epistemologische Entwicklungen spielen eine Rolle, wenn, um frühzeitig Eingang in den Forschungsdiskurs zu finden, erste Resultate in Preprints als vorläufige Ergebnisse veröffentlicht werden. Im Folgenden stellen wir erste Überlegungen zu diesem Komplex vor, die in laufender Forschungsarbeit systematisch ausgearbeitet werden sollen.

Zu Beginn wird darauf einzugehen sein, wie Ergebnisse im Rahmen der Heuristik argumentativ legitimiert werden (1). Als Fazit hiervon formulieren wir, dass die Geschichtswissenschaft sich verstärkt visuellen Darstellungsformen ihrer epistemologischen Verfahrensweisen öffnen muss. Diese mediale Verschiebung resultiert aus einer sich schleichend, mehr implizit als explizit verändernden Epistemologie der historischen Wissenschaften besonders im Hinblick auf eine Post-Chronologie und dies führt zu einer Neuordnung der Argumentationslogik (2). In einem weiteren Schritt arbeiten wir heraus, wie in den Visual Analytics solch komplexe Zusammenhänge bereits mit Visualisierungen repräsentiert und analysiert werden (3). Damit ist eine Grundlage gegeben, um abschließend in einer theoretischen Argumentation die genauen Rahmenanforderungen an eine visuelle Heuristik speziell für die Geschichtswissenschaft in einem ersten Aufschlag zu entwickeln (4).

1 Warum ist Heuristik zentral für den Wissenschaftsbetrieb im frühen 21. Jahrhundert?

Unsere Überlegungen gehen davon aus, dass wir in der Anfangsphase unserer jeweiligen akademischen Entwicklung, die eine Generation auseinanderliegt, die Korrelation von Öffentlichkeit und dem Verfassen einer Promotion unterschiedlich erlebt haben: Der Promotionsgutachter von SH riet ihm nachdrück-

lich, niemandem vom Gegenstand seiner Doktorarbeit und von seinen Fortschritten zu erzählen. Nicht zumindest, bevor diese fertig ausgearbeitet und veröffentlicht seien. CW hat während seiner Promotionsphase mehrfach an verschiedenen Orten seine laufenden Forschungen vorgestellt, er hat wissenschaftliche Kontakte im Kontext des Themas im In- und Ausland bis in die Vereinigten Staaten geknüpft. Er konnte dies selbst- und eigenständig tun – und wurde von seinem betreuenden Erstgutachter, SH, darin bestärkt und unterstützt.

Nun könnte man einfach sagen, dass das Wissenschaftssystem sich gewandelt hat. Es gewährt jungen Wissenschaftler:innen mehr Raum für eigene akademische Aktivitäten. Und ohnehin war nicht überall Anfang der 1990er Jahre die Situation für Doktorand:innen und junge Doktor:innen so, wie für SH, der eben bei einem besonders altmodischen Vertreter einer klassischen Ordinarien-universitätsauffassung als wissenschaftlicher Assistent arbeitete.

Man kann aber auch, und das ist die klügere Variante, anstatt alles als Ansammlung von Fehlern und Vorurteilen einer längst verstorbenen Generation anzusehen, nach den wissenschaftskulturell relevanten Veränderungen fragen, die sich in der Differenz der beiden Erfahrungen widerspiegeln. Man könnte dies so lesen, dass in einem (mittlerweile veralteten) System nur das fertige Ergebnis relevant ist für die Scientific Community insgesamt und für das Fortkommen in der eigenen Karriere. In der aktuellen Verfassung aber hat das System Interesse an Arbeitsberichten und am Austausch über Zwischenergebnisse und erwartete Resultate. Nach dieser Lesart interessierte sich zumindest im wissenschaftsöffentlichen Raum niemand für „noch nicht ganz Fertiges“. Von Interesse war das „Werk“ und das war eine abgeschlossen erforschte und verfasste Einheit. Noch vor der Jahrtausendwende wandelte sich dies grundlegend.

Eine weitere Beobachtung stützt dies. Um uns Professor:innen an unserer Universität dem neuen Präsidium vorzustellen, gab es den Vorschlag, wir sollten Listen der erfolgreichen Drittmiteleinwerbungen anlegen: Hiernach sollten wir uns selbst nicht mehr über Ergebnisse präsentieren, sondern darüber, dass man relativ zu Beginn einer Forschungstätigkeit Menschen mit Verfügungsmacht über Finanzen hat überzeugen können, Geld für die Durchführung des Forschungsvorhabens zu investieren. Die individuelle Forschungsleistung wird zunehmend über die erfolgreiche Drittmiteleinwerbung bewertet und Forschende werden über dieses Kriterium gerankt.

Beide Geschichten haben etwas gemeinsam. Im Prozess des Forschens verliert die Phase der Veröffentlichung von Endergebnissen ihre zentrale Bedeutung. Die Phase der Formulierung von erwarteten Ergebnissen wird zunehmend wichtiger. Diese Phase, in der mit einem geringen eingesetzten Zeitbudget und mit begrenztem Wissen aufgrund der meist nur rudimentär erfolgten Quellen-

sichtung und Literaturrecherche erwartete Ergebnisse formuliert werden, die zumindest als hoch wahrscheinlich angesehen werden können, nennen wir Heuristik. Wir meinen damit etwas mehr und anderes als Droysen, der unter Heuristik noch im Wesentlichen das Auffinden des passenden Quellenkorpus verstand: „Die Heuristik schafft uns die Materialien zur historischen Arbeit herbei; sie ist die Bergmannskunst, zu finden und ans Licht zu holen.“¹ Dies geschah noch vor dem Hintergrund, dass man im Historismus glaubte, aus den Quellen, wären sie nur kritisch auf ihre Passung zur Vergangenheit geprüft, würde die wahre Geschichte „herausprudeln“, die dann dort durch Interpretationen zu ergänzen war, wo Bewertungen vorzunehmen oder Überlieferungslücken zu füllen waren. Heute sehen wir in Quellen eher Spuren, die nicht nur quellenkritisch geprüft, sondern auch mit theoretischem Instrumentarium analysiert werden müssen.

Die gesteigerte Bedeutung der Heuristik im Wissenschaftsbetrieb bedeutet, dass hier mehr symbolisches Kapital liegt, als dies früher der Fall war. Dazu passt auch die Beobachtung, dass unsere Formate immer kürzer und schneller kommunizierbar werden. Wer bloggt, verzichtet in der Regel auf einen aufwendigen Nachweisapparat, da das Blogformat grundsätzlich auf knappe Darlegungen und Einblicke abzielt. Gleichzeitig soll eine direkt anschließende überregionale oder gar globale Diskussion angeregt werden. So kann aus der allgemeinen Aufwertung heuristischer Prozesse mehr symbolisches Kapital geschlagen werden. Themen und Forschungsfragen können auf diese Weise auch frühzeitig „besetzt“ werden, um im Spiel der Aufmerksamkeitsökonomie wahrgenommen zu werden.

Wenn nun die frühzeitige Einwerbung von Finanzmitteln und damit die Anerkennung von Heuristiken zu Forschungsvorhaben zunehmend wichtiger wird im Wissenschaftssystem, so bedarf es größerer methodologisch gesicherter Nachvollziehbarkeit, sonst verbleiben wir im Kontext klassischer Begriffe wie „Erkenntnisinteresse“ und „Deutungshorizont“. Diese Heuristik muss die aktuelle Vorstellung, wie Wissen verfasst sein sollte, widerspiegeln. Hier setzt unsere zweite Voraussetzung an: die (digitale) Visualität als Medium der Heuristik.

¹ Johann Gustav Droysen: *Historik: Historisch-kritische Ausgabe Bd. 1*, hg. v. Peter Leyh (Stuttgart: Frommann-Holzboog, 1977), 400.

2 Was hat eine Heuristik der Zukunft mit post-chronologischen Netzwerken zu tun?

Es mag im ersten Moment seltsam erscheinen, dass angesichts unseres Zieles, Geschichte zu schreiben, wir nun das Thema Visualität aktivieren. Wenn wir Ergebniserwartungen in der Phase der Heuristik auf der Basis begrenzter Informationen mit hoher Transparenz und Überzeugungskraft formulieren wollen, müssen wir fragen, wie die Geschichte am Ende geschrieben werden soll – was bedeutet zu fragen, wie die einzelnen Elemente zusammengebunden werden. Und damit ist zu fragen, wie Wissen (im 21. Jahrhundert zeitgemäß) zu organisieren ist. Und dass die mediale Organisation von Wissen (mit-)konstitutiv ist für die Geschichtswissenschaft, wissen wir nicht nur wegen Hayden Whites narratologischen Beitrags zu Handlungsstrukturen in der Geschichtsdarstellung.²

Um dies zu klären, gehen wir einen Umweg über die klassische, sprich historistische Heuristik.

Nach dieser Auffassung des historischen Arbeitsprozesses, wie man sie bei Droysen findet³, passiert Folgendes:

1. Man hat eine historische Frage.
2. Es werden passende Quellen gesucht.
3. Diese werden mittels quellenkritischer Methode kritisch daraufhin überprüft, welche Fakten aus ihnen herausgezogen werden können.
4. Diese Fakten werden in zeitlicher Abfolge, sprich „*chronologisch*“ aneinandergereiht.
5. Das war es dann schon.

Das klingt nach wenig, aber es war lange Zeit der Kern historiographischen Selbstverständnisses, weil die chronologische Reihe einen entscheidenden Mehrwert enthält: Jedes Element, das in einer solchen Reihe steht, hat einen direkten Bezug auf das Vorgehende und das Folgende. Wir Menschen neigen dazu, eine solche Reihung als kausale Bedingung oder zumindest als Abfolge eines kohärenten Sachverhaltes zu lesen – und wenn wir das nicht können, es für dadaistischen Unsinn zu halten. „In Brasilien schlägt ein Schmetterling mit den Flügeln. In New York tobt ein Wirbelsturm.“⁴ Das halten wir für Unsinn, bis

2 Hayden V. White, *Metahistory: The Historical Imagination in Nineteenth-Century Europe* (Baltimore: Johns Hopkins Univ. Pr., 1973). Allgemein zur Narrativitätsdebatte: Stefan Haas, „Fiktionalität in den Geschichtswissenschaften,“ in *Fiktionalität: Ein interdisziplinäres Handbuch*, hg. v. Tobias Klauk und Tilmann Köppe (Berlin: de Gruyter, 2014), 516–532.

3 Droysen, *Historik*, 67 ff.

4 Edward N. Lorenz, *The Essence of Chaos* (London: UCL Pr., 1993), 181–184.

wir uns chaostheoretisch die Zwischenschritte vorstellen können. Und dann lesen wir es als Kausalität.

In der Geschichtswissenschaft tun wir das auch beziehungsweise taten wir das im Zeitalter des Historismus. Eine chronologische Reihe von Ereignissen lesen wir als eine (schwach) kausale Entwicklung. Schwach kausal, weil die Geisteswissenschaften des 19. Jahrhunderts immer mit einer gewissen Kontingenz gerechnet haben: Es ist folgerichtig, dass es so geschah, aber ein „großer Mann“ hätte auch richtungsändernd eingreifen können. Die Reihenfolge der Ereignisse und das chronologische Erzählen sind die meist implizite Kausalerklärung für das, was in der Vergangenheit geschah. Deswegen ist Quellensammeln Heuristik und erlaubt es, eine erste Hypothese zu formulieren, wenn man das Material gesammelt, kritisch geprüft und das Ergebnis der Prüfung chronologisch in Textform (!) aneinandergereiht hat.

Es gibt also eine Kongruenz zwischen dem bevorzugten Medium der Geschichtswissenschaft, dem geschriebenen Text, und der die Geschichtswissenschaft tragenden Vorstellung (zeitlicher) Kausalität – und kulminiert im Begriff der Chronologie. Es handelt sich eben um den Logos der Geschichte (doppelter Wortsinn) und deswegen wird *Geschichtsschreibung* und *Geschichtswissenschaft* so oft synonym eingesetzt.

Das alles hat natürlich viel mit Hayden Whites erzähltheoretischem Verständnis zu tun. Aber über ihn hinausgehend kann man fragen, ob das im 21. Jahrhundert so noch ausreicht. Haben wir wirklich das Bild einer Geschichte (doppelter Wortsinn) als monolineare Abfolge von Einzelelementen?⁵

Der ältere von uns beiden Autoren hat seine Assistentenzeit in der Sozial- und Wirtschaftsgeschichte verbracht und dieser Bereich war der erste, der anfang, vergleichend zu verfahren: Wenn man ein Buch schreibt über Industrialisierung in Europa (um es nicht allzu groß werden zu lassen), schreibt man dies dann als klar nummerierte Abfolge der Ereignisse entlang einer Zahlenreihe? Wohl kaum, denn die Entwicklungen in verschiedenen Ländern können ähnlich oder analog ablaufen, aber häufig zeitlich versetzt; in England mit einigen Jahrzehnten Vorlauf zu Ländern wie Belgien und Schweden. Würden wir dann in unserem Buch für jedes Land wieder neu nennen, wann die Spinnmaschine erstmals auftritt? Zumal sie nicht immer neu erfunden, sondern importiert oder gestohlen wird, es also Wechselwirkungen, Transfergewinne oder -verluste gibt. Es ist somit nicht nur eine bloße Aneinanderreihung (wie in der klassischen politischen Ereignisgeschichte), sondern ein komplexes Gewebe von sich

⁵ Konzepte von Zeitlichkeit sind bis heute immer wieder für die Geschichtsforschung und -schreibung diskutiert worden. Als jüngerer Beitrag hervorzuheben ist Achim Landwehr, *Diesseits der Geschichte: Für eine andere Historiographie* (Göttingen: Wallstein, 2020).

bedingen, sich abgrenzen, neu erfinden oder verbessern, ausspionieren und kopieren, aber doch auch modifizieren. Manchmal wird etwas vergessen und es kommt Generationen später zu einer Neuanknüpfung. Dann wird dies in eine neue Zeit geholt, ohne sich langsam durch kontinuierliche Rezeption angepasst zu haben. Diese Geschichte ist kein kontinuierlicher Fluss, sondern ein Re-entry, das die historische Linie abkürzt, ähnlich wie ein Wurmloch.

Für eine Wirtschaftsgeschichte der Industrialisierung – wie für viele andere Geschichten – ist nicht die Linie die passende Erzählstruktur, sondern das Netzwerk oder Multilinearität ein Geflecht von Bedingungen, Abhängigkeiten und Bezügen. Eine solche Netzwerkgeschichte ist nicht als klassischer Text realisierbar, sondern als Bild: weil das Bild es erlaubt, genau diese Hyperstruktur zu realisieren. Wer heute Forschung vorstrukturieren möchte, muss dies, um es zeitgemäß zu tun, in einem Medium formulieren, das Netzwerkstrukturen darstellen kann: Und das sind nicht jene, die diachron funktionieren, sondern die, die Synchronität darstellen können.⁶ Und deswegen arbeiten wir an einer *visuellen* Heuristik. Wir schließen damit an Visualisierungspraktiken an, die sich in den Digital Humanities (DH) bis heute ausdifferenziert entwickelt haben. Dies soll im nächsten Schritt deutlich werden.

3 Heuristische Verfahren der Visual Analytics als Innovationstreiber

In den DH spielen Visualisierungen eine große Rolle, um komplexe Zusammenhänge – wie etwa zeitliche Beziehungen in Forschungsdaten – zu beforschen. Sie fungieren nicht als reine Illustrationen, sondern als Verfahrensmittel: Sie ermöglichen es, informative Muster in Datensätzen zu erkennen, um sodann weitergehende Analysen anzustellen. Unter dem Label „Visual Analytics“ haben wir es mit einem Forschungsfeld zu tun, dessen Diskurs derzeit auf mehreren Ebenen stattfindet:

1. Mikroebene: Forschungsprojekte demonstrieren und besprechen den Einsatz ihrer Visualisierungen als Analysehilfsmittel.

⁶ Unsere Überlegungen speisen sich aus verschiedenen bild- und medien- sowie technikwissenschaftlichen Ansätzen, siehe: Stefan Haas, „Vom Schreiben in Bildern: Visualität, Narrativität und digitale Medien in den historischen Wissenschaften,“ *Zeitenblicke: Digitale Medien und Wissenschaftskulturen* 5/3 (2006). Zugriff am 17.05.2022, http://www.zeitenblicke.de/2006/3/Haas/index_html.

2. Mesoebene: Visualisierungsverfahren beziehungsweise -tools werden zum Gegenstand der Diskussion.
3. Metaebene: Die wissenschaftstheoretische Auseinandersetzung nimmt sich visueller (meist gleichzeitig interaktiver) Datenrepräsentation als Phänomen an. Im Fokus steht die Frage, wie und weshalb Visualisierungen das Methodenspektrum der DH bereichern.

Für den heuristischen Kontext ist besonders der dritte Bereich, die Methodenreflexion und -kritik auf der Metaebene relevant. Visualisierungen ermöglichen anders als Listen, Tabellen oder weitere, eher lineare Darstellungsweisen einen synchronen Überblick über Datenrepräsentationen und regen auf diese Weise zu einem vertiefenden Zugriff auf erkennbare Strukturen und Muster an. Dieser analytische Mehrwert hat dazu geführt, dass Visualisierungen zum Teil unter den Schlagworten „Logik des Bildlichen“ und „ikonischer Vernunft“⁷ diskutiert werden. Gemeint sind dabei einerseits diagrammatische Formen abstrakter Datenrepräsentationen,⁸ etwa als Netzwerkgraphen. Andererseits sind Repräsentationen ganzer Forschungsobjekte angesprochen, beispielsweise, wenn viele Bilddigitalisate in einem zweidimensionalen Koordinatensystem nach ihrer Farbgebung angeordnet werden, um hier Ähnlichkeiten und Unterschiede zwischen allen Bildern nachvollziehen zu können, dabei gleichzeitig die Bilder als solche präsent zu halten. Beispiele dieser Art hat Lev Manovich als „media visualization“⁹ besprochen. Im Gegensatz zu nüchternen, vermeintlich „objektiven“ Datenrepräsentationen hat vor allem Johanna Drucker für die Geisteswissenschaften solche eingefordert, welche die Gemachtheit der Visualisierung und die dabei zum Tragen kommende Perspektivität der Forschenden in den Vordergrund rücken. Derartige „humanistic visualizations“¹⁰ markieren etwa Bereiche in den Visualisierungen, für die zwar keine Daten vorliegen, für die die Existenz von Daten aber begründeterweise angenommen werden kann. Man denke nur an historische Quellen, die uns aufgrund von Archivschäden heute nicht mehr zur Verfügung stehen, von deren früherer Existenz wir aber gesichert wissen. Aus heuristischer Perspektive ist eine grobe Abbildung derartig blinder Flecken gemeinsam mit den empirisch erhobenen Daten besonders viel-

7 Martina Heßler und Dieter Mersch (Hg.), *Logik des Bildlichen: Zur Kritik der ikonischen Vernunft* (Bielefeld: transcript, 2009).

8 Sybille Krämer und Christina Ljungberg (Hg.), *Thinking with Diagrams: The Semiotic Basis of Human Cognition* (Boston: de Gruyter Mouton, 2016).

9 Lev Manovich, *Cultural Analytics* (Cambridge, MA: The MIT Press, 2020), 203–206, 223–244.

10 Johanna Drucker, *Graphesis: Visual Forms of Knowledge Production* (Cambridge, MA: Harvard Univ. Pr., 2014), 135–192.

versprechend. Schließlich können damit sowohl erste Beobachtungen und Forschungshypothesen angestellt werden, als auch transparent gemacht werden, auf welcher Grundlage diese heuristischen Schritte erfolgen und wie umgrenzt diese Grundlage ist.

Diese Auflistung heuristischer Praktiken in den Visual Analytics ist selbstverständlich nicht erschöpfend. Das bisher Dargestellte verdeutlicht jedoch bereits, wie Visualisierungen als „tools of scholarly inquiry and narrative“¹¹ wirken können, wie es David J. Staley fasst.

Wir folgen seinem Grundverständnis, wenn wir Visualisierungen als räumliche Organisation bedeutungstragender Informationen verstehen, um systematische Untersuchungen voranzutreiben.¹² Dabei liefert Staley neben wenigen anderen Beiträgen zum Einsatz visueller Medien in der historischen Forschung erste Ansätze.¹³ Eine tiefer gehende theoretisch-methodologische Fundierung heuristischer Einsatzmöglichkeiten für Visualisierungen fehlt allerdings bislang für unsere Disziplin. Zwar führen zahlreiche Forschungsprojekte Analysen mithilfe von Visualisierungen durch und deren instrumenteller Wert wird zumeist durchaus anerkannt. Allerdings erfolgt der Gebrauch größtenteils eher „unbeschwert“.¹⁴ Eine reflektierte Auseinandersetzung unter fachspezifischer Einschärfung erscheint vor allem deswegen wesentlich, weil die instrumentelle Funktion von Visualisierungen als Analysehilfsmittel nur dann gezielt ausgespielt werden kann, wenn sie den geschichtswissenschaftlichen Erkenntnisinteressen gerecht werden. So manche, aber nicht jede Visualisierungsmethode ist für die Geschichtsforschung im gleichen Maße instruktiv wie für andere digitale Geisteswissenschaften.

Welche Gestaltungskriterien müssen Visualisierungen also erfüllen, um die beschriebenen heuristischen Effekte theoretisch untermauert und reflektiert im Forschungsalltag anwenden zu können?

Gerade die letzte Frage richtet sich gegen die erwähnte „unbeschwerte“ Verwendung von Visualisierungen, indem es zunächst um die medialen Eigenschaften und Potenziale visueller Repräsentation geht, woraus neue Fragen erwachsen: Was bedeutet die synchrone Abbildung bedeutungstragender Ele-

¹¹ So David J. Staley, *Computers, Visualization, and History: How New Technology Will Transform our Understanding of the Past* (Armonk: Sharpe, 2014), xiii.

¹² Staley, *Computers*, 36.

¹³ Haas, *Schreiben in Bildern*.

¹⁴ So hat es Noah Bubenhofer analog für den Gebrauch von Visualisierungen in der Linguistik formuliert. Noah Bubenhofer, „Visual Linguistics: Plädoyer für ein neues Forschungsfeld,“ in *Visualisierung sprachlicher Daten: Visual Linguistics – Praxis – Tools*, hg. v. Noah Bubenhofer und Marc Kupietz (Heidelberg: Heidelberg Univ. Publ., 2018), 25–26.

mente (zum Beispiel Knoten und Kanten in einem Netzwerk) für die Wahrnehmung von Betrachter:innen? Wie funktioniert eine „Grammatik des Bildlichen“¹⁵, die anderen Regeln folgt als schriftsprachliche Grammatik? Hier wird in einem ersten Schritt nach dem Möglichkeitsraum gefragt, den das Visualisieren medial anbietet. Im zweiten Schritt knüpft die Frage an, wie man sich dieser Möglichkeiten für ein Forschungsinteresse gezielt bedienen kann. Mit anderen Worten liefert eine medientheoretische Fundierung ein klareres Bild von Visualisierungspotenzialen, womit reflektiertere heuristische Verwendungsweisen möglich werden.

Im Folgenden soll ein erstes Angebot in diese Richtung gemacht werden. Dabei werden allen voran Visualisierungen von Netzwerken als vielversprechend für die Visual Heuristics in der Geschichtswissenschaft vorgestellt.

4 Netzwerkgraphen als heuristische Werkzeuge der Geschichtswissenschaft

Wenn Visualisierungen Muster im Datenbestand erkennen lassen, dann stehen diese für Korrelationen zwischen Objekten. Diese visuell repräsentierten Korrelationen lassen sich in einzelne Aussagen übersetzen und damit sprachlich analytisch wiedergeben. Mit diesem Verständnis gehen wir also von keiner scharfen Trennung zwischen bildlicher und sprachlicher Modalität aus. Vielmehr folgen wir einem multimodalen Ansatz, der bildlinguistisch informiert ist. Hiernach sind bildliche Darstellungen als Propositionskomplexe aufzufassen.¹⁶

Solch ein Netzwerk an Aussagen wird durch vorausgegangene Datenerhebungen gewonnen, ehe es visuell repräsentiert wird. Dabei hängen die angewandten Erhebungsparameter stets von menschlichen Entscheidungen ab; die Propositionskomplexe erwachsen also einer Perspektivität und implizieren Konzepte, die in sie eingeschrieben worden sind. Gleiches gilt für die Entscheidung, in welcher genauen Form sie visuell repräsentiert werden sollen,¹⁷ beziehungsweise wie die Visualisierung modelliert werden soll.¹⁸

¹⁵ Gunther Kress und Theo van Leeuwen, *Reading Images: The Grammar of Visual Design* (London: Routledge, 2021).

¹⁶ Vgl. Franziska Große, *Bild-Linguistik: Grundbegriffe und Methoden der linguistischen Bildanalyse in Text- und Diskursumgebungen* (Frankfurt a. M.: Lang, 2011), 118–22.

¹⁷ Vgl. Drucker, *Graphesis*.

¹⁸ Zur Modellierung von Daten und Visualisierungen siehe: Zoe Schubert, Jan Wieners und Øyvind Eide, „Visuelle modellbasierte Experimente in den (digitalen) Geisteswissenschaften,“

Wenn man nun die durch die Visualisierung repräsentierten einzelnen Aussagen zusammen- beziehungsweise in Beziehungen zueinander setzt, können narrative Strukturen aufgespürt werden: Nehmen wir etwa an, Objekt A ist immer dann zu beobachten, wenn die Objekte B und C auftreten, ansonsten nicht. Stehen diese Objekte für bestimmte Ereignisse, können wir nicht nur ein zeitliches Nacheinander ablesen, sondern hierbei ein Bedingungsverhältnis vermuten und überprüfen, wie in den vorausgegangenen Abschnitten bereits angesprochen. Dies gilt auch für weitere Konstellationen im Datenbestand, wie Häufungen, Asymmetrien, komplett isolierte Objekte oder Ähnliches. Etwas verkürzend ausgedrückt repräsentiert die Visualisierung in solchen Fällen „implizite Erzählungen“. Allerdings sind sie aufgrund der bildlichen Modalität „nur vage und unterdeterminiert“¹⁹ wiedergegeben; der genaue Inhalt der Zusammenhänge ist nicht sogleich erkennbar. Er muss analytisch herausgelesen werden und dieses Herauslesen kann als eine heuristische Praxis verstanden werden.

Netzwerkgraphen korrespondieren nun besonders deutlich mit diesem Phänomen, weil die einzelnen Objekte als Knoten individuell repräsentiert werden und die Kanten die semantischen Verbindungen zwischen ihnen strukturell eindeutig wiedergeben. Man kann hier von einer ikonischen Abbildung dieser propositionalen Strukturen sprechen.²⁰ Aus einer heuristischen Sicht ist dies besonders wertvoll, geht es für die Formulierung von Forschungshypothesen doch um das Erkennen bedeutungsvoller Zusammenhänge, die es im Anschluss sprachlich wohlformuliert auszubuchstabieren gilt. Wenn man es auf eine verkürzte Formel bringen will, dann spiegelt ein Netzwerkgraph die Kohärenzstruktur inhaltlicher Bezüge in besonders direkter Weise wider.

Dieses Verständnis steht im Einklang mit Gunther Kress' und Theo van Leeuwens Konzept der „analytical structures“, die beide Pioniere der Multimodalitätsforschung in ihrem Standardwerk „Reading Images“ erläutern: „Analytical structures“ sind visuell repräsentierte Strukturen, in denen Bildelemente in Be-

in *Fabrikation von Erkenntnis – Experimente in den Digital Humanities*, hg. v. Manuel Burghardt et al. (Luxembourg: Melusina Press, 2021–2022. doi: 10.17175/sb005_007.

19 Hartmut Stöckl, „Sprache-Bild-Texte lesen: Bausteine zur Methodik einer Grundkompetenz,“ in *Bildlinguistik: Theorien – Methoden – Fallbeispiele*, hg. v. Hans-Joachim Diekmannshenke, Michael Klemm und Hartmut Stöckl (Berlin: Erich Schmidt Verlag, 2011), 45–70.

20 Vgl. Linda Freyberg, „Iconicity as Simultaneous Plurality – Beyond the Restraints of Formal Knowledge Organization Systems,“ in *Charles S. Peirce: Bridging the Disciplinary Boundaries of Natural Sciences and Humanities*, hg. v. Elize Bisanz (Berlin: Peter Lang, 2019), 123–150 sowie Dies. *Ikonizität der Information: Die Erkenntnisfunktion struktureller und gestalteter Bildlichkeit in der digitalen Wissensorganisation* (Berlin: Humboldt Universität zu Berlin, 2021).

ziehungen zu etwas Größerem dargestellt werden („meronymical relations“).²¹ Kress und van Leeuwen sehen in einem visualisierten Netzwerk konkreter eine „connected analytical structure“²². Die Beziehungen von Elementen zum „großen Ganzen“ und die Verbindung der Elemente untereinander sind hier also das Wesentliche.

Wertet man einen Netzwerkgraphen heuristisch aus, haben die Ergebnisse zwar immer nur einen ersten Annäherungswert; allerdings können auf dieser Basis deutlich transparenter und damit rationaler als mit klassischen hermeneutischen Daten- und Quelleninterpretationen oder impliziten Vergleichsverfahren Forschungshypothesen generiert werden. Schließlich wird durch die überblicksartige und interaktiv ansteuerbare Visualisierung nicht nur das Erkennen interessanter Zusammenhänge erleichtert. Auch der Verweis auf diese Zusammenhänge wird nachvollziehbarer. Bei Forschungsfragen mit umfangreichem Datenmaterial werden Forschungshypothesen und transparente Offenlegung überhaupt erst mithilfe solcher Visualisierungen möglich. Als ein passendes Format bieten sich hier etwa Data Stories an, in denen Daten vorgestellt, kommentiert, erste Interpretationen angeboten und Visualisierungen eingebunden werden können.²³

Innerhalb der visualisierten impliziten Erzählungen interessieren Historiker:innen je nach Forschungsfrage verschiedene Zusammenhänge, die nicht für alle weiteren Geisteswissenschaften gleichermaßen von Bedeutung sind. Hier wären beispielsweise Netzwerke von historischen Akteur:innen, Diskursen, politischen Entscheidungen und anderen Untersuchungsobjekten zu nennen. Auch Argumentationsstrukturen in beforschten historischen Diskursen werden interessant, in denen Einzelaussagen zueinander in (kausale, kontingente, begründende, negierende usw.) Beziehungen gesetzt werden. Kommen wir auf die postchronologischen komplexen Zeitrelationen zurück, wirkt das heuristische Potenzial von Netzwerkgraphen besonders vielversprechend: Transferleistungen in transnationalen Bezügen, Wechselwirkungen unterschiedlicher Wirklichkeitssegmente wie Gesellschaft, Politik, Kultur, Technik usw. Sie können derart komplex ausfallen, dass für ihre Untersuchung monolineare Abläufe kaum tragen. Netzwerkartige Visualisierungen stehen dem mit ihrer Fähigkeit,

²¹ Kress und van Leeuwen, *Reading Images*, 76.

²² Ibid., 91.

²³ Siehe etwa Anna Feigenbaum und Aria Alamalhodaie, *The Data Storytelling Workbook* (London, New York: Routledge, 2020). Ein Beispiel kann gegeben werden mit der Data Story von Fabio Fromme et al., *Sammeln, Beschreiben, Kategorisieren – Vom Umgang mit menschlichen Überresten zu Beginn der modernen Wissenschaft* (Bielefeld: SFB 1288, Practices of Comparing, 2020).

räumlich komplexer und überblicksartig abbilden zu können, entgegen. Dies gilt allein schon durch die Abbildbarkeit von gleichzeitigen, eben nicht sukzessiven Wechselwirkungen, komplexen Abhängigkeiten bis zum Re-entry.

5 Resümee: Visuelle Heuristik für die Geschichtswissenschaft im 21. Jahrhundert

Im heutigen Wissenschaftsbetrieb ist eine gewachsene Bedeutung für die Kommunikation vorläufiger Forschungsergebnisse und der Prospektion erwartbarer Resultate zu beobachten. Im Vergleich zu vergangenen Jahrzehnten bleibt die Publikation endgültiger Forschungsergebnisse zwar wesentlich für die Teilnahme am Wissenschaftsdiskurs, doch der Präsentation von „nicht Fertigem“ kommt ein gesteigertes Gewicht zu. Dafür lassen sich verschiedene Gründe angeben, die in vielen Fällen mit der Generierung symbolischen Kapitals durch Forschende zusammenhängen: (1) Forschung beruht mehr denn je auf der Drittmittelfinanzierung. Vor diesem Hintergrund ist es zu einem maßgebenden Leistungskriterium geworden, Gutachter:innen von Forschungsvorhaben und -zielen zu überzeugen. (2) Erfolg für die eigene Karriere bemisst sich auch darin, wie schnell Themen und Ansätze mit ersten Positionierungen „besetzt“ werden. Schnelle Publikationsformen und der zeitige Austausch mit der Wissenschaftscommunity werden dafür bemüht. (3) Neben die Finanzierungs- und Karriereaspekte treten methodische und epistemische Gesichtspunkte. In zunehmend interdisziplinären Vorhaben (nicht nur) digitaler Forschung, wird der Austausch über Zwischenergebnisse, Methoden und Daten bedeutsam, um aus unterschiedlichen methodischen und disziplinären Richtungen Rückmeldungen zu erhalten sowie an den interdisziplinären Forschungsdiskurs anzuschließen.

Allen drei vorgenannten Punkten liegt die Forschungsphase der Heuristik zugrunde. In breiter wissenschaftstheoretischer Definition handelt es sich dabei um die Formulierung erwarteter beziehungsweise als hoch wahrscheinlich angesehener Resultate, nachdem mit beschränktem Zeitbudget und begrenztem Vorwissen eine erste unvollständige Quellen- und Literatursichtung erfolgt ist. Wenn der Forschungsphase der Heuristik heute aber eine gesteigerte Bedeutung zukommt, dann stellt sich die Frage, wie heuristische Aussagen rational abgesichert werden können. Wie kann das oftmals bemühte „erwartbare Ergebnis“ derart fundiert und plausibilisiert werden, dass wir über allzu allgemeine Bemerkungen über „Erkenntnisinteresse“ und „Deutungshorizont“ hinauskommen?

Für ein Antwortangebot haben wir uns auf zeitliche Zusammenhänge konzentriert, auf die sich die Geschichtswissenschaft zentral richtet und die sich nach klassischem heuristischen Verständnis vor allem in chronologischen Abfolgen erschöpfen. Diese suggerieren kausale Prozesse oder zumindest klar kohärente Entwicklungsschritte innerhalb historischer Entwicklungen. Inzwischen fragen Historiker:innen allerdings zunehmend nach der Gleichzeitigkeit von Prozessen, nach Transfer, Verflechtung oder Re-entry – nach Zusammenhängen, die nicht dem Ordnungsprinzip linear-chronologischer Abfolgen gehorchen und sich der Logik kausaler Ketten entziehen. Non-Linearität erweist sich als angemesseneres Denkmuster. Visualisierte Netzwerke dienen sich dann als heuristische Werkzeuge an, da mit ihnen postchronologische, netzwerkartige Zusammenhänge synchron und hiermit besonders eindeutig wiedergegeben werden können. So spielen Netzwerkgraphen neben anderen Visualisierungsformen in den Visual Analytics auch seit Längerem eine bedeutende Rolle, um Resultate komplexer Datenanalysen mithilfe eines interaktiven und bildlichen Interfaces auszuwerten. Der heuristische Zugriff kann hiermit besonders fruchtbar gelingen, weil die Komplexität der strukturell repräsentierten Zusammenhänge erhalten bleibt, dabei aber Überblick und Orientierung für die Identifikation interessanter Korrelationen und Muster gestiftet werden. Solche Muster können erkenntnisfördernde Beobachtungen hervorbringen, die in die spätere historiografische Erzählung eingehen. Hier kann, metaphorisch, von „impliziten“ Erzählungen innerhalb der Visualisierung gesprochen werden, die es „herauszulesen“ und zu explizieren gilt.

Während ein solches Vorgehen in den Visual Analytics an der Tagesordnung ist, erfolgt sie oftmals eher „unbeschwert“, ohne tiefergehende Reflexion der semiotischen Eigenschaften von Visualisierungen. Wir haben aus diesem Grund an interdisziplinäre Beiträge zur Multimodalitätsforschung angeschlossen, um herauszustellen, aus welchen medialen Qualitäten heraus visualisierte Netzwerke als analytische Werkzeuge fungieren können: Die bildliche Qualität von Netzwerkgraphen zeigt sich im Wesentlichen in der ikonischen Repräsentation von Zusammenhängen, genauer gesagt in der direkten Abbildung von Kohärenzstrukturen. Eine solche strukturelle Übersicht sagt zwar noch wenig über die semantischen Gehalte der Zusammenhänge aus, aber sie regt zu einer genaueren Ergründung ebendieser an. Die bildliche Darstellung symbolisiert einen Komplex an Aussagen (Propositionskomplex), welche die einzelnen Zusammenhänge sprachlich ausbuchstabieren. Eine derartige theoretische Fundierung hat den Vorteil, dass sie den Rahmen an Gestaltungsmöglichkeiten medientheoretisch absteckt, damit hiernach Forschungsvorhaben sich dessen bedienen können, um ebenjene Visualisierungsdesigns zu entwickeln, die zum Forschungsdesign passen. Im Fall von Geschichtsforschung, die den epistemi-

schen Anspruch auf Ergründung postchronologischer und weiterer non-linearer Zusammenhänge erhebt, identifizieren wir Netzwerkgraphen als einen solchen passenden Visualisierungstyp.

Im Rahmen der geschichtswissenschaftlichen Heuristik bieten sich Netzwerkgraphen jedoch nicht allein zur Analyse an. Historiker:innen können auch ihre (Hypo-)Thesen, vorläufigen Resultate und Ergebnisprospektionen stärker rationalisieren, wenn sie bei ihren Explikationen auf entsprechende Teile der visualisierten Zusammenhänge verweisen. Die derart angesprochenen Konstellationen und Muster bleiben dabei im Gesamtzusammenhang des ganzen Netzwerkes sichtbar. Endgültige Forschungsergebnisse mögen in ihrer historiografischen Form zwar weiterhin nicht netzwerkartig aufgebaut sein, sondern höchstens multilinear – auch komplexe Wissensangebote verfügen über eine zwar verwobene, aber wohlstrukturierte argumentative und narrative Form.²⁴ Doch im Stadium der Heuristik erleichtern visualisierte Netzwerke erstens die Hypothesenbildung und stellen sie zweitens transparenter dar, als dies derzeit zu meist der Fall ist.

Bibliographie

- Bubenhofer, Noah. *Visual Linguistics: Plädoyer für ein neues Forschungsfeld*. In *Visualisierung sprachlicher Daten: Visual Linguistics – Praxis – Tools*, hg. v. Noah Bubenhofer und Marc Kupietz, 25–62. Heidelberg: Heidelberg Univ. Publ., 2018.
- Droysen, Johann Gustav. *Historik: Historisch-kritische Ausgabe Bd. 1*, hg. v. Peter Leyh. Stuttgart: Frommann-Holzboog, 1977.
- Drucker, Johanna. *Graphesis: Visual Forms of Knowledge Production*. Cambridge, MA: Harvard Univ. Pr. (metaLABprojects), 2014.
- Feigenbaum, Anna und Aria Alamalhodaie. *The Data Storytelling Workbook*. London, New York: Routledge, 2020.
- Freyberg, Linda. „Iconicity as Simultaneous Plurality – Beyond the Restraints of Formal Knowledge Organization Systems“. In *Charles S. Peirce: Bridging the Disciplinary Boundaries of Natural Sciences and Humanities*, hg. v. Elize Bisanz, 123–150. Berlin: Peter Lang (Peirce Studies, 10), 2019.
- Freyberg, Linda. *Ikonizität der Information: Die Erkenntnisfunktion struktureller und gestalteter Bildlichkeit in der digitalen Wissensorganisation*. Berlin: Humboldt-Universität zu Berlin 2021.

²⁴ Vgl. Christian Wachter, *Geschichte digital schreiben: Hypertext als non-lineare Wissensrepräsentation in der Digital History* (Bielefeld: transcript, 2021) sowie Christian Wachter, „Publishing Complexity in the Digital Humanities“, *magazén: International Journal for Digital and Public Humanities* 1 (2, 2021), 103–118. doi: 10.30687/mag/2724-3923/2021/03/004.

- Fromme, Fabio, Pascal Gehlhaar, Jonas Penningroth und Hagen Salzmann. *Sammeln, Beschreiben, Kategorisieren – Vom Umgang mit menschlichen Überresten zu Beginn der modernen Wissenschaft*. Bielefeld: SFB 1288, Practices of Comparing, 2020. doi: 10.4119/unibi/2946762.
- Große, Franziska. *Bild-Linguistik: Grundbegriffe und Methoden der linguistischen Bildanalyse in Text- und Diskursumgebungen*. Frankfurt a. M.: Lang 2011.
- Haas, Stefan. „Vom Schreiben in Bildern: Visualität, Narrativität und digitale Medien in den historischen Wissenschaften.“ *Zeitenblicke: Digitale Medien und Wissenschaftskulturen* 5/3 (2006). Zugriff am 17.05.2022, http://www.zeitenblicke.de/2006/3/Haas/index_html.
- Haas, Stefan. „Fiktionalität in den Geschichtswissenschaften“. In *Fiktionalität: Ein interdisziplinäres Handbuch*, hg. v. Tobias Klauk und Tilmann Köppe, 516–532. Berlin: de Gruyter, 2014.
- Heßler, Martina und Dieter Mersch (Hg.). *Logik des Bildlichen: Zur Kritik der ikonischen Vernunft*. Bielefeld: transcript, 2009.
- Krämer, Sybille und Christina Ljungberg (Hg.). *Thinking with Diagrams: The Semiotic Basis of Human Cognition*. Boston: de Gruyter Mouton, 2016.
- Kress, Gunther und Theo van Leeuwen. *Reading Images: The Grammar of Visual Design*. Third Edition. London: Routledge, 2021.
- Landwehr Achim. *Diesseits der Geschichte: Für eine andere Historiographie*. Göttingen: Wallstein, 2020.
- Lorenz, Edward N. *The Essence of Chaos*, 179–182. London: UCL Pr., 1993.
- Manovich, Lev. *Cultural Analytics*. Cambridge, MA: The MIT Press, 2020.
- Schubert, Zoe, Jan Wieners und Øyvind Eide, „Visuelle modellbasierte Experimente in den (digitalen) Geisteswissenschaften.“ In *Fabrikation von Erkenntnis – Experimente in den Digital Humanities*, hg. v. Manuel Burghardt, Lisa Dieckmann, Timo Steyer, Peer Trilcke, Niels Walkowski, Joëlle Weis und Ulrike Wuttke. Luxembourg: Melusina Press, 2021–2022. doi: 10.17175/sb005_007.
- Staley, David J. *Computers, Visualization, and History: How New Technology Will Transform our Understanding of the Past*. Second Edition. Armonk: Sharpe 2014.
- Stöckl, Hartmut. *Sprache-Bild-Texte lesen: Bausteine zur Methodik einer Grundkompetenz*. In *Bildlinguistik: Theorien – Methoden – Fallbeispiele*, hg. v. Hans-Joachim Diekmannshenke, Michael Klemm und Hartmut Stöckl, 45–70. Berlin: Erich Schmidt Verlag 2011.
- Wachter, Christian. *Geschichte digital schreiben: Hypertext als non-lineare Wissensrepräsentation in der Digital History*. Bielefeld: transcript 2021.
- Wachter, Christian. „Publishing Complexity in the Digital Humanities.“ *magazén: International Journal for Digital and Public Humanities* 1 (2, 2021), 103–118. doi: 10.30687/mag/2724-3923/2021/03/004.
- White, Hayden V. *Metahistory: The Historical Imagination in Nineteenth-Century Europe*. Baltimore: Johns Hopkins Univ. Pr., 1973.

Moritz Feichtinger

Materialität und Praxis

Zur Analyse historischer Datenverarbeitung am Beispiel des Vietnamkrieges

Abstract: Genuinely digital archival materials require a new set of analytical and critical approaches. They display seemingly paradoxical characteristics, some of which they share with conventional sources, but some of which differ significantly from them. In addition to their content, born-digital sources have certain formal properties and are characterized by the materiality of the respective processing and storage media. Using the example of a database system from the Vietnam War, the article discusses the specific requirements but also the potentials of an analysis of historical data processing and original digital sources, taking into account their material properties and the socio-technical practice of their collection, processing and storage.

Keywords: databases, born digital sources, hermeneutics, source criticism, Vietnam War computer history, archiving

Zusammenfassung: Originär digitale Quellen erfordern ein neues Instrumentarium der Quellenkritik und Analyse. Sie weisen scheinbar paradoxe Eigenschaften auf, die sie zum Teil mit herkömmlichen Quellen teilen, die sich von diesen zum Teil aber auch deutlich unterscheiden. Neben ihrem inhaltlichen Gehalt haben sie bestimmte formale Eigenschaften und sind von der Materialität der jeweiligen Verarbeitungs- und Speichermedien gekennzeichnet. Am Beispiel eines Datenbanksystems aus dem Vietnamkrieg diskutiert der Beitrag die spezifischen Anforderungen aber auch Potentiale einer Analyse historischer Datenverarbeitung und originär digitaler Quellen unter Berücksichtigung ihrer materiellen Eigenschaften und der sozio-technischen Praxis ihrer Erhebung, Verarbeitung und Speicherung.

Schlagwörter: Datenbanken, Born-digital Quellen, Hermeneutik, Quellenkritik, Vietnamkrieg, Computer-Geschichte, Archivierung

Im Jahr 2011 formulierte der Zeithistoriker Kiran Klaus Patel die Frage „wer archiviert eigentlich die SMS von Angela Merkel?“¹ Obwohl diese Frage bereits zwei Jahre später überraschend von Edward Snowden beantwortet wurde, bleiben Patels Appelle an die Geschichtswissenschaften aktuell: Patel verweist auf die Relevanz digitaler Kommunikation zum Verständnis und der Erforschung der jüngsten Vergangenheit, auf die Dringlichkeit, dieser Relevanz in der archivalischen Sammlungs- und Bewahrungspraxis zu entsprechen und appelliert schließlich (implizit) an die eigene Zunft, sich auf die methodischen Herausforderungen im Umgang mit einer scheinbar neuen Quellengattung einzulassen.

Heute, zehn Jahre später, ist die Digital History aus den Diskussionen des Faches nicht mehr wegzudenken. Auffällig ist jedoch, dass Arbeiten zur Zeitgeschichte und speziell solche, die dezidiert originär digitale Quellen untersuchen, nach wie vor eher rar sind. Dies ist keinesfalls nur einem methodischen Konservativismus der Disziplin geschuldet, sondern hat auch mit dem Charakter digitaler Quellen zu tun. Ihre Erforschung stellt die Geschichtswissenschaft vor erhebliche und bislang ungelöste methodische und analytische Probleme, die im Folgenden anhand von drei Paradoxien vorgestellt werden. Dem schließt sich eine kurze Skizze vielversprechender theoretisch-konzeptioneller Zugänge zu originär digitalen Quellen (digital born) an, deren Potential daraufhin am Beispiel eines Datenbanksystems aus dem Vietnamkrieg illustriert wird.

1 Drei Paradoxien originär digitaler Quellen: Überfluss und Mangel, Stabilität und Volatilität, Virtualität und Materialität

Der Digital History mangelt es nicht an Manifesten und euphorischen Programmatiken, in denen ihre Innovationskraft und die durch sie ermöglichte Erweiterung der Perspektiven und breitere empirische Fundierung geschichtswissenschaftlicher Forschung beschworen wird.² Zudem sind solche Texte von der Gewissheit getragen, dass die Grundtendenz zur fortschreitenden Digitalisierung auch zu einer quantitativen Steigerung von digitalen Quellen führen wird. An-

¹ Kiran Klaus Patel, „Zeitgeschichte im digitalen Zeitalter. Neue und alte Herausforderungen,“ *Vierteljahreshefte für Zeitgeschichte* 59/3 (2011): 331–351.

² Jo Guldi und David Armitage, *The History Manifesto* (Cambridge, UK: Cambridge University Press, 2014); Shawn Graham, Ian Milligan und Scott Weingart, *Exploring Big Historical Data: The Historian's Macroscopic* (London: Imperial College Press, 2016).

gesichts der seit der Jahrtausendwende laufenden großen Projekte der Retrodigitalisierung durch Archive und Bibliotheken sowie des erleichterten Online-Zugangs zu diesen Beständen entspricht die oft formulierte Wahrnehmung eines Überflusses an möglichen Quellen durchaus der Alltagserfahrung praktizierender Historiker:innen.³

In krassem Gegensatz steht diese Gewissheit zu Warnungen vor einem informationellen „schwarzen Loch“ digitaler Amnesie, ja, einem digitalen „dunklen Zeitalter“, wie sie seit der Jahrhundertwende vermehrt zu vernehmen sind.⁴ Die Archivwissenschaft befasst sich bereits seit den 1990er Jahren intensiv mit dieser Problematik, wohl aus der praktischen Erfahrung, dass damals digitale Dokumente meist ausgedruckt an Archive übergeben wurden (und damit als interpoliert und selektiv gelten konnten), oder dass Archive digitale Quellen und Artefakte stetig in neue Formate überführen mussten, was ebenfalls einer Interpolation gleichkommen konnte.⁵ Außerdem gewann in der Langzeitbewahrung auch die Selektion der zu bewahrenden Zeugnisse angesichts einer steigenden Menge von digitalen Aktivitäten und Hinterlassenschaften an Bedeutung.⁶

Archiv-, Bibliotheks- und Geschichtswissenschaften versuchen dieser Problemlage durch die gemeinsame Entwicklung von Standards zur Bewahrung von und des Zugangs zu originär digitalen Quellen zu begegnen.⁷ Letztlich aber

3 Roy Rosenzweig, „Scarcity or Abundance? Preserving the Past in a Digital Era,“ Roy Rosenzweig Center for History and New Media, <http://rrchnm.org/essay/scarcity-or-abundance-preserving-the-past-in-a-digital-era/> Zugriff am 19.02.2022; Ian Milligan, *History in the Age of Abundance? How the Web Is Transforming Historical Research* (Montreal: McGill-Queen's University Press, 2019).

4 Oliver Kiechle, „Digitalia fundamentalis – Speicher und Zugriff,“ *Diskrete Werte*, 07.10.2015, <https://digitalia.hypothesen.org/21>. Zugriff am 19.02.2022.

5 Einer der frühesten und nach wie vor umtriebigsten Mahner vor einem Verlust des digitalen Kulturerbes ist der ehemalige RAND-Forscher Jeff Rothenberg, siehe: Jeff Rothenberg, „Avoiding Technological Quicksand: Sections 1–3,“ <https://www.clir.org/pubs/reports/rothenberg/introduction/#longevity>; Jeff Rothenberg, „Preserving Authentic Digital Information,“ <https://www.clir.org/pubs/reports/pub92/rothenberg/>. Zugriff am 19.02.2022.

6 Die rapide technische Obsoleszenz digitaler Speichermedien und Plattformen beschäftigt seit den 1990ern naturwissenschaftliche Forschungszentren ebenso wie Privatpersonen, siehe etwa die Warnungen des ehemaligen Google Chefs Vint Cerf vor dem Verlust privater Korrespondenzen und Erinnerungen oder die entsprechenden Vorkehrungen, die am internationalen Forschungsverbund CERN dagegen unternommen werden: <https://www.theguardian.com/technology/2015/feb/13/google-boss-warns-forgotten-century-email-photos-vint-cerf>; <https://cerncourier.com/a/data-preservation-is-a-journey/>. Zugriff am 19.02.2022.

7 Der älteste und am weitesten verbreitete Standard, das seit 1997 erarbeitete Open Archival Information System (OAIS) Referenzmodell, legte den Grundstein für eine einheitliche Bewahrungs- und Beschreibungspraxis, und hat seit seiner Einführung 2002 zahlreiche Erweiterungen und Ergänzungen erfahren. Ein Überblick findet sich bei: Erin Baucom, „A Brief History of

verweisen solche Anstrengungen auf die enormen und dieser Quellenform inhärenten Anforderungen: das exponentielle quantitative Wachstum möglicher Quellen; der damit einhergehende Mehraufwand in der Anreicherung und Beschreibung der Bestände mit entsprechenden Metadaten; und nicht zuletzt die technischen, personellen und auch energetischen Schwierigkeiten, der Obsoleszenz historischer Systeme und Medien langfristig und nachhaltig zu begegnen.⁸ Originär digitale Quellen sind also von der paradoxen Situation gekennzeichnet, dass gleichzeitig ein Überfluss und ein Mangel zu herrschen scheint.

Ein zweites Paradox, das digital born Quellen inhärent ist, betrifft ihre gleichzeitige Flüchtigkeit (oder Volatilität) und Stabilität. Im Zusammenhang mit der bereits angeschnittenen Obsoleszenz und der andauernden Suche nach Strategien zur Langzeitarchivierung beschreibt etwa Paul Dourish dieses paradoxe Verhältnis:

On one hand, digitization and digital information are often associated with preservation and archival practices; think, for example, of efforts to digitize ancient manuscripts to make them available even as material on which they are written crumbles, or, on a more mundane level, of the services that promise to “preserve” your family members by digitizing old photographs, film, and videotape. Yet the digital devices on which these digital

Digital Preservation,“ *Mansfield Library Faculty Publications* 31 (2019), https://scholarworks.umt.edu/cgi/viewcontent.cgi?article=1033&context=ml_pubs. Zugriff am 19.02.2022; sowie Thorsten Ries und Gábor Palkó, „Born-Digital Archives,“ *International Journal of Digital Humanities* 1, 1 (April 2019), 1–11, doi: 10.1007/s42803-019-00011-x; Frank Kell, „LINKAGE – Digitale Gegenwart und Zukunft Historischer Forschung. Die Ziele der Konsortialinitiative 4Memory,“ <https://blog.historikerverband.de/2020/11/11/linkage-digitale-gegenwart-und-zukunft-historischer-forschung/>. Zugriff am 19.02.2022.

⁸ Aus archivalischer Perspektive, siehe: Michael Friedewald und Timo Leimbach, „Computer-software als digitales Erbe. Probleme aus der Sicht der Technikgeschichte,“ in *Neues Erbe. Aspekte, Perspektiven und Konsequenzen der digitalen Überlieferung*, hg. v. Caroline Robertson-von Trotha und Robert Hauser (Karlsruhe: KIT Scientific Publishing, 2011), 201–219; Siegrid Schieber, „Digitale Archivierung im Hessischen Landesarchiv,“ *ABI Technik*, 37/2 (2017): 93–102, doi: 10.1515/abitech-2017-0022; Andrea Hänger, „Die Mühsal der Zugänglichkeit. Methodische Herausforderungen beim Aufbau virtueller Lesesäle,“ *Zeithistorische Forschungen/Studies in Contemporary History*, Online-Ausgabe, 17/1 (2020): 179–189, doi: 10.14765/zzf.dok-1763; Christian Keitel, „Digitale Archivalien,“ *Südwestdeutsche Archivalienkunde JAHR?*, <https://www.leo-bw.de/themenmodul/sudwestdeutsche-archivalienkunde/querschnittsartikel/digitale-archivalien>; Gerald Maier, „Die Bedeutung der Archive für Forschungsdaten in der Geschichtswissenschaft,“ *Blog Historikerverband*, 11.11.2020, <https://blog.historikerverband.de/2020/11/11/die-bedeutung-der-archive-fuer-forschungsdaten-in-der-geschichtswissenschaft/>. Zugriff am 19.02.2022.

materials are dependent are not themselves eternal or even particularly long-lived. Computer-systems from even just a few years ago are already obsolete and aging.⁹

Diese Paradoxie setzte sich von den Medien und Formaten fort hin zu den Dateien und Daten selbst, wie wir von unserem täglichen Umgang mit Schreibprogrammen und Emails wissen: Unsere Texte und Korrespondenzen existieren in mehreren Stadien, vom Entwurf bis zur ersten publikationsfähigen Version durchlaufen sie mehrfache Durchgänge der Veränderung. Auch Websites, Chatverläufe und natürlich allen voran Datenbanken sind einer stetigen Erneuerung und Aktualisierung unterworfen, eine Eigenschaft, die der Etymologie des Begriffs „Datum“ als etwas Gegebenes, Festes, zuwiderläuft.¹⁰ Damit haben Datenbanken und die in ihnen enthaltenen Informationen einen inhärent provisorischen und volatilen Charakter, der sowohl ihre Archivierung als auch ihre geschichtswissenschaftliche Analyse stets mit der Frage konfrontiert, ob es sich bei einer Quelle um einen „Schnappschuss“ oder um den „letzten Stand“ handelt, und was dies jeweils für ihre Interpretation bedeutet. Mehr noch: viele digital born Quellen sind nicht nur dynamisch, sondern auch interaktiv. Vor allem bei Datenbanken, aber auch bei anderen originär digitalen Artefakten ist zu konstatieren, dass einzelne Datensätze und Datenpunkte überhaupt erst durch Interaktion – etwa durch Abfrage – entstehen.¹¹

Im Widerspruch zur Flüchtigkeit originär digitaler Quellen steht ihre bemerkenswerte Stabilität, die sich der einfachen Kopierbarkeit und damit potentiellen Loslösung vom jeweiligen konkreten Speicher verdankt. Die meisten digital born Quellen existieren bereits in mehreren Kopien, noch bevor sie bewusst vervielfältigt oder gespeichert werden. Matthew Kirschenbaum diskutiert die erstaunliche Stabilität und Permanenz digital-elektronischer (Text-)Dateien eingehend:

Creating a file and saving it to a hard drive does not yield a simple one-to-one correspondence between the document (or file of whatever type) and its record on the disk. First,

⁹ Paul Dourish, *The Stuff of Bits: An Essay on the Materialities of Information* (Cambridge, Mass.: The MIT Press, 2017), 62.

¹⁰ Zum Begriff siehe: Daniel Rosenberg, „Daten vor Fakten,“ in *Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie*, hg. v. Ramón Reichert (Bielefeld: transcript-Verlag, 2014), 133–156.

¹¹ In der vom BKA unter Horst Herold aufgebauten Datenbank zur Suche und Bekämpfung der RAF wären demnach nicht die einzelnen Einträge, wie das Register sämtlicher amtlich gemeldeter Kfz-Kennzeichen von historischem Interesse, sondern die Strategie der Beamten bei ihrer Durchsuchung und Abfrage, die jedoch der Datenbank selbst äußerlich sind, siehe: David Gugerli und Hannes Mangold, „Betriebssysteme und Computerfahndung: Zur Genese einer digitalen Überwachungskultur,“ *Geschichte und Gesellschaft* 42/1 (2016): 144–174.

word processors and other productivity software routinely include an auto-save function that writes a snapshot of an open file to the disk at set intervals. [...] This phenomenon is sometimes known as “ambient data”, the term emphasizing the way in which records accumulate on a file system absent the intervention of any single, conscious (human) agency.¹²

Dieser Prozess der Vervielfältigung durch identische Kopien setzt sich auf bewusster Ebene durch Backup-Routinen klügerer Computer-Nutzer:innen fort sowie auf eher unbewusster Ebene durch die zahllosen Kopien, die eine simple Email auf ihrem Weg durch das Netz auf allen beteiligten Servern generiert. Diese Vervielfältigung und die durch sie bewirkte Stabilisierung erschöpft sich nicht im (semantischen) Inhalt von Dateien, sondern erfasst auch eine Myriade in ihnen enthaltener Metadaten wie Zeitstempel und Log-files. Originär digitale Quellen sind also von der paradoxen Qualität gekennzeichnet, dass sie gleichzeitig als ephemere, flüchtig und volatil beschrieben werden können, aber auch als stabil und persistent.

Ein drittes Paradox, das digital born Quellen zu eigen ist, liegt in ihrer gleichzeitigen Virtualität und Materialität begründet. In die Alltagssprache eingegangene Metaphern wie „virtueller Raum“ oder „cloud“ suggerieren geradezu metaphysische Qualitäten digitaler Informationen, wie Ubiquität und Unabhängigkeit von Energie, Arbeit, Speicherformaten oder Speichermedien. Diese idealistische Wahrnehmung wird jedoch zunehmend in Frage gestellt und mitunter scharfer Kritik unterzogen durch Arbeiten, die auf die infrastrukturellen Voraussetzungen der Digitalisierung und des Internets verweisen, oder ihre ökologischen Kosten betonen und ihre Grundlagen in einer globalen sowie innergesellschaftlichen ökonomischen Ungleichheit und ausbeuterischen Arbeitsverhältnissen herausarbeiten.¹³ Abgesehen von dieser notwendigen Einbettung digitaler Objekte in die physische Um-Welt wirft die Materialität digitaler Medien bedeutsame Fragen für Archiv- und Geschichtswissenschaft auf. Denn wenn-

¹² Matthew G. Kirschenbaum, *Mechanisms: New Media and the Forensic Imagination*, (Cambridge, Mass: MIT Press, 2012), 51.

¹³ Siehe beispielhaft: Lisa Parks und Nicole Starosielski (Hg.), *Signal Traffic: Critical Studies of Media Infrastructures, The Geopolitics of Information* (Urbana: University of Illinois Press, 2015); Dirk van Laak, *Alles im Fluss: die Lebensadern unserer Gesellschaft – Geschichte und Zukunft der Infrastruktur* (Frankfurt a. M.: S. Fischer, 2018); Michael Homberg, „Mensch/Mikrochip. Die Globalisierung der Arbeitswelten in der Computerindustrie 1960 bis 2000 – Fragen, Perspektiven, Thesen“, *Vierteljahrshefte für Zeitgeschichte* 66/2 (2018): 267–293; Nathan Ensmenger, „The Environmental History of Computing“, *Technology and Culture* 59/4S (2018): 7–33; Robert Markley, „Boundaries: Mathematics, Alienation, and the Metaphysics of Cyberspace“, in *Virtual Realities and their Discontents*, hg. v. Dems. (Baltimore MD: Johns Hopkins University Press 1996), 55–78.

gleich digitale Objekte nicht notwendig an individuelle Computersysteme und Speichermedien gebunden sind (im Sinn von „ein ganz bestimmtes System oder mein persönlicher Computer/USB-Stick, diese oder jene Festplatte oder Rechenzentrum“), so können sie doch notwendigerweise nur in und auf spezifischen Speichern und Systemen existieren und funktionieren. Originär digitale Quellen sind wie alle digitalen Objekte stets mindestens an einem spezifischen Ort physisch und materiell präsent und funktional von einem spezifischen informationellen System zu ihrer Verarbeitung und Darstellung abhängig. Wie Matthew Kirschenbaum in seiner Analyse der „Grammatologie des Hard-Drives“ herausgearbeitet hat, handelt es sich auch bei digital gespeicherten Informationen letztlich um „Inschriften“, also physische und materielle Repräsentationen auf einer Trägersubstanz. Diese kann vorliegen etwa in Form von elektromagnetischen Polarisierungen auf einer Festplatte, CD-ROM oder Diskette, oder in den festgehaltenen elektrischen Spannungen in den Zuständen der einzelnen Halbleiter in einem Flash-Speicher, wie etwa in USB-Sticks oder SD-Karten.¹⁴ Damit unterscheiden sich die modernsten Speichermedien in ihren Funktionsprinzipien und ihrer Materialität letztlich nicht von den Lochkarten der Frühzeit der Informations- und Computertechnologie, in denen die jeweiligen Informationen eingestanzte waren. Jedes Speichermaterial ist dabei durch Verfall und Verlust gefährdet – wenngleich in verschiedener Weise und in verschiedenem Tempo – indem es physisch degradiert oder entsprechende Hardware zur Auslese nicht mehr zur Verfügung steht.¹⁵ Selbst wenn es gelingt, einem obsoleten Speichermedium historische Datensätze zu entreißen, ist damit keineswegs sichergestellt, dass diese dann sinnvoll ausgewertet oder auch nur gelesen werden können. Denn die oben schon thematisierte Obsoleszenz betrifft nicht nur Hardware, sondern auch Software. Dieses Problem betrifft keineswegs nur digital born Quellen und entsprechende Anwendungen aus der Frühzeit des Computerzeitalters. Schon Datenbankanwendungen, Textprogramme und Webanwendungen aus den frühen 1990er Jahren lassen sich nur unter größten Mühen les-

14 Einzelne Bits auf elektromagnetischen Speichermedien wie Festplatten lassen sich sogar mit einem Elektronenrastermikroskop erkennen und darstellen, vgl. Kirschenbaum, *Mechanisms*, 62.

15 Ein anschauliches Beispiel dafür sind die geradezu heroischen Bemühungen der Beschäftigten der Bundesbeauftragten für die Stasiunterlagen, die auf Magnetbändern und Magnetplatten gespeicherten digitalen Hinterlassenschaften der Stasi und anderer DDR Behörden zu erschließen, siehe: Birgit Salamon, „Die archivfachliche Arbeit an den MfS-Geheimdienstunterlagen – Fragen und Herausforderungen“, *Der Archivar* 55/3 (2002): 203, <https://www.bstu.de/archiv/fachbeitraege/die-archive-der-bundesbeauftragten-fuer-die-stasiunterlagen-bstu-die-archivfachliche-arbeit-an-den-mfs-geheimdienstunterlagen-fragen-und-herausforderungen/>. Zugriff am 19.02.2022.

bar machen, geschweige denn ihr historischer Anwendungszusammenhang und ihr Aussehen wiederherstellen.¹⁶ Originär digitale Quellen haben somit die paradoxe Eigenschaft, dass sie zwar einerseits abstrakte und digitale Repräsentationen von Informationen darstellen, gleichzeitig aber ihre Materialität und ihre Gebundenheit an spezifische Hard- und Software die archivalische Bewahrung und geschichtswissenschaftlichen Analyse maßgeblich beeinflussen. Das jeweilige physische Substrat eines digitalen Objekts, seine jeweilige funktional-informationelle Umgebung, und letztlich auch sein historisch-sozialer Kontext sind jedoch zugleich die Schlüssel zu seiner Kritik und Analyse als geschichtswissenschaftliche Quelle.

2 Historische Praxeologie und „material turn“: Konzeptionelle und methodische Zugänge

Der ehemalige Leiter der Abteilung für elektronische Dokumente am US-amerikanischen Nationalarchiv, Kenneth Thibodeau, hat der Archivwissenschaft und der Geschichtswissenschaft eine konzise analytische Definition digitaler Objekte zur Verfügung gestellt. Nach dieser vielzitierten Definition hat jedes digitale Objekt spezifische physikalische, logische und konzeptuelle Eigenschaften:

All digital objects are entities with multiple inheritance; that is, the properties of any digital object are inherited from three classes. Every digital object is a physical object, a logical object, and a conceptual object, and its properties at each of those levels can be significantly different. A *physical* object is simply an inscription of signs on some physical medium. A *logical* object is an object that is recognized and processed by software. The *conceptual* object is the object as it is recognized and understood by a person, or in some cases recognized and processed by a computer application capable of executing business transactions.¹⁷

16 Die Bemühungen des Marbacher Literaturarchivs, den in jeder Hinsicht obskuren Nachlass Friedrich Kittlers zu erschließen, künden von den vielfältigen Schwierigkeiten eines solchen Unterfangens im Umgang mit born-digitals. Der späte Kittler war Linux-Enthusiast und begeisterter Autodidakt im Umgang mit Computern, weshalb seine Verzeichnis- und Ablagelogik einzigartig ist und er auch mit selbst geschriebenen Programmen arbeitete. Neben diesen qualitativen Herausforderungen sorgt zugleich der schiere Umfang des Nachlasses mit gut einem Terrabyte für anhaltende Beschäftigung in Marbach: <https://www.mww-forschung.de/born-digitals>. Zugriff am 19.02.2022.

17 Kenneth Thibodeau, „Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years,“ *CLIR and the Library of Congress, The State of Digital preserva-*

In eine ganz ähnliche Richtung entwickelte Matthew Kirschenbaum seine Überlegungen zur Materialität digitaler Artefakte. Er unterscheidet zwischen einer formalen und einer forensischen Materialität. Erstere entspricht dabei der Klasse der logischen Objekte in der Definition von Thibodeau, also den (temporären) Formen und Eigenschaften, die digitale Objekte im Hinblick und im Austausch mit der jeweiligen Software und dem sie umgebenden informationellen System haben müssen. Die forensische Materialität entspricht dem, was Thibodeau die physikalische Ebene nennt, also die manifeste Markierung auf einer physischen Substanz.¹⁸

Von großer theoretischer Bedeutung ist auch die von Niels Brügger eingeführte analytische Differenzierung zwischen (retro-)digitalisierten Quellen, born digital und re-born digital Quellen. Vor allem die letzte Gattung hat weitreichende analytische Konsequenzen, da sie auf den Umstand verweist, dass digital born Quellen, welche zu Forschungszwecken aus ihrem Entstehungskontext extrahiert und zur Analyse aufbereitet und angereichert werden, letztlich mit den Originalen nicht mehr identisch sind.¹⁹

Wichtige Impulse verdankt die Diskussion um die Materialität digitaler Objekte schließlich Paul Dourishs Untersuchung der Auswirkungen von Materialität(en) auf den Inhalt von Daten. In Fallstudien zu Tabellenkalkulationsprogrammen, Datenbanken und der Infrastruktur des Internets zeigt Dourish, dass sich die logische beziehungsweise formale Materialität von Daten und ihre jeweiligen Eigenschaften durch jeweils spezifische Anforderungen und Möglichkeiten informationelle, aber auch soziale Praktiken determinieren, ihnen also im Sinne Bruno Latours *agency*, oder doch zumindest Eigensinn zugesprochen werden kann.²⁰

Originär digitale Quellen erfordern einen quellenkritischen Zugang wie alle anderen Quellen auch, also eine Auseinandersetzung mit ihrem Entstehungskontext, der mit ihnen möglichen Interaktionen sowie ihrer jeweiligen Rolle und Wirkmächtigkeit in Diskursen und politischen Dynamiken. Das häufigste informationelle Ökosystem historischer digitaler Quellen sind Datenbanken. In einem Arrangement von Klassen und Kategorien, Spalten und Zeilen sowie de-

tion: An International Perspective, April 2002, <https://www.clir.org/pubs/reports/pub107/thibodeau/>. Zugriff am 19.02.2022.

18 Kirschenbaum, *Mechanisms*, 13–17.

19 Niels Brügger, *The Archived Web. Doing History in the Digital Age* (Cambridge, MA: MIT Press, 2018), 21–23.

20 Dourish, *The Stuff of Bits*, 202–206; Edwin Sayes, „Actor-Network Theory and Methodology: Just What Does It Mean to Say That Nonhumans Have Agency?“ *Social Studies of Science* 44/1 (2014): 134–149.

ren Sortierung wurden und werden die meisten Daten aggregiert, gespeichert und verarbeitet. Von der einfachen Tabelle bis zu komplexen relationalen Datenbanken und modernen NoSQL-Systemen gehören Datenbanken zu den wichtigsten Infrastrukturen und Ordnungssystemen des Informationszeitalters.²¹ In der Auswahl und Benennung von Attributen und Kategorien einer jeweiligen Datenbank sind wichtige konzeptionelle Vorannahmen und Entscheidungen enthalten, die den einzelnen Datenpunkten einen Sinnzusammenhang verleihen. Diese Entscheidungen für eine gewisse Repräsentation der Welt und gegen mögliche andere, können ihrerseits wiederum Rückwirkungen auf die (reale) Welt haben, indem sie bestimmte Wahrnehmungen und Interpretationszusammenhänge gegenüber anderen begünstigen. Diese sinnstiftende Wirkung von Datenbanken hat Rob Kitchin konzipiert beschrieben:

Databases and data infrastructures are not simply neutral, technical means of assembling and sharing data; they are not merely products that store captured data about the world, but are bundles of contingent and relational processes that do work in the world. They are complex sociotechnical systems that are embedded within a larger institutional landscape of researchers, institutions and corporations, constituting essential tools in the production of knowledge, governance and capital.²²

Für die Geschichtswissenschaft und insbesondere die an digital born Quellen interessierte Digital History ergibt sich daraus eine zwingende quellenkritische Notwendigkeit, sich eingehender mit Datenbanken zu befassen.²³

Datenbanken funktionieren und existieren nur auf Grundlage und im Verbund mit ihrem jeweiligen Betriebssystem, dem *Database Management System* (DBMS). Datenbanken sind somit nicht nur reine Speichermedien für Daten, sie konstituieren Informationen oftmals erst durch ihre Abfrage und die Prinzipien

21 Zur Geschichte und Funktionsweise der gebräuchlichsten Datenbanksysteme siehe: David Gugerli, „Die Welt als Datenbank. Zur Relation von Softwareentwicklung, Abfragetechnik und Deutungsautonomie,“ in *Daten. Nach Feierabend. Zürcher Jahrbuch für Wissensgeschichte* 3, hg. v. David Gugerli et al. (Zürich/Berlin: Diaphanes, 2007), 11–36; Marcus Burkhardt, *Digitale Datenbanken: eine Medientheorie im Zeitalter von Big Data* (Bielefeld: Transcript, 2015); Markus Krajewski, *ZettelWirtschaft. Die Geburt der Kartei aus dem Geiste der Bibliothek* (Berlin: Kulturverlag Kadmos, 2002). Zum Begriff und Konzept des „Informationszeitalters“: James R. Beniger, *The Control Revolution: Technological and Economic Origins of the Information Society* (Cambridge, Mass.: Harvard University Press, 1997).

22 Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & their Consequences* (Los Angeles: SAGE Publications, 2014), 51

23 Pascal Föhr, *Historische Quellenkritik im Digitalen Zeitalter*, Dissertation, Universität Basel 2018, doi: 10.5451/unibas-006805169, hier vor allem 47.

ihrer Ablage- und Aggregationsroutinen.²⁴ Damit ergeben sich mehrere analytische Perspektiven zu ihrer Erforschung und Historisierung: Databasing kann verstanden werden als epistemologischer Zugang zur Ordnung der Welt, der aus der Unterteilung von Eindrücken und Erfahrungen in signifikante und unwichtige Signale besteht, deren Unterteilung in diskrete Einheiten und deren Einordnung in ein Klassifikationssystem aus Attributen, Kategorien und Werten.²⁵ Databasing ist zugleich eine (oftmals kollektive) sozio-technische Praktik, in der die epistemologisch-konzeptionelle Vorarbeit der Datafizierung im Medium digitaler Datenbanksysteme realisiert und operationalisiert wird. Diese Praktik besteht zunächst aus der Eingabe, Manipulation und Abfrage der Daten, dehnt sich aber auch auf die Diskussionen der beteiligten Menschen untereinander und ihrer oftmals nicht reibungslosen Interaktion mit den Maschinen aus.²⁶ Nicht zuletzt kann Databasing als Versuch verstanden werden, mithilfe der Objektivitätssuggestion computer-generierter Informationen und großer Zahlenreihen Evidenz zu produzieren und im Diskurs als strategische Ressource einzusetzen.²⁷

Neuere Strömungen der Computergeschichte wenden sich zunehmend hin zu sozialgeschichtlich inspirierten Studien zur Interaktion humaner historischer Akteur:innen mit den jeweiligen Systemen und damit weg von Meistererzählungen über geniale (meist männliche) Erfinder und Pioniere und den von ihnen geschaffenen Technologien.²⁸ Nicht zufällig geraten durch diese Perspektivverschiebung vermehrt Frauen und ihr Beitrag zur Entwicklung moderner Informa-

24 Zu Entwicklung und Geschichte von DBMS siehe: Thomas Haigh, „How Data Got its Base: Information Storage Software in the 1950s and 1960s,“ in: *IEEE Annals of the History of Computing* 31/ 4 (Oktober 2009): 6–25, doi: 10.1109/MAHC.2009.123; Thomas J. Bergin und Thomas Haigh, „The Commercialization of Database Management Systems, 1969–1983,“ *IEEE Annals of the History of Computing* 31/4 (Oktober 2009): 26–41, doi: 10.1109/MAHC.2009.107.

25 Bianca Priel und Daniel Houben, „Einführung. Soziologische Perspektiven auf die Datafizierung der Gesellschaft,“ in *Datengesellschaft: Einsichten in die Datafizierung des Sozialen*, hg. v. Dens. (Bielefeld: Transcript, 2018), 7–32, hier vor allem 18.

26 Ein Beispiel einer historischen Analyse von Databasing-Praktiken ist: Rebecca M. Lemov, *Database of Dreams: the Lost Quest to Catalog Humanity* (New Haven: Yale University Press, 2015); Zu praxeologischen Ansätzen in der Zeitgeschichte allgemein siehe: Lucas Haasis und Constantin Rieske (Hg.), *Historische Praxeologie: Dimensionen vergangenen Handelns* (Paderborn: Ferdinand Schöningh, 2015).

27 Theodore M. Porter, „Quantification and the Accounting Ideal in Science,“ *Social Studies of Science* 22/4 (1992): 633–651; Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge, Mass.: Harvard University Press, 2011).

28 Programmatisch formuliert bei: Thomas Haigh, „Introducing the Early Digital,“ in *Exploring the Early Digital*, hg. v. Dens. (Cham: Springer International Publishing, 2019), 1–18, doi: 10.1007/978-3-030-02152-8_1.

tions- und Kommunikationstechnologien in den Blick.²⁹ Zudem entfaltet der Blick weg vom Computer und hin zum ihn bedienenden und wartenden Personal (und – seit der Durchsetzung des Personal Computers – auch zu privaten Anwender:innen) ein Panorama der Arbeitswelt und Kultur im Informationszeitalter, das weit über die Labore und Büros des Silicon Valley mit ihren obligaten Yogamatten und Tischtennisplatten hinausreicht.³⁰

Um die Funktionsweise historischer Informationsverarbeitungssysteme und die Interaktion daran beteiligter Personengruppen besser verstehen zu können, ist es mitunter hilfreich, historische Systeme zu reaktivieren und historische Praktiken durch Reenactment nachzuvollziehen. Diese Vorgehensweise immunisiert gegen teleologische Rückprojektionen, in denen historische Phänomene zu bloßen Vorläufern und Stationen auf dem Weg zur Gegenwart (und damit ihrer Vollendung) degradiert werden.³¹ Eine direkte Auseinandersetzung mit der historischen Technik ermöglicht zugleich Einsichten in die für ihr Funktionieren notwendigen Arbeiten und die Grenzen mit solcher Technik möglichen Datenverarbeitung. Insbesondere Forschungen zur Frühgeschichte der Digitalisierung profitieren dabei vom Austausch mit technischen Sammlungen und Museen, um ein Verständnis für Computer zu entwickeln, die keine der heute gängigen Schnittstellen wie grafische Benutzeroberflächen und Computer-Maus, ja nicht einmal Bildschirme aufwiesen.³² Das Potential der hier vorgestellten theoretischen Zugänge soll nun im Folgenden anhand einer konkreten Fallstudie skizziert werden.

29 Marie Hicks, *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*, History of Computing (Cambridge, MA: MIT Press, 2018); Janet Abbate, *Recoding Gender: Women's Changing Participation in Computing* (Cambridge, Mass.: MIT Press, 2017).

30 David N. Pellow und Lisa Sun-Hee Park, *The Silicon Valley of Dreams: Environmental Injustice, Immigrant Workers, and the High-Tech Global Economy*, Critical America (New York: New York University Press, 2002); Nathan Ensmenger, *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise* (Cambridge, Mass.: MIT Press, 2010); für neuere Perspektiven auf die Computernutzung in Westdeutschland siehe: Frank Bösch (Hg.), *Wege in die digitale Gesellschaft: Computernutzung in der Bundesrepublik 1955–1990* (Göttingen: Wallstein Verlag, 2018).

31 Thomas Haigh, Mark Priestly und Chrispin Rope haben mit ihrer Studie zur Funktionsweise (bzw. deren häufiges Ausbleiben) des „ersten modernen Computers“ ENIAC vorbildlich demonstriert, dass ein detaillierter Nachvollzug und eine entsprechende „dichte Beschreibung“ eine große Zahl weit verbreiteter Mythen aufzudecken und zu dekonstruieren vermag: Thomas Haigh, Mark Priestley und Crispin Rope, *ENIAC in Action: Making and Remaking the Modern Computer* (Cambridge MA: MIT Press, 2018).

32 Haigh, „Introducing the Early Digital“; Lori Emerson, *Reading Writing Interfaces: From the Digital to the Bookbound* (Minneapolis: University of Minnesota Press, 2014).

3 Fragmente einer Kriegsmaschine

Während des Vietnamkriegs versuchten die Amerikaner und ihre Verbündeten, ihre technische Überlegenheit gegenüber der hochmotivierten und von der Lokalbevölkerung unterstützte Guerillabewegung des Vietcong auszuspielen. Auch Computer, Datenverarbeitungssysteme und Datenbanktechnologie kamen dabei zum Einsatz. Die Überreste dieses frühen Versuchs, mithilfe von elektronischer Datenverarbeitung komplexe soziale Dynamiken zu erfassen und zu modellieren, werden in verschiedenen Formaten vom US-Nationalarchiv aufbewahrt. An ihnen lässt sich zeigen, wie die oben genannten Eigenschaften originär digitaler Quellen, vor allem ihre formal-logische und physische Materialität, für ihre Analyse und Kritik fruchtbar gemacht werden können.

Als der 1961 ins Amt gekommene US-Präsident John F. Kennedy den Ford-Manager Robert S. McNamara zum neuen Verteidigungsminister ernannte, war dies für viele Beobachter:innen, aber auch für den Berufenen selbst eine Überraschung. Der Überraschungseffekt erfasste ebenso die militärische Führung der USA, denn der neue Minister begann sogleich mit einer rigorosen Umstrukturierung der Streitkräfte, die alsbald als revolutionär bezeichnet wurde.³³ Unter dem Eindruck des rasch eskalierenden Kalten Krieges und der thermonuklearen Bedrohung setzte McNamara seine Kompetenz und Macht ein, um dem US-Militär nicht nur strategische Überlegenheit zu sichern, sondern auch Effizienz und durchaus zugleich finanzielle Rentabilität. Bei der Durchsetzung dieser Ziele gegen den nicht unerheblichen Widerstand der militärischen Elite unterstützte ihn ein Stab aus ehemaligen Weggefährten aus der Harvard Business School, die in der Behörde bald als „whiz kids“ bezeichnet wurden (was sich sinngemäß mit „Klugscheißer“ übersetzen ließe). Ähnlich rigoros wie McNamara zu Beginn seiner Amtszeit Armee und Luftwaffe einer strengen Kosten-Nutzen-Kontrolle unterworfen und strategische Machtprotektion logistisch interpretiert hatte, machte er auch den Vietnamkrieg 1967 zur Chefsache, nachdem das seit 1965 stetig intensivierte Engagement der US-Armee nicht zur erwünschten Stabilisierung des südvietnamesischen Regimes und der Eindämmung der kommunistischen Guerilla geführt hatte. Als typischer Manager seiner Zeit und glühender Verfechter des *Operations Research* und der Systemanalyse verlangte er ein „objektives“ Bild der Verhältnisse, auf das künftige politische und militärische Ent-

³³ Charles R. Shrader, *History of Operations Research in the United States Army, V. 2, 1961–1973* (U.S. Department of Defense, U.S. Army Center of Military History, 2008), 37–68; Charles A. Stevenson, *Warriors and Politicians. US Civil-Military Relations Under Stress* (New York: Routledge, 2006), 153–164.

scheidungen gegründet sein sollten.³⁴ Wenig verwunderlich sollte diese Objektivität mit den modernsten Methoden der Datensammlung und -analyse – also mit modernen Computern – erreicht werden. Schon vor 1967 setzte die US-Armee in Vietnam Computer ein, etwa zur Verwaltung ihrer Logistik oder der Abwicklung der Soldzahlungen und Urlaubsansprüche der Truppe. Auch mit aus heutiger Sicht bizarren Ansätzen wurde experimentiert, etwa bei der Verkabelung des Ho-Chi-Minh-Pfades mit Urin- und Geräuschsensoren, um Bombardierungen auf die Nachschubliefereien des Vietcong automatisiert auszulösen.³⁵ In der Konsequenz eskalierend wirkten sich Versuche einer statistischen Betrachtung und Steuerung des Kriegsgeschehens im berühmten „body count“ aus, durch den der militärische Druck auf die kommunistische Guerilla quantifizierbar gemacht werden sollte, wo Informationen zu ihrer Truppenstärke fehlten.³⁶

Die seit 1967 eingerichteten Systeme zur statistischen Erfassung der „Pazifizierung“ der südvietnamesischen Landgebiete zeichneten sich demgegenüber durch ihren holistischen Anspruch und ihre (sozial-)wissenschaftliche Begleitung aus.³⁷ Um die Aufstandsbekämpfung der südvietnamesischen Armee und ihres amerikanischen Alliierten aus der ewigen Wiederkehr eines Katz-und-Maus-Spiels zu befreien, das sich in wirkungslosen „search und destroy“ Missionen erschöpfte, und um der breiten Unterstützung der Guerilla durch die Landbevölkerung auch politisch zu begegnen, richtete Präsident Lyndon B. Johnson 1967 eine zivil-militärische Super-Behörde unter Leitung des US-Bot-

34 Joy Rohde, *Armed with Expertise: The Militarization of American Social Research During the Cold War* (Ithaca: Cornell University Press, 2013); Sharon Weinberger, *The Imagineers of War: The Untold Story of DARPA, the Pentagon Agency that Changed the World* (New York: Vintage Books, 2018); Torben Lütjen, „Vom ‚Gospel of Efficiency‘ zum ‚War of Ideas‘. Zum Verhältnis von Wissenschaft, Politik und Ideologien in den Vereinigten Staaten,“ *Archiv für Sozialgeschichte* 50 (2010): 373–394.

35 Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, Mass: MIT Press, 1996), 3–6; Seymour J. Deitchman, „The ‚Electronic Battlefield‘ in the Vietnam War,“ *The Journal of Military History* 72/3 (2008): 869–887; Oliver Belcher, „Sensing, Territory, Population: Computation, Embodied Sensors, and Hamlet Control in the Vietnam War,“ *Security Dialogue* 50/5 (Oktober 2019): 416–436.

36 Gregory A. Daddis, *No Sure Victory: Measuring U.S. Army Effectiveness and Progress in the Vietnam War* (New York: Oxford University Press, 2011), 95–106.

37 Siehe dazu allgemein: Stefan Kaufmann, „Die Wissensformierung der ‚counterinsurgency‘ im Vietnamkrieg,“ *Traverse* 16/3 (2009): 37–52; Claus Pias, „Mit dem Vietcong rechnen. Der Feind als Gestalt und Kunde,“ in *Freund, Feind & Verrat. Das politische Feld der Medien*, hg. v. Cornelia Epping-Jäger, Torsten Hahn und Erhard Schüttelpelz (Köln: DuMont, 2004), 157–83; Antoine J. Bousquet, *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity* (New York: Columbia University Press, 2009), 137–139.

schafters Robert Komer ein, das Büro für *Civil Operations and Revolutionary Development Support* (CORDS).³⁸ Hier sollten die Informationen aus den verschiedenen Quellen der US-Armee (*Military Assistance Command, Vietnam*, MACV), des Verteidigungsministeriums, der CIA, der *Agency for International Development* (AID) und der Südvietnamesischen Regierung und Streitkräfte gebündelt und koordiniert werden. Voraussetzung für eine effektive Koordination der Aufstandsbekämpfung war ein umfassendes Wissen über die Landbevölkerung, deren „Herzen und Köpfe“ zu gewinnen waren. Dieses Wissen sollte durch ein computerisiertes Daten-Verarbeitungssystem gewährleistet werden, das *Hamlet Evaluation System*, HES. Dieses im Januar 1967 in Betrieb genommene Datenbanksystem enthielt einen Index sämtlicher rund 12 000 Dörfer und Weiler unter der Kontrolle des südvietnamesischen Regimes und seiner Alliierten, inklusive demographischer Informationen und geografischer Koordinaten, Einschätzungen zur Sicherheitslage und Indikatoren zur jeweiligen sozio-ökonomischen Situation, Selbstverwaltung, Entwicklungspotentiale und so weiter. Zusätzlich zu diesen Grunddaten erhoben speziell geschulte Teams regelmäßig Daten zu den politischen Einschätzungen, Zukunftserwartungen und Moralvorstellungen der Bevölkerung, anhand derer die Wirksamkeit einzelner Maßnahmen militärischer Repression, psychologischer Kriegsführung und sozioökonomischer Entwicklungsinitiativen evaluiert und somit gezielter dosierbar gemacht werden sollten. Einige der renommiertesten Sozialwissenschaftler:innen der USA beteiligten sich an der Konzeption und Operationalisierung dieses Großversuchs in angewandter Sozialforschung. Angeschlossene Teilsysteme wie das psychographisch orientierte *Pacification Attitudes Analysis System* (PAAS) oder die mit Vietcong-Deserteuren durchgeführten Interviews im Rahmen des „Chieu Hoi“-Programms (Offene Arme) flossen später in Publikationen beteiligter Wissenschaftler:innen und in die Fachdiskurse der quantitativen Sozialwissenschaften sowie der Formulierung der Modernisierungstheorie ein.³⁹ Ein auf Grundlage der Erfahrungen mit der Aufstandsbekämpfung in Vietnam entwickeltes System zur Früherkennung potentieller (kommunistisch inspirierter) Aufstände in den Vereinigten Staaten selbst, das „Project Camelot“, provozierte schließlich den

38 Richard A. Hunt, *Pacification: The American Struggle for Vietnam's Hearts and Minds* (Boulder: Westview Press, 1995); Christian Tripodi, „Pacification in Vietnam 1964–72,“ in *The Unknown Enemy: Counterinsurgency and the Illusion of Control*, hg. v. Dems. (Cambridge: Cambridge University Press, 2020), 112–137.

39 Michael E. Latham, *Modernization as Ideology: American Social Science and „Nation Building“ in the Kennedy era* (Chapel Hill: University of North Carolina Press, 2000); Nils Gilman, *Mandarins of the Future: Modernization Theory in Cold War America* (Baltimore: Johns Hopkins University Press, 2007).

Protest kritischer Forscher und der politisierten Studentenorganisationen der späten 1960er Jahre.⁴⁰ Dieser kritischen Einordnung der im Vietnamkrieg eingesetzten Systeme und auch der sie begünstigenden technokratischen, kaltblütigen Mentalität hat sich die Geschichtswissenschaft später größtenteils angeschlossen.⁴¹ Allerdings befasste sich die historische Forschung bislang vor allem mit der Rolle computerbasierter Analysen in der strategischen Entscheidungsfindung der US-Streitkräfte und weniger mit dem Funktionieren, geschweige denn der Praxis und Materialität der Datenverarbeitung.

Die Datenverarbeitung des HES bestand aus einem mehrstufigen Prozess, in dessen Verlauf die Datensätze auf verschiedenen Medien gespeichert und verarbeitet wurden, sich ihr Informationsgehalt durch Aggregation und Prozessierung mitunter erheblich änderte und an dem eine ganze Reihe sehr verschiedener historischer Akteure beteiligt waren. Am Beginn standen auf Karton gedruckte Fragebögen. Die Fragebögen und ihre Auswertungsmatrix waren von einem Team aus Sozialwissenschaftlern um den Politologen Ithiel de Sola Pool entwickelt worden, der durch die Computer-basierte Berechnung von Wahlergebnissen Bekanntheit erlangt hatte.⁴² Die Karten bestanden aus insgesamt 136 Fragen zu 18 Indikatoren aus den Bereichen Politik, Sicherheit, Wirtschaft, Verwaltung und Gesundheit, die jeweils in einer fünfstufigen Skala zu bewerten waren. Teams aus jeweils einem amerikanischen Offizier (*District Senior Advisor*, DSA) und einigen vietnamesischen Assistenten und Übersetzern wurden in jedem der 256 militärischen Sub-Distrikte monatlich ausgeschickt, um die Situa-

40 Zu „Camelot“ sowie der Beteiligung angewandter Sozialwissenschaften an der Counterinsurgency in Vietnam siehe: Mark Solovey und Hamilton Cravens, *Cold War Social Science: Knowledge Production, Liberal Democracy, and Human Nature* (Basingstoke GB: Palgrave Macmillan, 2014); Stuart W. Leslie, *The Cold War and American Science: The Military-Industrial-Academic Complex at MIT and Stanford* (New York: Columbia University Press, 1993); Irving Louis Horowitz, *The Rise and Fall of Project Camelot: Studies in the Relationship between Social Science and Practical Politics*, Rev. ed. (Cambridge, Mass: MIT Press, 1974); Janeen M Klinger, *Social Science and National Security Policy Deterrence, Coercion, and Modernization Theories*. (New York: Palgrave 2019); Joy Rohde, *Armed with Expertise: The Militarization of American Social Research during the Cold War* (Ithaca: Cornell University Press, 2013); Yasha Levine, *Surveillance Valley – The Secret Military History of the Internet* (New York: Public Affairs, 2018), 67.

41 James William Gibson, *The Perfect War: Technowar in Vietnam* (New York: Atlantic Monthly Press, 2000); Thomas C Thayer, *War Without Fronts* (Naval Institute Press, 2016); Oliver Belcher, „Data Anxieties: Objectivity and Difference in Early Vietnam War Computing,“ in *Algorithmic Life: Calculative Devices in the Age of Big Data*, hg. v. Louise Amoore und Volha Piotukh (New York: Routledge, 2016), Kap. 7.

42 Jill Lepore, *IF THEN: How the Simulmatics Corporation Invented the Future* (New York: Live-right, 2020).

tion in allen Dörfern und Weilern zu erfassen und auf den Karten einzutragen.⁴³ Am Ende des Monats musste der DSA die Ergebnisse seiner Erhebungen inklusive zusätzlicher Bemerkungen auf einem vorgedruckten Formblatt eintragen und an das Provinz-Hauptquartier weiterleiten, wo die Formblätter gesammelt, auf Vollständigkeit und Kohärenz überprüft und schließlich nach Saigon in das CORDS-Hauptquartier weitergeleitet wurden. Dort wurden die Blätter von der *Reports and Analysis Division* (RAD) erneut geprüft und sortiert.⁴⁴ In der *Data Management Agency* des Oberkommandos wurden die Formblätter schließlich von angestellten Vietnamesinnen auf Lochkarten gestanzt und dann in die Computer eingespeist. Auf Grundlage dieser monatlich aktualisierten Daten berechneten Analysten des RAD (zum Teil Militärs und Geheimdienstler, aber auch zivile Sozialwissenschaftler und Angestellte privater Datenverarbeitungs-Firmen unter Regierungsvertrag) Veränderungen und Prognosen zur politisch-militärischen Situation. Die Resultate wurden schließlich in den monatlich erscheinenden „Southeast Asia Analysis Reports“ zusammengefasst, auf riesigen Karten geplottet und der gesamte Monatsreport zusätzlich in mehrfacher Ausführung auf Endlospapier ausgedruckt.⁴⁵ Die schiere Menge der im HES verarbeiteten Daten war für damalige Verhältnisse beispiellos: 9 000 Lochkarten wurden monatlich gestanzt, 10 Millionen Einträge enthielt das Masterfile auf mehreren Magnetbändern und 90 000 Seiten umfassten die Reports. Insgesamt produzierte das Oberkommando der US-Streitkräfte in Vietnam sagenhafte sechs Tonnen bedrucktes Papier, weshalb der Vietnamkrieg bis heute als einer der am besten dokumentierten militärischen Konflikte des 20. Jahrhunderts gilt.⁴⁶ Dieses gewaltige und komplizierte Unterfangen war an verschiedenen Stationen des Prozederes anfällig für Fehler und Verluste, die nicht nur an der mitunter eingeschränkten Zuverlässigkeit der Befragten und an der Erhebungspraxis, oder an der allgemeinen Unübersichtlichkeit des Guerilla-Krieges lagen, sondern auch an der jeweiligen Materialität der Speichermedien. Die Frage-Karten für die einzelnen Dörfer schafften es zum Beispiel oftmals nicht bis ins Provinzhauptquartier, etwa wenn sie bei Gefechten verloren gingen, durch das Dschungelklima, Regen oder Matsch zerstört oder durch Nagetierfraß beschä-

⁴³ US National Archives and Records Administration, College Park, Maryland (NARA), MACV/CORDS, Pacification Study Group, HES Briefing manuscript 1972, ARC ID 4489248, Box 31.

⁴⁴ NARA/CORDS, Reports & Analysis, Control Data Corporation HES/70 Description, ARC ID 6087884, Box 4.

⁴⁵ NARA, MACV/CORDS, Information Report February 1973, ARC ID 4526870; NARA, MACV/CORDS, Southeast Asia Analysis Reports, October 1969 – February 1970, ARC ID 4506183.

⁴⁶ Erwin R. Brigham, „Pacification Measurement in Vietnam: The Hamlet Evaluation System,“ *Military Review*, May 1970.

digd wurden.⁴⁷ Das System erwies sich außerdem oft als zu statisch, um die Dynamik des Krieges und seiner Konsequenzen zu erfassen. So wussten sich die Analysten nicht anders zu behelfen, als die Millionen von Binnenflüchtlingen unter der Kategorie ihrer Herkunftsdörfer zu erfassen, was das Gesamtbild oftmals bis ins Absurde verzerrte.⁴⁸ Entsprechende Revisionen des Systems in den Jahren 1970 und 1971 führten zu weiteren Problemen. So sollte 1970 die Genauigkeit des Systems verbessert werden, indem weitere Fragen auf die Dorf-Karten aufgenommen wurden. Diese erweiterte Matrix passte dann allerdings nicht mehr auf eine einzige Lochkarte, was zu einer Verdoppelung der benötigten Menge an Lochkarten und damit zu weiteren potentiellen Fehlern und Ungenauigkeiten führte.⁴⁹ Eine Revision der Verarbeitungsroutinen und damit der logisch-formalen Materialität der Datensätze zog zudem stets eine aufwändige Neuprogrammierung des Systems nach sich, da Abfrage-Codes neu geschrieben werden mussten, wie ein mit diesen Arbeiten befasster Programmierer resigniert zu Protokoll gab:

The HES/70 system has been advertised as a variable, flexible, easily changed system that can accept new questions and models with little or no trouble. This is true only of the system programs involved in generating the data base. The Edit-A, Edit-B, update, scoring, scoring merge and derived processor can be revised very easily by generating a new QTAB and making a few small program modifications. The problem starts with the programs that were written to interrogate the data base. In most cases the programs were written 'quick and dirty' in order to satisfy the customer's need for reports with the idea that the programs would be rewritten at a later date utilizing the QTAB and copy FDs. If the programs are modified to accept the revised data base what will happen when the second revision takes place, modification of modifications. Each time a program is altered some of the original smooth flow is lost. After several modifications even the author has trouble following the program.⁵⁰

Die Computer selbst schließlich, raumfüllende IBM 360/40 Maschinen, waren enorm wartungsintensiv und fielen zudem durch Feuchtigkeit und häufige

47 NARA, MACV/CORDS, MACV/AR, Synopsis of Advisor Remarks, Military Region2, HES Report 30 September 1970, ARC ID 6087884, Box 3.

48 NARA, MACV/CORDS, Reports and Analysis Directorate-Analysis Division, Analysis Report Concepts of HES Reaggregation for 1971, 14 December 1970, ARC ID 6087884, Box4, folder 472-10112-4-3.

49 NARA, MACV/CORDS, MACCORDS-Rad, Revised Punchcard Instructions, 25 February 1971, ARC ID 6087884, Box 3.

50 NARA, MACV/CORDS Analysis Division, Control Data Corporation: HES/70 Re-aggregation – Resource-Implications, 7 September 1970, ARC ID 6087884, Box 5; siehe auch: NARA, MACV/CORDS, Control Data Corporation, Working-paper HES/70 Re-aggregation Proposal, 14 December 1970, ARC ID 6087884, Box 3.

Stromausfälle oft tagelang aus. Ein ehemaliger IBM-Programmierer, Dan Feltham, berichtete von den schwierigen Betriebsbedingungen in den Rechenzentren in Saigon:

The USAID computer rooms had water pipes and air ducts running across the low ceilings and early on there was no raised floor that normally covered computer cables. Anyone walking through the area had to duck, bob, and weave as they picked their way around the machines and stepped over the thick cables. Whenever we had to add some new IBM equipment, figuring out where to put it was like solving a large jigsaw puzzle [...]. At another nearby machine installation, a 407 tabulating machine suffered an extended outage due to rats chewing through the power cables. The CE repairing that machine had to make up an imaginative outage code for his work report. Computers are said to have 'bugs' but they aren't supposed to have rats.⁵¹

Auch die Bandspulen, die das Magnetband mit dem Masterfile bewegten, brannten durch und nicht alle Magnetbänder waren der hohen Belastung und dem tropischen Klima gleich gut gewachsen.⁵²

Die materiellen Metamorphosen der im Vietnamkrieg generierten Datensätze und ihre jeweiligen formal-logischen, semantischen und historischen Konsequenzen ziehen sich weiter durch die Überlieferungsgeschichte dieses außergewöhnlichen Quellenbestandes, die einer eigenen Studie würdig wäre. Eine kartellrechtliche Klage von vier US-amerikanischen Computerherstellern gegen die scheinbare Bevorzugung des marktdominierenden Giganten IBM bei der Vergabe von Regierungsaufträgen im Jahr 1969 führte dazu, dass ein Gericht die Zerstörung sämtlicher elektronischer Dokumente verbot. Während in den leitenden Bürokration in Regierung und Militär wenig Verständnis für den wissenschaftlichen Wert und die Geschichtsmächtigkeit aberhunderter Meter nur von Maschinen lesbarer Magnetbänder vorherrschte (und auch die offizielle Politik des Nationalarchivs vorsah, wichtige Dokumente auszudrucken und die Originale zu vernichten), empfahl eine von der RAND-Corporation durchgeführte Studie deren Bewahrung für künftige historische Forschungen. Dieser Empfehlung wurde letztlich entsprochen und die weit verstreuten Bestände in das US-Nationalarchiv überführt. Ein dort ehemals mit diesen Beständen befasster Archivar, Donald Fisher Harrison, berichtete allerdings 1988, dass die Forschenden Zurückhaltung bei der Nutzung dieser Bestände zeigten.⁵³ Neben dem schlechten Ruf

51 Dan E. Feltham, *When Big Blue Went to War. The History of the IBM Corporation's Mission in Southeast Asia During the Vietnam War*, (Bloomington: Abott Press, 2012), 79.

52 NARA, MACV/CORDS, Inventory Control Center/Directorate of Data Systems, IBM360/40 and Spectr/70 trouble reports [kein Datum], ARC ID 5708184, Box 3.

53 Donald Fisher Harrison, „Computers, Electronic Data, and the Vietnam War,“ in: *Archivaria* 26 (1988), 18–32.

der Daten als unzuverlässig und propagandistisch habe dies vor allem an ihrem Format gelegen: Die Daten waren für eine bestimmte Software konfiguriert und formatiert (logische/formale Materialität), ein DBS namens *National Information Processing System* (NIPS). Um die Nutzung zu erleichtern und auch die bis dato auf Magnetbändern (physische Materialität) gespeicherten Dateien leichter transportabel und übertragbar zu machen, begann das US Nationalarchiv 1978 einige der als am wichtigsten eingeschätzten Dateien in Software-unabhängige Formate zu übertragen (re-born digitals).

Diese zwei Manifestationen digitaler Quellen – zum einen born, zum anderen re-born digitals – unterscheiden sich voneinander in mehr als nur ihrem Format: Der Bitstream der ursprünglichen NIPS-Dateien enthielt in den vorderen Sektoren Codes des DBMS, die zur Dateiverwaltung notwendig waren, wie etwa Timestamps, File Maintenance Statements und andere Kontrollinformationen für die Programmierer. Im Prozess des Transfers in modernere Formate wurden diese Sektionen entfernt und in die Dokumentation ausgelagert, jedoch nur teilweise. So lässt sich nach der Migration zwar der ursprüngliche konzeptuell/semantische Inhalt der Dateien erschließen, nicht aber ihr logisch-formaler.

4 Fazit

Von der Erhebung der Daten in den Dörfern in Weilern Südvietnams bis zu ihrer Übertragung in die Zentralcomputer des Armeekommandos in Saigon und schließlich in die Depots und Lesesäle des Nationalarchivs durchliefen die Informationen des HES mehrere Stationen und wurden dabei oftmals auf neue physische Speichermedien übertragen, die wiederum jeweils spezifische materielle Eigenschaften und spezifische formale Anforderungen hatten. Am Beispiel des HES lässt sich illustrieren, wie durch den Nachvollzug des Lebenszyklus historischer Daten durch ihre mehrfachen konzeptuellen, logischen und physischen Transformationen hindurch die jeweils beteiligten historischen Akteur:innen hervortreten.

Eine gründliche quellenkritische Auseinandersetzung mit dem Datenbanksystem, seinen Revisionen und der jeweiligen Determiniertheit der Daten durch ihre konzeptuelle, logische und physische Materialität offenbart die Unzulänglichkeit und den konstruierten Charakter des Systems, wirft aber zugleich auch interessante Schlaglichter auf die Denkweise und das Erkenntnisinteresse der damaligen Akteur:innen. Angesichts der enormen Bedeutung, welche die Erfahrungen mit dem HES-System für die US-amerikanische empirische Sozialfor-

schung und die Weiterentwicklung von Datenbanktechnologien hatten, erscheint die Auseinandersetzung mit der Praktik des „Databasing“ und seinen ideen- und wissensgeschichtlichen Auswirkungen daher als ein produktiver Zugang zu dieser schwierigen Quelle.

Die Analyse der HES-Datenbank und der damit verbundenen Systeme versprechen durch seine spezifische Überlieferungsgeschichte zugleich weiterreichende Erkenntnisse für die archivalische Bewahrung und geschichtswissenschaftliche Erforschung historischer Datenbanken: Auf der Ebene der Archive spricht die Erfahrung mit diesem System dafür, eine möglichst umfassende Bewahrungspraxis anzustreben, also nicht nur historische Datensätze der Forschung zugänglich zu machen, sondern wenn möglich die vollständige Umgebung etwa in Form von emulierten Systemen inklusive extensiver Dokumentation zu bewahren.⁵⁴ Die Geschichtswissenschaft und hier wiederum insbesondere die Digital History wäre gut beraten, sich in engen Austausch mit Zugängen wie der experimentellen Medienarchäologie zu begeben, um ihre eigentliche Kernkompetenz, die Kritik und Kontextualisierung von historischen Quellen, auch im digitalen Zeitalter nutzbar zu machen.⁵⁵

Bibliographie

Abbate, Janet. *Recoding Gender: Women's Changing Participation in Computing* (Cambridge, Mass.: MIT Press, 2017).

54 Matthew Kirschenbaum et al., „Digital Materiality: Preserving Access to Computers as Complete Environments“, *University of California Digital Library* (2009), <https://escholarship.org/uc/item/7d3465vg>. Zugriff am 24.05.2022; Jens-Martin Loebel, *Lost in Translation: Leistungsfähigkeit, Einsatz und Grenzen von Emulatoren bei der Langzeitbewahrung digitaler multimedialer Objekte am Beispiel von Computerspielen* (Glückstadt: Hülsbusch, 2014); Dirk von Suchodoletz, *Funktionale Langzeitarchivierung digitaler Objekte – Erfolgsbedingungen des Einsatzes von Emulationsstrategien* (Freiburg im Breisgau: Albert-Ludwigs-Universität, 2008), URN: nbn:de:0008-2008070219.

55 Mareike König, „Geschichtswissenschaft und Archive im digitalen Zeitalter: Chancen, Risiken und Nebenwirkungen“, *Der Archivar* 3 (2020): 245–251, https://www.archive.nrw.de/sites/default/files/media/files/Archivar_2020-3-_Internet.pdf. Zugriff am 21.02.2022. Gerben Zaagsma, „On Digital History“, *BMGN – Low Countries Historical Review* 128/4 (2013), doi: 10.18352/bmgn-lchr.9344; Andreas Fickers, „Hands-on! Plädoyer für eine experimentelle Medienarchäologie“, *Technikgeschichte* 82/1 (2015): 67–85; Andreas Fickers, „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?“, *Zeithistorische Forschungen/Studies in Contemporary History* 17/1 (2020): 157–168, doi: 10.14765/zzf.dok-1765.

- Baucom, Erin. „A Brief History of Digital Preservation,“ *Mansfield Library Faculty Publications* 31 (2019), https://scholarworks.umd.edu/cgi/viewcontent.cgi?article=1033&context=ml_pubs. Zugriff am 21.02.2022.
- Belcher, Oliver. „Data Anxieties: Objectivity and Difference in Early Vietnam War Computing.“ In *Algorithmic Life: Calculative Devices in the Age of Big Data*, hg. v. Louise Amoore und Volha Piotukh (New York: Routledge, 2016).
- Belcher, Oliver. „Sensing, Territory, Population: Computation, Embodied Sensors, and Hamlet Control in the Vietnam War,“ *Security Dialogue* 50/5 (Oktober 2019): 416–436.
- Beniger, James R. *The Control Revolution: Technological and Economic Origins of the Information Society* (Cambridge, Mass.: Harvard University Press, 1997).
- Bergin, Thomas J. und Thomas Haigh. „The Commercialization of Database Management Systems, 1969–1983,“ *IEEE Annals of the History of Computing* 31/4 (Oktober 2009): 26–41, doi: 10.1109/MAHC.2009.107.
- Bösch, Frank (Hg.). *Wege in die digitale Gesellschaft: Computernutzung in der Bundesrepublik 1955–1990* (Göttingen: Wallstein Verlag, 2018).
- Bousquet, Antoine J. *The Scientific Way of Warfare: Order and Chaos on the Battlefields of Modernity* (New York: Columbia University Press, 2009), 137–139.
- Brigham, Erwin R. „Pacification Measurement in Vietnam: The Hamlet Evaluation System,“ *Military Review*, May 1970.
- Brügger, Niels. *The Archived Web. Doing History in the Digital Age* (Cambridge, MA: MIT Press, 2018), 21–23.
- Burkhardt, Marcus. *Digitale Datenbanken: eine Medientheorie im Zeitalter von Big Data* (Bielefeld: Transcript, 2015).
- Daddis, Gregory A. *No Sure Victory: Measuring U. S. Army Effectiveness and Progress in the Vietnam War* (New York: Oxford University Press, 2011), 95–106.
- Deitchman, Seymour J. „The ‚Electronic Battlefield‘ in the Vietnam War,“ *The Journal of Military History* 72/3 (2008): 869–887.
- Desrosières, Alain. *The Politics of Large Numbers: A History of Statistical Reasoning* (Cambridge, Mass.: Harvard University Press, 2011).
- Dourish, Paul. *The Stuff of Bits: An Essay on the Materialities of Information* (Cambridge, Mass.: The MIT Press, 2017), 62.
- Edwards, Paul N. *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, Mass: MIT Press, 1996), 3–6.
- Emerson, Lori. *Reading Writing Interfaces: From the Digital to the Bookbound* (Minneapolis: University of Minnesota Press, 2014).
- Ensmenger, Nathan. *The Computer Boys Take Over: Computers, Programmers, and the Politics of Technical Expertise*, (Cambridge, Mass.: MIT Press, 2010).
- Ensmenger, Nathan. „The Environmental History of Computing,“ *Technology and Culture* 59/4S (2018): 7–33.
- Feltham, Dan E. *When Big Blue Went to War. The History of the IBM Corporation’s Mission in Southeast Asia During the Vietnam War* (Bloomington: Abott Press, 2012).
- Fickers, Andreas. „Hands-on! Plädoyer für eine experimentelle Medienarchäologie,“ *Technikgeschichte* 82/1 (2015): 67–85.
- Fickers, Andreas. „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?,“ *Zeithistorische Forschungen/Studies in Contemporary History* 17/1 (2020): 157–168, doi: 10.14765/zzf.dok-1765.

- Fisher Harrison, Donald. „Computers, Electronic Data, and the Vietnam War,“ *Archivaria* 26 (1988), 18–32.
- Föhr, Pascal. *Historische Quellenkritik im Digitalen Zeitalter*, Dissertation, Universität Basel 2018, doi: 10.5451/unibas-006805169.
- Friedewald, Michael und Timo Leimbach, „Computersoftware als digitales Erbe. Probleme aus der Sicht der Technikgeschichte,“ *Neues Erbe. Aspekte, Perspektiven und Konsequenzen der digitalen Überlieferung*, hg. v. Caroline Robertson-von Trotha und Robert Hauser (Karlsruhe: KIT Scientific Publishing, 2011), 201–219.
- Gibson, James William. *The Perfect War: Technowar in Vietnam* (New York: Atlantic Monthly Press, 2000).
- Gilman, Nils. *Mandarins of the Future: Modernization Theory in Cold War America* (Baltimore: Johns Hopkins University Press, 2007).
- Graham, Shawn, Ian Milligan und Scott Weingart, *Exploring Big Historical Data: The Historian's Macroscope* (London: Imperial College Press, 2016).
- Gugerli, David. „Die Welt als Datenbank. Zur Relation von Softwareentwicklung, Abfragetechnik und Deutungsautonomie.“ In *Daten. Nach Feierabend. Zürcher Jahrbuch für Wissensgeschichte* 3, hg. v. David Gugerli et al. (Zürich/Berlin: Diaphanes, 2007), 11–36.
- Gugerli, David und Hannes Mangold, „Betriebssysteme und Computerfahndung: Zur Genese einer digitalen Überwachungskultur,“ *Geschichte und Gesellschaft* 42/1 (2016): 144–174.
- Guldi, Jo und David Armitage. *The History Manifesto* (Cambridge, UK: Cambridge University Press, 2014).
- Haasis, Lucas und Constantin Rieske (Hg.). *Historische Praxeologie: Dimensionen vergangenen Handelns* (Paderborn: Ferdinand Schöningh, 2015).
- Haigh, Thomas, Mark Priestley und Crispin Rope. *ENIAC in Action: Making and Remaking the Modern Computer* (Cambridge MA: MIT Press, 2018).
- Haigh, Thomas. „How Data Got its Base: Information Storage Software in the 1950s and 1960s,“ *IEEE Annals of the History of Computing* 31/ 4 (Oktober 2009): 6–25, doi: 10.1109/MAHC.2009.123.
- Haigh, Thomas. „Introducing the Early Digital,“ in *Exploring the Early Digital*, hg. v. Dems. (Cham: Springer International Publishing, 2019), 1–18, doi: 10.1007/978-3-030-02152-8_1.
- Hänger, Andrea. „Die Mühsal der Zugänglichkeit. Methodische Herausforderungen beim Aufbau virtueller Lesesäle,“ *Zeithistorische Forschungen/Studies in Contemporary History* 17/1 (2020): 179–189, doi: 10.14765/zzf.dok-1763.
- Hicks, Marie. *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*, History of Computing (Cambridge, MA: MIT Press, 2018).
- Homberg, Michael. „Mensch/Mikrochip. Die Globalisierung der Arbeitswelten in der Computerindustrie 1960 bis 2000 – Fragen, Perspektiven, Thesen,“ *Vierteljahrshefte für Zeitgeschichte* 66/2 (2018): 267–293.
- Horowitz, Irving Louis. *The Rise and Fall of Project Camelot: Studies in the Relationship between Social Science and Practical Politics*, Rev. ed. (Cambridge, Mass: MIT Press, 1974).
- Hunt, Richard A. *Pacification: The American Struggle for Vietnam's Hearts and Minds* (Boulder: Westview Press, 1995).
- Kaufmann, Stefan. „Die Wissensformierung der ‚counterinsurgency‘ im Vietnamkrieg,“ *Traverse* 16/3 (2009): 37–52.

- Keitel, Christian. „Digitale Archivalien,“ *Südwestdeutsche Archivalienkunde* <https://www.leo-bw.de/themenmodul/sudwestdeutsche-archivalienkunde/querschnittsartikel/digitale-archivalien>. Zugriff am 20.02.2022.
- Kell, Frank. „LINKAGE – Digitale Gegenwart und Zukunft Historischer Forschung. Die Ziele der Konsortialinitiative 4Memory,“ <https://blog.historikerverband.de/2020/11/11/linkage-digitale-gegenwart-und-zukunft-historischer-forschung/>. Zugriff am 20.02.2022.
- Kiechle, Oliver. „Digitalia fundamentalis – Speicher und Zugriff,“ *Diskrete Werte*, 07.10.2015, <https://digitalia.hypotheses.org/21>. Zugriff am 19.02.2022.
- Kirschenbaum, Matthew, et al. „Digital Materiality: Preserving Access to Computers as Complete Environments,“ *University of California Digital Library* (2009), <https://escholarship.org/uc/item/7d3465vg>. Zugriff am 24.05.2022.
- Kirschenbaum, Matthew G. *Mechanisms: New Media and the Forensic Imagination*, (Cambridge, Mass: MIT Press, 2012).
- Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures & their Consequences* (Los Angeles: SAGE Publications, 2014).
- Klinger, Janeen M. *Social Science and National Security Policy Deterrence, Coercion, and Modernization Theories*. (New York: Palgrave 2019).
- König, Mareike. „Geschichtswissenschaft und Archive im digitalen Zeitalter: Chancen, Risiken und Nebenwirkungen,“ *Der Archivar* 3 (2020): 245–251, https://www.archive.nrw.de/sites/default/files/media/files/Archivar_2020-3-_Internet.pdf. Zugriff am 19.02.2022.
- Krajewski, Markus. *ZettelWirtschaft. Die Geburt der Kartei aus dem Geiste der Bibliothek* (Berlin: Kulturverlag Kadmos, 2002).
- Laak, Dirk van. *Alles im Fluss: Die Lebensadern unserer Gesellschaft. – Geschichte und Zukunft der Infrastruktur* (Frankfurt a. M.: S. Fischer, 2018).
- Latham, Michael E. *Modernization as Ideology: American Social Science and „Nation Building“ in the Kennedy Era* (Chapel Hill: University of North Carolina Press, 2000).
- Lemov, Rebecca M. *Database of Dreams: The Lost Quest to Catalog Humanity* (New Haven: Yale University Press, 2015).
- Lepore, Jill. *IF THEN: How the Simulmatics Corporation Invented the Future* (New York: Liveright, 2020).
- Leslie, Stuart W. *The Cold War and American Science: The Military-Industrial-Academic Complex at MIT and Stanford* (New York: Columbia University Press, 1993).
- Levine, Yasha. *Surveillance Valley – The Secret Military History of the Internet* (New York: Public Affairs, 2018).
- Loebel, Jens-Martin. *Lost in Translation: Leistungsfähigkeit, Einsatz und Grenzen von Emulatoren bei der Langzeitbewahrung digitaler multimedialer Objekte am Beispiel von Computerspielen* (Glückstadt: Hülsbusch, 2014).
- Lütjen, Torben. „Vom ‚Gospel of Efficiency‘ zum ‚War of Ideas‘. Zum Verhältnis von Wissenschaft, Politik und Ideologien in den Vereinigten Staaten“ *Archiv für Sozialgeschichte* 50 (2010): 373–394.
- Maier, Gerald. „Die Bedeutung der Archive für Forschungsdaten in der Geschichtswissenschaft,“ *Blog Historikerverband*, 11.11.2020, <https://blog.historikerverband.de/2020/11/11/die-bedeutung-der-archive-fuer-forschungsdaten-in-der-geschichtswissenschaft/>. Zugriff am 19.02.2022.
- Markley, Robert. „Boundaries: Mathematics, Alienation, and the Metaphysics of Cyberspace,“ in *Virtual Realities and their Discontents*, hg. v. Doms. (Baltimore MD: Johns Hopkins University Press 1996), 55–78.

- Milligan, Ian. *History in the Age of Abundance? How the Web Is Transforming Historical Research* (Montreal: McGill-Queen's University Press, 2019).
- Parks, Lisa und Nicole Starosielski (Hg.), *Signal Traffic: Critical Studies of Media Infrastructures, The Geopolitics of Information* (Urbana: University of Illinois Press, 2015).
- Patel, Kiran Klaus. „Zeitgeschichte im digitalen Zeitalter. Neue und alte Herausforderungen,“ *Vierteljahreshefte für Zeitgeschichte* 59/3 (2011): 331–351.
- Pellow, David N. und Lisa Sun-Hee Park, *The Silicon Valley of Dreams: Environmental Injustice, Immigrant Workers, and the High-Tech Global Economy*, Critical America (New York: New York University Press, 2002).
- Pias, Claus. „Mit dem Vietcong rechnen. Der Feind als Gestalt und Kunde,“ in *Freund, Feind & Verrat. Das politische Feld der Medien*, hg. v. Cornelia Epping-Jäger, Torsten Hahn und Erhard Schüttelpelz (Köln: DuMont, 2004), 157–83.
- Porter, Theodore M. „Quantification and the Accounting Ideal in Science,“ *Social Studies of Science* 22/4 (1992): 633–651.
- Prietl, Bianca und Daniel Houben, „Einführung. Soziologische Perspektiven auf die Datafizierung der Gesellschaft,“ in *Datengesellschaft: Einsichten in die Datafizierung des Sozialen*, hg. v. Dens. (Bielefeld: Transcript, 2018), 7–32.
- Ries, Thorsten und Gábor Palkó, „Born-Digital Archives.“ *International Journal of Digital Humanities* 1,1 (April 2019): 1–11, doi: 10.1007/s42803-019-00011-x.
- Rohde, Joy. *Armed with Expertise: The Militarization of American Social Research during the Cold War* (Ithaca: Cornell University Press, 2013).
- Rosenberg, Daniel. „Daten vor Fakten,“ in *Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie*, hg. v. Ramón Reichert (Bielefeld: transcript-Verlag, 2014), 133–156.
- Rosenzweig, Roy. „Scarcity or Abundance? Preserving the Past in a Digital Era,“ Roy Rosenzweig Center for History and New Media, <http://rrchnm.org/essay/scarcity-or-abundance-preserving-the-past-in-a-digital-era/>. Zugriff am 20.02.2022.
- Rothenberg, Jeff. „Avoiding Technological Quicksand: Sections 1–3“, <https://www.clir.org/pubs/reports/rothenberg/introduction/#longevity>. Zugriff am 19.02.2022.
- Rothenberg, Jeff. „Preserving Authentic Digital Information,“ <https://www.clir.org/pubs/reports/pub92/rothenberg/>. Zugriff am 19.02.2022.
- Salamon, Birgit. „Die archivfachliche Arbeit an den MfS-Geheimdienstunterlagen – Fragen und Herausforderungen,“ *Der Archivar* 55/3 (2002): 203, <https://www.bstu.de/archiv/fachbeitraege/die-archive-der-bundesbeauftragten-fuer-die-stasiunterlagen-bstu-die-archivfachliche-arbeit-an-den-mfs-geheimdienstunterlagen-fragen-und-herausforderungen/>. Zugriff am 19.02.2022.
- Sayes, Edwin. „Actor-Network Theory and Methodology: Just What Does It Mean to Say That Nonhumans Have Agency?,“ *Social Studies of Science* 44/1 (2014): 134–149.
- Schieber, Siegrid, „Digitale Archivierung im Hessischen Landesarchiv,“ *ABI Technik*, 37/2 (2017): 93–102, doi: 10.1515/abitech-2017-0022.
- Shrader, Charles R. *History of Operations Research in the United States Army, V. 2, 1961–1973* (U.S. Department of Defense, U. S. Army Center of Military History, 2008), 37–68.
- Solovey, Mark und Hamilton Cravens. *Cold War Social Science: Knowledge Production, Liberal Democracy, and Human Nature* (Basingstoke GB: Palgrave Macmillan, 2014).
- Stevenson, Charles A. *Warriors and Politicians. US Civil-Military Relations Under Stress* (New York: Routledge, 2006), 153–164.

- Suchodoletz, Dirk von. *Funktionale Langzeitarchivierung digitaler Objekte – Erfolgsbedingungen des Einsatzes von Emulationsstrategien* (Freiburg im Breisgau: Albert-Ludwigs-Universität, 2008), URN: urn:nbn:de:0008-2008070219.
- Thayer, Thomas C. *War Without Fronts* (Naval Institute Press, 2016).
- Thibodeau, Kenneth. „Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years,“ *CLIR and the Library of Congress, The State of Digital preservation: An International Perspective*, April 2002, <https://www.clir.org/pubs/reports/pub107/thibodeau/>. Zugriff am 20.02.2022.
- Tripodi, Christian. „Pacification in Vietnam 1964–72.“ In *The Unknown Enemy: Counterinsurgency and the Illusion of Control*, hg. v. Dems., 112–137. Cambridge: Cambridge University Press, 2020.
- Weinberger, Sharon. *The Imagineers of War: The Untold Story of DARPA, the Pentagon Agency that Changed the World* (New York: Vintage Books, 2018).
- Zaagsma, Gerben. „On Digital History,“ *BMGN – Low Countries Historical Review* 128/ 4 (2013), doi: 10.18352/bmgn-lchr.9344.

Melanie Althage

Potenzziale und Grenzen der Topic-Modellierung mit *Latent Dirichlet Allocation* für die Digital History

Abstract: Topic Modeling is a widespread method in Digital Humanities and Digital History in order to examine the thematic structure of extensive text collections, such as to trace themes and motifs in novels or publication trends in (scientific) journals. Although it represents a large family of algorithms, *Latent Dirichlet Allocation* (LDA), first introduced in 2003, is the de facto standard for these purposes. However, the original goals behind the development of LDA differ significantly from today's application scenarios in the humanities and historical sciences. The article therefore provides an overview of the development context and the assumptions underlying the LDA procedure and, on this basis, discusses the potentials and limitations relevant for historical research. Limitations are inherent in the algorithm, especially with regard to the temporality of historical data. Accordingly, the article pleads for a stronger reception of further developments in the field of Topic Modeling beyond LDA.

Keywords: digital history, topic modeling, latent Dirichlet allocation, methodological criticism, natural language processing, machine learning

Zusammenfassung: Topic Modeling ist in den Digital Humanities und der Digital History eine verbreitete Methode zur Untersuchung der inhaltlich-thematischen Struktur umfangreicher Textkollektionen, etwa um Themen und Motive in Romanen oder Publikationstrends in (Fach-)Zeitschriften nachzuvollziehen. Obgleich es sich um eine große Familie von Algorithmen handelt, stellt die 2003 erstmals vorgestellte *Latent Dirichlet Allocation* (LDA) hierfür den Quasi-Standard dar. Allerdings weichen die ursprünglichen Ziele, die hinter der Entwicklung von LDA standen, von den heutigen Anwendungsszenarien deutlich ab. Der Beitrag zeichnet daher überblickshaft den Entwicklungskontext und die dem Verfahren zugrundeliegenden Annahmen nach und erörtert auf dieser Basis die für geschichtswissenschaftliche Fragestellungen relevanten Potenziale und Grenzen. Insbesondere hinsichtlich der Temporalität historischer Daten sind dem Algorithmus Limitierungen inhärent. Der Beitrag plädiert entsprechend dafür, die Weiterentwicklungen im Bereich Topic Modeling über LDA hinaus stärker zu rezipieren.

Schlagwörter: Digital History, Topic Modeling, Latent Dirichlet Allocation, Methodenkritik, Natural Language Processing, Machine Learning

1 Einleitung

Die informationstechnologischen Entwicklungen seit dem 20. Jahrhundert sowie der sogenannte *computational turn* veränderten die Art und Weise der geistes- und geschichtswissenschaftlichen Forschungstätigkeit grundlegend. Nicht nur stehen immer mehr digitalisierte und genuin digitale Quellen nahezu zeit- und ortsunabhängig zur Verfügung, auch das Set an Werkzeugen und Methoden wandelt und erweitert sich sukzessive. Eine solche Methode ist Topic Modeling. Sie zielt darauf, statistisch signifikante Sprachgebrauchsmuster in umfangreichen Textkorpora zu identifizieren und Dokumente zu gruppieren, die diese Muster miteinander teilen.¹ Seit rund 15 Jahren wird Topic Modeling in den digitalen Geisteswissenschaften eingesetzt, um beispielsweise die Entwicklung von thematischen Trends in fachwissenschaftlichen Publikationsorganen² oder auch den Status der Digital Humanities als Disziplin³ zu untersuchen. Auch in der Digital History findet die Methode vermehrt Anwendung, etwa zur Identifikation und Analyse von Diskursen oder Publikationstrends in umfangreichen historischen Quellenkorpora.⁴

Wird ein genauerer Blick auf die konkret eingesetzten Verfahren geworfen, zeigt sich, dass die erstmals 2003 vorgestellte und seither weiterentwickelte La-

¹ Für eine allgemeine Einführung siehe David M. Blei, „Probabilistic Topic Models: Surveying a Suite of Algorithms that Offer a Solution to Managing Large Document Archives,“ *Communications of the ACM* 55/4 (2012): 77–84, doi: 10.1145/2133806.2133826.

² Exemplarisch: David Mimno, „Computational Historiography: Data Mining in a Century of Classics Journals,“ *ACM Journal on Computing and Cultural Heritage* 5/1 (2012): 3:1–3:19, doi: 10.1145/2160165.2160168.

³ Jüngst: Jan Luhmann und Manuel Burghardt, „Digital Humanities – A Discipline in its Own Right? An Analysis of the Role and Position of Digital Humanities in the Academic Landscape,“ *Journal of the Association for Information Science and Technology* (2021): 1–24, doi: 10.1002/asi.24533.

⁴ Exemplarisch: Anselm Küsters und Elisa G. Moreno, „Mining ‚Die Zeit‘: A Structural Topic Model Analysis of Spain’s Image in the German Media 1946–2009,“ *UPIER Working Paper* 16 (2019): 1–38; Estelle Bunout und Milan van Lange, „Nibbling at Text: Identifying Discourses on Europe in a Large Collection of Historical Newspapers Using Topic Modelling“ (21.06.2019), <https://www.c2dh.uni.lu/thinking/nibbling-text-identifying-discourses-europe-large-collection-historical-newspapers-using>. Zugriff am 06.12.2021.

tent Dirichlet Allocation (LDA)⁵ am häufigsten eingesetzt wird. Das liegt unter anderem daran, dass der Algorithmus durch gebrauchsfertige und gut dokumentierte Tools wie Mallet⁶ oder (informelle) Programmierutorials und Softwarebibliotheken hochverfügbar und erschlossen ist. Zwar finden sich im erweiterten Feld der Digital Humanities Untersuchungen, die auf die Herausforderungen hinweisen, die mit diesem hochgradig parametrisierenden Verfahren einhergehen und entsprechende Workflows für die produktive Integration in den Forschungsprozess vorschlagen,⁷ analoge Studien aus dezidiert geschichtswissenschaftlicher Perspektive fehlen dagegen weitestgehend oder greifen lediglich einzelne Aspekte kursorisch heraus.⁸

Zur Schließung dieser Lücke möchte der vorliegende Artikel beitragen und die Potenziale und Grenzen der Adaptierbarkeit von Topic Modeling mit LDA aus historischer Perspektive beleuchten. Eine Auseinandersetzung mit den methodologisch-epistemologischen Auswirkungen sowie eine Sensibilisierung für die vielfältigen Konfigurationsmöglichkeiten des Algorithmus ist gerade dann besonders wichtig, wenn auf Standardeinstellungen oder gebrauchsfertige Werkzeuge zurückgegriffen wird. Insbesondere sind die Limitierungen für historische Fragestellungen zu diskutieren, die vor allem darin bestehen, dass die Topic-Modellierung mit LDA die Temporalität der Daten nicht berücksichtigt. Zunächst wird ein knapper Blick auf den Entwicklungskontext und die dem Verfahren zugrundeliegenden Annahmen geworfen. Anschließend werden die Eigenschaften der Topic-Modellierung mit LDA diskutiert und die Frage aufgeworfen, was Topics eigentlich sind und was die Arbeit mit Topics für die digitale historische Forschungsarbeit bedeutet. Als Anwendungsbeispiel hierfür dienen exemplarische Befunde aus dem Topic Modeling der zwischen 1996 und Juni

5 David M. Blei, Andrew Y. Ng und Michael I. Jordan, „Latent Dirichlet Allocation,“ *Journal of Machine Learning Research* 3 (2003): 993–1022.

6 Andrew K. McCallum, „MALLET: A Machine Learning for Language Toolkit,“ 2002, <http://mallet.cs.umass.edu>. Zugriff am 06.12.2021.

7 Siehe unter anderem Inna Uglanova und Evelyn Gius, „The Order of Things: A Study on Topic Modeling of Literary Texts“, *CEUR Workshop Proceedings* 2723 (2020): 57–76, <http://ceur-ws.org/Vol-2723/long7.pdf>. Zugriff am 06.12.2021.

8 Aufschlüsse aus geschichtswissenschaftlicher Perspektive bieten: Peter Andorfer, „Turing Test für das Topic Modeling: Von Menschen und Maschinen erstellte inhaltliche Analysen der Korrespondenz von Leo von Thun-Hohenstein im Vergleich,“ *Zeitschrift für digitale Geisteswissenschaften* 2 (2017), doi: 10.17175/2017_002; Martin Fechner und Andreas Weiß, „Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts,“ *Zeitschrift für digitale Geisteswissenschaften* 2 (2017), doi: 10.17175/2017_005; Lino Wehrheim, „Economic History Goes Digital: Topic Modeling the Journal of Economic History,“ *Cliometrica* 13 (2019): 83–125, hier: 89–96, doi: 10.1007/s11698-018-0171-7.

2019 auf dem Fachkommunikationsportal H-Soz-Kult⁹ publizierten Tagungsberichte.¹⁰

2 Topic Modeling

Wer sich erstmals mit der Methode Topic Modeling auseinandersetzt, wird mit hoher Wahrscheinlichkeit zunächst über LDA stolpern; in den Digital Humanities wird es häufig synonym für Topic Modeling gebraucht. Allerdings handelt es sich bei der Methode eigentlich um eine große Familie von Algorithmen.¹¹ Sie werden seit den 2000er-Jahren in unterschiedlichsten Kontexten eingesetzt, um umfangreiche Textsammlungen über das sogenannte *Clustern* von kookkurrenten Termen automatisiert zu beschreiben und zu klassifizieren.

Die computergestützten Anfänge reichen zurück bis etwa in die 1980er Jahre und entstammen dem Forschungsfeld der generativen probabilistischen Modelle. Hierbei wird angenommen, dass beobachtbare Variablen mit unbeobachteten *latenten*, also nicht unmittelbar zu erfassenden Parametern auf eine bestimmte Art und Weise interagieren, was im Grunde für die Generierung des Datensatzes verantwortlich ist.¹² Folgen wir diesem methodischen Ansatz, gehen wir gewissermaßen davon aus, dass Autor:innen zur Textproduktion auf ein Set von Themen zurückgreifen, die sich über ein spezifisches Vokabular präsentieren lassen. Aus diesen themenspezifischen Vokabularen bedienen sich Autor:innen dann zu unterschiedlichen Anteilen. Mit der Topic-Modellierung sollen nun die verborgenen „Themenfelder“, die initial als Bausteine der

9 H-Soz-Kult. Fachinformation und Kommunikation für die Geschichtswissenschaften, <https://www.hsozkult.de/>.

10 Die Modellierungsergebnisse stammen aus dem laufenden Promotionsvorhaben der Autorin, Arbeitstitel: „Mining the Historian’s Web – Methodenkritische Reflexion quantitativer Verfahren zur Analyse genuin digitaler Quellen am Beispiel der historischen Fachkommunikation“.

11 Dazu zählen unter anderem die *Non-negative Matrix Factorization*: Daniel D. Lee und H. Sebastian Seung, „Learning the Parts of Objects by Non-negative Matrix Factorization,“ *Nature* 401 (1999): 788–791, doi: 10.1038/44565, oder auch das *Structural Topic Modeling*: Margaret E. Roberts et al., „The Structural Topic Model and Applied Social Science,“ *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation* (2013), und andere mehr.

12 Blei, „Probabilistic Topic Models,“ 79–80; Ike Vayansky und Sathish A. P. Kumar, „A Review of Topic Modeling Methods,“ *Information Systems* 94 (2020): 1–15, hier: 2, doi: 10.1016/j.is.2020.101582.

Textproduktion dienen, aus dem fertigen Text rekonstruiert werden.¹³ Es ist dabei ein möglicher Ansatz unter vielen, der dazu dient, die Komplexität umfangreicher Datenmengen wie Textkollektionen zu reduzieren, um die potenziell relevanten Informationen auszuheben.¹⁴

Ursprünglich ging es darum, die Limitierungen von Vorgängerverfahren wie *Term Frequency – Inverse Document Frequency*, *Latent Semantic Analysis/Indexing* oder *probabilistic Latent Semantic Analysis/Indexing* für die Beschreibung und Exploration unbekannter Datenmengen zu überwinden. Mit der *Non-negative Matrix Factorization* 1999 und LDA 2003 wurden dann Verfahren vorgestellt, die es über die Identifikation statistisch signifikanter Wortgebrauchsmuster ermöglichten, die Dokumente einer Kollektion mittels eines unüberwachten Modellierungsverfahrens zusammenzufassen.¹⁵ Die heutigen komplexen Anwendungsszenarien sowie die Besonderheiten historischer Fragestellungen spielten im Entwicklungsprozess allerdings keine Rolle und so scheinen die Erkenntnisziele von Geistes- und Geschichtswissenschaftler:innen sich insofern von denen der Entwickler:innen entfernt zu haben, als es nicht allein um die Klassifikation im Sinne optimierter Suchheuristiken, sondern auch um die Prägung neuer Perspektiven auf die Quellen geht.¹⁶

Grundsätzlich ist die Integration von Methoden wie Topic Modeling in den Werkzeugkasten der Historiker:innen vielversprechend. Ein zentraler Vorteil, der generell für computationale Methoden gilt, besteht in der systematischen Verarbeitung der Daten sowie der Skalierbarkeit des Forschungsdesigns. Während in „analogen“ Forschungskontexten für gewöhnlich Stichproben und Fallbeispiele ausgewählt und exemplarisch untersucht werden, lassen sich digitale Methoden bei entsprechender Rechenleistung auf beliebig große Quellenkorpora und Untersuchungszeiträume anwenden. Für Topic Modeling kommt hinzu, dass es sich um ein sogenanntes „unüberwachtes“ Verfahren handelt, das heißt, die Modellierungsergebnisse speisen sich allein aus den Inputdaten und beruhen nicht auf zuvor mit bestimmten Annahmen und Theorien definierten

13 Blei, „Probabilistic Topic Models,“ 78, Anm. a auf 78; Blei, „Topic Modeling,“ Abs. 8, 12–13.

14 Blei, Ng und Jordan, „Latent Dirichlet Allocation,“ 993–994; Wehrheim, „Economic History,“ 93.

15 Siehe überblickshaft zu dieser Entwicklung Blei, Ng und Jordan, „Latent Dirichlet Allocation,“ 993–995, 1000–1002; Vayansky und Kumar, „A Review“, 2; Rubayyi Alghamdi und Khalid Alfalqi, „A Survey of Topic Modeling in Text Mining,“ *International Journal of Advanced Computer Science and Applications* 6/1 (2015): 147–153, hier: 147–149, doi: 10.14569/IJACSA.2015.060121.

16 Vgl. Benjamin M. Schmidt, „Words Alone: Dismantling Topic Models in the Humanities,“ *Journal of Digital Humanities* 2/1 (2012), <http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/>, Abs. 7, 9. Zugriff am 06.12.2021.

Kategorien.¹⁷ Das eröffnet die Möglichkeit, umfangreiche historische Quellenkorpora heuristisch und explorativ zu erschließen und dabei potenziell unbekannte und verborgene thematische Muster zu entdecken, die Quellen referenzieren, die sonst möglicherweise nicht berücksichtigt worden wären.¹⁸ Ein weiterer Vorteil besteht in der Sprachunabhängigkeit der Verfahren, wo andere Methoden spezieller Sprachmodelle oder Lexika bedürften. Topic Modeling kann prinzipiell auf beliebige Quellenbestände angewendet werden, wenngleich die besonderen Eigenschaften bestimmter Sprachen, Textgattungen oder Sprachstufen mit ganz eigenen Herausforderungen einhergehen.¹⁹

3 Latent Dirichlet Allocation

Wie eingangs angeführt, handelt es sich bei LDA bis dato um einen der populärsten Topic-Modeling-Ansätze. Als probabilistischem Modell liegt ihm die Annahme zugrunde, dass sich Dokumente aus einer *latenten*, also verborgenen thematischen Struktur zusammensetzen, die sich aus der jeweils spezifischen Kombination von Wörtern ablesen lässt. Da diese Kombination verborgen ist, muss sie aus der Häufigkeitsverteilung und Kookkurrenz der Wörter abgeleitet werden. Für diese Aufgabe wird die nach dem deutschen Mathematiker Peter Gustav Lejeune Dirichlet (1805–1859) benannte *Dirichlet*-Wahrscheinlichkeitsverteilung eingesetzt.²⁰ Dabei ist LDA von zwei Annahmen motiviert: 1) Es gibt ein festes Set von Wortverwendungsmustern, das heißt Topics, die in den Dokumenten statistisch gehäuft gemeinsam vorkommen, und 2) jedes Dokument der Kollektion weist diese Topics zu unterschiedlichen Anteilen auf.²¹ Dementsprechend werden die Quellen mittels LDA zum einen als Set von Topics mitsamt ihrer proportionalen Anteile an den Dokumenten repräsentiert, was üblicher-

17 Vgl. Andorfer, „Turing Test,“ Abs. 62; Bunout und van Lange, „Nibbling at Text,“ Abs. 11, 13; Wehrheim, „Economic History,“ 94.

18 Vgl. zu dieser Einschätzung Wehrheim, „Economic History,“ 94–95.

19 Vgl. *ibid.*, 92; für ein Verfahren, das beispielsweise den Herausforderungen der Mehrsprachigkeit begegnet, siehe David Mimno et al., „Polylingual Topic Models,“ *EMNLP 2009* (2009), 880–889.

20 Allen B. Riddell, „How to Read 22198 Journal Articles: Studying the History of German Studies with Topic Models,“ in *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*, hg. v. Matt Erlin und Lynne Tatlock (New York: Boydell & Brewer, 2014), 91–114, hier: 100. Die statistisch-mathematischen Grundlagen und Funktionsweisen des Algorithmus werden ausführlich beschrieben in Blei, Ng und Jordan, „Latent Dirichlet Allocation“; Blei, „Probabilistic Topic Models,“ 78–82.

21 Blei, „Topic Modeling,“ Abs. 5.

weise als sogenannte *Document-Term-Matrix* formalisiert wird, also als Tabelle, die je Dokument die Topic-Gewichtung dokumentiert. Zum anderen wird als *Output* eine nach Relevanz geordnete Auflistung des Vokabulars je Topic generiert.

Ermittelt werden diese Mixturen über einen randomisierten, generativen Prozess, der von einer Reihe von nutzerseitigen Einstellungen abhängt: Nach einer im Vorhinein festzulegenden Anzahl an Topics (zum Beispiel 80) weist der Algorithmus in einem ersten Schritt einerseits die einzelnen Wörter der Dokumente zufällig den Topics zu und andererseits jedem Dokument eine zufällige Topic-Mixtur. Es liegt nun in der Hand der Nutzer:innen festzulegen, in wie vielen Wiederholungen (*Iterationen*) durch den Vergleich der dokumentenbasierten Kookkurrenzen die gebildeten Wortcluster algorithmisch optimiert werden sollen. Mit jeder weiteren Iteration wird die Wort-zu-Topic- und Topic-zu-Dokument-Verteilung verfeinert, wobei irgendwann ein praktikables Maximum erreicht wird, das es auszutariieren gilt.²²

Tab. 1: Exemplarischer Vergleich von zwei LDA-Implementierungen auf Basis der zwischen 1996 und Juni 2019 auf H-Soz-Kult publizierten Tagungsberichte.

Mallet-LDA	Gensim-LDA
digital projekt datum digitale edition digitali- sierung online möglichkeit internet historisch stellen information neue datenbank tech- nisch portal neu nutzen digital_humanities datenbanken	digital digitalisierung projekt edition datum technisch digital_edition online herausforder- ung standard portal text tools dps format soft- ware stellen möglichkeit werkzeug plattform
erinnerung geschichte vergangenheit erinne- rungskultur historisch gedächtnis erinnern holocaust umgang gedenken opfern ereignis zeitzeugen national auseinandersetzung nar- rativ erfahrung jahr öffentlich denkmal	gedenkstätte Erinnerung Ort Lager Opfer holo- caust Vortrag Denkmal Gedenken Geschichte jüdisch zeigen erinnern stellen Zwangsarbeit täter ehemalg häftling workshop historisch

Dabei ist LDA allerdings nicht gleich LDA. Die Berechnung aller möglichen Topic-Konstellationen im Verhältnis zum Gesamtkorpus ist mathematisch komplex und daher nur durch eine Annäherung zu erreichen.²³ Diesbezüglich lassen sich die Implementierungen von LDA in zwei Kategorien einteilen: 1) solche,

²² Vgl. Blei, „Probabilistic Topic Models,“ 78–82; Wehrheim, „Economic History,“ 86–89.

²³ Blei, „Probabilistic Topic Models,“ 81.

die *Variational-Bayes-Inference*-Ansätze²⁴ und 2) solche, die *Gibbs sampling*²⁵ nutzen.²⁶ Während hier ein ausführlicher Vergleich der statistischen Grundlagen zu weit führen würde,²⁷ ist die Erwähnung wichtig, um darauf hinzuweisen, dass bei der Wahl des Modellierungsverfahrens nicht nur die Frage zu berücksichtigen ist: „LDA – ja oder nein?“ sondern auch: „Welche konkrete Ausführung von LDA?“ In geisteswissenschaftlichen Anwendungskontexten ist etwa das Softwarepaket Mallet etabliert. Es bietet bei ausreichender Datenmenge und -qualität robuste und semantisch interpretierbare Ergebnisse und wurde auch für das hier verwendete Fallbeispiel verwendet (siehe die exemplarischen Topics in Tabelle 3 im Anhang²⁸). Die Pythonbibliothek Gensim bietet demgegenüber einen performanteren Algorithmus für sehr große Datenmengen.²⁹ Bei identischer Datenbasis unterscheiden sich die resultierenden Modelle deutlich. Grund dafür ist, dass Gensim *Variational Bayes* zur Ermittlung der Topic-Dokument- beziehungsweise Wort-Topic-Distributionen nutzt und Mallet demgegenüber *Gibbs sampling*. Zwar ermitteln die Algorithmen zum Teil einander ähnelnde Wortcluster (siehe Tabelle 1), allerdings würde ihre spezifische Wort-Zusammenstellung abweichende Schwerpunkte der mit den Topics referenzierten Dokumente nahelegen. Derzeit fehlen Studien, die sich mit den epistemologi-

24 Nach Blei, Ng und Jordan, „Latent Dirichlet Allocation,“ insbes. 1003–1007.

25 Nach Thomas L. Griffiths und Mark Steyvers, „Finding Scientific Topics,“ *Proceedings of the National Academy of Sciences of the United States of America* 101, 1 (2004): 5228–5235, hier: insbes. 5229–5230, doi: 10.1073/pnas.0307752101.

26 Blei, „Probabilistic Topic Models,“ 81; Jordan Boyd-Graber, David Mimno und David Newman, „Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements,“ in *Handbook of Mixed Membership Models and Their Applications*, hg. v. Edoardo M. Airoldi et al. (Boca Raton, Florida: CRC Press, 2014), 225–254, hier: 231.

27 Siehe etwa Boyd-Graber, Mimno und Newman, „Care and Feeding,“ 231–233; sowie den systematischen Vergleich in Christine K. Mulunda et al., „Review of Trends in Topic Modeling Techniques, Tools, Inference Algorithms and Applications,“ *ISCM* (2018): 28–37, hier: 33–34, doi: 10.1109/ISCM.2018.8703231.

28 Das in Auswahl in Tabelle 3 repräsentierte Topic-Modell besteht aus 80 Topics, die in 6 000 Iterationen und mit einer Hyperparameter-Optimierung nach festgesetzten Intervallen gebildet wurden. Die Daten wurden mit freundlicher Genehmigung von H-Soz-Kult über ein Python-Skript aus der HTML-Struktur des Webauftritts ausgelesen und als CSV-Datei tabellarisch strukturiert abgespeichert. Insgesamt wurden für den Zeitraum 1996 bis Juni 2019 8 040 Tagungsberichte heruntergeladen, davon bilden 6 740 Tagungsberichte das mehrheitlich deutschsprachige Analysekorpus. Die Volltexte des Teilkorpus bestehen aus 12 062 497 Token (durchschnittlich 1 789 Token pro Dokument).

29 Zur Dokumentation: Gensim, models.lidamodell – Latent Dirichlet Allocation, <https://radimrehurek.com/gensim/models/lidamodell.html>. Zugriff am 06.12.2021. Zu Gensim siehe auch: Radim Řehůřek und Petr Sojka, „Software Framework for Topic Modelling with Large Corpora,“ *Proceedings of the LREC Workshop on New Challenges for NLP Frameworks* (2010): 45–50.

schen und methodologischen Konsequenzen dieser Abweichungen und in der Folge unterschiedlichen Forschungsergebnissen auseinandersetzen.

Unabhängig von der konkreten Implementierung handelt es sich bei LDA zudem anders als bisweilen zu lesen³⁰ um ein hochgradig parametrisierendes Verfahren. Auch wenn davon gesprochen wird, Topic-Modeling-Verfahren seien im Vergleich zu manuellen Kategorisierungen beispielsweise „objektiv“, da das Modellierungsergebnis allein auf Basis der eingespeisten Daten gebildet wird,³¹ darf das nicht darüber hinwegtäuschen, dass nicht zuletzt die konkrete Anwendung eingebettet ist in einen komplexen Forschungsprozess. Dieser wird vom jeweiligen Wissenshorizont der Forschenden genauso geprägt wie von den entsprechenden forschungstheoretischen Annahmen des jeweiligen Vorhabens, was sich auf die Aufbereitung der Daten und die Wahl des finalen Modells auswirkt. Insbesondere letzteres ist mit einem zeitintensiven Evaluationsprozess verbunden, der die Ergebnisse unterschiedlicher Parametereinstellungen berücksichtigen und reflektieren muss und oftmals darauf zielt, das bestmöglich interpretierbare Modell zu finden.³² Insofern Topic Modeling also von Menschen eingesetzt wird, die mit der Anwendung konkrete Forschungsfragen verbinden und die Daten entsprechend aufbereiten, ist auch die Topic-Modellierung nicht neutral, wohl aber systematisch.

Während LDA in diesem Sinne zwar grundsätzlich auf unvorbereitete Textdaten angewendet werden kann, wird die Qualität des Ergebnisses durch einige strukturierende Vorverarbeitungsschritte (*Preprocessing*) aus dem Bereich des *Natural Language Processing* deutlich verbessert. Wie groß der Einfluss dieser Vorbereitung auf die Textbasis ist, zeigt die Gegenüberstellung der Wortanzahl des hier verwendeten Fallbeispiels vor- und nachher: Zu Beginn enthält das aus 6 740 Tagungsberichten bestehende Textkorpus etwa 12 062 497 Token, das sind durchschnittlich 1 789 Token pro Dokument. Am Ende des Preprocessing waren es noch 5 751 616 Token zu durchschnittlich 853 pro Tagungsbericht, die zur Modellierung der Topics herangezogen wurden.³³

³⁰ Siehe Andorfer, „Turing Test“, Abs. 14.

³¹ So etwa in Wehrheim, „Economic History“, 94.

³² Anna Shadrova macht im Kontext der vielfältigen Parameter auf die Gefahr des *Confirmation Bias* beziehungsweise von Apophänie aufmerksam, siehe dies., „Topic Models Do Not Model Topics: Epistemological Remarks and Steps Towards Best Practices“, *Journal of Data Mining and Digital Humanities* (2021), doi: 10.46298/jdmdh.7595, 16–17.

³³ Das Preprocessing der mehrheitlich deutschsprachigen Tagungsberichte umfasste: 1) die Tokenisierung und Entfernung aller Satz- und numerischen Zeichen (Der Ausschluss numerischer Werte rechtfertigt sich durch den Umstand, dass in den Tagungsberichten Zahlen häufig in Bezug auf Veranstaltungen oder Anmerkungen genannt werden, sie also mehrheitlich keinen Beitrag zur thematischen Struktur der Beiträge leisten), 2) Lemmatisierung (inkl. manueller

Die Ermittlung der adäquaten Anzahl zu generierender Topics ist eine der offenkundigsten Herausforderungen, da sie der inhaltlichen Diversität des entsprechend aufbereiteten Korpus Rechnung tragen muss.³⁴ LDA bietet keine Automatisierungsfunktion zur Ableitung der optimalen Topic-Anzahl aus den Daten.³⁵ Es gibt verschiedene mathematische Messgrößen wie die *Topic Coherence*, die den Evaluationsprozess durch automatisierte Vergleichswerte zur Bewertung der semantischen Qualität der Wortcluster unterstützen können.³⁶ Gleichwohl bleibt es letztlich die Interpretationsleistung der Forschenden, die Granularität des Topic-Modells anhand der Fragestellung festzulegen.

Zu wenig Beachtung finden in der geistes- und geschichtswissenschaftlichen Forschungsliteratur aktuell noch die sogenannten Hyperparameter *Alpha* und *Beta*, die Einfluss auf die Topic-Dokument- beziehungsweise Wort-Topic-Verteilungen haben.³⁷ Die Konfiguration der Hyperparameter, die bei der Nutzung (gebrauchsfertiger) Werkzeuge oft durch Default-Werte abgenommen wird, beeinflusst das Verteilungsprofil des Modells deutlich: Grundsätzlich lässt sich zwischen symmetrischen und asymmetrischen Hyperparametern unterscheiden. Bei ersterem wird davon ausgegangen, dass die das distributionelle Profil beeinflussenden Werte gleichverteilt sind, letzterer eröffnet dagegen die Möglichkeit, dass manche Topics bezogen auf das Gesamtkorpus stärker ausgeprägt sein können als andere. Analoges gilt für die Gewichtung der Terme je Topic.³⁸

Anpassung), 3) Part-of-Speech-Tagging (nur Adjektive, Verben und Nomen wurden dem Modell zugrunde gelegt), 4) Lowercasing, 5) Entfernung korpuspezifischer Terme, 6) Bildung von Bi- und Trigrammen.

34 Eine niedrige Zahl von Topics führt zu tendenziell heterogenen Clustern, demgegenüber liefert eine sehr hohe Anzahl zum Teil redundante und spezifische Wortgruppen, die unter Umständen eine sinnvolle Interpretation erschweren, siehe dazu Christof Schöch, „Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama“, *Digital Humanities Quarterly* 11/2 (2017), <http://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html>, Abs. 20, Anm. 7. Zugriff am 06.12.2021; Fechner und Weiß, „Topic Modeling“, Abs. 26.

35 Für einen algorithmischen Ansatz, die Topics aus den Daten abzuleiten, siehe Yee W. Teh, Michael I. Jordan und David M. Blei, „Hierarchical Dirichlet Processes“, *Journal of the American Statistical Association* 101 (2006): 1566–1581, doi: 10.1198/016214506000000302.

36 Siehe Daniel Maier et al., „Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology“, *Communication Methods and Measures* 12, 2–3 (2018): 93–118, hier: 102–103, doi: 10.1080/19312458.2018.1430754; Uglanova und Gius, „Order of Things“, insb. 60.

37 Siehe zur Relevanz Hanna M. Wallach, David Mimno und Andrew McCallum, „Rethinking LDA: Why Priors Matter“, *Advances in Neural Information Processing Systems* 22 (NIPS 2009) (2009): 1973–1981.

38 Vgl. Vayansky und Kumar, „A Review“, 3; Wehrheim, „Economic History“, 89; Schöch, „Topic Modeling“, Abs. 20; Christof Schöch, „Topic Modeling with MALLET: Hyperparameter

Da LDA ein randomisiertes Verfahren ist, wird ihm gelegentlich eine mangelnde Reproduzierbarkeit der Forschungsergebnisse vorgeworfen.³⁹ Eigentlich würde jeder Modellierungsdurchlauf bei einem zufälligen Startpunkt im Korpus mit der Bildung der Wort-Topic- und Topic-Dokumenten-Verteilungen beginnen. Dadurch weichen die Ergebnisse verschiedener Durchläufe auch bei gleichen Einstellungen immer voneinander ab. Während Zufallsfaktoren nicht gänzlich ausgeschlossen werden können, ist es doch in einigen Implementierungen wie Mallet möglich, die zufällige Initialisierung durch die Festlegung eines konkreten Startwerts, dem sogenannten *random seed*, zu steuern.⁴⁰ Damit können bei identischen Parametern die Modellierungsergebnisse exakt repliziert werden. Allerdings ist auch diese Einstellung nicht konsequenzenlos. Schon die Abweichung um eine Position, von 2764 auf 2765, ruft voneinander abweichende Modelle hervor (siehe Tabelle 2), wobei festzustellen ist, dass die Wortcluster einander ähneln und die Unterschiede sich vor allem auf Ebene des Wortrangs ergeben.⁴¹

Tab. 2: Exemplarischer Vergleich von zwei Topics nach Modellierungsdurchläufen mit zwei verschiedenen *random seeds*.

Seed	2764	2765
Topic	erinnerung geschichte vergangenheit erinnerungskultur historisch gedächtnis erinnern holocaust umgang gedenken opfern ereignis zeitzeugen national auseinander- setzung narrativ erfahrung jahr öffentlich denkmal	erinnerung geschichte vergangenheit historisch erinnerungskultur holo- caust erinnern opfern umgang gedächtnis erfahrung gedenkstätte gedenken thema ereignis öffentlich jahr auseinander- setzung ort zeitzeugen

Neben den Konfigurationsoptionen limitieren einige im Algorithmus inhärente Eigenschaften den Aussagewert der Modellierungsergebnisse. Eine das Ergebnis stark beeinflussende Annahme ist beispielsweise die sogenannte *bag-of-words*-Annahme, nach der die Reihenfolge der lexikalischen Einheiten für die Ermittlung der semantischen Strukturen keine Rolle spielt.⁴² Auch berücksichtigt der

Optimization,“ *The Dragonfly’s Gaze. Computational Analysis of Literary Texts* (14.11.2016), <https://dragonfly.hypotheses.org/1051>. Zugriff am 06.12.2021.

39 So etwa in Bunout und Lange, „Nibbling at Text,“ Abs. 16–18.

40 Der in dieser Arbeit verwendete Seed „2764“ wurde durch einen Zufallsgenerator ermittelt, um eine subjektive Beeinflussung auszuschließen, was selbst wiederum eine methodisch relevante Entscheidung ist.

41 Zu einer vergleichbaren Erkenntnis kommt Schöch, „Topic Modeling,“ Abs. 14.

42 Blei, Ng und Jordan, „Latent Dirichlet Allocation,“ 994.

Algorithmus nicht den Aufbau einer Dokumentensammlung. Die jeweiligen Texte sind für den Algorithmus austauschbar.⁴³ Dies ist insbesondere bei historisch gewachsenen Korpora ein Manko. Denn das Topic-Modell mit LDA beinhaltet keine zeitliche Dimension oder eine äquivalente natürliche Ordnung; es ist statisch.⁴⁴ In der Praxis werden die erzeugten Modelle daher üblicherweise – so auch hier – um weitere Metadaten wie das Erscheinungsdatum eines Textes ergänzt.⁴⁵ Dadurch lässt sich zwar ermitteln, wie die Wahrscheinlichkeitswerte im zeitlichen Verlauf variieren, nicht aber wie sich dies auf der Ebene der lexikalischen Einheiten widerspiegelt. Zudem wurde dieser Algorithmus anhand normierter wissenschaftlicher Texte entwickelt und getestet.⁴⁶ Bei literarischen oder historischen Quellen, die oftmals einen heterogenen Sprachgebrauch und eine nicht-normierte Rechtschreibung und Stilistik aufweisen, kann sich dieser Umstand entsprechend im Modellierungsergebnis abzeichnen und besondere Vorverarbeitungsschritte notwendig machen.⁴⁷

4 Topics sind keine Themen

Abschließend bleibt noch die Frage zu diskutieren, wie der Begriff „Topic“ im Kontext von LDA verstanden wird und was sein Wesen für die historische Erkenntnisproduktion bedeutet. Die Anwendung von Topic Modeling in geistes- und geschichtswissenschaftlichen Forschungskontexten ist für gewöhnlich durch zwei Annahmen motiviert: erstens, dass die Topics kohärent sind und folglich die einzelnen, einem Topic zugewiesenen Begriffe etwas gemeinsam haben; zweitens, dass ein in zwei Dokumenten in etwa gleichwertig vorkommendes Topic in beiden Kontexten dieselbe Bedeutung trägt.⁴⁸ Sind diese Annahmen zutreffend und was genau ist ein Topic? Es ist statistisch-mathematisch

⁴³ Ibid.; Blei, „Probabilistic Topic Models,“ 82.

⁴⁴ Vgl. Riddell, „How to Read,“ 108.

⁴⁵ Exemplarisch *ibid.*, 102–107; Wehrheim, „Economic History,“ 100; Jeri Wieringa setzte sich in einer Blogserie mit den gebräuchlichen Berechnungsweisen zur Topic-Modellierung anhand der Zeitachse auseinander, siehe dazu den Auftaktpost: dies., „Ways to Compute Topics over Time: Part 1,“ *From Data to Scholarship. Musings on History, Religion, and Computers* (21.06.2017), <http://jeriwieringa.com/2017/06/21/Calculating-and-Visualizing-Topic-Significance-over-Time-Part-1/>. Zugriff am 06.12.2021.

⁴⁶ Siehe Blei, Ng und Jordan, „Latent Dirichlet Allocation,“ 1007–1014.

⁴⁷ Siehe zur Relevanz der Vorverarbeitung literarischer Korpora Uglanova und Gius, „Order of Things“.

⁴⁸ Schmidt, „Words Alone,“ Abs. 3.

formuliert zunächst nicht mehr oder weniger als eine Wahrscheinlichkeitsverteilung über die gesamte Textkollektion, die beschreibt, welche Wörter eines fixen Gesamtvocabulars gehäuft gemeinsam vorkommen. Aus diesen gemeinsamen Häufigkeiten ergeben sich dann Wortgruppen, die zu unterschiedlichen Anteilen in den einzelnen Dokumenten der Textsammlung vertreten sind. Der Begriff „Topic“ vermag damit zwar Assoziationen an beispielsweise die Topik oder „Topoi“ zu wecken, jedoch sind damit keine epistemologischen Behauptungen über die Wahrscheinlichkeiten des gemeinsamen Vorkommens hinaus verbunden.⁴⁹ Werden Topics allerdings, wie in den Geisteswissenschaften üblich, als semantische Felder⁵⁰ oder Diskurse⁵¹ interpretiert, dann geschieht das auf Grundlage des Konzepts der distributionellen Semantik, die zahlreichen Methoden aus dem Bereich der Computerlinguistik zugrunde liegt.⁵²

Gemäß der distributionellen Semantik ergibt sich die Bedeutung lexikalischer Einheiten aus ihrer Vorkommenshäufigkeit und zwar insbesondere aus ihrem gemeinsamen Vorkommen mit anderen lexikalischen Einheiten in einem spezifischen Kontext. Das heißt, dass Wörter, die beispielsweise im selben Dokument auftreten, dazu tendieren, zum selben Themenfeld zu gehören. Diese Häufigkeitsbeziehungen zwischen Wörtern werden über Koordinaten in einem Raum numerisch repräsentiert, dem *vector space*.⁵³ Auf diese Weise werden textuelle Daten für den ‚semantisch blinden Computer‘⁵⁴ auf einer Bedeutungssebene operationalisier- und vergleichbar. Insofern wird angenommen, dass sich auch historische Forschungsthemen durch ein spezifisches Set von Begriffen ausdrücken, die sich in den einzelnen Quellen niederschlagen und über die automatisierte Mustererkennung extrahiert werden können.

Welche Typen von Topics mit der Modellierung möglich sind, veranschaulicht Tabelle 3 im Anhang, die exemplarisch 15 von 80 Topics aus einem Modell der zwischen 1996 und Juni 2019 auf dem Fachinformationsportal H-Soz-Kult veröffentlichten Tagungsberichte auflistet. Es ergeben sich im Wesentlichen drei verschiedene Topic-Typen, die unterschiedliche Eigenschaften der Textkollektion repräsentieren:

⁴⁹ Blei, Ng und Jordan, „Latent Dirichlet Allocation,“ Anm. 1 auf 996.

⁵⁰ So z. B. Andrew Piper, *Enumerations: Data and Literary Study* (Chicago und London: The University of Chicago Press, 2018): 67; Bunout und Lange, „Nibbling at Text,“ Abs. 16.

⁵¹ Fechner und Weiß, „Topic Modeling,“ Abs. 3, 7.

⁵² Vgl. Piper, *Enumerations*, 13; Schöchl, „Topic Modeling,“ Abs. 14.

⁵³ Blei, „Topic Modeling,“ Abs. 9; Piper, *Enumerations*, 13–14.

⁵⁴ Nach Silke Schwandt, „Digitale Methoden für die Historische Semantik: Auf den Spuren von Begriffen in digitalen Korpora,“ *Geschichte und Gesellschaft* 44 (2018): 107–134, hier: 108, 133.

1. *Metatopics* bezeichnen solche Wortcluster, die vor allem etwas über die Textart und ihre formale Gestaltung aussagen. Sie enthalten sprachliche Elemente, die das Korpus als Ganzes miteinander verbinden, weswegen sie in der Regel hohe Wahrscheinlichkeitswerte aufweisen. So enthält Topic 46 der Tagungsberichte Terme, die im Umfeld der Tagungs- und Vortragspraxis Verwendung finden. Es spiegelt folglich stilistische und gattungsspezifische Eigenschaften der Quelle wider. Zwar sind Metatopics für die inhaltliche Auswertung zu vernachlässigen, hinsichtlich der formalen Einordnung des Untersuchungsgegenstands vermögen sie indes interessante Einblicke zu liefern.⁵⁵
2. *Noisy Topics* sind gekennzeichnet durch informationelles Rauschen. Da ein wesentlicher Vorzug der Topic-Modellierung darin besteht, die Regelmäßigkeiten in der sprachlichen Struktur textueller Daten aufzudecken,⁵⁶ bilden sich auch Cluster heraus, die bestimmte Phänomene auf der Textoberfläche und weniger semantische Zusammenhänge repräsentieren. Im Falle von H-Soz-Kult clustern zum Beispiel Dokumente aus der Frühphase, in denen aufgrund der Textkodierung keine Umlaute verwendet wurden (Topic 19) oder auch solche, die fremdsprachige Partikel enthalten und damit sprachraumspezifische Topics generieren (Topic 67). Solche Cluster zeigen die sprachliche Heterogenität der Daten auf und können Ansatz für weitere Bereinigungs-schritte sein.
3. *Semantische Wortfelder* sind die Art von Topics, die wir uns von Topic-Modellen erhoffen. Sie repräsentieren idealerweise kohärente Themenfelder, die sich sinnvoll interpretieren und in den Dokumenten nachweisen lassen. Für die Tagungsberichte konnten 71 solcher Cluster generiert werden.

Die unterschiedlichen Clustertypen verdeutlichen, dass diese Methode nicht automatisch Themen aus den Textdaten extrahiert, sondern in erster Linie statistisch auffällige Wortverwendungsmuster. Diese können abhängig von der Datengrundlage wie Themen oder Diskurse, aber auch als literarische Motive interpretiert werden, dürfen aber nicht darüber hinwegtäuschen, dass zugleich

⁵⁵ Auch in anderen Forschungsprojekten sind Metatopics, die in Bezug auf die Fragestellung keine semantische Qualität haben, regelmäßig dominant vertreten, siehe etwa Schöch, „Topic Modeling“, Abs. 23. Oft finden sie zugunsten einer Konzentration auf semantische Cluster indes keine Erwähnung.

⁵⁶ Uglanova und Gius, „Order of Things“, 71–72.

andere, strukturelle und stilistische sowie anderweitige Sprachphänomene möglich sind.⁵⁷

Schauen wir uns eines der vermeintlich semantisch kohärenten Topics genauer an: Topic 40. Die Top-20 Wörter des Topics verweisen in ihrer Zusammensetzung auf das Themenfeld „Digitalität“ (siehe Abbildung 1 sowie Tabelle 3 im Anhang). Wie bereits erläutert, wurden dem Modell entsprechende Metadaten hinzugefügt, um eine Auswertung des zeitlichen Verlaufs zu ermöglichen. Sowohl anhand der durchschnittlichen jährlichen Topic-Verteilung als auch der Entwicklung der Beitragszahlen (siehe Abbildung 1) zeigt sich, dass nach einem frühen Peak 2003 die Auseinandersetzung mit Themen rund um die Digitalität im Verhältnis zu anderen Topics im Tagungsgeschehen zuzunehmen scheint.

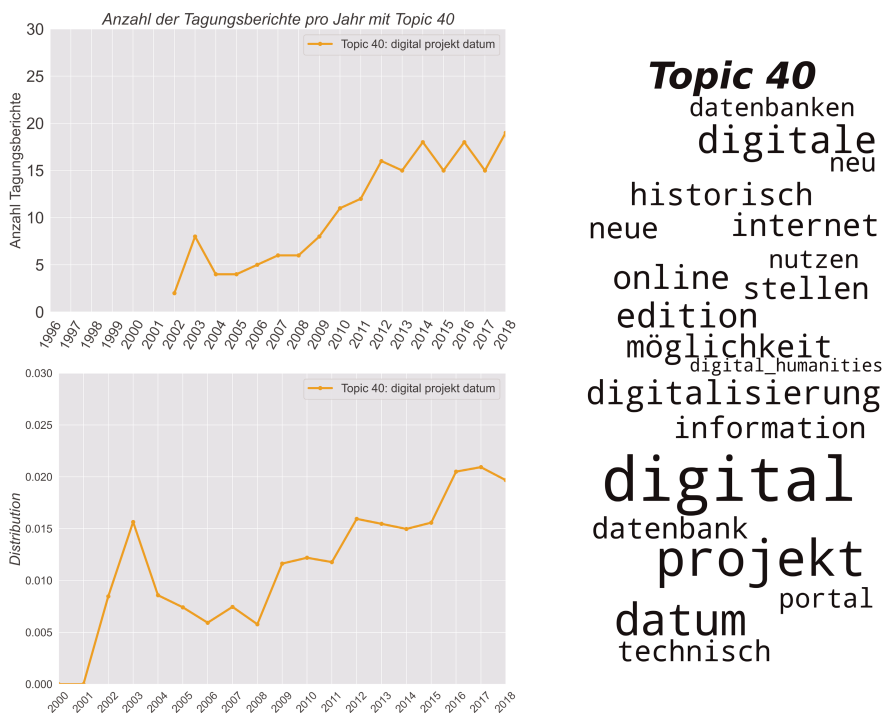


Abb. 1: Entwicklung von Topic 40 in den Tagungsberichten anhand der jährlichen Beitragsanzahl (Schwellenwert: 15 %) und durchschnittlichen jährlichen Verteilung (Schwellenwert: 5 %).

⁵⁷ Siehe auch die Ausdifferenzierung von potenziellen Störquellen in Topics in Boyd-Graber, Mimno und Newman, „Care and Feeding,“ 234–237.

Anhand des Wortclusters lässt sich erahnen, dass Projekte, Editionen, Datenbanken, aber auch das Internet allgemein sowie Portale den Diskurs rundum die Digitalität auf Tagungen bestimmen, weniger im Vergleich dagegen die Digital Humanities. Aber wir erinnern uns: Topics bestehen aus Wörtern mit einer festen, unveränderlichen Wahrscheinlichkeit, das heißt, die lexikalische Komposition der Topics ist immer gleich, unabhängig davon, ob es einem Dokument aus dem Jahr 2003 oder 2018 zugewiesen wurde. Die diachrone Analyse der Verteilung von Topic 40 ist weniger eine Untersuchung eines distinkten Themas als vielmehr eine der sich verändernden Ausprägung der Wahrscheinlichkeitswerte der Dokument-Topic-Zuweisungen, also, ob mehr oder weniger beziehungsweise dominantere oder rezessivere Terme eines Clusters in den Dokumenten gemeinsam auftreten. Über die konkrete Zusammensetzung dieser Terme für den jeweiligen Zeitschnitt und die Dokumente ist damit allerdings nichts gesagt. Annahmen über die Bedeutung und Kohärenz eines Topics auf Basis der Top-Wörter allein sind daher kritisch und bergen die Gefahr, sich fehlleiten zu lassen.⁵⁸ Eine erste Durchsicht der gruppierten Dokumente offenbart etwa, dass die inhaltliche Spannbreite bei den Tagungsberichten sehr weitgefächert ist und Themen umfasst wie den Stellenwert des Internet für die historische Zunft, die Etablierung von Infrastrukturprojekten und Datenbanken, die Öffnung des Editions- und Zeitschriftenwesens für neue Medien und Technologien sowie Fragen nach den Möglichkeiten eines barrierefreieren Zugangs zu Institutionen des kulturellen Erbes beinhaltet, aber auch Auseinandersetzungen mit der Digitalisierung an sich bis hin zu den Digital Humanities und schließlich der Digital History. Diese in den Dokumenten offenbar werdende Heterogenität wird über statische Topic-Modelle nicht sichtbar.

Hinzu kommt, dass prinzipiell alle Dokumente und alle Wörter eines Korpus zu jeder Zeit mit jedem Topic verknüpft sind oder, um es mit den Worten Andrew Pipers zu sagen: „Each topic contains within it the seeds of every other topic.“⁵⁹ Das Gewicht der Topics für die einzelnen Dokumente variiert indes entsprechend des spezifischen Wortgebrauchs in den einzelnen Dokumenten beträchtlich, wobei die meisten Werte gegen 0 Prozent tendieren. Für eine Zeitreihenanalyse sind die Modellierungsergebnisse, die ohnehin schon eine Komplexitätsreduktion der Forschungsressource darstellen, daher über Schwellenwerte und Ähnliches zu normalisieren, was die relativen Verteilungsergebnisse noch einmal modifiziert. In der Regel lassen sich Topics überhaupt erst ab etwa 5 Prozent begrifflich in den Dokumenten nachweisen, ab 15 Prozent schlagen sie sich

⁵⁸ Vgl. Schmidt, „Words Alone,“ Abs. 5–6.

⁵⁹ Piper, *Enumerations*, 74.

dann auch inhaltlich nieder.⁶⁰ Je nach gesetztem Schwellenwert und Normalisierungsworkflow unterscheidet sich folglich die Aussagekraft der probabilistischen Verteilungen. Was tatsächlich untersucht wird, ist der musterhafte Sprachgebrauch, der erst ab einer gewissen Dichte thematische Rückschlüsse erlaubt.

Doch auch die Beurteilung dieser Verteilungen selbst, die ja nicht selten Basis diachroner Analysen sind, ist schwierig. Dass die Topic-Gewichtung ab- oder zunimmt, muss noch nicht bedeuten, dass das Topic an sich an Relevanz verliert oder gewinnt. Topics werden üblicherweise auf Basis der ersten zehn bis 30 Wörter gelabelt, um sie handhabbar zu machen. Hier wurde mit „Digitalität“ ein recht allgemeines Etikett gewählt unter das sich Vieles subsumieren lässt. Die Distributionen beschränken sich für die Klassifikation aber nicht nur auf diese Top-Wörter und so kann es sein, dass ein Dokument mit einer Topic-Distribution von 30 Prozent⁶¹ ebenso das Themenfeld „Digitalität“ abdeckt wie eines mit 51 Prozent.⁶² Der Unterschied ergibt sich vor allem aus dem Gewicht der Wörter für das Topic. Beim manuellen Klassifizieren wäre eine solche qualitative Unterscheidung wahrscheinlich gar nicht vorgekommen, entscheidend wäre, dass überhaupt eine Auseinandersetzung mit „Digitalität“ stattfindet. Wiederum zeigt sich, dass die inhaltlichen Schwerpunkte von Dokumenten mit gleicher Distribution kontextabhängig deutlich variieren können.⁶³ Insofern ist es fraglich, was eine Zeitreihenanalyse der prozentualen Anteile im Rahmen einer Untersuchung thematischer Konjunkturen aussagt, weswegen hier zusätzlich die Entwicklung der absoluten Beitragszahlen mit Topic 40 angeführt wurde (siehe Abbildung 1).

Ohne eine Berücksichtigung der tatsächlichen Inhalte der Dokumente durch *Close Reading* ist eine diachrone Analyse von Topics allein auf Basis der Top-Wörter und arbiträrer Schwellenwerte in ihrem Aussagewert begrenzt, auch, weil das Verfahren ursprünglich nicht darauf ausgerichtet war. Entspre-

60 Andrew Piper stellt in Bezug auf seine Analyse von Romanen des 18. bis frühen 20. Jahrhunderts fest, dass ein Topic erst ab einem Schwellenwert von 20 Prozent tatsächlich inhaltlich im Dokument festgestellt werden könne, siehe *ibid.*, 76–77. Wehrheim, „Economic History,“ 112 erfasst Topics dagegen als „substantial“ ab einem Wert von 10 Prozent. Das richtige Maß ist also stets in Abhängigkeit von der jeweiligen Quelle zu treffen. Hier wurde für die Normalisierung ein Schwellenwert von 5 Prozent angesetzt, um Veränderungen im Verteilungsprofil abbilden zu können.

61 <https://www.hsozkult.de/conferencereport/id/tagungsberichte-4510>. Zugriff am 06.12.2021.

62 <https://www.hsozkult.de/conferencereport/id/tagungsberichte-6455>. Zugriff am 06.12.2021.

63 Siehe etwa die Tagungsberichte <https://www.hsozkult.de/conferencereport/id/tagungsberichte-189> und <https://www.hsozkult.de/conferencereport/id/tagungsberichte-193> mit einer Distribution von jeweils 38 Prozent. Zugriff jeweils am 06.12.2021

chend des ursprünglichen Entwicklungsziels ist es dagegen sehr gut als exploratives, heuristisches Mittel geeignet, um in der Vielzahl von Dokumenten potenziell für eine Fragestellung relevante Quellen zu identifizieren, die dann aber einer detaillierteren Analyse bedürfen. Für Forschungsansätze, die Topic Modeling für temporale Untersuchungen anwenden möchten, erscheint es vielversprechend, Evolutionsmodelle wie *Dynamic Topic Modeling* mit Blick auf ihre Adaptierbarkeit für geschichtswissenschaftliche Forschungsvorhaben umfangreicher zu erforschen.⁶⁴ Auf diese Weise wäre es möglich, die semantischen Wortfelder selbst zu historisieren und zu untersuchen, wie sich der mit einem Themenfeld einhergehende spezifische Sprachgebrauch wandelt. Aber auch generell sind die epistemologischen Dimensionen von Topics und ihren probabilistischen Verteilungswerten systematischer zu erforschen wie es Anna Shadrova jüngst aus linguistischer Perspektive kritisch getan hat.⁶⁵

5 Fazit

Algorithmen wie LDA gewähren einen spezifischen und neuartigen Blickwinkel auf historische Quellen und stellen eine vielversprechende Erweiterung für den Werkzeugkasten der Historiker:innen dar. Besonders geeignet ist die Topic-Modellierung mit LDA etwa für heuristische Zwecke, um im Vorfeld einer detaillierten Analyse oder in Kombination mit *Close-Reading*-Ansätzen erste systematische und sprachunabhängige Einsichten in umfangreiche Quellenbestände zu erhalten. Allerdings geht das Verfahren mit einigen Herausforderungen einher: So hängen die möglichen Forschungsergebnisse unmittelbar von der Vorbereitung der Textdaten sowie der Konfiguration der Modellparameter ab und sind mit einem komplexen Evaluationsprozess hinsichtlich der Qualität, Reproduzierbarkeit und Zuverlässigkeit des Modells verbunden. Für geschichtswissenschaftliche Fragestellungen besonders relevant ist zudem der Umstand, dass im Modellierungsprozess selbst beispielsweise die Kontext- und Zeitgebundenheit der Quellen keine Berücksichtigung findet. Schließlich erweist sich die dokumentenunabhängige Interpretation der Topics an sich als schwieriges Unterfangen. Die retrospektive Gleichsetzung von Topics mit kohärenten Themen ver-

⁶⁴ Siehe David M. Blei und John D. Lafferty, „Dynamic Topic Models,“ *Proceedings of the 23rd International Conference on Machine Learning* (2006): 113–120, doi: 10.1145/1143844.1143859, sowie in Anwendung etwa Jo Guldi, „Parliament’s Debates about Infrastructure: An Exercise in Using Dynamic Topic Models to Synthesize Historical Change,“ *Technology and Culture* 60/1 (2019): 1–33, doi: 10.1353/tech.2019.0000.

⁶⁵ Shadrova, „Topic Models.“

schleiert die vielfältige Gestalt der Wortcluster, die zwar semantische Beziehungen abbilden, je nach probabilistischer Ausprägung aber auch zu Fehleinschätzungen über die Inhalte und ihre Entwicklung führen kann oder andere Phänomene des Wortgebrauchs repräsentiert.

Obgleich sich die Topic-Modellierung mit LDA aufgrund der verfügbaren Softwaretools und -bibliotheken grundsätzlich anbietet, ist stets im Einzelfall für die jeweilige Quellengrundlage und Fragestellung die Eignung zu prüfen. Tiefergehend zu erforschen ist, bis zu welchem Grad sich geschichtswissenschaftliche Forschungsthemen und insbesondere -perspektiven anhand eines spezifischen Wortgebrauchs nachweisen und letztlich automatisiert untersuchen lassen. Es erscheint empfehlenswert, in den deutschsprachigen Geistes- und Geschichtswissenschaften über LDA als Standard hinauszugehen und die internationalen und disziplinübergreifenden Entwicklungen stärker aufzunehmen und auch in Software-Umgebungen umzusetzen.⁶⁶ Denn die Wahl des konkreten Modellierungsverfahrens entscheidet darüber, welche Aussagen sich über die sprachliche Struktur einer Dokumentensammlung treffen lassen.

Bibliographie

- Andorfer, Peter. „Turing Test für das Topic Modeling: Von Menschen und Maschinen erstellte inhaltliche Analysen der Korrespondenz von Leo von Thun-Hohenstein im Vergleich.“ *Zeitschrift für digitale Geschichtswissenschaften* 2 (2017). doi: 10.17175/2017_002.
- Alghamdi, Rubayyi und Khalid Alfalqi. „A Survey of Topic Modeling in Text Mining.“ *International Journal of Advanced Computer Science and Applications* 6/1 (2015): 147–153. doi: 10.14569/IJACSA.2015.060121.
- Blei, David M., Andrew Y. Ng und Michael I. Jordan. „Latent Dirichlet Allocation.“ *Journal of Machine Learning Research* 3 (2003): 993–1022.
- Blei, David M. und John D. Lafferty. „Dynamic Topic Models.“ *Proceedings of the 23rd International Conference on Machine Learning* (2006): 113–120. doi: 10.1145/1143844.1143859.
- Blei, David M. und John D. Lafferty. „A Correlated Topic Model of Science.“ *The Annals of Applied Statistics* 1/1 (2007): 17–35. doi: 10.1214/07-AOAS114.

⁶⁶ Siehe für Verfahren, die die Temporalität von Datenkollektionen berücksichtigen: Xuerui Wang und Andrew McCallum, „Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends“, *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2006): 424–433, doi: 10.1145/1150402.1150450; Blei und Lafferty, „Dynamic Topic Models“. Zur verbesserten Untersuchung von Beziehungen zwischen Topics werden dynamische Ansätze weiterentwickelt, siehe zum Beispiel grundlegend David M. Blei und John D. Lafferty, John, „A Correlated Topic Model of Science“, *The Annals of Applied Statistics* 1/1 (2007): 17–35, doi: 10.1214/07-AOAS114.

- Blei, David M. „Probabilistic Topic Models: Surveying a Suite of Algorithms that Offer a Solution to Managing Large Document Archives.“ *Communications of the ACM* 55/4 (2012): 77–84. doi: 10.1145/2133806.2133826.
- Boyd-Graber, Jordan, David Mimno und David Newman. „Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements.“ In *Handbook of Mixed Membership Models and Their Applications*, hg. v. Edoardo M. Airolidi, David M. Blei, Elena A. Erosheva und Stephen E. Fienberg, 225–254. Boca Raton und Florida: CRC Press, 2014.
- Bunout, Estelle und Milan van Lange. „Nibbling at Text: Identifying Discourses on Europe in a Large Collection of Historical Newspapers Using Topic Modelling“ (21.06.2019), <https://www.c2dh.uni.lu/thinkering/nibbling-text-identifying-discourses-europe-large-collection-historical-newspapers-using>. Zugriff am 06.12.2021.
- Fechner, Martin und Andreas Weiß. „Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts.“ *Zeitschrift für digitale Geschichtswissenschaften* 2 (2017). doi: 10.17175/2017_005.
- Griffiths, Thomas L. und Mark Steyvers. „Finding Scientific Topics.“ *Proceedings of the National Academy of Sciences of the United States of America* 101, 1 (2004): 5228–5235. doi: 10.1073/pnas.0307752101.
- Guldi, Jo. „Parliament’s Debates about Infrastructure: An Exercise in Using Dynamic Topic Models to Synthesize Historical Change.“ *Technology and Culture* 60/1 (2019): 1–33. doi: 10.1353/tech.2019.0000.
- Küsters, Anselm und Elisa G. Moreno. „Mining ‚Die Zeit‘: A Structural Topic Model Analysis of Spain’s Image in the German Media 1946–2009.“ *UPIER Working Paper* 16 (2019): 1–38.
- Lee, Daniel D. und Sebastian H. Seung. „Learning the Parts of Objects by Non-negative Matrix Factorization.“ *Nature* 401 (1999): 788–791. doi: 10.1038/44565.
- Luhmann, Jan und Manuel Burghardt. „Digital Humanities – A Discipline in its Own Right? An Analysis of the Role and Position of Digital Humanities in the Academic Landscape.“ *Journal of the Association for Information Science and Technology* (2021): 1–24. doi: 10.1002/asi.24533.
- Maier, Daniel, A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, G. Heyer, U. Reber, T. Häussler, H. Schmid-Petri und S. Adam. „Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology.“ *Communication Methods and Measures* 12/2–3 (2018): 93–118. doi: 10.1080/19312458.2018.1430754.
- McCallum, Andrew Kachites. „MALLET: A Machine Learning for Language Toolkit.“ 2002, <http://mallet.cs.umass.edu>. Zugriff am 06.12.2021.
- Mimno, David, Hanna M. Wallach, Jason Naradowsky, David A Smith und Andrew McCallum. „Polylingual Topic Models.“ *EMNLP 2009* (2009): 880–889.
- Mimno, David. „Computational Historiography: Data Mining in a Century of Classics Journals.“ *ACM Journal on Computing and Cultural Heritage* 5/1 (2012): 3:1–3:19. doi: 10.1145/2160165.2160168.
- Mulunda, Christine K., Peter W. Wagacha und Lawrence Muchemi. „Review of Trends in Topic Modeling Techniques, Tools, Inference Algorithms and Applications.“ *ISCM 2018* (2018): 28–37. doi: 10.1109/ISCM.2018.8703231.
- Piper, Andrew. *Enumerations: Data and Literary Study*. Chicago und London: The University of Chicago Press 2018.
- Řehůřek, Radim und Petr Sojka. „Software Framework for Topic Modelling with Large Corpora.“ *Proceedings of the LREC Workshop on New Challenges for NLP Frameworks* (2010): 45–50.

- Riddell, Allen Beye. „How to Read 22 198 Journal Articles: Studying the History of German Studies with Topic Models.“ In *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*, hg. v. Matt Erlin und Lynne Tatlock, 91–114. New York: Boydell & Brewer, 2014.
- Roberts, Margaret E., Dustin Tingley, Brandon M. Stewart und Edoardo M. Airoldi. „The Structural Topic Model and Applied Social Science.“ *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation* (2013).
- Schmidt, Benjamin M. „Words Alone: Dismantling Topic Models in the Humanities.“ *Journal of Digital Humanities* 2/1 (2012), <http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/>. Zugriff am 06.12.2021.
- Schöch, Christof. „Topic Modeling with MALLET: Hyperparameter Optimization.“ *The Dragonfly's Gaze. Computational Analysis of Literary Texts* (14.11.2016), <https://dragonfly.hypotheses.org/1051>. Zugriff am 06.12.2021.
- Schöch, Christof. „Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama.“ *Digital Humanities Quarterly* 11/2 (2017), <http://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html>. Zugriff am 06.12.2021.
- Schwandt, Silke. „Digitale Methoden für die Historische Semantik: Auf den Spuren von Begriffen in digitalen Korpora.“ *Geschichte und Gesellschaft* 44 (2018): 107–134.
- Shadrova, Anna. „Topic Models Do Not Model Topics: Epistemological Remarks and Steps Towards Best Practices.“ *Journal of Data Mining and Digital Humanities* (2021). doi: 10.46298/jdmdh.7595.
- The, Yee W., Michael I. Jordan und David M. Blei. „Hierarchical Dirichlet Processes.“ *Journal of the American Statistical Association* 101 (2006): 1566–1581. doi: 10.1198/016214506000000302.
- Uglanova, Inna und Evelyn Gius. „The Order of Things: A Study on Topic Modeling of Literary Texts.“ *CEUR Workshop Proceedings* 2723 (2020): 57–76. <http://ceur-ws.org/Vol-2723/long7.pdf>. Zugriff am 06.12.2021.
- Vayansky, Ike und Sathish A. P. Kumar. „A Review of Topic Modeling Methods.“ *Information Systems* 94 (2020): 1–15. doi: 10.1016/j.is.2020.101582.
- Wallach, Hanna M., David Mimno und Andrew McCallum. „Rethinking LDA: Why Priors Matter.“ *Advances in Neural Information Processing Systems* 22 (NIPS 2009) (2009): 1973–1981.
- Wang, Xuerui und Andrew McCallum. „Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends.“ *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2006): 424–433. doi: 10.1145/1150402.1150450.
- Wehrheim, Lino. „Economic History Goes Digital: Topic Modeling the Journal of Economic History.“ *Cliometrica* 13 (2019): 83–125. doi: 10.1007/s11698-018-0171-7.
- Wieringa, Jeri. „Ways to Compute Topics over Time: Part 1.“ *From Data to Scholarship. Musings on History, Religion, and Computers* (21.06.2017), <https://jeriwieringa.com/2017/06/21/Calculating-and-Visualizing-Topic-Significance-over-Time-Part-1/>. Zugriff am 06.12.2021.

Anhang

Tab. 3: Topics in den auf H-Soz-Kult zwischen 1996 und 2019 veröffentlichten Tagungsberichten (Auswahl). Topics, die nicht zu den semantischen Wortfeldern zählen, sind kursiv hervorgehoben.

Nr.	Die 20 für das Topic wahrscheinlichsten Wörter	Verteilung
10	erinnerung geschichte vergangenheit erinnerungskultur historisch gedächtnis erinnern holocaust umgang gedenken opfern ereignis zeitzeugen national auseinandersetzung narrativ erfahrung jahr öffentlich denkmal	8,1 %
13	französisch frankreich paris französische de deutsch deutschen_französisch le beitrag deutschland_frankreich pariser la deutschland les beziehung deutschen_französische franzose untersuchen jean land	2,5 %
19	<i>fuer waehrend schliesslich ueber universitaet koennen zj verhaeltnis oefentlichkeit haetten fuehrte preussen beitraege berlin grossen bevoelkerung wuerden koenne referat zeuge_jehovas</i>	1,3 %
22	antik römisch rom griechisch archäologisch zeit archäologie spätantike römische quelle stadt athen kaiser chr lassen antike jahrhundert_chr römer spätantiken jh_chr	4,6 %
34	deutsch holocaust jude bevölkerung opfern deutsche besatzung widerstand ss deutschland zweite_weltkrieg sowjetisch besetzen kollaboration konferenz wehrmacht verbrechen kriegten ghetto nationalsozialistischen	3,5 %
35	jüdisch jude jüdische antisemitismus israel judentum antisemitisch deutsch geschichte deutschland christlich jüdisch_gemeinde leben israelisch religiös deutschen_jüdisch palästina gemeinde holocaust antisemitische	2,8 %
40	digital projekt datum digitale edition digitalisierung online möglichkeit internet historisch stellen information neue datenbank technisch portal neu nutzen digital_humanities datenbanken	4,5 %
43	nationalsozialismus ns deutsch nationalsozialistischen politisch hitler widerstand volksgemeinschaft nationalsozialist nsdap dritte_reich weimarer_republik ns_zeit ns_regime kontinuierität nachkriegszeit deutsche nationalsozialistische forschung jugendbewegung	4,7 %
46	<i>vortrag stellen zeigen beispiel tagung jahrhundert frage beitrag verschiedenen bedeutung untersuchen unterschiedlich deutlich stehen gehen sektion widmen entwicklung rollen aspekt</i>	82,4 %
49	text schrift texten jahrhundert autor handschrift werken antik buch edition de rezeption mittelalterlich inhalt wissen übersetzung werk quelle buchen sprache	5,7 %
50	religiös gott mittelalterlich mittelalter heilige jahrhundert christlich ritual christus text kirche tod tradition göttlich religion heilig theologisch person christliche liturgie	5,0 %

Tab. 3 (fortgesetzt)

Nr.	Die 20 für das Topic wahrscheinlichsten Wörter	Verteilung
52	europäisch europa national international europäische land deutschland deutsch transnationalen transfer eu internationale vergleichen transnatio- nale konzept ebene frankreich staat integration stark	7,9 %
62	identität kultur politisch kulturell jahrhundert national nation gesellschaft sprache deutsch diskurs gruppe kulturelle mythos konstruktion ethnisch religiös beispiel vorstellung geschichte	10,8 %
67	<i>of_the and_the university_of</i> konferenz usa university world london ger- man european modern amerikanisch social international history europe studies john national historical	2,5 %
76	global kolonial china europäisch afrika welt japan indien global- geschichte chinesisch globalisierung indisch kolonie kolonialismus euro- pa sklaverei koloniale japanisch afrikanisch missionar	4,9 %

Digitale Lehre, Wissenschaftskommunikation und Public History

Lara Mührenberg, Ute Verstegen

Christliche Archäologie im *Inverted Classroom* und auf *virtueller Exkursion*

Abstract: Since 2018, the innovative teaching projects „CA 2.x – Christian Archaeology in the inverted classroom“ and „Virtual Excursions“ have been developed at FAU’s chair of Early Christian Archaeology. Their aim is to improve teaching and to increase the attractiveness of the subject. On the basis of a target group analysis, concepts have been developed that focus on the students of „Generation Z“ and take their needs and reception habits seriously. While videos on typical research topics were produced as part of the inverted classroom project, 360° images and videos of important monuments were created for the virtual excursions. Both projects follow a modular structure. This not only allows us to use the material in almost any course type, it can also be updated and expanded with little effort.

Keywords: Early Christian Archaeology, research in teaching and learning, inverted classroom, virtual reality, Generation Z, teaching video

Zusammenfassung: Seit 2018 werden in der Christlichen Archäologie in Erlangen die innovativen Lehrprojekte „CA 2.x – Christliche Archäologie im Inverted Classroom“ und „Virtual Excursions“ entwickelt. Sie haben zum Ziel, die Lehre zu verbessern und die Studienattraktivität des Faches zu steigern. Auf der Basis einer eigens zusammengestellten Zielgruppenanalyse sind Konzepte erarbeitet worden, die die Studierenden der „Generation Z“ fokussieren und ihre Bedarfe und Rezeptionsgewohnheiten ernst nehmen. Während im Rahmen des Inverted Classroom-Projektes unter anderem Videos zu Forschungsthemen des Faches gedreht worden sind, sind für die virtuellen Exkursionen 360°-Aufnahmen bedeutender Monumente entstanden. Beide Projekte sind in einer modularen Struktur erstellt worden. Diese ermöglicht es nicht nur, das Material in nahezu allen Lehrveranstaltungen zu verwenden, es kann auch mit geringem Arbeitsaufwand aktualisiert und erweitert werden.

Schlagwörter: Christliche Archäologie; Lehr-Lern-Forschung; Inverted Classroom; Virtual Reality; Generation Z; Lehrvideo

Wie kann ein archäologischer Einführungskurs für Studierende strukturiert sein, die keinen Computer mehr besitzen, Texte nur am Smartphone lesen, Frontalunterricht wünschen und die Dozierenden filmen wollen, um keine Mitschrift anfertigen zu müssen? Seit etwa drei Jahren strömt die neue *Generation Z*¹ an die Hochschulen und fordert die Lehrenden heraus. Vor uns stehen die ersten echten *digital natives*,² die eine rein analoge Welt nicht mehr kennen und virtuelle Realitäten als Teil ihrer Wirklichkeit erleben. Studien attestieren dieser Generation zudem eine ausgeprägte Konsumhaltung. Schon vor der Covid19-Pandemie, die zu einem starken Anstieg digitaler Lehrangebote an den Universitäten geführt hat, hat sich die Frage gestellt: Was tun?

Seit 2018 reagieren unterschiedliche Projekte am Lehrstuhl für Christliche Archäologie der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) auf diese Frage. Im Folgenden werden zwei dieser Lehrprojekte vorgestellt.

1 Was ist Christliche Archäologie?

Das Fach Christliche Archäologie beschäftigt sich mit den materiellen Zeugnissen des frühen Christentums in den geographischen Gebieten des ehemaligen Römischen Reiches rund um das Mittelmeer und darüber hinaus wie beispielsweise entlang der Seidenstraßen. Zeitliche Forschungsschwerpunkte liegen auf der Spätantike, dem Frühmittelalter sowie in mittel- und spätbyzantinischer Zeit. Neben archäologischen Funden und Befunden werden auch die kulturellen Kontexte frühchristlicher Lebenswelten berücksichtigt und erforscht. Zum fachlichen Grundlagenwissen gehören Kenntnisse der zentralen Monumente und Denkmäler der Spätantike und des frühen Christentums sowie der wichtigsten biblischen und außerbiblischen Textquellen. Die Christliche Archäologie steht somit chronologisch zwischen benachbarten Fächern wie der Klassischen und Provinzialrömischen Archäologie und der Kunstgeschichte. Enge inhaltliche Anbindungen und fachlicher Austausch bestehen zu den genannten Fächern ebenso wie zur Theologie und Geschichtswissenschaft. An der FAU ist das Fach organisatorisch in den Fachbereich Theologie eingegliedert und in unterschiedliche Studiengänge eingebunden, zum Beispiel Archäologische Wis-

1 *Generation Z: Metastudie über die kommende Generation* (2016): 4, hg. v. Triple-A-Team-AG, Zugriff am 30.06.2021, http://tripleateam.ch/wp-content/uploads/2016/06/Generation_Z_Meta studie.pdf.

2 Philippe Wamphler, *Generation „Social Media“: Wie digitale Kommunikation Leben, Beziehungen und Lernen Jugendlicher verändert* (Göttingen: Vandenhoeck & Ruprecht, 2014), 22.

senschaften, Mittelalter- und Frühe Neuzeitstudien, Evangelische Theologie, Kulturgeschichte des Christentums. Christliche Archäologie ist nicht als eigenes Fach studierbar, und der Bestandteil „christlich“ in der Denomination des Teilfachs führt bei Archäologiestudierenden zu Beginn ihres Studiums immer wieder zu Fehleinschätzungen der Fachinhalte.

2 CA 2.x – Christliche Archäologie im *Inverted Classroom*

Seit April 2018 ist das Projekt CA 2.x – *Christliche Archäologie im Inverted Classroom* am Lehrstuhl für Christliche Archäologie der FAU angesiedelt. Es setzt sich aus didaktischer Perspektive mit den Auswirkungen der digitalen Transformation unserer Gesellschaft auf das Lernverhalten der Studierenden und mit den daraus resultierenden neuen Anforderungen an eine zeitgemäße Hochschullehre auseinander. Entwickelt und erprobt wurde das Konzept für den propädeutischen Kurs *Einführung in die Christliche Archäologie*, der turnusmäßig jedes Wintersemester angeboten und mit einer Klausur abgeschlossen wird. Diese Lehrveranstaltung wird von Studierenden unterschiedlichster Fachrichtungen besucht, zum Beispiel der Archäologie und Kunstgeschichte sowie der Evangelischen Theologie. Das Projekt reagiert zum einen auf stark gestiegene Studierendenzahlen bei einer gleichbleibenden Anzahl von Dozierenden und die damit verbundene Undurchführbarkeit der bisherigen Lehrformate, denn die zuvor vorgesehenen Kurzreferate von Studierenden hatten sich bei einer Gruppengröße mit 50 oder 60 Personen zu einem didaktisch wenig sinnvollen Vortragsstakkato entwickelt. Zum anderen greift es die veränderten Lern- und Rezeptionsgewohnheiten der neuen Studierendengeneration (Generation Z) auf, die auch dazu geführt haben, dass grundlegende Kompetenzen des geisteswissenschaftlichen Arbeitens – wie das eigenständige Leseverstehen komplexer Texte – nicht mehr vorausgesetzt werden können.

Das Lehrprojekt verfolgt das Konzept des *Inverted Classroom*³ und fußt auf drei Säulen: Erstens werden zentrale und klausurrelevante Inhalte durch Lehr-

³ Vgl. dazu zunächst die grundlegende Literatur von und für Jürgen Handke, dem Pionier der digitalen Hochschullehre im deutschsprachigen Raum: Jürgen Handke, *Patient Hochschullehre: Vorschläge für eine zeitgemäße Lehre im 21. Jahrhundert* (Marburg: Tecum, 2014); Ders., *Handbuch Hochschullehre digital: Leitfaden für eine moderne und mediengerechte Lehre* (Marburg: Tecum, 2015); *Inverted classroom and Beyond: Lehren und Lernen im 21. Jahrhundert*, hg. v. Eva-Marie Großkurth und Jürgen Handke (Marburg: Tecum, 2016); *Inverted Classroom: Past,*

videos vermittelt, die den Studierenden zur Vorbereitung der jeweiligen Sitzung auf dem universitätseigenen E-Learning-Portal und zudem über den institutseigenen YouTube-Kanal INVESTIGATIO_CA bereitgestellt werden.⁴ Diese können im eigenen Lerntempo, orts- und zeitunabhängig sowie beliebig oft angesehen werden. Zweitens werden die Studierenden in ihrer Selbstlernphase unterstützt, indem auf dem E-Learning-Portal nicht nur die Aufgabenstellungen bereitgestellt werden, unter denen die Videos angesehen werden sollen, sondern auch weiterführende Texte, Hintergrundinformationen und Quellentexte. Drittens sind die (in der Covid19-Pandemie virtuellen) Präsenzphasen vollkommen neu konzipiert worden und werden nun dafür genutzt, das zuvor erarbeitete Wissen in aktivierenden Lehrformaten zu überprüfen, zu diskutieren und gegebenenfalls zu vertiefen sowie die fachrelevanten Skills der Studierenden (zum Beispiel Abstraktionsfähigkeit sowie Identifizieren, Extrahieren und Systematisieren von Information⁵) zu optimieren.

Grundlage des Lehrvideokonzepts bilden hinsichtlich der Medienpraxis und des Lernverhaltens der Studierenden die Ergebnisse einer eigens zusammengestellten Zielgruppenanalyse⁶ sowie ein Meinungsbild, das 2018 unter den

present and future, hg. v. Sabrina Zeaiter und Jürgen Handke (Marburg: Tecum, 2020). Eine ertragreiche Zusammenfassung bietet die Festschrift für Jürgen Handke: *Pioniergeist, Ausdauer, Leidenschaft: Festschrift zu Ehren von Prof. Dr. Jürgen Handke*, hg. v. Sabrina Zeaiter und Peter Franke (Marburg: Universitätsbibliothek, 2020), doi: 10.17192/es2020.0018. Darüber hinaus vgl. für den Fokus diversitätsorientierter Lehre: Eva Nolte und Karsten Morisse, „Inverted Classroom: Eine Methode für vielfältiges Lernen und Lehren?“, in *Teaching Trends 2018: Die Präsenzhochschule und die digitale Transformation. Digitale Medien in der Hochschullehre*, Bd. 7, hg. v. Susanne Robra-Bissantz et al. (Münster: Waxmann, 2019), 105–112.

⁴ Sandra Schön und Martin Ebner, *Gute Lernvideos: So gelingen Web-Videos zum Lernen!* (Norderstedt: Books on Demand, 2013), 12; YouTube-Kanal INVESTIGATIO_CA, https://www.youtube.com/channel/UCaD687K1-gMm5PXf_VHjfcA.

⁵ Der Historiker Philippe Rygiel hat auf die Wichtigkeit dieser zentralen Kompetenzen historisch-geisteswissenschaftlichen Arbeitens auch im digitalen Kontext und bezogen auf den Umgang mit digital vorliegenden Quellen verwiesen, vgl. Philippe Rygiel, *Historien à l'âge numérique: Essai* (Villeurbanne: Presses de l'Enssib, 2017), 176.

⁶ Die Zielgruppenanalyse basiert auf folgenden Studien: Roman Herzog Institut e.V. (Hg.), *Neue Werte, neue Gesellschaft, neue Arbeitswelt? Die Generation von Morgen* (2016), Zugriff am 30.06.2021, <https://www.romanherzoginstitut.de/publikationen/detail/die-neue-generation.html>; Education Group (Hg.), *Oö. Jugend-Medien-Studie 2017: Das Medienverhalten der 11- bis 18-Jährige. Medienverhalten der Jugendlichen aus dem Blickwinkel der Jugendlichen* (2017), Zugriff am 30.06.2021, https://www.edugroup.at/fileadmin/DAM/Innovation/Forschung/Dateien/JugendMedienStudie_ZusFassung_2017.pdf; Medienpädagogischer Forschungsverbund Südwest (mpfs) (Hg.), *JIM 2017: Jugend, Information, (Multi-) Media Basisstudie zum Medienumgang 12- bis 19-Jähriger in Deutschland* (2017), Zugriff am 30.06.2021, https://www.mpfs.de/fileadmin/files/Studien/JIM/2017/JIM_2017.pdf; Shell (Hg.), *Shell-Jugendstudie 2015: Zusammenfassung* (2015), Zugriff am 30.06.2021, <https://www.shell.de/medien/medienstudie/2015>.

Teilnehmer:innen der Einführungsveranstaltung durchgeführt worden ist. Die Ergebnisse haben gezeigt, dass die Studierenden zunehmend der sogenannten „Generation Z“⁷ angehören: Durch die Nutzung digitaler Medien ist eine Veränderung des kognitiven Profils erkennbar, das sich merklich von demjenigen der Vorgängergenerationen unterscheidet.⁸ Es stellt sich heraus, dass etwa Fähigkeiten wie Geduld, Genauigkeit oder Reflexionsvermögen abnehmen und zudem ein nicht-linearer Lesefluss bevorzugt wird.⁹ Dies führt dazu, dass sowohl eine veränderte Methodik zum Erreichen der Lernziele als auch andere Schwerpunktsetzungen erforderlich sind, um die neue Studierendengeneration an die Arbeitsweisen geisteswissenschaftlicher Disziplinen heranführen zu können. Alle rezipierten Studien sowie das erhobene Meinungsbild sind zu dem Ergebnis gekommen, dass Lehrvideos zum Selbststudium und zur Vorbereitung seitens der Generation Z sehr erwünscht sind.¹⁰ Der Videokonsum erfolgt dabei in

menfassung (Hamburg 2015), Zugriff am 30.06.2021, https://www.bundestag.de/resource/blob/632026/7558ce0d8a172b47f78c770dfb35d6ac/KDrs-19-28-38_Broschuere-Shell-Jugendstudie-data.pdf. Neuere Studien der letzten Jahre bestätigen den Trend, der sich bereits 2015/16 abgezeichnet hat.

7 Ausführlich zur Generation Z im Bezug zur Hochschule äußert sich Scholz. Christian Scholz, „Wie ‚ticken‘ Studierende der Generation Z?“ in *Neue Studierendengeneration, neue Herausforderungen: Fachübergreifende Modelle, Module, Maßnahmen. Eine Dokumentation der Fachtagung der Studierendenakademie der Heinrich-Heine-Universität Düsseldorf*, hg. v. Holger Ehlert (Düsseldorf: düsseldorf university press, 2017), 15–26.

8 Andreas Belwe und Thomas Schutz, *Smartphone geht vor: Wie Schule und Hochschule mit dem Aufmerksamkeitskiller umgehen können* (Bern: Hep, ¹2014), 26–27.

9 Scholz, *Generation Z*, 21. Wampfler, *Generation „Social Media“*, 121. Vgl. zum Leseverhalten auch Emese Csobanka, „The Z Generation“, *Acta Technologica Dubnicae* 6/2 (2016): 63–76, doi: 10.1515/atd-2016-0012. Sowie grundsätzlich zum weniger fokussierten Lesen am Bildschirm: Gerhard Lauer, *Lesen im digitalen Zeitalter* (Darmstadt: wbg Academic, 2020): 119, Zugriff am 30.06.2021, https://files.wbg-wissenverbindet.de/Files/Article/ARTK_ZOA_1024585_0001.pdf. Allerdings ist zu betonen, dass bislang empirische Studien, die potenzielle Unterschiede im Leseverhalten zwischen berufsbedingtem, akademischen Lesen und Freizeitlesen berücksichtigen, fehlen, ebenso wie solche, die in einem vergleichbaren Setting das unterschiedliche Lesen desselben Textes am Bildschirm und in Printversion untersuchen würden. Vgl. zu diesem Desiderat: Axel Kuhn und Svenja Hagenhoff, „Kommunikative statt objektzentrierte Gestaltung: Zur Notwendigkeit veränderter Lesekonzepte und Leseforschung für digitale Lesemedien“, in *Lesen X.O: Rezeptionsprozesse in der digitalen Gegenwart*, hg. v. Sebastian Böck, Julian Ingelmann, Kai Matuszkiewicz und Friederike Schruhl (Göttingen: Vandenhoeck & Ruprecht): 27–45.

10 Marc Calmbach et al. (Hg.), *Wie ticken Jugendliche 2016? Lebenswelten von Jugendlichen im Alter von 14 bis 17 Jahren in Deutschland* (Wiesbaden: Springer, 2016), 26.

hohem Maße via Smartphone,¹¹ was eine Anpassung von Bildaufbau und Ästhetik an diese mediale Praxis notwendig macht.

3 Lehrvideoformate

Die Videos von CA 2.x werden in einer modularen Struktur erstellt.¹² Im „Hauptvideo“ (HV) werden in einer Länge von max. 15–20 Minuten auf einer wissenschaftlich anspruchsvollen und abstrakten Ebene die Themen der Lehrveranstaltung verhandelt. Die HVs werden thematisch durch „Micro-Teaching-Videos“ (MTV) ergänzt, die in max. fünf Minuten in freierer optischer Präsentation Definitionen, Sachverhalte und Termini vermitteln. Eine weitere Videogruppe bilden Videos „von Studierenden für Studierende“, die in Zusammenarbeit mit Studierenden entstehen und sich inhaltlich mit relevanten christlich-archäologischen Themen oder Objekten beschäftigen. Durch die modulare Struktur ist es für die Studierenden möglich, die Videos entsprechend ihres Bedarfs auszuwählen und zu kombinieren. Da die Lehrvideos nicht nach Lehrveranstaltungssitzungen eines spezifischen Kurses, sondern nach thematischen Einheiten der Lehrinhalte gegliedert sind, ist ihre Nachhaltigkeit und freie Kombination erhöht.¹³ Durch die modulare Struktur werden außerdem nicht-lineare Sehgewohnheiten aufgegriffen, wodurch auf das veränderte kognitive Profil der Generation Z reagiert werden kann. Durch die MTVs und die Videos „von Studierenden für Studierende“ wird die bei der Zielgruppe ausgeprägte Rezeption von Lern- und Wissensvideos auf Plattformen wie YouTube (und deren Filmästhetik) ernst genommen und für die Lehre nutzbar gemacht.¹⁴

11 Achim Berg (Hg.), *Kinder und Jugendliche in der digitalen Welt* (Berlin, 2019): 4, Zugriff am 30.06.2021, https://www.bitkom.org/sites/default/files/2019-05/bitkom_pk-charts_kinder_und_jugendliche_2019.pdf.

12 So kann auf die verkürzte Aufmerksamkeitsspanne der Studierenden reagiert werden: Wampfler, *Generation „Social Media“*, 22. Sowie Belwe und Schutz, *Smartphone*, 77–78.

13 Die Videos kommen auch in anderen Lehrveranstaltungen, beispielsweise bei Seminaren oder Übungen, gewinnbringend zum Einsatz.

14 Bei der Videoplattform YouTube handelt es sich seit Jahren um das beliebteste Internetangebot der Generation Z. Medienpädagogischer Forschungsverbund Südwest (mpfs) (Hg.), *JIM 2020: Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger* (2021), Zugriff am 30.06.2021, https://www.mpfs.de/fileadmin/files/Studien/JIM/2020/JIM-Studie-2020_Web_final.pdf. Vgl. bereits Medienpädagogischer Forschungsverbund Südwest, *JIM 2017*, 32. Junge Erwachsene informieren sich bevorzugt auf YouTube über Wissenschaftsthemen: Wissenschaft im Dialog gGmbH (Hg.), *Wissenschaftsbarometer 2020: Ergebnisse nach Subgruppen* (Berlin, 2020), Zugriff am 30.06.2021, <https://www.wissenschaft-im-dialog.de/file>

Die HVs bieten über die reine Vermittlung von Fakten die Möglichkeit, größere Themengebiete in ihrer Komplexität zu verdeutlichen. Neben der inhaltlichen Tiefe ist hierbei stets das Ziel, künstlerisch anspruchsvolle Videos zu erstellen, die auch filmisch überzeugen. Bisher ist ein HV zur frühchristlichen Sakralarchitektur entstanden, bei dem Greenscreenaufnahmen und Realfilmsequenzen gedreht werden konnten, sowie Passagen mit zahlreichen Abbildungen und Fotografien kombiniert worden sind.



Abb. 1: Ute Versteegen mit der Replik einer spätantiken Öllampe am Set des HVs „Sakralarchitektur“ in der Friedenskirche in Potsdam (Foto: Falk Nicol, 2018).

Die Studierenden sollen in diesem Video beispielsweise lernen, dass sinnliche Raumeindrücke für die Interpretation archäologischer Monumente unerlässlich, zugleich aber kulturell und zeitlich gebunden sind. Eine andere gestalterische Lösung wurde für die filmische Umsetzung der Fachgeschichte gefunden: Fiktive Interviews in der Ästhetik des Schwarzweißfilms porträtieren ausgewählte Fachvertreter:innen der letzten Jahrtausende, wobei durch die humorvolle Wahl der Drehorte und Requisiten bewusst eine Brechung mit fachlicher „Held:innenverehrung“ erzeugt worden ist.



Abb. 2: Josef Strzygowski (dargestellt von Sebastian Watta) und Joseph Wilpert (dargestellt von Alexander Proksch) im Gespräch mit Lara Mührenberg (Foto: Falk Nicol, Stella Enge und Luis Fensel, 2019).

Über das umfassende HV hinaus, das die historischen Persönlichkeiten vorstellt und ihre Forschungen in einen zeit- und wissenschaftsgeschichtlichen Kontext einbettet, sind zu allen Forschenden Langfassungen der Interviews abrufbar.

Die MTVs heben sich in Stil und Sprache deutlich von den HVs ab: Sie werden im Whiteboard-Stil erstellt. Für die Videos sind eigens Comic-Zeichnungen sowie ein Designkonzept erarbeitet worden, die einen unverwechselbaren Stil ermöglichen.



Abb. 3: Die Heilige Helena, das Maskottchen des YouTube-Kanals INVESTIGATIO_CA, und Kaiser Konstantin (Zeichnungen: Lara Mührenberg).

Sprache, Intonation und Geschwindigkeit orientieren sich an der bei der Zielgruppe beliebten YouTube-Ästhetik.¹⁵ In 3–5 Minuten können so fachrelevante Termini, Fakten, Daten und so weiter zu ausgewählten Themen vermittelt werden.



Abb. 4: In einem Video der Studierenden Milena Bergmann und Marin A. Frank erläutert der spätantike Bischof Maximian die liturgischen Ausstattungselemente frühchristlicher Kirchenbauten (Foto: Milena Bergmann und Marin A. Frank)

Die Videos „von Studierenden für Studierende“ erstellen die Studierenden mit Unterstützung durch die Lehrenden und technisch versierte Hilfskräfte. Im Sommersemester 2019 ist eigens eine Lehrveranstaltung angeboten worden, in welcher die Studierenden relevante Grundlagenkenntnisse wie das Erstellen von Drehbüchern, die Fokussierung der Zielgruppe, das Formulieren von Lernzielen, das Schreiben von Texten sowie das Agieren vor der Kamera erlernen konnten. In ihren Videos befassen sich die Studierenden mit ganz unterschiedlichen Themen, beispielsweise antiken und mittelalterlichen Mauertechniken oder der liturgischen Ausstattung frühchristlicher Kirchenbauten.

¹⁵ Vgl. zur Erstellung geeigneter Lernvideos: Schön und Ebner, *Lernvideos*. Sowie Jan Ullmann, „Entwicklung von Erklärvideos für einen Englisch Selbstlernkurs im Rahmen des ‚Flipped Classroom‘ Prinzips“ (Dissertation, LMU München, 2018), doi: 10.5282/edoc.22645.

4 Präsenzphasen

Eine besondere Herausforderung für die Lehre ist, dass nicht nur die Dozierenden, sondern auch Kommiliton:innen, die sich in höheren Semestern befinden oder bereits älter sind, nicht der Generation Z angehören. Für sie können digitale Methoden, die sich an der Ästhetik und am kognitiven Profil der Generation Z orientieren, ungewohnt sein. Daher ist das Inverted Classroom-Konzept zunächst nur in der Einführungsveranstaltung angeboten worden, in welcher die jüngsten Studierenden anzutreffen sind. Eine weitere Herausforderung ist, dass die Angehörigen der Generation Z zwar ausnahmslos mit Smartphones ausgestattet sind, mit denen sie alle Dateiformate (Videos, PDFs und so weiter) ansehen,¹⁶ aber zu Beginn des Studiums – zumindest vor der aktuellen Covid19-Pandemie-Situation – keinen Laptop besaßen oder nutzten, was unter anderem bei der Erstellung von Hausarbeiten und Referatspräsentationen ein Problem darstellte. Zudem ist festzustellen, dass die Studierenden nur ungern selbst im Seminar aktiv werden möchten. Das Inverted Classroom-Konzept bietet durch die Neukonzeptionierung der Präsenzphasen¹⁷ nun aber die Möglichkeit, die Studierenden aus ihrer Komfortzone zu holen und sie zu aktiver Mitarbeit zu ermuntern. Sogar in der Videokonferenz-Lehre unter den Bedingungen der Covid19-Pandemie, in der viele Lehrende ausgeschaltete Kameras und Mikros beklagt haben, hat dies sehr gut funktioniert.

Alle Maßnahmen zielen darauf ab, die Studierenden durch die Nutzung digitaler Methoden für die Auseinandersetzung mit Fachinhalten der Christlichen Archäologie zu begeistern und zugleich grundlegende Techniken der geisteswissenschaftlichen Arbeit zu vermitteln.¹⁸ Da nach dem ersten Testlauf des Konzepts festgestellt werden musste, dass viele Studierende unvorbereitet zu den

16 Vgl. dazu: Medienpädagogischer Forschungsverbund Südwest, *JIM 2017*, 26.

17 Ausführlich zur Konzeption und Umsetzung der Präsenzphasen: Lara Mührenberg und Ute Verstegen, „CA 2.x – Christliche Archäologie im inverted classroom: Ein Beitrag zur videobasierten digitalen Lehre an der Friedrich-Alexander-Universität Erlangen-Nürnberg,“ in Zeaiter und Handke (Hrsg.), *Inverted Classroom*, 165–177. Sowie Lara Mührenberg, „CA 2.x – Christliche Archäologie im inverted classroom,“ in *Vom Untergrund ins Internet: Die Institute für Ur- und Frühgeschichte, Klassische Archäologie und Christliche Archäologie mit dem gemeinsamen Studiengang ‚Archäologische Wissenschaften‘ an der Friedrich-Alexander-Universität Erlangen-Nürnberg – Stand 2018/19*, hg. v. Doris Mischka et al. (Erlangen: Friedrich-Alexander-Universität, 2019): 65–67.

18 Dazu Scholz, *Generation Z*, 21–23. Gerade in Hinblick auf die Rezeption wissenschaftlicher Texte benötigen die Studierenden Unterstützung, da der nicht-lineare Lesefluss der Studierenden es ihnen erschwert, lange und komplexe Fachtexte zu verstehen. Vgl. dazu Belwe und Schutz, *Smartphone*, 81. Sowie Wampfler, *Generation „Social Media“*, 121.

Präsenzsitzungen erschienen sind, ist ein kurzer E-Selbsttest eingeführt worden, der in Vorbereitung zu den Sitzungen zu absolvieren ist.¹⁹

Durch die ineinandergreifende Struktur aller Bestandteile des Lehrkonzepts, die Ausrichtung auf die Lernziele (Wissen und Kompetenzen) und die detaillierte Ermittlung und Einbeziehung der Bedarfe der Studierenden ist es möglich, in der Klausur neben reinen Fakten in offenen Fragen Wissen und Kompetenzen abzuprüfen. Die Ergebnisse der Klausuren zeigen, dass durch die stärkere Aktivierung in den Präsenzveranstaltungen Anwendungswissen und Transferleistungen der Kursteilnehmenden besser werden und nicht nur auswendig gelerntes Faktenwissen wiedergegeben werden kann. Die Rückmeldung der Studierenden zeigt, dass die kurzen MTVs besonders beliebt sind und diese wiederholt abgerufen werden. Zum Ende des Wintersemesters 2019/20 ist zudem eine Evaluation durchgeführt worden.²⁰ Auch diese Ergebnisse waren sehr positiv und zeigen, dass die Kombination aus Präsenzphasen und eigenständiger Vorbereitung auch aus Sicht der Studierenden einen Gewinn darstellt.²¹ Es hat sich gezeigt, dass das Konzept vor allem die jüngeren Studierenden stark anspricht. Vor allem die Studienanfänger:innen favorisieren klar eine Wissensvermittlung über kurze Videoformate gegenüber längeren Videos oder Textressourcen. Im Rahmen der pandemiebedingten Online-Lehre wird das *Inverted Classroom*-Konzept im Verhältnis zur reinen Videokonferenz-Lehre besonders positiv reflektiert.²²

5 Virtual Excursions

Bereits vor dem Beginn der Covid19-Pandemie, die zu einem Ausbau der digitalen Lehrangebote geführt hat, ist am Lehrstuhl für Christliche Archäologie ein

19 Die Idee, die Studierenden vor der Präsenzphase auf ihren Selbstlernerfolg hin zu prüfen, stammt im Ursprung von Handke: Zeaiter und Franke, *Pioniergeist*, 93–100.

20 Hierfür ist das Inventar zur Evaluation von Blended Learning (IEBL) des Leibniz-Instituts verwendet worden. Dieses basiert auf dem hinlänglich erprobten und psychometrisch validierten Heidelberger Inventar zur Lehrveranstaltungsevaluation in seiner revidierten Fassung (HILVE-II). Das IEBL selbst wurde psychometrisch überprüft (n=122).

21 Zu vergleichbaren Ergebnissen kommen auch Claudia Mertens et al., „To flip or not to flip?“, Empirische Ergebnisse zu den Vor- und Nachteilen des Einsatzes von *Inverted-Classroom*-Konzepten in der Lehre,“ in *Hochschullehre als reflektierte Praxis: Fachdidaktische Fallbeispiele mit Transferpotenzial*, hg. v. Tobias Schmohl and Kieu-Anh To (Bielefeld: Bertelsmann, 2019): 13–28.

22 Die Studierenden haben in der mündlichen Feedbackbesprechung zur Lehre unter Bedingungen der Covid19-Pandemie äußerst positive Rückmeldung zu diesem Lehrformat gegeben.

zweites digitales Lehrprojekt geplant worden: Seit Sommer 2020 wird das Projekt *Virtual Excursions* verfolgt.²³ Dieses Lehrprojekt hat zum Ziel, auf innovative Weise das Lehrangebot durch virtuelle Exkursionen mithilfe von 360°-Aufnahmen zu verbessern und zu ergänzen. Zum fachlichen Grundlagenwissen der Christlichen Archäologie gehören Kenntnisse der zentralen Monumente und Denkmäler der Spätantike und des frühen Christentums, welche in der Lehre durch visuelle Repräsentationen der entsprechenden Denkmäler im Seminar oder durch Exkursionen an ihre (Stand-)Orte erworben werden. Den optimalen Lernerfolg für das Kennenlernen und Verstehen dieser Denkmäler erzielt erfahrungsgemäß der persönliche Besuch der Originale mittels Exkursionen zu Museen oder an antike Stätten. Die eigene Sinneswahrnehmung und die körperliche Erfahrung ermöglichen es, die Denkmäler in ihren Größendimensionen und räumlichen Kontexten zu erfahren und zu analysieren.²⁴ Dies erzielt auch eine bessere Memorierbarkeit als das Betrachten von Abbildungen in Büchern oder von Fotos auf PowerPoint-Präsentationen.

Eine vergleichbare Erfahrung eröffnen 360°-Aufnahmen, in welche man mittels einer *Virtual Reality*-Brille eintreten kann. Es besteht die Option, sich zu drehen und auf- und abzuschauen und auf diese Weise einen Raumeindruck zu erhalten, Entfernungen abzuschätzen, die Lage bestimmter Bauteile zueinander zu ermessen, Blickachsen zu verstehen und anderes. All dies ist durch zweidimensionale Pläne und Abbildungen nur eingeschränkt möglich und erfordert jahrelange Einübung. Die Option, im Seminar einen Kirchenraum oder eine Grabkammer virtuell zu bereisen, bereichert die Lehre grundsätzlich und erleichtert den Studierenden das Verständnis. Zudem können die Studierenden so orts- und zeitunabhängig auf die Monumente zugreifen, was ein Gewinn für die

23 Die wissenschaftliche Auseinandersetzung mit dem Einsatz von *Virtual Reality* in der Lehre ist in den letzten Jahren sprunghaft angestiegen. Grundlegend sei auf folgende Literatur verwiesen: Gökçe Akçayır und Carrie Demmans Epp, *Designing, Deploying, and Evaluating Virtual and Augmented Reality in Education* (Hershey, Pennsylvania: IGI Global, 2021); Dong Hwa Choi, Amber Dailey-Hebert und Judi Simmons Estes, *Emerging Tools and Applications of Virtual Reality in Education* (Hershey, Pennsylvania: IGI Global, 2016). Eine Einführung in Funktionsweisen und Anwendungsmöglichkeiten bietet: Zeynep Taççin, *Virtual and Augmented Reality: An Educational Handbook* (Cambridge: Cambridge Scholars Publishing, 2020).

24 Zur Konstituierung von Räumen durch ihre Besucher:innen vgl. Martina Löw, *Raumsoziologie* (Frankfurt a. M.: Suhrkamp, 2017). Analysen zum Potenzial der Erschließung von Räumen durch Exkursionen als Lehrformat stammen bislang vor allem aus der Geographie: Kerstin Neeb, *Geographische Exkursionen im Fokus empirischer Forschung: Analyse von Lernprozessen und Lernqualitäten kognitivistisch und konstruktivistisch konzeptionierter Schülerexkursionen* (Weingarten: Selbstverlag des Hochschulverbandes für Geographie und ihre Didaktik, 2012); Anne-Kathrin Lindau und Tom Renner, „Räume durch geographische Exkursionen und Fragen erschließen“, *Hallesches Jahrbuch für Geowissenschaften* 41 (2018): 63–76.

Nachbereitung der Seminarsitzungen, die Vorbereitung von Referaten, das Lernen für Prüfungen und vor allem für die Erarbeitung der Denkmälerkenntnis, die für unser Fach außerordentlich wichtig ist, darstellt.

Ziel des Projektes ist es nicht, reale Reisen an archäologische Stätten abzuschaufen, denn die Wahrnehmung eines Raumes oder Ortes mit allen dem Menschen zur Verfügung stehenden Sinnen, ist (noch) nicht ersetzbar. Vielmehr soll die Möglichkeit geschaffen werden, ein Monument nach Bedarf in simulierter Dreidimensionalität in Augenschein nehmen zu können. Ein gewichtiger Vorteil virtueller Reisen ist dabei zudem, dass sie eine klimaneutrale sowie kosten- und zeitgünstige Alternative darstellen. Darüber hinaus erlauben sie, für den Fall unvorhersehbarer Krisen (wie das Corona-Virus oder beispielsweise eine Verschlechterung der weltpolitischen Lage), welche die Durchführbarkeit von Studium und Lehre einschränken, vorzusorgen.

Für die Erstellung eines Lehrkonzeptes konnte auf die Forschungen und Erfahrungen des Projekts *CA 2.x – Christliche Archäologie im Inverted Classroom* zurückgegriffen und aufgebaut werden. Insbesondere die Erkenntnis, dass eine modulare Struktur dem kognitiven Profil der Generation Z entgegenkommt, hat auch hier Berücksichtigung gefunden, so dass die mit 360°-Videos und -Fotos dokumentierten Bauten ausgerichtet nach individuellen Bedürfnissen besucht werden können. Das Einsatzszenario von Digitalpanoramen in der kunsthistorischen und archäologischen Lehre ist nicht neu.²⁵ Durch den Markteintritt entsprechender 360°-Kameras und VR-Headsets zu erschwinglichen Konditionen in den Consumerbereich können nun aber neue Möglichkeiten ausgelotet werden, die wesentlich betrachterimmersiver sind als zuvor. Ein weiterer zentraler Punkt der Projektkonzeption ist daher, nicht nur „begehbare“ Standbilder anzufertigen.²⁶ Vielmehr sollen, einer realen Exkursion möglichst ähnlich, Lernsituationen erzeugt werden, in welche die Studierenden über einen Screen oder sogar mittels einer VR-Brille eintreten können. Analog zu realen Exkursionen werden sie in den Videos von Expert:innen durch die Monumente geführt. Diese zeigen Besonderheiten der Bauten auf und diskutieren fachliche Fragen direkt

²⁵ Ute Versteegen, „Abschied vom Dia – eine Reprise: Digitale Repräsentationen dreidimensionaler Bauten und Objekte in der kunsthistorischen Lehre,“ *Kunstchronik* 58 (2005): 347–348; Dies., „Vom Mehrwert digitaler Simulationen dreidimensionaler Bauten und Objekte in der architekturgeschichtlichen Forschung und Lehre“ (Vortrag gehalten auf dem XXIX. Deutschen Kunsthistorikertag, Regensburg, 16. März 2007), doi: 10.11588/artdok.00000473.

²⁶ Vgl. einen der wenigen ersten Erfahrungsberichte zu den Vorteilen der Nutzung von 360°-Videos: Laura J. Thompson et al., „Using 360-degree video for immersive learner engagement,“ *Journal of Extension* 56/5 (2018), Zugriff am 30.06.2021, <https://tigerprints.clemson.edu/joe/vol56/iss5/1/>.

am Bau und Objekt. Durch den Eindruck, Teil einer Exkursionsgruppe zu sein, soll erreicht werden, eine beinahe partizipative Situation zu erzeugen.

6 360°-Videos

Auch die 360°-Videos werden in einer modularen Struktur erstellt. Im September 2020 sind trotz der Covid19-Pandemie erste Drehaufnahmen in Trier und Köln möglich gewesen. Unter anderem sind 360°-Videos der sogenannten Konstantinbasilika beziehungsweise Palastaula in Trier und der Kirche St. Gereon in Köln entstanden. Jeweils sind typische Exkursionssequenzen an wechselnden Orten inner- und außerhalb des Baus nachempfunden worden. Dabei sind nicht nur die Ausführungen der Expertinnen, die durch die Räume geführt haben, eingefangen, sondern gezielt auch Lehrsituationen und Diskussionen erzeugt worden, wie sie üblicherweise bei realen Exkursionen entstehen. Nachträglich sind die Videos aufwändig bearbeitet worden; so sind beispielsweise eigens erstellte Pläne in die Videos eingefügt worden, um die Orientierung im Raum zu unterstützen und die Verortung innerhalb der baulichen Struktur zu ermöglichen.

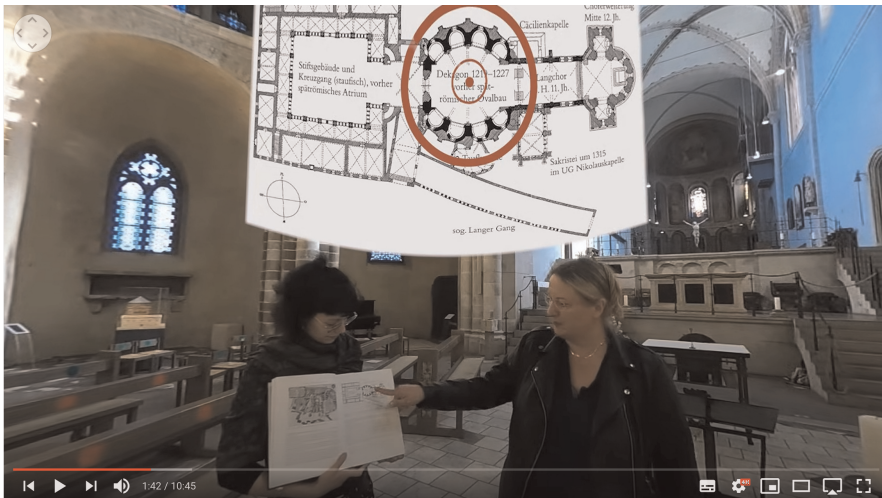


Abb. 5: Ute Versteegen führt durch den Kirchenbau St. Gereon in Köln. Der eingefügte Plan ermöglicht die Orientierung im Gebäude (Foto: Falk Nicol, Alissa Dittes und Luis Fensel, 2020).

Zusätzlich zu den geführten virtuellen Rundgängen sind klassische 360°-Standbilder an verschiedenen Standorten innerhalb und außerhalb der Gebäude erstellt worden, die eine eigenständige und bedarfsorientierte virtuelle Begehung der Orte ermöglichen.

Weitere 360°-Videos sollen in Zukunft im Rahmen realer Exkursionen zu Stätten und Denkmälern des frühen Christentums aufgenommen werden. Zudem erlauben die geringen Abmessungen der 360°-Kamera, sie auch auf Dienstreisen mitzunehmen und bei der Besichtigung relevanter Monumente relativ spontan ein Video zu drehen. Für die Nutzung über den persönlichen wissenschaftlichen Gebrauch hinaus ist hierfür jedoch das Einholen von Drehgenehmigungen im Vorfeld zu beachten.

7 Umsetzung in der Lehre

Aufgrund der Covid19-Pandemie mussten seit März 2020 alle geplanten Exkursionen des Lehrstuhls abgesagt werden. Da die Studierenden des Studienganges *Archäologische Wissenschaften* aber im Bachelorstudium Exkursionsmodule mit mindestens 10 ECTS nachweisen müssen, wurden im Wintersemester 2020/1 und im Sommersemester 2021 vor allem für die kurz vor dem Abschluss stehenden Studierenden virtuelle Exkursionen angeboten.²⁷ Im Winter wurde auf das lehrstuhleigene Material zurückgegriffen und virtuell nach Köln und Trier gereist. Die Lehrveranstaltung kombinierte die individuelle Erfahrung der virtuellen Expertinnenführungen in St. Gereon und in der Trierer Palastaula per VR oder YouTube-360°-Video mit Gruppenreferaten von Studierenden über andere archäologische Stätten in Trier, an denen unkommentierte 360°-Videos und -Standbilder aufgenommen worden waren (Porta Nigra, Barbarathermen, Kaiserthermen). Die Referate fanden über die Videokonferenz-Software Zoom statt, wobei die Studierenden zur Präsentation unterschiedliche Kombinationen aus PowerPoint und Online-Videosequenzen nutzten. Da die Studierenden selbst keine eigenen VR-Brillen besaßen, wurde eine Sitzung durchgeführt, in der eine Anleitung zum Basteln eines Cardboards für das eigene Smartphone gegeben

²⁷ Luise Drake et al., „Digitale Lehre in den archäologischen Wissenschaften an der Friedrich-Alexander-Universität Erlangen-Nürnberg – Erfahrungen nach einem Jahr Coronapandemie,“ *Blickpunkt Archäologie* 2 (2021): 103–113.

wurde. Ergänzend wurde durch eine Hilfskraft des Lehrstuhls zusätzlich ein Do-It-Yourself-Video erstellt und auf den YouTube-Kanal hochgeladen.²⁸



Abb. 6: Alissa Dittes erklärt im YouTube-Video, wie man aus dem Smartphone eine VR-Brille bastelt (Foto: Christopher Koch, 2021).

Die VR-Lernsituation variierende Präsenzsitzungen waren wegen der strengen Hygienevorschriften bis Herbst 2021 nicht möglich. Geplant sind vor allem Tests zu zwei Rezeptionssituationen: das gemeinsame Betrachten der VR-Videos im Seminarraum via Beamer und die gemeinsame Rezeption eines 360°-Standbilds via VR-Brille durch mehrere Studierende gleichzeitig im selben Raum. In beiden Lernsituationen soll der gemeinsame fachliche Austausch über den virtuell besuchten Ort gefördert werden. Ziel ist, durch diese Rezeptionssituationen zu erproben, wie nah eine Gruppe an das gemeinsame Lernerlebnis einer realen Exkursion herangeführt werden kann und welche spezifischen Erfordernisse für den Lernerfolg durch das VR-Erleben noch nicht erfüllt werden können.

Da viele Museen ihr Online-Angebot im Verlauf der Covid19-Pandemie stark ausgebaut haben und nun auch virtuelle Rundgänge anbieten,²⁹ führte die virtuelle Exkursion im Sommersemester 2021 in die Eremitage nach St. Petersburg.

²⁸ *DIY: VR-Brille basteln*, Zugriff am 30.06. 2021, <https://www.youtube.com/watch?v=Lk6KlgEFd70>.

²⁹ Vgl. Ellie King et al., „Digital Responses of UK Museum Exhibitions to the COVID-19 Crisis, March–June 2020,“ *Curator: The Museum Journal* (2021): 9, doi: 10.1111/cura.12413; Peter Maxwell, „The rise of the virtual gallery tour: What works and what doesn't (yet),“ *Frame* 8 (2020),

Die Lehrveranstaltung griff hierbei auf das durch das Museum selbst bereitgestellte Angebot eines virtuellen Rundgangs zurück.³⁰ Während einige Museen, die mit ihren digitalen Angeboten eine Vorreiterrolle auf dem internationalen Museumsparkett einnehmen, inzwischen virtuelle Touren anbieten, bei denen die Ausstellungsbereiche per *Indoor Mapping* mit LiDAR-Technologie 3D-digitalisiert wurden und deren Punktwolken von den Online-Nutzer:innen on the fly durchschritten werden können,³¹ nutzen viele Häuser weiterhin 360°-Panoramen. Auch diese lassen sich in der Lehre durchaus gewinnbringend einsetzen. Die 360°-Panoramen der musealen Räume der Eremitage umfassen beispielsweise auch Erschließungssituationen wie Treppen und Korridore, so dass der Weg von der Kasse bis zum gesuchten Museumsobjekt in seiner Vitrine sehr gut virtuell nachvollzogen werden kann. Für einen wesentlichen Teil der ausgestellten Objekte existieren in den VR-Rundgängen der Eremitage zudem Verlinkungen von der Vitrine in den digitalen Sammlungskatalog, der zwar öffentlich noch wenige Meta-Informationen, dafür aber zum Beispiel passable Fotos der Objekte enthält.

Die virtuelle Museumsexkursion nach St. Petersburg, welche ebenfalls als Online-Veranstaltung via Zoom stattfand, zielte wiederum darauf, verschiedene, möglichst partizipative Lehr-/Lernszenarien anhand des zur Verfügung stehenden Online-Angebots zu erproben. Neben den virtuellen Referaten vor den Vitrinen erschlossen sich die Studierenden die Museumsräume auch in Kleingruppen in sogenannten Breakout-Rooms, das sind Gruppenarbeitsräume der Software Zoom, und ‚schlenderten‘ gemeinsam durch den virtuellen Rundgang. Dieses Format wurde von den Studierenden in der Abschlussevaluation sehr begrüßt. Die Evaluation hat darüber hinaus klar gezeigt, dass sich die Studierenden das Format virtueller Exkursionen auch nach der Covid19-Pandemie als ergänzendes Format im Lehrportfolio wünschen. Positiv hervorgehoben wurde die Möglichkeit, auf diese Weise Museen und Ausgrabungsstätten bereisen zu können, die aus politischen, konservatorischen oder finanziellen Gründen nicht zugänglich sind oder deren Besuch durch eine Gruppe mit einer besonders schlechten Klimabilanz verbunden wäre. Ebenfalls geschätzt wurde die Mög-

Zugriff am 30.06.2021, <https://www.frameweb.com/article/the-rise-of-the-virtual-gallery-tour-what-works-and-what-doesnt-yet>.

30 The State Hermitage Museum, Zugriff am 30.06.2021, <https://srv1.hermitagemuseum.org/wps/portal/hermitage/panorama>.

31 Beispielsweise das Deutsche Museum in München, <https://virtualtour.deutsches-museum.de/>. Beispiele für archäologische Museen sind das Musée national d'histoire et d'art in Luxemburg, <https://my.matterport.com/show/?m=S1wvMgypEXN> und das Anadolu Medeniyetleri Müzesi in Ankara, <https://matterport.com/de/media/3414?industry%5B8%5D=8&industry%5B6%5D=6>, Zugriff am 30.06.2021.

lichkeit, sich künftig durch die Nutzung der virtuellen Rundgänge besser auf einen konkreten Museumsbesuch vorbereiten zu können und bereits den Weg zu den gesuchten Museumsobjekten im Vorfeld virtuell abgehen zu können.

Um bestehende museale Angebote für universitäre virtuelle Exkursionen mit Studierenden noch besser nutzen zu können, wären in Zukunft Szenarien von Interesse, welche zum einen die Kommunikation mit den Fachkurator:innen in den Museen durch ein Zuschalten in die Online-Konferenz einbeziehen und zum anderen die Kontextualisierung von Einzelobjekten über ihr Vitrinendasein hinaus erweitern könnten.³² Links auf 3D-Digitalisate der Sammlungsobjekte aus dem virtuellen Rundgang heraus, wie sie das Harvard Museum of the Ancient Near East bereits anbietet,³³ könnten beispielsweise mit maßstabsgerechten Einblendungen von 3D-Vergleichobjekten aus anderen Sammlungen verbunden werden.

8 INVESTIGATIO_CA

Der YouTube-Kanal INVESTIGATIO_CA³⁴ des Lehrstuhls ist 2019 online gegangen. Im Sinne des Konzepts von Open Educational Resources wird christlich-archäologisches Wissen hier einer breiten Öffentlichkeit innerhalb und auch außerhalb der Universitäten zugänglich gemacht. Ein fachfremdes Publikum, Lehrer:innen, Pfarrer:innen und Dozierende anderer Hochschulen können somit ebenfalls die erstellten Videos nutzen. Interessant sind die Videos aufgrund der gemeinsamen Lehrinhalte für Fächer wie die Kunstgeschichte, die Theologie oder die Klassische Archäologie. Sämtliche Filme des Projektes CA 2.x – *Christliche Archäologie im Inverted Classroom* sowie erste Videos aus dem Projekt *Virtual Excursions* sind auf YouTube verfügbar. Auf diese Weise kann auch die öffentliche Sichtbarkeit des sogenannten Kleinen Fachs Christliche Archäologie erhöht werden.³⁵

³² Vgl. die Anforderungskriterien in: Caitlin Chien Clerkin und Bradley L. Taylor, „Online Encounters with Museum Antiquities,“ *American Journal of Archaeology* 125/1 (2021): 175, doi: 10.3764/aja.125.1.0165.

³³ Harvard Museum of the Ancient Near East, Zugriff am 30.06.2021, <https://my.matterport.com/show/?m=bS17YNKCggy&referrer=https://hmane.harvard.edu>.

³⁴ INVESTIGATIO_CA: https://www.youtube.com/channel/UCaD687K1-gMm5PXf_VHjfcA.

³⁵ Zur Christlichen Archäologie im Rahmen der Diskussion um sogenannte Kleine Fächer: Ute Verstegen, *Kleine Fächer im Porträt: Christliche Archäologie* (2019), Zugriff am 30.06.2021, <https://www.kleinefaecher.de/beitraege/blogbeitrag/christliche-archaeologie.html>.

Die Kanalanalysen der Gesamtlaufzeit seit Juni 2019, die YouTube zur Verfügung stellt, stellen hierbei eine wichtige Information dar, mit deren Hilfe das entwickelte Lehrkonzept hinsichtlich der Anschlussfähigkeit an die Zielgruppe kritisch überprüft werden kann. Die Klickzahlen und Kommentare zeigen deutlich, dass Videos, die auch für schulische Lehrinhalte – beispielsweise im Religions-, Latein- oder Kunstunterricht – relevant sind, besonders häufig angesehen werden. Neben der Anzahl der Aufrufe ist auch die Anzeigedauer der Videos aufschlussreich. Es zeigt sich, dass die Videos in der Regel nicht länger als für eine Dauer von zwei oder drei Minuten angesehen werden. Die Altersgruppe der 18–24-Jährigen stellt die größte Gruppe der Nutzer:innen dar (gefolgt von der Gruppe der 45–54-Jährigen). Überraschenderweise überwiegt trotzdem die durchschnittliche Nutzung eines Computers als digitales Endgerät (60,2%) gegenüber dem Smartphone (29,3%). Eine eingehende Analyse zeigt jedoch, dass dies auf eine Entwicklung zurückzuführen sein dürfte, die mit der Covid19-Pandemie in Zusammenhang steht: Seit Beginn der Online-Lehre an Universitäten und Schulen hat sich der Zugriff über Computer (außerhalb der Schul- und Semesterferien) teilweise verdreifacht. Eine interessante Beobachtung ist hierbei, dass die durchschnittliche Wiedergabedauer der Smartphonennutzenden (2,15 Min.) deutlich hinter der der Computernutzenden (2,30 Min.) zurückbleibt. Peaks können in den Zeiträumen verzeichnet werden, in denen mit Prüfungsvorbereitungen zu rechnen ist.

9 Fazit

Die Situation der Covid19-Pandemie stellte Lehrende ebenso wie Studierende nicht nur vor eine enorme Herausforderung. Aus jetziger Perspektive hat sie sich auch als ein Katalysator für die Weiterentwicklung und die Akzeptanz neuer digitaler Lehrformate erwiesen. Mit der Entwicklung der Lehrkonzepte der videogestützten *Inverted Classroom*-Lehre und der virtuellen Exkursionen hatten der Lehrstuhl Christliche Archäologie der FAU bereits vor der Covid19-Pandemie begonnen. Durch die Anpassung an die Erfordernisse der pandemiebedingten Online-Lehre haben sich diese Formate auch als sehr tragfähig für zunächst nicht geplante digitale Lehrsituationen erwiesen. Die Evaluationen der Formate in Präsenz und in der Online-Lehre zeigen, dass bei den Studierenden der FAU eine hohe Akzeptanz und Offenheit für die neu eingeführten Lehrformen besteht. Auch für die Zeit nach der Pandemie wird eine Weiterführung und ein Ausbau der Online-Angebote und der Lehrveranstaltungen mit virtuellem Anteil gewünscht und bestehen bleiben.

Bibliographie

- Akçayır, Gökçe und Carrie Demmans Epp. *Designing, Deploying, and Evaluating Virtual and Augmented Reality in Education*. Hershey, Penn.: IGI Global, 2021.
- BITKOM (Hg.). *Digitale Schule – vernetztes Lernen: Ergebnisse repräsentativer Schüler- und Lehrerbefragungen zum Einsatz digitaler Medien im Schulunterricht*. Berlin 2015. Zugriff am 30. Juni, 2021. <https://www.bitkom.org/sites/default/files/file/import/BITKOM-Studie-Digitale-Schule-2015.pdf>.
- Belwe, Andreas und Thomas Schutz. *Smartphone geht vor: Wie Schule und Hochschule mit dem Aufmerksamkeitskiller umgehen können*. Bern: Hep, ¹2014.
- Berg, Achim (Hg.). *Kinder und Jugendliche in der digitalen Welt*. Berlin 2019. Zugriff am 30. Juni, 2021. https://www.bitkom.org/sites/default/files/2019-05/bitkom_pk-charts_kinder_und_jugendliche_2019.pdf.
- Calmbach, Marc, Silke Borgstedt, Inga Borchard, Peter M. Thomas und Berthold B. Flaig, Hgg. *Wie ticken Jugendliche 2016? Lebenswelten von Jugendlichen im Alter von 14 bis 17 Jahren in Deutschland*. Wiesbaden: Springer, 2016.
- Choi, Dong H., Amber Dailey-Hebert und Judi Simmons Estes, *Emerging Tools and Applications of Virtual Reality in Education*. Hershey, Penn.: IGI Global, 2016.
- Clerkin, Caitlin Chien und Bradley L. Taylor. Online Encounters with Museum Antiquities. *American Journal of Archaeology* 125/1 (2021): 165–175, doi: 10.3764/aja.125.1.0165.
- Csobanka, Emese. „The Z Generation“. *Acta Technologica Dubnicae* 6/2 (2016): 63–76, doi: 10.1515/atd-2016-0012.
- Drake, Louise, Fabian Gapp, Corinna Reinhardt, Luise Sauer, Laura Stiller, Ute Verstegen und Franziska Wanka. „Digitale Lehre in den archäologischen Wissenschaften an der Friedrich-Alexander-Universität Erlangen-Nürnberg – Erfahrungen nach einem Jahr Coronapandemie.“ *Blickpunkt Archäologie* 2 (2021): 103–113 (im Druck).
- Education Group (Hg.). *Oö. Jugend-Medien-Studie 2017: Das Medienverhalten der 11- bis 18-jährigen. Medienverhalten der Jugendlichen aus dem Blickwinkel der Jugendlichen*. 2017. Zugriff am 30.06.2021. https://www.edugroup.at/fileadmin/DAM/Innovation/Forschung/Dateien/JugendMedienStudie_ZusFassung_2017.pdf.
- Großkurth, Eva-Marie und Jürgen Handke (Hg.). *Inverted classroom and Beyond: Lehren und Lernen im 21. Jahrhundert*. 4. ICM-Fachtagung an der Philipps-Universität Marburg. Marburg: Tecum, 2016.
- Handke, Jürgen. *Patient Hochschullehre: Vorschläge für eine zeitgemäße Lehre im 21. Jahrhundert*. Marburg: Tecum, 2014.
- Handke, Jürgen. *Handbuch Hochschullehre digital: Leitfaden für eine moderne und mediengerechte Lehre*. Marburg: Tecum, 2015.
- Hofhues, Sandra, Mandy Schiefner-Rohs, Sandra Aßmann und Taiga Brahm (Hg.). *Studierende – Medien – Universität: Einblicke in studentische Medienwelten*. Münster, New York: Waxmann, 2020. Zugriff am 30.06.2021, <https://www.waxmann.com/index.php?eID=download&buchnr=4049>.
- King, Ellie, M. Paul Smith, Paul F. Wilson und Mark A. Williams. „Digital Responses of UK Museum Exhibitions to the COVID-19 Crisis, March – June 2020.“ *Curator: The Museum Journal* 2021: 1–18. doi: 10.1111/cura.12413.
- Kuhn, Axel und Svenja Hagenhoff, „Kommunikative statt objektzentrierte Gestaltung: Zur Notwendigkeit veränderter Lesekonzepte und Leseforschung für digitale Lesemedien,“ in

- Lesen X.0: Rezeptionsprozesse in der digitalen Gegenwart, hg. v. Sebastian Böck, Julian Ingelmann, Kai Matuszkiewicz und Friederike Schruhl (Göttingen: Vandenhoeck & Ruprecht): 27–45.
- Lauer, Gerhard. *Lesen im digitalen Zeitalter*. Darmstadt: wbg Academic, 2020. Zugriff am 30.06.2021. https://files.wbg-wissenverbindet.de/Files/Article/ARTK_ZOA_1024585_0001.pdf.
- Landau, Anne-Kathrin und Tom Renner. „Räume durch geographische Exkursionen und Fragen erschließen.“ *Hallesches Jahrbuch für Geowissenschaften* 41 (2018): 63–76.
- Löw, Martina. *Raumsoziologie*. Frankfurt a. M.: Suhrkamp, 2017.
- Medienpädagogischer Forschungsverbund Südwest (mpfs) (Hg.). *JIM 2017. Jugend, Information, (Multi-) Media Basisstudie zum Medienumgang 12- bis 19-Jähriger in Deutschland*. 2017. Zugriff am 30.06.2021. https://www.mpfs.de/fileadmin/files/Studien/JIM/2017/JIM_2017.pdf.
- Medienpädagogischer Forschungsverbund Südwest (mpfs) (Hg.). *JIM 2020: Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger*. 2021. Zugriff am 30.06.2021. https://www.mpfs.de/fileadmin/files/Studien/JIM/2020/JIM-Studie-2020_Web_final.pdf.
- Mertens, Claudia, Fabian Schumacher, Oliver Böhm-Kasper und Melanie Basten. „To flip or not to flip?“ Empirische Ergebnisse zu den Vor- und Nachteilen des Einsatzes von Inverted-Classroom-Konzepten in der Lehre.“ In *Hochschullehre als reflektierte Praxis. Fachdidaktische Fallbeispiele mit Transferpotenzial*, hg. v. Tobias Schmohl und Kieu-Anh To, 13–28. Bielefeld: Bertelsmann, 2019.
- Mührenberg, Lara. „CA 2.x – Christliche Archäologie im inverted classroom.“ In *Vom Untergrund ins Internet: Die Institute für Ur- und Frühgeschichte, Klassische Archäologie und Christliche Archäologie mit dem gemeinsamen Studiengang „Archäologische Wissenschaften“ an der Friedrich-Alexander-Universität Erlangen-Nürnberg – Stand 2018/19*, hg. v. Doris Mischka, Andreas Grüner, Corinna Reinhardt, Thorsten Uthmeier und Ute Verstegen, 65–67. Erlangen: Friedrich-Alexander-Universität, 2019.
- Mührenberg, Lara und Ute Verstegen. „CA 2.x – Christliche Archäologie im inverted classroom: Ein Beitrag zur videobasierten digitalen Lehre an der Friedrich-Alexander-Universität Erlangen-Nürnberg.“ In *Inverted Classroom: Past, present and future. 8. ICM Konferenz am 26. und 27. Februar 2019 in Marburg*, hg. v. Sabrina Zeaiter und Jürgen Handke, 165–177. Marburg: Tecum, 2020.
- Neeb, Kerstin. *Geographische Exkursionen im Fokus empirischer Forschung: Analyse von Lernprozessen und Lernqualitäten kognitivistisch und konstruktivistisch konzeptionierter Schülerexkursionen*. Zugl.: Gießen, Univ., Diss., 2010. Geographiedidaktische Forschungen 50. Weingarten: Selbstverl. des Hochschulverbandes für Geographie und ihre Didaktik, 2012.
- Nolte, Eva und Karsten Morisse. „Inverted Classroom: Eine Methode für vielfältiges Lernen und Lehren?“ In *Teaching Trends 2018: Die Präsenzhochschule und die digitale Transformation*. Digitale Medien in der Hochschullehre 7, hg. v. Susanne Robra-Bissantz und Oliver J. Bott, 105–112. Münster: Waxmann, 2019.
- Roman Herzog Institut e.V. (Hg.). *Neue Werte, neue Gesellschaft, neue Arbeitswelt? Die Generation von Morgen*. 2016. Zugriff am 30.06.2021. <https://www.romanherzoginstitut.de/publikationen/detail/die-neue-generation>.
- Rygiel, Philippe. *Historien à l'âge numérique: Essai*. Villeurbanne: Presses de l'Esssib, 2017. Zugriff am 30.06.2021. <https://books.openedition.org/pressesenssib/6303>.

- Scholz, Christian, „Wie ‚ticken‘ Studierende der Generation Z?“ In *Neue Studierendengeneration, neue Herausforderungen: Fachübergreifende Modelle, Module, Maßnahmen. Eine Dokumentation der Fachtagung der Studierendenakademie der Heinrich-Heine-Universität Düsseldorf*, hg. v. Holger Ehlert, 15–26. Düsseldorf: düsseldorf university press, 2017.
- Schön, Sandra und Martin Ebner. *Gute Lernvideos: So gelingen Web-Videos zum Lernen!* Nordestedt: Books on Demand, 2013.
- Shell (Hg.). *Shell-Jugendstudie 2015: Zusammenfassung*. Hamburg 2015. Zugriff am 30.06.2021. https://www.bundestag.de/resource/blob/632026/7558ce0d8a172b47f78c770dfb35d6ac/KDrs-19-28-38_Broschuere-Shell-Jugendstudie-data.pdf.
- Taçgîn, Zeynep. *Virtual and Augmented Reality: An Educational Handbook*. Cambridge: Cambridge Scholars Publishing, 2020.
- Thompson, Laura J., Brian Krienke, R. B. Ferguson und Joe Luck. „Using 360-degree video for immersive learner engagement.“ *Journal of Extension* 56/5 (2018). Zugriff am 30.06.2021. <https://tigerprints.clemson.edu/joe/vol56/iss5/1/>.
- Triple-A-Team-AG, Hg. *Generation Z: Metastudie über die kommende Generation*. 2016. Zugriff am 30.06.2021. http://tripleateam.ch/wp-content/uploads/2016/06/Generation_Z_Meta studie.pdf.
- Ullmann, Jan. „Entwicklung von Erklärvideos für einen Englisch Selbstlernkurs im Rahmen des ‚Flipped Classroom‘ Prinzips“. Diss., LMU München, 2018, doi: 10.5282/edoc.22645.
- Verstegen, Ute. „Abschied vom Dia – eine Reprise. Digitale Repräsentationen dreidimensionaler Bauten und Objekte in der kunsthistorischen Lehre“. *Kunstchronik* 58 (2005): 347–348.
- Verstegen, Ute. „Vom Mehrwert digitaler Simulationen dreidimensionaler Bauten und Objekte in der architekturgeschichtlichen Forschung und Lehre.“ Vortrag auf dem XXIX. Deutschen Kunsthistorikertag, Regensburg, 18. März 2007, doi: 10.11588/artdok.00000473.
- Verstegen, Ute. *Kleine Fächer im Porträt: Christliche Archäologie* 2019. Zugriff am 30.06.2021. <https://www.kleinefaecher.de/beitraege/blogbeitrag/christliche-archaeologie.html>.
- Wampfler, Philippe. *Generation „Social Media“: Wie digitale Kommunikation Leben, Beziehungen und Lernen Jugendlicher verändert*. Göttingen: Vandenhoeck & Ruprecht, 2014.
- Wissenschaft im Dialog gGmbH (Hg.). *Wissenschaftsbarometer 2020: Ergebnisse nach Subgruppen*. Berlin, 2020. Zugriff am 30.06.2021. https://www.wissenschaft-im-dialog.de/fileadmin/user_upload/Projekte/Wissenschaftsbarometer/Dokumente_20/20_WiD_Tabel lenband_Wissenschaftsbarometer2020_final.pdf.
- Zeaiter, Sabrina und Jürgen Handke (Hrsg.). *Inverted Classroom: Past, present and future*. 8. ICM Konferenz am 26. und 27. Februar 2019 in Marburg. Marburg: Tecum, 2020.
- Zeaiter, Sabrina und Peter Franke (Hrsg.). *Pioniergeist, Ausdauer, Leidenschaft: Festschrift zu Ehren von Prof. Dr. Jürgen Handke*. Marburg: Universitätsbibliothek, 2020, doi: 10.17192/es2020.0018.

René Smolarski, Hendrikje Carius, Marcus Plaul

Perspectives and Challenges of Historical Research with Citizen Participation: A Critical Reflection on the Example of “Cinema in the GDR”

Abstract: Participatory research formats have recently gained in importance in historical research. This is particularly reflected in the increase in historically oriented citizen science projects. However, this development has so far received little attention in the relevant research literature and, especially in the German-speaking world, remains at best a niche phenomenon. Thus, the aim of this article is to encourage academic debate on the relevance of citizen science in historical science, with a focus on digital-based approaches. The article begins with the definition of citizen science and then examines its positioning in humanities and historical science, which is followed by a discussion of the potentials and limitations of participatory approaches based on the research project “Cinema in the GDR”. The paper draws on cross-disciplinary and discipline-specific studies, initial project-side evaluations, and experiences from the project implementation. Finally, the article asks under which prerequisites digital-based citizen science approaches could gain long-term acceptance as a research method in historical science.

Keywords: citizen science, digital citizen science, crowdsourcing, public history, cinema history, history of everyday life in the GDR, knowledge and science communication

Zusammenfassung: In der geschichtswissenschaftlichen Forschung haben in jüngster Zeit partizipative Forschungsformate stark an Bedeutung gewonnen. Dies spiegelt sich vor allem in einer Zunahme historisch orientierter Citizen Science-Projekte wider. Diese Entwicklung findet bislang jedoch kaum Beachtung in der einschlägigen Forschungsliteratur und führt vor allem im deutschsprachigen Raum allenfalls ein Nischendasein. Ziel des vorliegenden Beitrags ist daher, die fachwissenschaftliche Debatte über die Tragfähigkeit bürgerbeteiligter Forschung in den Geschichtswissenschaften anzuregen und rückt dabei insbesondere digitalbasierte Zugänge in den Fokus. Ausgehend von einer Definition des Citizen Science-Begriffs und dessen Verortung in den Geistes- und Ge-

schichtwissenschaften werden anhand des Forschungsprojektes “Kino in der DDR” die Potenziale und Grenzen partizipativer Ansätze herausgearbeitet. Der Beitrag stützt sich sowohl auf disziplinübergreifende als auch einige wenige disziplingebundene Untersuchungen, erste projektseitige Evaluationen und Erfahrungen aus der Projektumsetzung. Abschließend wird über die Frage diskutiert, unter welchen Voraussetzungen sich digitalbasierte Citizen Science-Ansätze nachhaltig als Forschungsmethode in den Geschichtswissenschaften etablieren könnten.

Schlagwörter: Citizen Science, digitale Citizen Science, Crowdsourcing, Public History, Kinogeschichte, Alltagsgeschichte der DDR, Wissen, Wissenschaftskommunikation

1 Introduction

First coined in the mid-1990s in the Anglo-Saxon world, the label of *citizen science*¹ and related terms such as citizen social science, citizen humanities or crowdsourcing (which is more limited with regard to the degree of participation) have been used to summarise approaches that aim to involve an interested public in the scientific process of discovery. Since their beginnings, these formats have continued to differentiate their objectives in the sense of a public engagement with science². In addition to forms of dialogical exchange about science (for example science cafés, citizens’ conferences and round tables), they also include opportunities of direct cooperation and collaboration in order to generate new scientific knowledge.³ Beyond their use in the natural sciences, they have recently also found their way into the historical sciences – with a close interrelation to research methods and perspectives such as public history, digital history or participatory research in general.⁴ Consequently, this corresponds

1 Cf. Alan Irwin, *Citizen Science: A Study of People, Expertise and Sustainable Development* (London: Routledge, 1995).

2 Cf. Jack Stilgoe, Simon J. Lock and James Wilsdon, “Why Should We Promote Public Engagement with Science?” *Public Understanding of Science. Special Issue: Public Engagement in Science* 23/1 (2014): 4–15, doi: 10.1177/0963662513518154.

3 Cf. Birte Fährlich, “Wissenschaftsevents zwischen Popularisierung, Engagement und Partizipation,” in *Forschungsfeld Wissenschaftskommunikation*, ed. Heinz Bonfadelli et al. (Wiesbaden: Springer VS, 2015), 170–171.

4 For the topic of citizen science in the humanities, see authors such as Kristin Oswald and René Smolarski, “Einführung: Citizen Science in Kultur und Geisteswissenschaften,” in *Bürger*

with the increasing significance of public science within a knowledge society in which the communication and transfer of knowledge acts as a central component of research. Seen in this context, science and the public sphere are understood to be in a dialogue-oriented relationship, one that goes beyond the unilateral social information and educational function of research. These efforts are accompanied in a special way by the new possibilities of digital technologies and the open science paradigm (*open citizen science*⁵). In their combination, both components appear as conditional factors of a citizen science oriented towards a digitally-based interlocking of science and society. On the one hand, technical innovations and new digital media make it easier for citizen scientists to participate in citizen science projects. On the other hand, digital infrastructures and technologies for data analysis and processing provide the basis for a more efficient and sustainable handling of the collected data.⁶ In the meantime, a multitude of projects that function exclusively as web-based digital citizen science have evolved.⁷

In the German context, citizen science has established itself through initiatives such as the *GEWISS Consortium of Bürger schaffen WISSEN*, especially in the areas of natural sciences.⁸ However, even if it is possible in the historical sciences to refer to traditional lines of voluntary engagement by citizens in historical associations or historical workshops,⁹ the prerequisites, basic conditions

Künste Wissenschaft – Citizen Science in Kultur und Geisteswissenschaften, ed. Kristin Oswald and René Smolarski (Guttenberg: Computus, 2016) or Cord Arendes, “Historiker als ‘Mittler zwischen den Welten’? Produktion, Vermittlung und Rezeption historischen Wissens im Zeichen von Citizen Science und Open Science,” *Heidelberg Jahrbücher Online* 2 (2017): 19–58, doi: 10.17885/heup.hdjbo.2017.23691.

5 Cf. Katrin Vohland and Claudia Göbel, “Open Science and Citizen Science als symbiotische Beziehung? Eine Gegenüberstellung von Konzepten,” *TATuP. Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis* 26/1–2 (2017): 18–24, doi: 10.14512/tatup.26.1-2.18.

6 Cf. Rick Bonney et al., “Next Steps for Citizen Science,” *Science* 343/6178 (2014): 1436–1437, doi: 10.1126/science.1251554.

7 Cf. Tobias Füchslin, “What are you folding for? Nutzungsmotive von Citizen Science Online Games und ihre Lerneffekte,” *merzWissenschaft – Magazine for Medienpädagogik* (2016): 110–122, doi: 10.5167/uzh-131685.

8 The consortium of Citizens Create Knowledge is funded by the German Federal Ministry for Education and Research (BMBF) and coordinated by the Museum of Natural History Berlin – Leibniz Institute for Evolution and Biodiversity Research (MfN), as well as the Helmholtz Centre for Environmental Research (UFZ).

9 Cf. authors such as Karen Bruhn, Thorge Petersen and Swantje Piotrowski, “Offen, vernetzt und grenzenlos. Zu den Chancen und Potentialen der Digital Humanities in der Regionalgeschichte,” in *Regionalgeschichte. Potentiale des historischen Raumbezugs*, ed. Nina von Gallion et al. (Göttingen: Vandenhoeck & Ruprecht, 2021); Andrea Stieldorf, “Geschichtsvereine: Ihre Bedeutung für die Landesgeschichte,” in *Landesgeschichte und public history*, ed. Arnd

and potentials of the digitally based citizen science approaches have only been explored in a rudimentary way in the humanities and especially the historical sciences. This situation requires further systematic methodical-epistemological reflection and an interdisciplinary exchange.¹⁰ In addition, this finding applies to the definition of citizen science as it is also formulated in the GEWISS Consortium's Green Paper on Citizen Science Strategy 2020. According to this paper, citizen science is understood in the broader sense as the participation of the (non-expert) public in the various phases of the scientific work process under adherence to scientific standards.¹¹ This approach addresses every degree in the participation of citizens engaged outside of the university. These can range from cooperation, collaboration and co-production as the most frequent forms of participation (data collection and processing) to active participation in the research process in the sense of its co-design (inclusion in the development of research questions and data evaluation).¹² In the ultimately highly differentiated citizen science practice of historically oriented research institutes, professional societies, associations/organisations and collection-holding institutions, citizen science projects are primarily found in the fields of archaeology, preservation of historical monuments, genealogy, local and regional history, philately and even art history, cultural history and the history of knowledge. In the sense of crowdsourcing, the most frequent possibilities for participation in various areas of the research process are the provision of historical sources (for example Europeana 1914–1918¹³), image and text classification (for example Artigo¹⁴ and

Reitemeier (Ostfildern: Thorbecke, 2020), 103–112. Also see Thorsten Hoppe-Hartmann, “Citizen Science – Ein neues Konzept für die Geschichtswissenschaft oder ein alter Hut?”, #historisierend. Karls Geschichte(n), Last update on 26/09/2018, <https://geschichten.hypotheses.org/48>, accessed July 20, 2021.

10 A collected volume called *Citizen Science in the Geschichtswissenschaften* on this topic is in preparation. It will be published in 2022 in the open access series *DH&CS. Schriften des Netzwerks für digitale Geisteswissenschaften und Citizen Science*.

11 “Citizen science describes the participation of persons in scientific processes without being bound to this area of science through an institution. This participation may include the short-term collection of data and range up to an intensive investment of leisure time to become more deeply absorbed in a research topic together with scientists and/or other volunteers. [...] The important factor is [...] the adherence to scientific standards, which primarily means transparency with regard to the methods of data collection and public discussions of the results.” Cf. GEWISS, *Grünbuch Citizen Science Strategie 2020* (Berlin: Bürger schaffen Wissen – Wissen schafft Bürger project, 2016), 13. urn:nbn:de:101:1-20160621985. (Translated by the authors).

12 Cf. Melissa V. Eitzel et al., “Citizen Science Terminology Matters: Exploring Key Terms,” *Citizen Science: Theory and Practice* 2/1 (2017): 1, doi: 10.5334/cstp.96.

13 Europeana 1914–1918, <http://europeana1914-1918.eu/>.

14 Artigo, <https://www.artigo.org/>.

Tag.Check.Score¹⁵), transcription and annotation work (for example Transcribe Bentham, Transcribe Europeana 1914–1918, e-manuscripta or Letters 1916–1923¹⁶), as well as further forms of evaluation processes for historical collections and data (for example the provision of contemporary historical knowledge in the form of eyewitness interviews).

Against this outlined background and in view of the currently existing empirical findings from ongoing historically oriented citizen science projects, the article asks about the potential of citizen science as a method within the digital historical sciences and explores the challenges and possible boundaries of this approach in historical research. At the same time, the viability of citizen science approaches that extend beyond crowdsourcing is subjected to debate with the functional conditions of humanistic citizen science communication. A special focus is directed towards the conceptual questions regarding cooperation, interaction and the negotiation of roles between the individual actors (community building, expectations and the culture of appreciation). The basis for these considerations is the example of an interdisciplinary project that is establishing a cross-project citizen science platform for the humanities and has been located at the University of Erfurt since 2019. Within the scope of the pilot project on the everyday history of the GDR, *Kino in der DDR – Rezeptionsgeschichte “von unten”*¹⁷, a modular infrastructure is being developed that can be applied to various project contexts and very specific questions. In terms of the methodology, the approaches of digital humanities and citizen science, as well as the specialised methods of communication science, computer science and historical sciences are intertwined.

¹⁵ Fraunhofer, https://www.fokus.fraunhofer.de/go/tag_check_score, accessed July 20, 2021.

¹⁶ Transcribe Bentham, <https://blogs.ucl.ac.uk/transcribe-bentham/>, Transcribathon, <https://transcribathon.com/en/>, <https://www.e-manuscripta.ch/>, Letters 1916–1923, <http://letters1916.maynoothuniversity.ie/>.

¹⁷ The project is funded from 2019 to 2022 by the Thüringer Aufbaubank. Cf. <https://projekte.uni-erfurt.de/ddr-kino>, accessed July 20, 2021. Hendrikje Carius, Christiane Kuller, Patrick Rössler and René Smolarski, “Development of a Cross-Project Citizen Science Platform for the Humanities,” in *Digital Humanities Austria 2018: Empowering Researchers*, ed. Marlene Ernst et al., 79–82, doi: 10.1553/dha-proceedings2018s79.

2 “The wisdom of the many”: Potentials of research with citizen participation in the historical sciences

The potentials of citizen science and especially the crowdsourcing-based approaches have already been worked out for the humanities and social sciences from various sides.¹⁸ The emphasis here is always on the cooperation of institutionalised and non-institutionalised science offering an added value for both the scientific participants and public concerns and concrete application issues.¹⁹ This also applies to historical research.

From the viewpoint of university-based historical sciences, the integration of an interested non-institutionalised public into the scientific process at various levels of citizen scientist participation – from crowdsourcing to citizen science – promises not only an integration of previously inaccessible historical source material²⁰ or analysis, organisation and transcription, but also the creation of new research perspectives and interdisciplinary problem-solving approaches.²¹

In the case of the “Cinema in the GDR” project currently conducted at the University of Erfurt, the focus is primarily on what Peter Finke somewhat pejoratively called “citizen science light”²², which primarily refers to various crowd-

18 A summary can be found in Claudia Göbel, Justus Henkel and Sylvi Mauermeister, *Kultur und Gesellschaft gemeinsam erforschen. Überblick und Handlungsoptionen zu Citizen Science in den Geistes- und Sozialwissenschaften* (Halle-Wittenberg: Institute for Higher Education Research (HoF) at the Martin Luther University of Halle-Wittenberg, 2020, 25–29).

19 Cf. Göbel, Henkel and Mauermeister, *Kultur*, 25.

20 In addition to the project “Cinema in the GDR” discussed here, main examples can be found in the area of philately: The philatelic source material upon which it is based generally only exist outside of the university research landscape and requires their own source-specific methodical approaches. For more information see Pierre Smolarski and René Smolarski, “Wissenschaftliches Stiefkind und amateurhafte Liebhaberei: Die Philatelie als historische Grundwissenschaft,” in *Die Historischen Grundwissenschaften heute. Tradition – Methodische Vielfalt – Neuorientierung*, ed. Étienne Doublier, Daniela Schulz and Dominik Trumpf, 95–119 (Köln/Weimar/Wien: Böhlau, 2020).

21 Cf. Oswald and Smolarski, “Einführung”, 14. An example of the social science perspective is Stefan Thomas, Susan Schröder and David Scheller, “Citizen Social Science – das Research Forum als partizipative Forschungsmethodik”, in *Forschendes Lernen – Lernendes Forschen: partizipative Empirie in Erziehungs- und Sozialwissenschaften*, ed. Sandra Eck (Weinheim/Basel: Beltz Juventa, 2019), 103–115.

22 Peter Finke, *Citizen Science. Das unterschätzte Wissen der Laien* (München: oekom, 2014), 41.

sourcing concepts. On the one hand, these include the availability and utilisation of previously unconsidered source materials – in this case, primarily contemporary images and press material from cinema buildings and cinema events – as well as the compilation of concrete data on the various types of cinema and film screening venues that were actively used in the territory of the former GDR between 1949 and 1990 (see Figure 1).

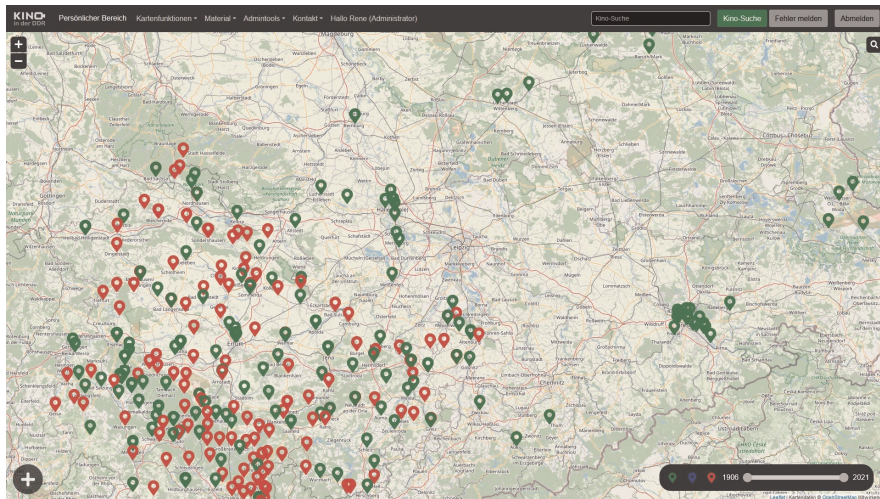


Fig. 1: Cinema locations recorded by the citizen scientists on the project page of “Cinema in the GDR” on the Citizen Science Platform (COSE) – status as of 31 May 2021.

Beyond these crowdsourcing components, the integration of non-institutional scientists also promises further potential for historical research in that accessibility, visibility and popularity of the respective research fields are increased through the high-profile implementation of digitally supported citizen science projects in particular. Citizen science is therefore always associated with the public communication of science. As already indicated, it is connected with the field of public history in the area of historical sciences as a result.²³

This public dimension of citizen science projects in particular offers the citizen scientists involved a high added value that extends beyond the mere “joy of

²³ Cf. René Smolarski and Sina Speit, “Was kann Public Interest Design aus Sicht der Geschichtswissenschaften sein?”, in *Was ist Public Interest Design?*, ed. Christoph Rodatz and Pierre Smolarski (Bielefeld: transcript, 2018), 95–112.

participating”²⁴. Especially participation in essentially historical sciences projects, which often meet with a great response from the public, promotes a critical approach to sources and better understanding for the complexity of social changes due to the accompanying teaching of methodical and theoretical competences.²⁵ Consequently, the public’s increased involvement beyond the realm of just crowdsourcing can also contribute to an examination of values and norms in society.²⁶ Above all, this means strengthening the social relevance of humanistic and/or historical knowledge beyond the boundaries of university-based historical science.²⁷

Furthermore, the communication of science in citizen science projects is not unidirectional. Instead, the university partners also receive direct feedback on their research questions and are sometimes confronted with critical questions that broaden their view of the topics that they are working on and may possibly even open up new perspectives.²⁸ For example, it also became apparent within the scope of the “Cinema in the GDR” project that certain aspects of the cinema in the GDR’s everyday history – such as the inclusion of film screening facilities at the companies and in the areas of the NVA (Nationale Volksarmee) had not been considered in the original concept of the project. Due to feedback from citizen scientists, it was possible to integrate these points into the course of the project.

Finally, the long-term positive effects of an intensified cooperation between university and non-university historical research should be mentioned here. After all, there is also hope for more sustainability and a younger generation for volunteer work and regional historical research beyond the direct added value for the individual projects. This occurs primarily through the networking of university historical research with the communities that work digitally such as the Verein für Computergenealogie (Association for Computer Genealogy, CompGen) and the various working groups under the umbrella of the Bund Deutscher Philatelisten (German Philatelic Association, BDPh).

Beyond mere crowdsourcing, the narratives associated with historical documents in society can also be addressed and the related issues and methodical

²⁴ Sonja Bettel, “Forschung zum Mitmachen”, *upgrade 1* (2016): 35.

²⁵ Cf. Oswald and Smolarski, “Einführung”, 10.

²⁶ Cf. Peter Finke, “Citizen Science und die Rolle der Geisteswissenschaften für die Zukunft der Wissenschaftsdebatte. Citizen Science in Kultur und Geisteswissenschaften”, in Oswald and Smolarski (eds.), *Bürger Künste Wissenschaft*, 32.

²⁷ Cf. Göbel, Henkel and Mauermeister, *Kultur*, 26, 60.

²⁸ Cf. Friederike Hendriks and Doris Niederhoff, “Was haben Wissenschaftler von Wissenschaftskommunikation? – Eine Interviewstudie” (article presented at the 10th Forum on Science Communication, Science in Dialogue, Braunschweig, 2017).

considerations integrated into the research as a result. This approach is especially relevant for the pilot project on GDR history since the legitimacy and recognition of different narratives is currently the subject of a debate that is particularly rife with conflict. The insights gained from the reflection on the dialogue between the scientific discipline and the social public are further developed within the project community on an interdisciplinary basis and flow into the technological and infrastructural implementations of the citizen science platform.

3 Digital project meets non-digital community: Challenges of a digital citizen science in the historical sciences

In addition to the added value and potentials of citizen science for science in general and the historical sciences in particular, the project initiators are also confronted with various challenges that must be taken into consideration when implementing the corresponding formats. For example, the citizen science approach is not quite uncontroversial in the academic discourse. In particular, critical voices question the potential attributed to it and are doubtful about its scientific usability and the quality of data collected with the help of lay researchers.²⁹

From the perspective of science, the question about the efficiency and effectiveness of inclusive citizen research also arises.³⁰ The integration of an interested public in the research process means considerable extra effort for the project collaborators. A mostly very heterogeneous group of citizen scientists must be acquired in a way that is effective for the public and appropriate for the target group, as well as motivated for the duration of the project. This requires a high degree of interaction between the scientists and the citizens. Lacking or inadequate communication, but also expectations that are too high on both sides, of science and lay researchers, can have a negative effect on the cooperation.

²⁹ Cf. Finke, *Citizen Science. Das unterschätzte Wissen der Laien*, 17.

³⁰ Cf. Nina Wicke, "Citizen Science – eine 'erfolgreiche' Entwicklung in der Wissenschaftskommunikation?," in *Intention und Rezeption von Wissenschaftskommunikation*, ed. Jutta Milde et al. (Köln: Herbert von Halem Verlag, 2021), 179.

In the critical discussion with citizen science, there are also complaints that the scientific debate has not yet extended beyond the abstract goal of public engagement with science³¹ and that there is a lack of concrete methods and guidelines on how a higher degree of participation and dialogue between the public and science could be achieved or even what the terms of participation and engagement actually mean.³² In addition, citizen science projects are also increasingly taking place on the internet in the course of technological change³³, and this also results in new requirements for a digital science communication – the use and impact of which has hardly been researched to date.³⁴

In view of the above-mentioned problem areas, the initiators of the “Cinema in the GDR” citizen science project also faced various challenges on multiple levels. Based on the main research question of reconstructing an everyday history of the motion picture in the GDR, the project is primarily aimed at persons who were socialised in the former GDR and who have conscious memories of cinema experiences in the GDR or can provide evidence relevant to the research such as diaries, programme booklets or photographs. In view of the resulting age structure of the citizens to be involved in the research process and the project’s chosen digital access via a virtual research environment, this requires a specific communication strategy that incorporates both non-digital and digital channels of knowledge and science communication.³⁵

With regard to the design of the media offers, deliberate attention was paid to science communication that is appropriate for the target audience. This is based not only on the media usage behaviour of the target group to be reached,³⁶ but also relies on corresponding openly, dialogue and participation-oriented formats in accordance with the requirements of the GEWISS Consortium.³⁷ For the first time, a non-university public was incorporated into the

31 Cf. Stilgoe, Lock and Wilsdon, “Public Engagement with Science”, 4–11.

32 Cf. Fähnrich, “Wissenschaftsevents”, 172–177.

33 Cf. Fuchslin, “Nutzungsmotive”, 110–111.

34 Cf. GEWISS, *Grünbuch Citizen Science Strategie 2020*, 24.

35 Cf. Anna-Rosa Haumann and René Smolarski, “Digital project meets analog community. Expectations and experience of a digital citizen science project on GDR history,” in *Austrian Citizen Science Conference 2020. 14–16 September 2020. Conference Proceedings* (2021), accessed July 20, 2021, <https://pos.sissa.it/393/010/pdf>.

36 The starting point for the online usage behaviour of the target groups to be addressed was formed by the results of the ARD/ZDF online study on the use of classes on the internet, as well as social media offers in the age groups of 50 years and above; cf. Natalie Beisch, Wolfgang Koch and Carmen Schäfer, “ARD/ZDF-Onlinestudie 2019: Mediale Internetnutzung und Video-on-Demand gewinnen weiter an Bedeutung,” *Media Perspektiven* 9 (2019): 382–383.

37 Cf. GEWISS, *Grünbuch Citizen Science Strategie 2020*, 24.

project activities just shortly after the research project's start in November 2019. This occurred in the form of a kick-off event in Erfurt with the aim of specifically stimulating the exchange between scientists and interested citizens. Within this context, a survey was conducted among the participants to identify further potentials for supporting the project communication. About one-third (32 percent) of the respondents wanted to have additional (on-site) events within the scope of the research project, and another 32 percent wished to be informed about the progress of the project by means of newsletters. About one-fifth of the respondents (22 percent) indicated that they wanted to use the project website as an information platform, in contrast with the nine percent who opted for social media, for example Twitter or Facebook.³⁸

Although the collected data is by no means representative, it shows that the adopted strategy of covering a relatively broad spectrum of communication offered in online and offline areas appears to be quite necessary and expedient. In addition to the classic channels of science communication such as press relations, events and workshops, this also includes online activities such as posts on Twitter and Facebook and preparing a blog to accompany the project³⁹. The latter already went online in October 2019, almost one year before the launch of the technical citizen science platform, in order to draw the attention of an interested public to the project at any early stage. Since then, the blog has been updated with new content on a regular basis. Among other things, it provides help on how to use the platform, information on events and historical contributions to the GDR's history of cinema and film. In addition, interim results of the project are published on the blog and it offers citizen scientists the possibility of presenting their own research findings or eyewitness accounts. The project's Twitter and Facebook channels make it possible to share the corresponding blog contributions with the community and successively expand the network of interested lay researchers. The offers initiated by the project leaders are aimed not only at potential citizen scientists, but also at possible multipliers of professional media work and existing citizen science communities in the German-

38 The brief survey was conducted during the kick-off event of the "Cinema in the GDR" project on 16th November 2019 in Erfurt. The original question was: "Through which channels would you prefer to stay in contact with the project? (multiple answers possible)" and got responses from 26 participants.

39 "Cinema in the GDR" project blog, University of Erfurt, <https://projekte.uni-erfurt.de/ddr-kino/>.

speaking countries (above all: Bürger schaffen Wissen⁴⁰, Österreich forscht⁴¹, Schweiz forscht⁴²).

Through its communication activities to date, the “Cinema in the GDR” citizen scientist platform has been able to register around 120 citizen scientists since its online launch in October 2020. They have made around 400 cinema entries as a group (status as of July 2021). In addition, there have been more than 70 inquiries by interested citizens who would like to contribute to the project as contemporary witnesses outside of the online platform. These include not only cinema-goers, but also projectionists, cashiers, cinema owners and filmmakers from the former GDR.

In order to further investigate the limits and possibilities of the digitally based citizen science approaches, it is necessary to increasingly evaluate corresponding formats. With regard to the “Cinema in the GDR” project, the initiators selected a evaluation process that accompanies the project. Through all project phases, it documents and analyses the viewpoints of both the project leaders and the citizen scientists. In addition to the direct feedback rounds on events and on the citizen science platform, this also includes quantitative and qualitative surveys, informal and formal discussion rounds with participants, as well as the letters submitted. Finally, by means of a systematic processing of the material, the objective is to examine to what extent the participatory “Cinema in the GDR” project can produce reliable results and whether the set expectations on the part of the scientists and lay researchers can be fulfilled.

4 Perspectives of digitally based citizen science approaches for the historical sciences

Digitally based citizen science approaches that go beyond crowdsourcing and involve the public in the scientific process appear to be quite interesting approaches for the historical sciences, based on the collected experiences in projects such as “Cinema in the GDR” or in other areas such as philately⁴³. However, a number of questions still need to be discussed for the localisation and perspective of citizen science in historical research. On the one hand, these are

⁴⁰ Bürger schaffen Wissen, <https://www.buergerschaffewissen.de>.

⁴¹ Österreich forscht, <https://www.citizen-science.at>.

⁴² Schweiz forscht, <https://www.schweizforscht.ch/>.

⁴³ Cf. Pierre Smolarski, René Smolarski and Silke Vetter-Schultheiss, ed., *Gezähnte Geschichte. Die Briefmarke als historische Quelle* (Göttingen: V&R unipress, 2019).

related to conceptual topics such as the interaction and negotiation of various roles between the actors involved (community building, expectations and culture of recognition) and the existing cooperation possibilities. On the other hand, these also relate to basic ethical-legal conditions. Moreover, discussion is required on what possibilities the existing research landscape offers at all for stimulating an interested public to systematically develop its own citizen scientist projects and/or develop its own historical questions on the source material acquired in the crowdsourcing projects. In addition, there is a need for a fundamental understanding of the premises along which digital infrastructures are conceptualised and implemented for citizen science projects. This also applies to the inclusion of citizen scientists in the development of not only historical questions but also the citizen science tools themselves. The design of digital infrastructures for citizen science projects is closely associated with the solutions generally required in digital historical sciences such as ensuring data quality and curation or generally considering FAIR data principles.⁴⁴ Interface design and user guidance require continuous processes of review and iterative development. If these components of digital citizen science infrastructure aim at a hermeneutically reflective reduction of complexity in the practice, the development of data literacy competences of all actors in digital citizen science is extremely essential. Further-reaching perspectives such as these may then provide the connecting links for actually establishing a structural change in the historical sciences with the systematic integration of citizens.⁴⁵

Bibliography

- Arendes, Cord. "Historiker als 'Mittler zwischen den Welten'? Produktion, Vermittlung und Rezeption historischen Wissens im Zeichen von Citizen Science und Open Science." *Heidelberger Jahrbücher Online* 2 (2017): 19–58. doi: 10.17885/heup.hdjbo.2017.23691.
- Beisch, Natalie, Wolfgang Koch, and Carmen Schäfer. "ARD/ZDF-Onlinestudie 2019: Mediale Internetnutzung und Video-on-Demand gewinnen weiter an Bedeutung." *Media Perspektiven* 9 (2019): 374–388.
- Bettel, Sonja. "Forschung zum Mitmachen." *upgrade* 1 (2016): 33–35.
- Bonney, Rick et al. "Next Steps for Citizen Science." *Science* 343/6178 (2014): 1436–1437. doi: 10.1126/science.1251554.
- Bruhn, Karen, Thorge Petersen, and Swantje Piotrowski. *Offen, vernetzt und grenzenlos. Zu den Chancen und Potentialen der Digital Humanities in der Regionalgeschichte*. In *Regio-*

⁴⁴ FORCE11, The FAIR Data Principles, <https://www.force11.org/group/fairgroup/fairprinciples>.

⁴⁵ Arendes, "Historiker", 51.

- nalgeschichte. Potentiale des historischen Raumbezugs*, edited by Nina von Gallion et al., 73–94. Göttingen: Vandenhoeck & Ruprecht, 2021.
- Carius, Hendrikje, Kuller, Christiane, Rössler, Patrick, and René Smolarski. “Development of a Cross-Project Citizen Science Platform for the Humanities.” In *Digital Humanities Austria 2018: Empowering Researchers*, edited by Marlene Ernst, Peter Hinkelmanns, Lina Maria Zangerl, and Katharina Zeppezauer-Wachauer, 79–82. doi: 10.1553/dha-proceedings2018s79.
- Eitzel, Melissa V. et al. “Citizen Science Terminology Matters: Exploring Key Terms.” *Citizen Science: Theory and Practice* 2/1 (2017): 1. doi: 10.5334/cstp.96.
- Fährnich, Birte. “Wissenschaftsevents zwischen Popularisierung, Engagement und Partizipation.” In *Forschungsfeld Wissenschaftskommunikation*, edited by Heinz Bonfadelli, Birte Fährnich, Corinna Luthje, Jutta Milde, Markus Rhomberg, and Mike S. Schäfer, 165–182. Wiesbaden: Springer VS, 2015.
- Finke, Peter. *Citizen Science. Das unterschätzte Wissen der Laien*. München: oekom, 2014.
- Finke, Peter. “Citizen Science und die Rolle der Geisteswissenschaften für die Zukunft der Wissenschaftsdebatte. Citizen Science in Kultur und Geisteswissenschaften.” In *Bürger Künste Wissenschaft. Citizen Science in Kultur und Geisteswissenschaften*, edited by Kristin Oswald and René Smolarski, 31–56. Gutenberg: Computus, 2016.
- Füchslin, Tobias. “What are you folding for? Nutzungsmotive von Citizen Science Online Games und ihre Lerneffekte.” *merzWissenschaft – Zeitschrift für Medienpädagogik* (2016): 110–122. doi: 10.5167/uzh-131685.
- GEWISS. *Grünbuch Citizen Science Strategie 2020*. Berlin: Bürger schaffen Wissen – Wissen schafft Bürger, 2016. urn:nbn:de:101:1-20160621985.
- Göbel, Claudia, Justus Henke, Sylvi Mauermeister, and Verena Plümpe. *Citizen Science jenseits von MINT: Bürgerforschung in den Geistes- und Sozialwissenschaften*. Halle-Wittenberg: Institute for Higher Education Research (HoF) at the Martin Luther University of Halle-Wittenberg, 2020.
- Göbel, Claudia, Justus Henke, and Sylvi Mauermeister. *Kultur und Gesellschaft gemeinsam erforschen. Überblick und Handlungsoptionen zu Citizen Science in den Geistes- und Sozialwissenschaften*. Halle-Wittenberg: Institute for Higher Education Research (HoF) at the Martin Luther University of Halle-Wittenberg, 2020.
- Haumann, Anna-Rosa, and René Smolarski. “Digital project meets analog community: Expectations and experiences of a digital citizen science project on GDR history.” In *Austrian Citizen Science Conference 2020. 14–16 September 2020. Conference Proceedings* (2021). Accessed July 20, 2021, <https://pos.sissa.it/393/010/pdf>.
- Hendriks, Friederike, and Doris Niederhoff. “Was haben Wissenschaftler von Wissenschaftskommunikation? – Eine Interviewstudie.” Article presented at the 10th Forum on Science Communication, Science in Dialogue, Braunschweig 2017.
- Hoppe-Hartmann, Thorsten. *Citizen Science – Ein neues Konzept für die Geschichtswissenschaft oder ein alter Hut?*, #historisierend. Karls Geschichte(n), Last update on 26/09/2018. Accessed July 20, 2021. <https://geschichten.hypotheses.org/48>.
- Irwin, Alan. *Citizen Science: A Study of People, Expertise and Sustainable Development*. London: Routledge, 1995.
- Oswald, Kristin, and René Smolarski. “Einführung: Citizen Science in Kultur und Geisteswissenschaften.” In *Bürger Künste Wissenschaft – Citizen Science in Kultur und Geisteswissenschaften*. Edited by Kristin Oswald and René Smolarski, 9–27. Gutenberg: Compu-

- tus, 2016. https://www.computus-druck.com/press/wp-content/uploads/2016/11/isbn_9783940598325.pdf.
- Smolarski, René and Sina Speit. "Was kann Public Interest Design aus Sicht der Geschichtswissenschaften sein?" In *Was ist Public Interest Design*, edited by Christoph Rodatz, and Pierre Smolarski, 95–112. Bielefeld: transcript, 2018.
- Smolarski, Pierre, René Smolarski, and Silke Vetter-Schultheiss, ed. *Gezähnte Geschichte. Die Briefmarke als historische Quelle*. Göttingen: V&R unipress, 2019.
- Smolarski, Pierre and René Smolarski, "Wissenschaftliches Stiefkind und amateurhafte Liebhaberei: Die Philatelie als historische Grundwissenschaft." In *Die Historischen Grundwissenschaften heute. Tradition – Methodische Vielfalt – Neuorientierung*, edited by Étienne Doublier, Daniela Schulz, and Dominik Trump, 95–119. Köln/Weimar/Wien: Böhlau, 2020.
- Stieldorf, Andrea. "Geschichtsvereine: Ihre Bedeutung für die Landesgeschichte. In *Landesgeschichte und public history*," edited by Arnd Reitemeier, 103–112. Ostfildern: Thorbecke, 2020.
- Stilgoe, Jack, Simon J. Lock, and James Wilsdon. "Why Should We Promote Public Engagement with Science?" *Public Understanding of Science. Special Issue: Public Engagement in Science* 23/1 (2014): 4–15. doi: 10.1177/0963662513518154.
- Surowiecki, James. *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Doubleday, 2004.
- Thomas, Stefan, Susan Schröder, and David Scheller. "Citizen Social Science – das Research Forum als partizipative Forschungsmethodik." In *Forschendes Lernen – Lernendes Forschen: partizipative Empirie in Erziehungs- und Sozialwissenschaften*, edited by Sandra Eck, 103–115. Weinheim/Basel: Beltz Juventa, 2019.
- Vohland, Katrin, and Claudia Göbel. "Open Science und Citizen Science als symbiotische Beziehung? Eine Gegenüberstellung von Konzepten." In *TATuP. Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis* 26/1–2 (2017): 18–24. doi: 10.14512/tatup.26.1-2.18.
- Wicke, Nina. "Citizen Science – eine 'erfolgreiche' Entwicklung in der Wissenschaftskommunikation?" In *Intention und Rezeption von Wissenschaftskommunikation*, edited by Jutta Milde, Ines C. Welzenbach-Vogel, and Maren Dern, 177–206. Köln: Herbert von Halem Verlag, 2021.

Mareike König und Paul Ramisch

Die twitternde Zunft

Historikertage auf Twitter (2012–2018)

Abstract: This article presents different perspectives on a corpus of tweets published under the respective conference hashtags during the four Historikertage from 2012 to 2018. The focus of the research is on Twitter practices and their development, the content of the tweets as well as networks of the tweeting historians in an academic conference situation over six years. For the first time, the historian community is subject of a Twitter study. Special features of the study are the relatively long period of time compared to other analyses, which is examined selectively on the basis of the Historikertage, as well as gender-specific evaluations made possible by open coding.

Keywords: historical research, Twitter, conference, Historikertag, content analysis, network analysis, digital methods

Zusammenfassung: Dieser Beitrag zeigt verschiedene Perspektiven auf ein Korpus von Tweets, die unter den jeweiligen Konferenz-Hashtags anlässlich der vier Historikertage von 2012 bis 2018 publiziert wurden. Im Mittelpunkt des Erkenntnisinteresses stehen Twitterpraktiken und ihre Entwicklung, Inhalte der Tweets sowie Netzwerke der twitternden historischen Zunft in einer wissenschaftlichen Tagungssituation über sechs Jahre hinweg. Erstmals wird damit in einer Twitterstudie die historische Fachcommunity in den Blick genommen. Besonderheiten der Studie sind der im Vergleich zu anderen Analysen verhältnismäßig lange Zeitraum, der punktuell anhand der Historikertage untersucht wird, sowie genderspezifische Auswertungen, die durch eine offene Kodierung ermöglicht werden.

Schlagwörter: Geschichtswissenschaft, Twitter, Tagung, Historikertag, Inhaltsanalyse, Netzwerkanalyse, digitale Methoden

Der alle zwei Jahre an wechselnden Orten stattfindende Historikertag ist mit durchschnittlich rund 3 500 Teilnehmenden einer der größten geisteswissenschaftlichen Kongresse in Europa. Die Tagung gilt als zentraler Ort der Zunft, an dem medien- und breitenwirksam Diskussionen geführt werden, die das

Fach in seinen inhaltlichen Ausrichtungen genauso prägen wie sein Bild in der Öffentlichkeit. Diskussion und Austausch stehen bei diesem Versammlungsformat im Mittelpunkt, wozu Sektionen, Einzelvorträge, Podiumsdiskussionen und Posterpräsentationen sowie – nicht weniger wichtig – Kaffeepausen und andere soziale Events gehören.

Als zusätzliche Austauschplattform sind für die „versammelte Zunft“¹ seit den 2010er Jahren soziale Medien hinzugekommen, unter denen der Kurznachrichtendienst Twitter (seit März 2006) eine besondere Rolle einnimmt.² Twitternde Tagungsteilnehmende publizieren in Echtzeit Kommentare und Eindrücke von der Tagung, diskutieren miteinander, posten Fotos und Links auf weiterführende Informationen und verabreden sich auf einen Kaffee oder ein „Twit-tagessen“ in den Pausen. Als kommunikativer „Rückkanal“³ ermöglicht Twitter die Partizipation der Zuhörer:innen und einen unmittelbaren Austausch, ohne laufende Vorträge zu stören. Dabei können auch Personen, die nicht physisch anwesend sind, interaktiv eingebunden werden.

Dieser Beitrag zeigt verschiedene Perspektiven auf ein Korpus von Tweets, die unter den jeweiligen Konferenz-Hashtags anlässlich der vier Historikertage von 2012 bis 2018 publiziert wurden. Im Mittelpunkt des Erkenntnisinteresses stehen Twitterpraktiken und ihre Entwicklung, Inhalte der Tweets sowie Netzwerke der twitternden historischen Zunft in einer wissenschaftlichen Tagungssituation über sechs Jahre hinweg.⁴

1 Matthias Berg et al., *Die versammelte Zunft. Historikerverband und Historikertage in Deutschland 1893–2000* (Göttingen: Wallstein, 2018).

2 Zu Twitter und seiner Entwicklung siehe Jean Burgess und Nancy K. Baym, *Twitter: A Biography* (New York: New York University Press, 2020). Allgemein zu sozialen Medien in der Wissenschaft: Mareike König, „Scholarly Communication in Social Media,“ in *Science Communication. Handbook of Communication Sciences HoCS*, 17, hg. v. Annette Leßmannmüller et al. (München: De Gruyter, 2020), 639–656.

3 Claire Ross et al., „Enabled Backchannel: Conference Twitter Use by Digital Humanists,“ *Journal of Documentation* 67 (2011): 214–237.

4 Präsentiert wird hier eine Auswahl an Ergebnissen. Eine vertiefte Darstellung der Methoden, rechtliche und ethische Überlegungen, weitere kommentierte Abbildungen und Graphen sowie die IDs der Tweets sind bei Zenodo, der Code bei GitHub hinterlegt, siehe: Mareike König und Paul Ramisch, *Historikertage auf Twitter (2012–2018). Datenreport und Datenset*, Zenodo 2022. doi: 10.5281/zenodo.6362301.

1 Vorgehensweise, Methoden und Grenzen der Studie

Daten aus sozialen Medien sind eine neuere Quellengattung, die seit mehreren Jahren eine starke Forschungsdynamik in verschiedenen Disziplinen ausgelöst haben.⁵ Aufgrund der relativ leicht zugänglichen Daten gehört Twitter dabei zu den am häufigsten untersuchten Plattformen.⁶ Ein beliebter Forschungsgegenstand im Bereich der wissenschaftlichen Nutzung von Twitter ist das Konferenztwittern.⁷ Die Untersuchungen zeigen unter anderem, dass sich Twitterpraktiken in den Disziplinen unterscheiden.⁸ In unserer Studie wird erstmals die historische Fachcommunity in den Blick genommen.

Bei der Arbeit mit Daten aus sozialen Medien stellen sich verschiedene Herausforderungen, zum Beispiel methodische Fragen beim Umgang mit großen Mengen an proprietären Daten, Kürze und Spezialjargon der Postings, Vielfalt und Komplexität der Inhalte sowie ethische und rechtliche Fragen im Hinblick auf das Sammeln, Aggregieren und das Wiedergeben von Inhalten und Metadaten.⁹ Für die vorliegende Studie wurden zitierte Tweets von Einzelpersonen anonymisiert, die Kommunikation der offiziellen Accounts von Einrichtungen da-

5 Siehe die Forschungsübersicht in Lori McCay-Peet und Anabel Quan-Haase, „What is Social Media and What Questions Can Social Media Research Help Us Answer?“, in *The SAGE Handbook of Social Media Research Methods*, hg. v. Luke Sloan und Anabel Quan-Haase (London: Sage Publications, 2017), 13–26.

6 Übersichten zu Methoden speziell bei Twitteranalysen zum Beispiel Manuel Burghardt, „Introduction to Tools and Methods for the Analysis of Twitter Data“, *Media Linguistics* 1 (2015): 74–91; Michael Zimmer und Nicholas John Proferes, „A Topology of Twitter Research: Disciplines, Methods, and Ethics“, *Aslib. Journal of Information Management* 66/3 (2014): 250–261.

7 Siehe zum Beispiel B. Hilary Smith Risser und Glenn Waddell, „Beyond the Backchannel: Tweeting Patterns After Two Educational Conferences“, *Educational Media International* 55/3 (2018): 202; Royce Kimmons und George Veletsianos, „Education Scholars’ Evolving Uses of Twitter as a Conference Backchannel and Social Commentary Platform“, *British Journal of Educational Technology* 47/3 (2016): 445–464. Eine Übersicht dazu auch bei Merja Mahrt et al., „Twitter in Scholarly Communication“, in *Twitter and Society*, hg. v. Katrin Weller et al. (New York: Peter Lang, 2014), 202–205.

8 Vgl. zum Beispiel B. Kim Holmberg und Mike Thelwall, „Disciplinary Differences in Twitter Scholarly Communication“, *Scientometrics* 2 (2014): 1027–1042; Mahrt et al., „Twitter in Scholarly Communication“, 204.

9 Zu rechtlichen und ethischen Fragen siehe Michael Beurskens, „Legal Aspects of Twitter Research“, in *Twitter and Society*, hg. v. Weller et al., 123–133; Matthew L. Williams et al., „Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users’ Views, Online Context and Algorithmic Estimation“, *Sociology* 51/6 (2017): 1149–1168.

gegen offen genannt.¹⁰ Eine Schwierigkeit stellten die technischen Entwicklungen bei Twitter seit 2012 dar, etwa die Verdoppelung der erlaubten Zeichenzahl pro Tweet im November 2017 auf 280 Zeichen sowie das Einführen von Retweets und Tweets als Zitate und damit das Sperren des Originaltweets für Änderungen beim Retweeten.

Das hier untersuchte Gesamtkorpus der Tweets stammt aus auf verschiedenen Wegen erstellten Teilkorpora (Tweetsammlungen über das Tool TAGS¹¹ sowie über die Twitter Premium API¹² mithilfe des RTweet-Packages in der Programmiersprache R beziehungsweise RSudio abgerufene Tweets). Diese wurden zusammengeführt und aufbereitet, um eine kohärente Datenbasis für die weiteren Analysen zu erhalten. Die unterschiedliche Erstellung der Korpora führt genau wie die erwähnten technischen Entwicklungen bei Twitter dazu, dass die Datenqualität teilweise variiert. Für diese Studie grundlegend ist zudem, dass ausschließlich Tweets berücksichtigt wurden, für die der Tagungshashtag der Historikertage und die Schreibvarianten verwendet wurden, also #histag, #hist-tag und #historikertag mit den jeweiligen Jahren in den Formaten 2018 und 18 angehängt. Dies dient dazu, die Studie auf die twitternde Zunft zu beschränken, die mit den Gepflogenheiten des Tagungshashtags vertraut ist, und nicht alle Tweets mit dem Wort „Historikertag“ in das Korpus aufzunehmen. Allerdings können dadurch relevante Tweets fehlen, bei denen der Tagungshashtag weggelassen wurde, sei es absichtlich oder aus Versehen: etwa wenn nur spezifische Sektionshashtags verwendet wurden oder wenn in Unterhaltungen oder bei sehr kritischen Tweets der Tagungshashtag weggelassen wurde, um einem Tweet weniger Sichtbarkeit zu geben und ihn nur mit der eigenen Follower-schaft zu teilen.¹³

In einem weiteren Schritt wurde das Twitterkorpus in mehreren Durchgängen und von mehreren Personen manuell kodiert.¹⁴ Kodierungen haben den Vorteil, für maschinelle Analysen vorzubereiten, ohne in den Datensatz einzugreifen. Sie sind jedoch anfällig für Uneinheitlichkeit bei der Vergabe der

10 Zu den rechtlichen und ethischen Überlegungen König und Ramisch, *Historikertage auf Twitter*.

11 TAGS – Twitter Archiving Google Sheet, <https://tags.hawksey.info/about/>.

12 Zu den beiden vorhandenen API, ihren Möglichkeiten und Beschränkungen siehe Burghardt, „Introduction“, 79.

13 Zu den Vor- und Nachteilen von Hashtag-Korpora gegenüber Stichwort-Korpora siehe Axel Bruns und Stefan Stieglitz, „Metrics for Understanding Communication on Twitter,“ in *Twitter and Society*, hg. v. Weller et al., 75.

14 Wir bedanken uns herzlich bei Elena Suárez Cronauer, Friederike Odenwald und Helen Bubinger für die Mitarbeit an der offenen Kodierung. Das iterative Vorgehen bei der Kodierung ist im Datenreport beschrieben, siehe König und Ramisch, *Historikertage auf Twitter*.

Codes, zumal bei kurzen Tweettexten. Zum einen wurden die insgesamt 1970 Accounts mit acht verschiedenen Gender- bzw. Typcodes kodiert (Mann, Frau, anonym, Einrichtung, Unternehmen, Zeitschriften, Bots, unbekannt). Zum anderen wurden die insgesamt 14 207 Tweets einer von sieben Kategorien zugeordnet, die in erster Linie die Funktion der Tweets beschreibt und weniger die Tweetinhalte.¹⁵ Zwar wird durch die vorgegebene qualitative Kategorisierung die Funktion der Tweets nicht unmittelbar aus unseren Daten selbst abgeleitet. Sie orientiert sich jedoch an einer ähnlichen Studie, die 2009 über das Tagungstwittern in der Digital Humanities-Community durchgeführt wurde, und schafft somit die Möglichkeit zum Vergleich.¹⁶

Tab. 1: Übersicht der vergebenen Kategorien

Code	Bedeutung	Beispiel
1	Kommentar zu Vortrag oder Veranstaltung	#histag12 schöner Abschluss des Historikertages mit einem runden Vortrag vom scheidenden Vorsitzenden des VHD „Gespaltene Gesellschaft, was heißt das denn? Die Alternative wäre dann eine homogene Gesellschaft, die kann es nicht geben.“ Leise Kritik am Konferenztitel von #HisTag18 seitens Wolfgang Schäuble?
2	Ressourcen und Links teilen	Twitter als Recherchemittel: Folien https://t.co/GaFXDg1R2k und Pad https://t.co/hhYEUTDyBR #histag16 #twitterlounge https://t.co/V7LRglFO6B Der Gärprozess hat einen offenen Brief an den @VHDtweets hervorgebracht: https://t.co/iULHgOQ4WL #HisTag18 #histocampstag
3	Gespräch/Diskussion	„@Twitteruser1: @Twitteruser2 Respekt. Dein Diss-Plakat ist echt nicht von schlechten Eltern! #histag12 http://t.co/2Z9QEAAA “ danke @Twitteruser1: @Twitteruser2 Dann braucht man aber auch Programmierer mit Grundverständnis für GW-Daten als Komplementär. #histag16 #digigw16

¹⁵ Ross et al., „Backchannel“, 219.

¹⁶ Vgl. die Kritik solcher Studien bei Matthias Meiler, „Wissenschaftliches Twittern. Linguistische Bestandsaufnahme und method(olog)ische Auslotung, insbesondere mit Fokus auf Konferenztweets“, *Linguistik Online* (2021): 87–111, doi: 10.13092/lo.106.7511.

Tab. 1 (fortgesetzt)

Code	Bedeutung	Beispiel
4	Notizen machen	„Wir haben ein Problem der Einflussnahme durch digitale Medien“, sagt Mergel als Ausgang für zunehmenden Populismus #histag16 #historydebate Objekte tragen Spuren ihrer Behandlung. Behältnisse und Installationen zeigen Lokalisierung in Handlungskontexten #histmat14 #HisTag14
5	Onlinepräsenz gestalten/Werbung (eigenes Panel, eigene Publikation)	Nach rundum ertragreichem und inspirierendem #histag12 wieder zuhause. Nächstes Mal nehme ich das richtige Ladekabel mit, um mitzutwittern Erstaunlich viele Smartphones hier, ob die auch alle twittern? #HisTag14
6	Organisatorisches und CfP	Die aktuellsten PROGRAMMÄNDERUNGEN für den #HisTag14 sind online einsehbar: http://t.co/PQyrjAvKUU Leider fällt der Vortrag „Twitter als Recherchemittel“ morgen in der #Twitterlounge um 14h wegen Erkrankung des Referenten aus #histag16
7	Unklar	#histag12 OK "si" natuerlich ... Und c p oob #histag16 jtp

In einem weiteren Schritt wurden nach einer Datenvorbereitung für die Netzwerkanalyse mit der freien Software Gephi alle Twitter-Nutzer:innen und deren Follower:innen- und Friends-Verbindungen betrachtet, von denen mindestens ein Tweet im Korpus enthalten ist, die an das Netzwerk angebunden sind und deren Accounts weder gelöscht noch auf privat umgestellt wurden.¹⁷ Und schließlich wurde eine inhaltliche Auswertung der Tweets mit dem webbasierten Dienst Voyant-Tools vorgenommen. Dabei wurde das Korpus nach der Da-

¹⁷ Drei Gruppen von Accounts wurden ausgefiltert: gelöschte Accounts (51), Accounts, die auf privat gestellt wurden (18), und Accounts, denen weniger als zwei Accounts aus dem weiteren Netzwerk folgen (246, teils Überschneidung mit gelöschten/privaten Accounts). Diese Barriere hilft sicherzustellen, dass die betrachteten Akteur:innen tatsächlich in dieses Netzwerk eingebunden sind. Somit untersuchen wir in der Netzwerkanalyse 1711 der insgesamt 1970 Accounts aus dem Korpus.

tenbereinigung¹⁸ auf der Wortebene analysiert und Worthäufigkeiten, chronologische Frequenzgraphen und Kookkurenzen untersucht. Hier zeigen sich schnell die Grenzen einer rein digitalen Inhaltsanalyse angesichts der kaum überschaubaren Menge an höchst unterschiedlichen Einzeltweets.

2 *Twitterstorians* und Praktiken der Twitternutzung

Anhand des so vorbereiteten Tweetkorpus wurden statistische Auswertungen mit der Programmiersprache R vorgenommen. In der Gesamtschau lässt sich eine starke Zunahme der Anzahl an twitternden Historiker:innen seit 2012 beobachten: Von 155 twitternden Accounts in Mainz 2012 stieg die Zahl der *Twitterstorians* auf 1 045 Accounts beim Historikertag in Münster 2018 an (Abbildung 1 und 2). Die Accountzahlen nahmen insbesondere bei Frauen deutlich zu, wobei in absoluten Zahlen die von Männern geführten Accounts 2018 nach wie vor überwiegen. Zu beachten ist, dass von den insgesamt 1 970 Accounts die große Mehrheit (1 282) selbst keine eigenen Originaltweets absetzte und nur deshalb Teil des Korpus sind, weil sie einen Tweet retweetet haben. 688 Accounts haben eigene Tweets gepostet, davon 392 mehr als einen. Insgesamt handelt es sich damit um einen spezifischen und kleinen Ausschnitt der Geschichtswissenschaft, der sich auf Twitter äußert. Eine qualitative Einordnung der Accounts nach Karrierestufe (Studierende, Promovierende, Post-Docs, Professor:innen) ist nur manuell möglich, wobei für eine Auswertung der Zeitpunkt des Twitterns berücksichtigt werden müsste. Sie ist in dieser Studie daher nicht erfolgt. Mit Blick auf das Gesamtkorpus lässt sich zum Zeitpunkt der Auswertung 2021 sagen, dass von den beteiligten Accounts von Historiker:innen (709 zu 515) nur rund 30 Accounts von (Junior-)Professor:innen geführt werden. Rückläufig ist über den Untersuchungszeitraum die Anzahl der anonymen Accounts, deren Anteil sich von 15,5 % auf 8,4 % fast halbiert. Ebenso wie die generell steigenden Accountzahlen spricht dies für eine zunehmend bewusste und strategische Verwendung von Twitter als Mittel der Fachkommunikation innerhalb der Geschichtswissenschaft.

18 Die Tweets wurden in Notepad++ für die Analysen vorbereitet, in dem Links, Accountnamen, Zahlen, Stoppwörter und anderes entfernt und häufige Wörter auf ihren Wortstamm zurückgeführt wurden. Außerdem wurde die Stoppwortliste in Voyant-Tools verwendet und iterativ angepasst.

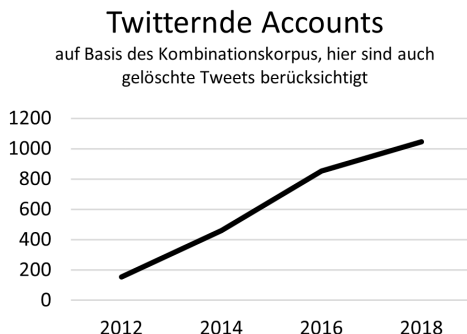


Abb. 1: Entwicklung der Anzahl der twitternden Accounts, n = 1970

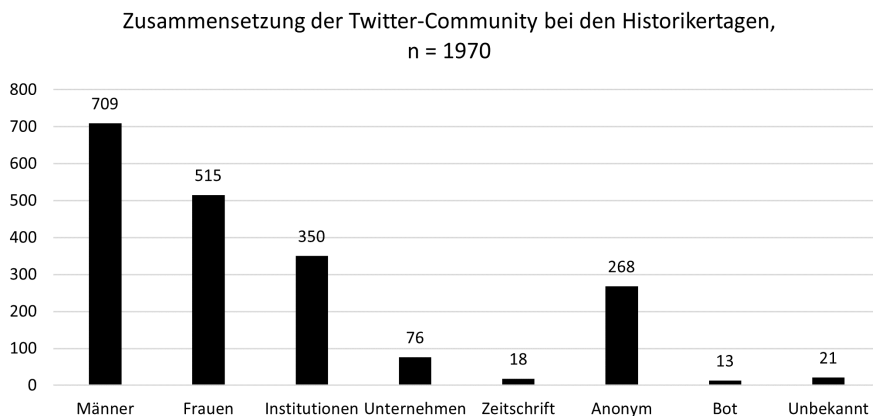


Abb. 2: Zusammensetzung der twitternden Accounts insgesamt

Eine steigende Anzahl an twitternden Historiker:innen bedeutet nicht zugleich eine im vergleichbaren Umfang steigende Anzahl an Tweets während der Historikertage, im Gegenteil: In absoluten Zahlen gehen diese nach einem starken Anstieg in Göttingen 2014 zurück und halbieren sich von dort bis Münster 2018 (Abbildung 3). Bezieht man Retweets in die Analyse mit ein, verschiebt sich das Bild: Bei den Gesamttweets ist der Knick erst 2016 zu sehen (Abbildung 4). Während die *early adopters* von Twitter in der Fachcommunity 2012 im Median rund zwei Tweets gepostet haben, war es 2018 im Median nur ein einzelner Tweet pro Account. Die Streuung ist dabei sehr groß: Einige Poweruser:innen haben verteilt über die vier Historikertage über 300 Tweets (inklusive Retweets) gepostet, so dass sich auch bei den *Twitterstorians* die für soziale Medien typi-

sche 90:9:1-Regel in Bezug auf die Produktion von Inhalten bestätigt.¹⁹ Nur drei Accounts haben dabei insgesamt über 200 Originaltweets gepostet, elf über 100.²⁰

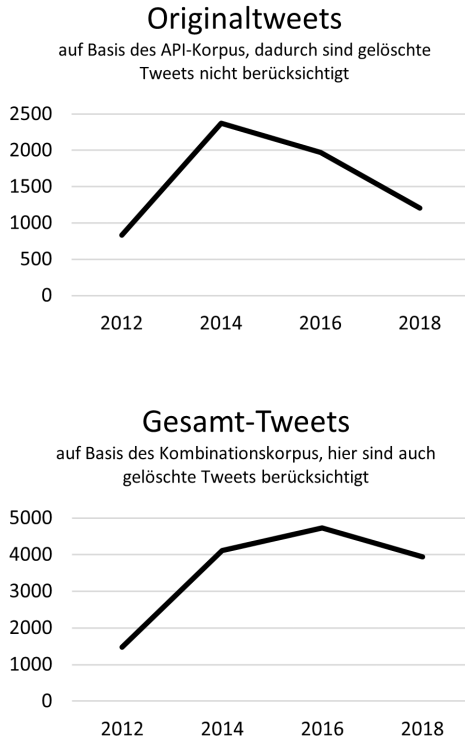


Abb. 3 und 4: Entwicklung der Tweetzahlen: Originaltweets und Gesamttweets (Originaltweets und Retweets)

Dennoch sind Historiker:innen während ihrer wichtigsten Tagung eine auf Twitter aktive Community, wie der Blick auf *Retweets* und *Likes* (bis November 2015:

¹⁹ Demnach bleiben 90 % der Nutzer:innen eher passiv in den sozialen Medien, 9 % posten ab und zu, während 1 % für den Großteil der Inhalte verantwortlich ist. Bruns und Stieglitz schlagen vor, Tweetkorpora nach diesen drei Gruppen aufgeteilt zu untersuchen, vgl. Bruns und Stieglitz, „Metrics“, 74.

²⁰ Siehe die Zahlen nach Gruppen pro Historikertag bei König und Ramisch, *Historikertage auf Twitter*.

Favs) während der Kongresse zeigt. Beides sind „soziale Akte“²¹. Doch während Likes privater Natur und weniger sichtbar²² sind und vor allem die Sympathie für den Inhalt eines Tweets anzeigen, gelten Retweets als bedeutendste Metrik, als Indikator für Resonanz und Aufmerksamkeit. Auffallend ist das Stagnieren der Retweets 2018 in absoluten Zahlen (Abbildung 5), was mit der niedrigeren Anzahl an Originaltweets zusammenhängen dürfte. Die Anzahl der Likes steigt dagegen fast konstant an (Abbildung 6). Während 2012 im Durchschnitt jeder Originaltweet zum Historikertag 0,9-mal favorisiert wurde, erhielten Originaltweets 2018 im Durchschnitt sieben Herzen. Mit Blick auf Retweets und Likes ist die Community an *Twitterstorians* über die Jahre also nicht passiver geworden, sie ist vielmehr anders aktiv. Die hohe Anzahl an Likes deutet darauf hin, dass die gesendeten Tweets gelesen werden und mit ihnen interagiert wird. Eine Hypothese ist, dass vor allem Likes, aber auch Retweets das Verfassen eigener Tweets in Teilen ersetzt haben. Erklärungen dafür liegen in den technischen Änderungen bei Twitter begründet, denn sie lassen sich allgemein auf der Plattform beobachten.²³ Zusätzlich könnte die größere Community und die steigende Anzahl an Follower:innen für mehr Sichtbarkeit der Tweets und damit auch für mehr Interaktion, aber auch einen höheren performativen Druck gesorgt haben. Dies und der spezielle soziale Kontext der Tagung könnte manche *Twitterstorians* besonders motiviert haben, einen Tweet zu produzieren, der häufig retweetet wird, während manche lieber die Ideen anderer verbreiten, als selbst zu twitern (siehe Abschnitt 6).



²¹ Johannes Paßmann, *Die soziale Logik des Likes. Eine Twitter-Ethnografie* (Frankfurt a. M.: Campus, 2018), 17.

²² Burgess und Baym, *Twitter*, 101.

²³ Zur Entwicklung der Retweets vgl. *ibid.*, 87, 92–97.

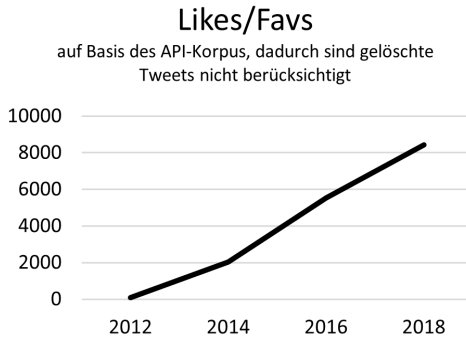


Abb. 5 und 6: Entwicklung der Retweets und Entwicklung der Likes

Die Entwicklung der Originaltweets mit Erwähnungen eines anderen Accounts (*Mentions*) zeigt in absoluten Zahlen bis 2016 eine starke Zunahme, dann einen starken Rückgang bis fast auf den Stand von 2012 (Abbildung 7). Die Kurve folgt aber nicht der Anzahl der Originaltweets, sondern schwankt bei den einzelnen Historikertagen, mit einem Höchststand 2012 von 38,8 %. In der Regel erhalten neben offiziellen Veranstaltungaccounts diejenigen die meisten *Mentions* und Antworten, die am häufigsten twittern,²⁴ was sich auch in unserem Korpus zeigt.

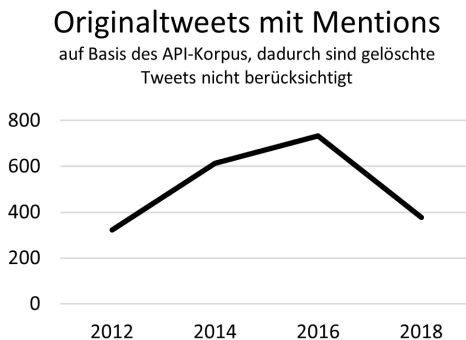


Abb. 7: Entwicklung der Tweets mit Erwähnungen anderer Accounts in absoluten Zahlen

Ein Blick auf die Antworten (*Replies*) im Korpus zeigt, dass diese in absoluten Zahlen wie prozentual gesehen zunehmen. Insgesamt haben 22,1 % (1 404) der Originaltweets mindestens eine Antwort bekommen, mindestens zwei Antwort-

²⁴ Smith Risser und Waddell, „Beyond the Backchannel,“ 202.

ten bekamen 5,6 % (356) der Tweets. Der Höchststand ist dabei 2018 erreicht, als 29,7 % der Tweets mindestens eine Antwort bekommen haben. Dies zeigt eine starke kommunikative Einstellung der twitternden Zunft während der Historikertage.

3 Funktionen der Tweets

Welche Funktion haben die zu den Historikertagen geposteten Tweets? Die Ergebnisse der Kodierung zeigen schwankende Zahlen bei den einzelnen Veranstaltungen (Abbildung 8), was nahe legt, dass hier – mit Ausnahme einiger Trends – weniger eine Entwicklung zu sehen ist, als kontextbezogenes Verhalten anlässlich der einzelnen Historikertage und ihrer spezifischen Dynamiken. Zu berücksichtigen ist außerdem die erwähnte Datenunsicherheit durch die manuelle Kodierung.

Zunächst zu den Trends: Bei allen vier Historikertagen zielten die Praktiken vor allem darauf, mit Tweets während der Tagung Notizen zu machen und Vorträge und Diskussionsbeiträge damit zu verbreiten und zu dokumentieren (Kategorie 4, blau). Diese Kategorie hält durchgängig über die Jahre den höchsten Anteil bei den Originaltweets, wobei dieser schwankt mit einem Tiefstwert von 23,7 % in 2012 und einem Höchstwert von 36,7 % in 2014. Zum Vergleich die Zahlen aus der DH-Community 2009: Auch dort entfielen die meisten Tweets auf die Kategorie 4, sie erreichten sogar einen Anteil von 43 % am Gesamtkorpus.²⁵

In diesen Bereich des gemeinsamen Konstruierens von Wissen gehören ebenso die Kategorien „Kommentare“ (Kategorie 1, dunkelgrün) sowie „Ressourcen und Links teilen“ (Kategorie 2, orange). Dass insgesamt nur rund ein Viertel (26,0 %) aller Originaltweets Kommentare oder Wertungen zum Vortrags- und Tagungsgeschehen beinhalten, mag überraschen, gilt Twitter doch als Ort, an dem jede:r die eigene Meinung kundtut. Im Vergleich zur Digital Humanities-Community 2009 zeigten sich die *Twitterstorians* dennoch meinungsfreudiger: Dort entfielen nur 3 % der Tweets in die Kategorie „Kommentare“.²⁶

²⁵ Ross et al., „Backchannel“, 224.

²⁶ Ibid.

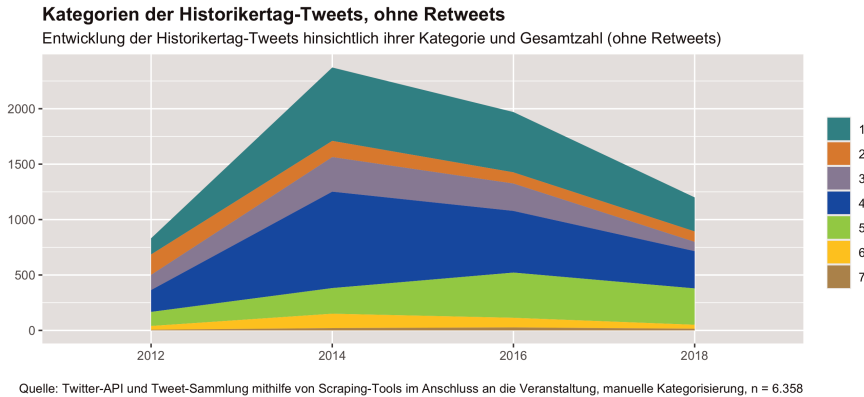


Abb. 8: Entwicklung der Auswertung der Funktionen von Tweets

Die Anzahl der Originaltweets, die Ressourcen und Links teilen (Kategorie 2, orange) blieb im Laufe der Jahre in absoluten Zahlen zwar gleich. Ihr Anteil am Gesamtkorpus ging jedoch stark zurück, mit einem regelrechten Einbruch zwischen 2012 (22,1 %) und 2014 (6,2 %). In Hamburg 2016 sank der Anteil auf 5,1 % und stieg dann in Münster 2018 auf 7,8 %. Dies bestätigt die statistische Auswertung der Anzahl der Originaltweets mit Links, die sich von 38,6 % in 2012 quasi halbiert auf 16,3 % in 2014 und sich dort bei den beiden folgenden Historikertagen etwa einpendelt.²⁷

Zu beobachten ist ebenso ein Rückgang der Gespräche (Kategorie 3, lila) unter dem Tagungshashtag von 16,5 % in Mainz 2012, mit einem Einbruch 2018 in Münster, als der Anteil auf nur 7,0 % absank. Da die über die Metadaten von Twitter ermittelten Replies jedoch auf 29,7 % ansteigen, bestätigt sich die These, dass bei Gesprächen der Tagungshashtag oftmals weggelassen wird, häufig bewusst, um das Gespräch weniger offiziell zu machen.²⁸ Dies könnte insbesondere in Münster eine Rolle gespielt haben (siehe Abschnitt 6).

²⁷ 2016 beinhalten 13,6 % der Originaltweets einen Link, 2018 sind es 16,3 %.

²⁸ Generell fehlen in Hashtag-Korpora zahlreiche Antworten, weil der Hashtag weggelassen wurde, vgl. Bruns und Stieglitz, „Metrics“, 75.

Weiter ist zu beobachten, dass Twitter auch in der historischen Fachcommunity zunehmend ein Ort der Eigenwerbung und Selbstdarstellung geworden ist.²⁹ Die Kategorie 5 „Onlinepräsenz gestalten und Werbung für eigenes Panel, eigene Publikation“ (helles Grün) umfasste 2012 in Mainz 15,3 %, sank 2014 auf 9,7 % ab, bevor der Anteil 2016 auf ein gutes Fünftel anstieg (20,7 %) und beim vorerst letzten Historikertag 2018 in Münster einen Höchststand von 27,3 % erreichte und damit die „Kommentare“ von Platz zwei im Kategorienschema verdrängte. Zu beachten ist dabei, dass Kategorie 5 sehr divers ist und witzig gemeinte Tweets genauso umfasst wie Hinweise zu eigenen Sektionen und Veranstaltungen, so dass es zu Fehlern in der Kodierung gekommen sein kann. Genau wie Gespräche dienen diese Tweets zugleich der Beziehungspflege und dem Aufbau von Netzwerken. Zum Vergleich: bei den Tagungen der Digital Humanities-Community 2009 entfielen weit weniger als halb so viele Tweets auf diese Kategorie (rund 10 %)³⁰, wobei offenbleibt, ob dieser Unterschied nur den allgemeinen Wandel der Praktiken oder auch communityspezifisches Verhalten spiegelt.

Wenig überraschend ist, dass wissenschaftliche Einrichtungen und Unternehmen in unserem Beispielkorpus ganz überwiegend Tweets zur Selbstpromotion (Kategorie 5) posten. Bei den Unternehmen erreicht diese Kategorie einen Spitzenwert: Dort fallen über die Hälfte ihrer Tweets (54,6 %) in diese Kategorie. Umgekehrt bedeutet dies zugleich, dass knapp über 45 % der Tweets nicht der Eigenwerbung dienen, was zeigt, dass sich die Kommunikation in den sozialen Medien auch bei Unternehmen deutlich von der üblichen Presse- und Öffentlichkeitsarbeit unterscheidet. Die Interaktion über Likes und Retweets mit den Tweets von Einrichtungen und Unternehmen ist generell hoch, was zeigt, dass sie eng mit der twitternden Zunft verbunden sind.

²⁹ Twitter hat sich über die Jahre zu einem „self-promotional environment“ gewandelt, so Burgess und Baym, *Twitter*, 83.

³⁰ Ross et al., „Backchannel“, 226.

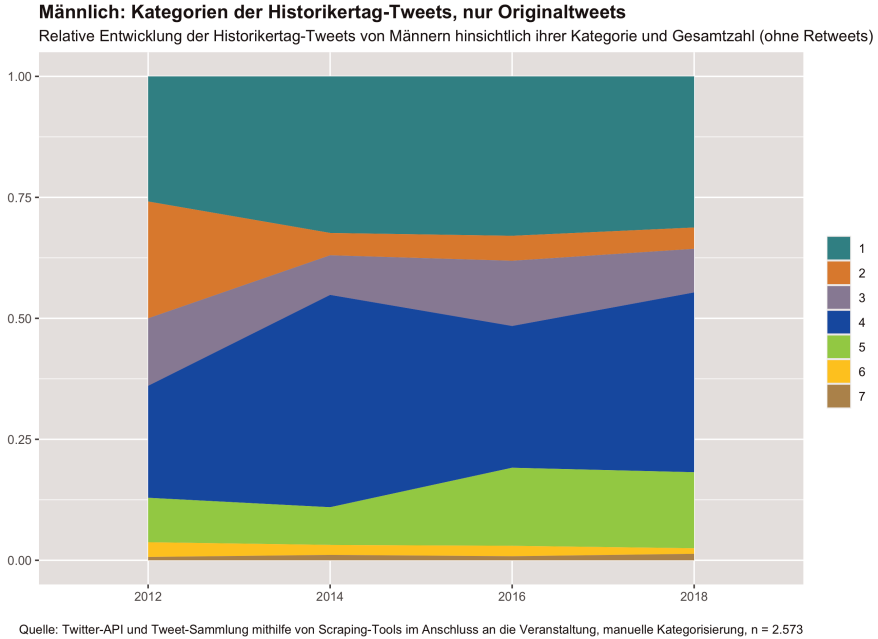


Abb. 9: Anteil der Funktionskategorien in den Tweets der von Männern geführten Accounts (von 100 %)

Interessant ist die Verteilung der Funktionen von Tweets bei den persönlichen Accounts: Während bei den Historikern „Notizen machen“ (37 %), „Kommentare posten“ (31,5 %) und „Onlinepräsenz gestalten“ (11,3 %) die drei wichtigsten Funktionen für die Tagungstweets darstellen (Abbildung 9), steht bei den Historikerinnen „Kommentare posten“ (30,4 %) an erster Stelle vor „Notizen machen“ (26,4 %) auf Platz zwei und „Gespräche“ (17,7 %) auf Platz drei (Abbildung 10). „Onlinepräsenz gestalten“ liegt bei den Historikerinnen nur auf Platz vier, auch wenn der Anteil mit 15,2 % höher ist als bei den Historikern. Ob diese Verteilung auf ein unterschiedliches Kommunikationsverhalten bei Twitter hindeutet, bleibt angesichts der Grenzen der Studie und der komplexen Zusammenhänge der Kommunikation schwer zu sagen.

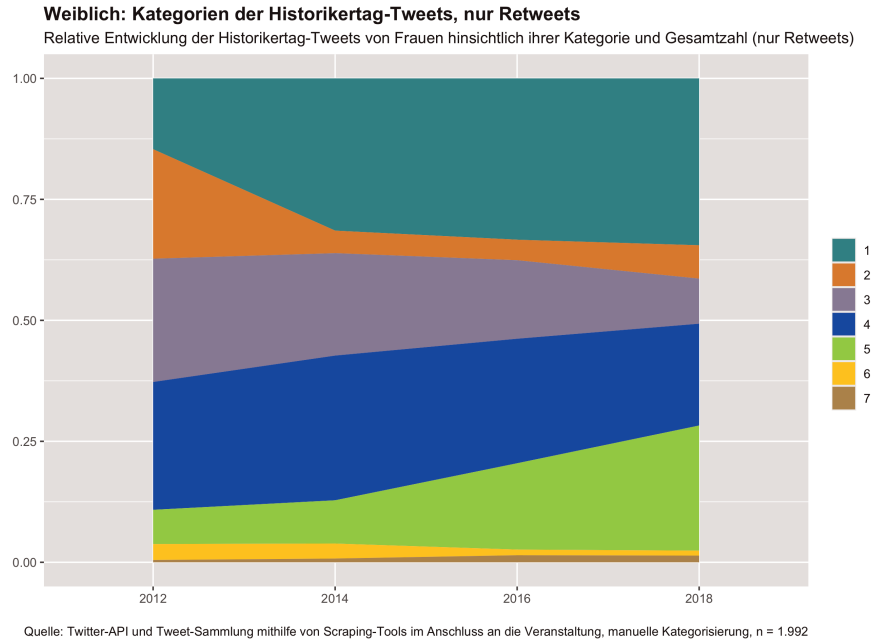


Abb. 10: Anteil der Funktionskategorien in den Tweets der von Frauen geführten Accounts (von 100 %)

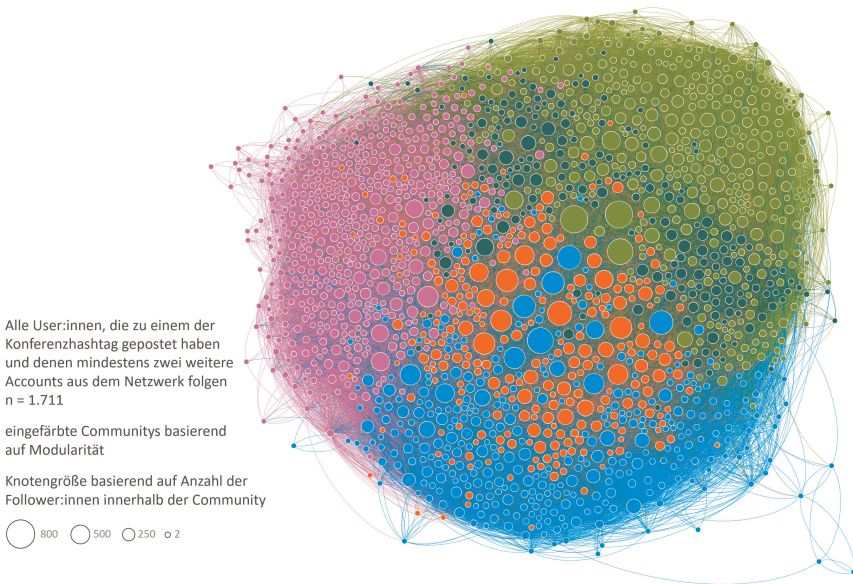
Organisatorische Fragen nehmen 4,5 % der Originaltweets im gesamten Korpus ein, mit einem Höchststand von 5,5 % beim Historikertag 2014 in Göttingen, was auch daran liegen dürfte, dass der Account @Historikertag 2014 besonders aktiv war und damit Gesprächsbereitschaft signalisierte. 1,1 % der Originaltweets ließen sich keiner Kategorie zuordnen (Kategorie 7, braun).

Insgesamt wird aus der statistischen Beschreibung und der Analyse der Funktionscodes eine Veränderung der Twitternutzung bei den Historikertagen von 2012 bis 2018 deutlich: Die noch wenigen *Twitterstorians* versuchen 2012 mit einer hohen Anzahl an Tweets pro Account eine breite Dokumentation des Historikertags zu erreichen. 38,6 % der Tweets sind mit Links angereichert, 22,1 % enthalten fachliche Hinweise und einschlägige Literatur als Serviceleistung für die Fachcommunity. Es gibt mehr Gespräche und Diskussionen unter dem Tagungshashtag. Die Aktivitäten verlagern sich in der Folge vom Posten eigener Tweets hin zum Lesen, Liken und Verbreiten der Tweets von anderen. Tweets mit Mentions und Antworten nehmen zu, Unterhaltungen unter dem Tagungshashtag gehen aber zurück und angesichts der größer werdenden Community steigen diese Zahlen nicht analog. Durch die Zunahme an mitlesenden Kolleg:innen und offizi-

ellen Accounts könnte der Druck zur Selbstdarstellung zugenommen haben, was unter anderem ein Grund sein könnte für das Sinken der Anzahl an Originaltweets und für die Zunahme an Tweets zur Selbstdarstellung.

4 Community-Netzwerke

Welche Cluster und Communities lassen sich innerhalb des Twitter-Netzwerks erkennen?³¹ Ein erster Blick auf die visuelle Darstellung des Graphen zeigt keine explizit abgetrennten Cluster und Communities. Die Größe der einzelnen Knoten basiert auf deren Eingangsgrad, hier also der Anzahl der Follower:innen innerhalb des Netzwerks. Anders als die Darstellung suggerieren mag, ist die Dichte des Netzwerks mit 0,044 nicht sehr groß (0 = keine Kanten, 1 = eine Kante zwischen allen untersuchten Knoten). 67 % der Accounts folgen weniger als 100 anderen Accounts aus dem Netzwerk und werden von weniger als 100 Accounts gefolgt.³²



31 Für die Netzwerkanalyse gilt n = 1711, vgl. Anm. 17.

32 Der durchschnittliche Clusterkoeffizient, das Maß für Untergruppen/Cliquen, ist mit 0,321 relativ hoch, der durchschnittlich kürzeste Pfad zwischen zwei Knoten beträgt 2,306, der längste kurze Pfad 6.

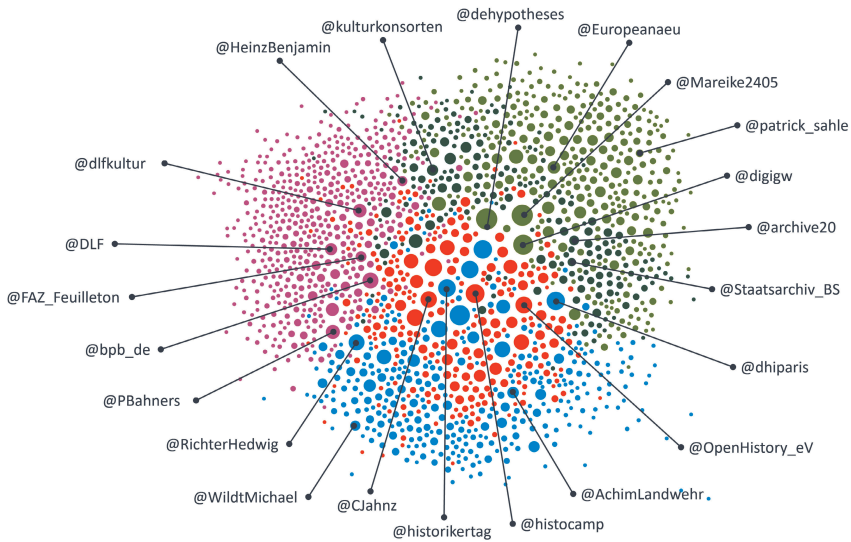
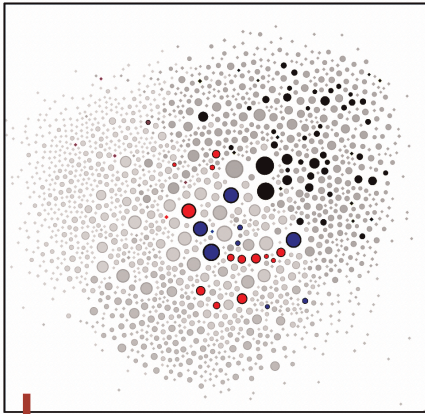


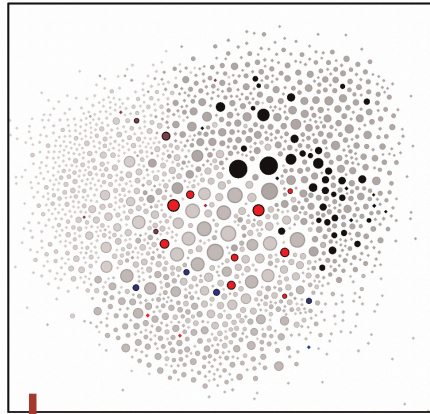
Abb. 11 und 12: Netzwerk-Graph eingefärbt hinsichtlich der auf Modularität basierenden Communities, links mit Follow-Verbindungen, rechts beispielhafte Kennzeichnung einzelner Akteur:innen

Die Modularität des Graphen offenbart eigene Sub-Communities, die sich jedoch nicht als eigene Cluster aus dem Gesamtgraphen herauslösen. Diese Sub-Communities sind zunächst durch Accounts mit hohen Followerzahlen bestimmt. Eine nähere Beschreibung kann durch das weitere Filtern der Daten erfolgen: Im oberen rechten Teil (grün eingefärbt) sind viele Akteur:innen aus dem Bereich der Digital History vertreten. In deren direkter Nachbarschaft befinden sich Akteur:innen aus dem Archivwesen. Dies bestätigt sich durch ein Filtern der Accounts hinsichtlich der Verwendung der Hashtags #digital und #archiv (Abbildung 13 und 14).³³ Im Zentrum sowie im unteren Teil des Graphen zeigen sich viele Akteur:innen aus der „traditionellen“ Geschichtswissenschaft, ohne auf den ersten Blick besondere Netzwerke hinsichtlich bestimmter thematischer Spezialisierungen zu bilden (in blau). Viele der orange eingefärbten Akteur:innen unterscheiden sich von der blau eingefärbten Sub-Community

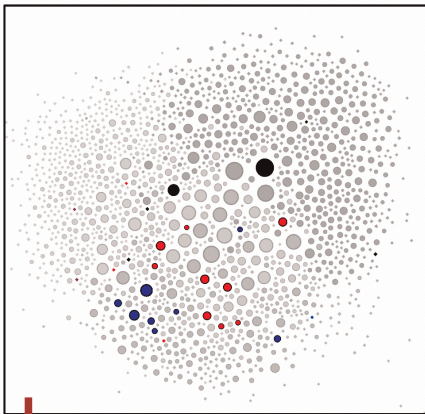
³³ Innerhalb dieses Netzwerks haben alle Gruppen einen Bezug zu Archiven, was sich im Posten von Tweets mit dem Wort „Archiv“ widerspiegelt. Als Hashtag verwendet zeigt sich aber ein starker Bezug zu Accounts aus dem Archivwesen. Diese Argumentation lässt sich analog auf die Digital-History-Akteur:innen und die Verwendung des Hashtags „digital“ übertragen.



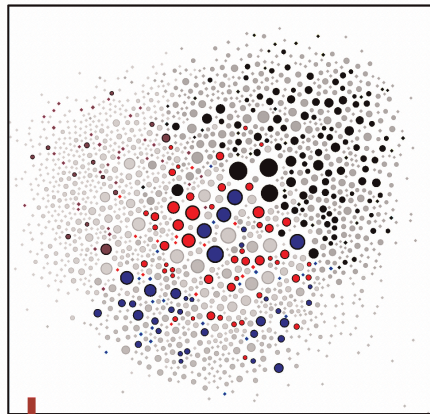
Accounts, die das
Hashtag #digital
benutzt haben (83)



Accounts, die das
Hashtag #archiv
benutzt haben (66)



Accounts, die das Wort
„kolonial“ oder „colonial“
getweetet haben (29)



Accounts, die das Wort
„sektion“ getweetet haben
(309)

Abb. 13–16: Darstellung des gefilterten Netzwerks nach Verwendung verschiedener Hashtags und Schlüsselworte

dadurch, dass sie bereits bei den Historikertagen 2012 und 2014 aktiv waren, wie die Analyse der Netzwerkdarstellung für die einzelnen Veranstaltungen zeigt. Sie bilden zusammen mit Akteur:innen aus der Digital History die *early adopters* beim Konferenztwittern anlässlich der Historikertage. Im linken Teil des Graphen sind Accounts aus dem Bereich Presse und Medien vertreten, dies lässt allgemein auf eine interessierte Öffentlichkeit schließen (in lila eingefärbt). Dieser Eindruck bestätigt sich durch ein Filtern des Graphen nach Accounts, die Wörter benutzen, die direkt auf das Vor-Ort-Tagungsgeschehen bezogen sind: Das Wort „sektion“ beispielsweise kommt in diesem Teil des Graphen kaum vor. Vermutlich ist aus dieser Teilgruppe nur ein geringer Teil an Personen während der Tagung tatsächlich vor Ort.

Es bleibt anzumerken, dass eine reine visuelle Analyse der komplexen Twitterdaten und Followernetzwerke nur begrenzt aussagekräftig ist, zumal Visualisierungen oft einen suggestiven Charakter haben.³⁴ In dieser Auswertung geben die Visualisierungen Hinweise auf Sub-Communities und laden zum weiteren explorativen Entdecken der Daten direkt in Gephi ein.

5 Perspektiven auf Inhalte der Tweets

Das für die Analyse mit Voyant-Tools vorbereitete Tweetkorpus enthält nach der Datenbereinigung 77 356 Wörter. Die Wortschatzdichte beträgt 0.217. Demnach kommt jedes fünfte Wort nur einmal vor. Im Vergleich verfügt das Teilkorpus der Tweets von Frauen (0.283) über eine etwas höhere Wortschatzdichte als das der Männer (0.265). Die Wortschatzdichte der Korpora von Personen ist damit insgesamt relativ hoch. Zu erwähnen ist, dass im Korpus 1 108 verschiedene Hashtags auftauchen, davon 508 nur einmal, was für die Kreativität der *Twitterstorians* beim Verfassen ihrer Tweets spricht.

Tab. 2: Top 15 der häufigsten Wörter (trunkiert) in den Originaltweets

Wort	Häufigkeit	Relative Frequenz
digigw14	349	4 239
digital	342	4 154
sektion	263	3 194
digigw16	220	2 672

³⁴ Vgl. Martin Grandjean, „A Social Network Analysis of Twitter: Mapping the Digital Humanities Community,“ *Cogent Arts & Humanities* 3/1 (2016), doi: 10.1080/23311983.2016.1171458.

Tab. 2 (fortgesetzt)

Wort	Häufigkeit	Relative Frequenz
gut	186	2 259
münster	169	2 052
historyslam	159	1 931
frage	149	1 810
twitter	122	1 482
uhr	109	1 324
göttingen	106	1 287
stand	106	1 287
hamburg	105	1 275
histmat14	102	1 239
spannend	97	178
panel	93	0,0011137992
gauck	92	0,0011018228
online	92	0,0011018228
thema	89	0,0010658938
gerade	88	0,0010539175

Mit Blick auf die Top 20 der meist verwendeten Begriffe im Gesamtkorpus von 2012–2018 zeigt sich, dass in den Tweets zu den Historikertagen digitale Themen klar im Vordergrund stehen (Tabelle 2). Am häufigsten genannt sind die spezifischen Sektionshashtags der AG Digitale Geschichtswissenschaft im Verband der Historiker und Historikerinnen Deutschlands (VHD), also digigw14, digigw16 sowie der Wortstamm „digital“. Wie schon in der Netzwerkanalyse gezeigt, waren es vor allem Digital Historians, die anlässlich der Historikertage getwittert haben, und zwar zu digitalen Themen. Dass dieses Korpus von einer Tagung stammt, von der live berichtet wird, ist ebenfalls deutlich zu sehen an häufig enthaltenen Begriffen wie „Sektion“, „Uhr“, „Workshop“, „Panel“, „Vortrag“, „Diskussion“, „gerade“.

Die Frequenzanalyse zeigt außerdem, dass der 2014 durchgeführte Historyslam ebenso wie die Sektion zur materiellen Kultur (#histmat) und die Clark-Krumeich-Debatte von 2014 über den Ausbruch des Ersten Weltkriegs (unterschiedliche Hashtags, daher hier nicht in den Top 15) zu den am häufigsten betwitterten Sektionen außerhalb von digitalen Themen gehören. Diese Sektionen wurden live gestreamt, was häufig den Effekt hatte, dass dazu mehr getwittert wurde (wobei bei #histmat nur fünf Accounts für 102 Tweets gesorgt haben).

Tab. 3 : Top 15 der häufigsten Wörter (trunkiert) in den Originaltweets der Männer

Wort	Häufigkeit	Relative Frequenz
digigw14	185	5 093
digital	139	3 827
sektion	106	2 918
histmat14	98	2 698
historyslam	91	2 505
gut	72	1 982
frage	70	1 927
kasino	68	1 872
gauck	56	1 542
agag	54	1 487
twitter	54	1 487
kokowi	48	1 321
münster	48	1 321
digigw16	47	1 294
diskussion	44	1 211

Tab. 4: Top 15 der häufigsten Wörter (trunkiert) in den Originaltweets der Frauen

Wort	Häufigkeit	Relative Frequenz
digigw16	124	5 725
digigw14	116	5 355
Digital	108	4 986
sektion	89	4 109
gut	59	2 724
münster	52	2 401
frage	44	2 031
spannend	43	1 985
twitter	38	1 754
hamburg	32	1 477
historyslam	32	1 477
panel	31	1 431
gerade	29	1 339
historydebate	29	1 339
online	29	1 339

Interessant ist der Vergleich zwischen den von Männern und den von Frauen geposteten Originaltweets im Hinblick auf die betwitterten Sektionen. Bei den Frauen sind außerdem die Adjektive „gut“ und „spannend“ prominent platziert (Tabelle 4). In absoluten Zahlen haben Männer zwar das Wort „gut“ häufiger verwendet: 72 zu 59. Aber der Wert der relativen Häufigkeit in Bezug zum jeweiligen Korpus ist bei den Frauen signifikant höher. Insgesamt überwiegen ganz deutlich positive Bewertungen von Vorträgen und Vortragenden unter den Tagungshashtags.³⁵ Allerdings führt eine Betrachtung der Tweets und Antworten auf weitere Tweets ohne Tagungshashtag, in denen deutlicher Kritik geäußert wurde. Die Erkenntnisgrenzen der rein digitalen Inhaltsanalyse von Tweets mit Voyant-Tools sind nicht zu übersehen. Sie erschöpfen sich zumeist in Einzelbeobachtungen, wie etwa, dass im Korpus der Frauen das Wort „spannend“ im Laufe der Zeit das Wort „gut“ als Bewertung abzulösen scheint. Die verschiedenen Visualisierungen laden jedoch zu einer explorativen Textanalyse im Wechsel mit einem *Close Reading* ein und zeigen weitere Fragen und Themen auf, für die eine vertiefte Analyse lohnend erscheinen.

Sonderfall 2018?

Tab. 5: Top 15 der häufigsten Wörter an Originaltweets im Teilkorpus 2018

Wort	Häufigkeit	Relative Frequenz
münster	161	6 955
digital	106	4 579
sektion	60	2 592
gut	39	1 685
panel	37	1 598
spannend	37	1 598
histocamptag	35	1 512
uhr	35	1 512
poster	31	1 339
doktorandenforum	30	1 296
resolution	30	1 296
thema	29	1 253
workshop	28	1 210
archivtag	27	1 166
gespaltene	27	1 166

³⁵ Von den insgesamt 186 verwendeten Tweets mit dem Wort „gut“ gibt es nur in neun Fällen die Kombination „nicht gut“.

Für das Teilkorpus des Historikertags von 2018 wird deutlich, dass digitale Themen zwar weiterhin am häufigsten in den Tweets behandelt werden (Tabelle 5). Doch fehlt die Sektion der AG Digitale Geschichte, und auch nicht-digitale Sektionen tauchen bei Twitter auf, wie etwa das überwiegend von jungen und auf Twitter aktiven Historiker:innen besuchte „Schnupperhistocamp“ (hashtag „histocamptag“). Das „Doktorandenforum“ ist laut Frequenzanalyse gleich häufig erwähnt wie die „Resolution“ des Verbandes, über die auch mit dem hashtag #vhdresolution (17 mal) getwittert wurde, das heißt in absoluten Zahlen spielte die Resolution mit 47 Originaltweets während des Historikertags bei Twitter keine hervorgehobene Rolle. Das zeigt ebenso der chronologische Verlauf der relativen Frequenzen, der Auftreten und Zunahme des Worts „Resolution“ auf Twitter ab Donnerstag nach der Mitgliederversammlung verdeutlicht (Abbildung 17).

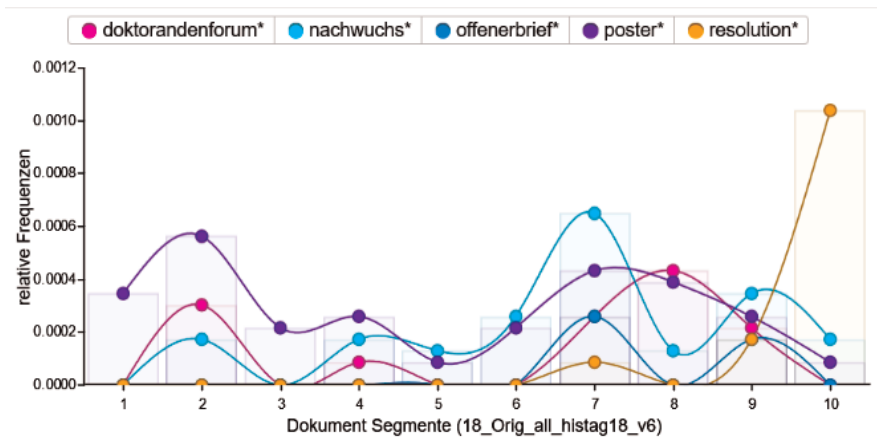


Abb. 17: Relative Frequenzen ausgesuchter Worte im Korpus 2018 in chronologischer Abfolge im Vergleich

Wie das explorative Analysieren in Voyant-Tools zeigt, stand 2018 bei Twitter ein anderes Thema im Vordergrund: die Postersession des Doktorandenforums und im weiteren Sinne der Umgang mit dem „Nachwuchs“ in der Wissenschaft. Die Tweets dazu unter dem Tagungshashtag waren zahlreicher, als es die Frequenzen in Abbildung 17 und Tabelle 5 nahelegen, da sie kein einheitliches Vokabular oder einheitliche Hashtags umfassten. Rund 130 Originaltweets mit hashtag #histag18 konnten wir dazu identifizieren (in einem Korpus von insgesamt 1429 Originaltweets), also knapp 10 %. Zahlreiche Diskussionsstränge zum Doktorandenforum und zum Umgang mit dem Nachwuchs wurden ohne Tagungshashtag publiziert und sollten dadurch vermutlich bewusst unter dem

Radar bleiben. Ein *Close Reading* der Tweets zeigt die Emotionalität des Themas: Das Doktorandenforum sorgte bei den Promovierenden für Unmut, weil die Veranstaltung in einem abseits gelegenen, nicht ausgeschilderten Raum stattfand, einige Poster zu Beginn nicht gedruckt vorlagen (wofür die Veranstalter zuständig waren) und die Promovierenden sich damit auf einen Nebenschauplatz gestellt fühlten. Dies ereignete sich am ersten Tag und das Thema zog sich damit durch die Veranstaltung.

Auffallend sind die hohen Retweet- und Like-Zahlen bei den kritischen Tweets, die für eine relative Dominanz des Themas sorgten. Neben der Ankündigung des offenen Briefs im Portal LISA, der 2018 wie auch im gesamten Korpus der am häufigsten geteilte Tweet war, zogen die kritischen Tweets von zwei Professoren die häufigsten Likes und Retweets auf sich, ein Zeichen für die Unterstützung für das Anliegen der Doktorand:innen durch Teile der Community wie auch für das hierarchische Gewicht der beiden Professorenaccounts. Von den Top zehn der am häufigsten gelikten Tweets 2018 waren die ersten vier mit thematischem Bezug zum Doktorandenforum. Zugleich fällt die Stille der Veranstalter und des Verbandes bei Twitter zum Thema Doktorandenforum auf: Weder der Tagungsaccount @historikertag noch der VHD antworteten bei Twitter auf die Vorwürfe.³⁶ Auch aufgrund dieses Schweigens dürften die Doktorand:innen den Kanal gewechselt und einen offenen Brief an den VHD im Portal LISA publiziert haben, der dort zahlreiche Kommentare erhielt.³⁷ Der Retweet der Ankündigung des Antwortbriefes des Verbandes war der einzige Tweet des VHD zum Thema.³⁸

Die Kontroverse um das Doktorandenforum könnte ein Grund sein, warum 2018 die Kommunikation auf Twitter zum Historikertag quantitativ rückläufig war. Zum einen – so ein Erklärungsversuch – könnten die beiden offiziellen Accounts mit ihrer im Vergleich zu den Vorjahren zurückhaltenden Twitterkommunikation (unbeabsichtigt) den Trend vorgegeben haben. Tagungen, bei de-

36 Der Account setzte während der Tagung ohnehin nur wenige Tweets ab, während er im Vorfeld der Tagung auf Twitter recht aktiv war. Insgesamt nur 47 Tweets wurden vom offiziellen Account unter dem Tagungshashtag gepostet. Zum Vergleich die Anzahl der Tweets des Accounts @Historikertag anlässlich der vorherigen Historikertage: 2012: 0, 2014: 196, 2016: 105.

37 „Gesplante Zunft? Welchen Stellenwert haben Promovierende für den Historiker_innenverband?“, *Portal LISA*, 27.9.2018, Zugriff am 27.7.2021, https://lisa.gerda-henkel-stiftung.de/offener_brief_historikertag18. Dazu die Antwort: „Antwort auf den offenen Brief der Doktorandinnen und Doktoranden“, *Portal LISA*, 28.9.2018, Zugriff am 27.7.2021, https://lisa.gerda-henkel-stiftung.de/antwort_auf_offenen_brief.

38 Der VHD postete zum Historikertag 2018 zwölf eigene Tweets und 19 Retweets (insgesamt 31 Tweets), war aber auch bei den vorhergehenden Veranstaltungen auf Twitter eher zurückhaltend (der Account existiert seit Dezember 2013), 2014: 20, 2016: 46 Tweets.

nen es offizielle Accounts gibt, die zum Twittern aufmuntern oder wie bei den Historikertagen 2014 und 2016 eine eigene Twitter Lounge haben, erhalten generell eine höhere Präsenz in den sozialen Medien. Zugleich drängte das Schweigen der offiziellen Accounts die Kontroverse in andere Kanäle – und zwar in die Blogosphäre – so dass die offizielle Diskussion dazu auf Twitter nachließ. Die sonst auf Twitter aktiven Doktorand:innen widmeten ihre Aufmerksamkeit anderen Inhalten und anderen Kanälen und hielten sich mit Tweets zum sonstigen Tagungsgeschehen zurück.³⁹ Die auf Twitter in Teilen beschädigte Atmosphäre regte darüber hinaus die Community nicht zum Tagungstwittern an. Weitere Gründe dürften ebenso eine Rolle für den Tweetrückgang gespielt haben: So zeigt ein Blick in die Zahlen der aktivsten Accounts, dass Personen, die bei einer Veranstaltung viel getwittert hatten, etwa weil sie vor Ort wohnen oder in die Organisation eingebunden waren, bei den Historikertagen danach sehr viel passiver waren. Auch andere Studien zeigen, dass die Twitternutzung ein und derselben Person bei einer Veranstaltung von Jahr zu Jahr unterschiedlich sein kann.⁴⁰ Es bleibt also abzuwarten, wie sich die Praktiken und Nutzungsformen bei den nächsten Historikertagen weiterentwickeln.

Bibliographie

- Antwort auf den offenen Brief der Doktorandinnen und Doktoranden, *Portal Lisa*, 28.9.2018. Zugriff am 27.7.2021. https://lisa.gerda-henkel-stiftung.de/antwort_auf_offenen_brief.
- Berg, Matthias, Olaf Blaschke, Martin Sabrow, Jens Thiel und Krijn Thijs. *Die versammelte Zunft. Historikerverband und Historikertage in Deutschland 1893–2000*. Göttingen: Wallstein, 2018.
- Beurskens, Michael. „Legal Aspects of Twitter Research.“ In *Twitter and Society*, hg. v. Katrin Weller, Axel Bruns, Jean Burgess und Merja Mahrt, 123–133. New York: Peter Lang, 2014.
- Bruns, Axel und Stefan Stieglitz. „Metrics for Understanding Communication on Twitter.“ In *Twitter and Society*, hg. v. Katrin Weller, Axel Bruns, Jean Burgess und Merja Mahrt, 69–82. New York: Peter Lang 2014.
- Burgess, Jean und Nancy K. Baym. *Twitter. A Biography*. New York: NY University Press, 2020.
- Burghardt, Manuel. „Introduction to Tools and Methods for the Analysis of Twitter Data.“ *Media Linguistics* 1 (2015): 74–91. http://10plus1journal.com/?page_id=268.
- „Gesplattene Zunft? Welchen Stellenwert haben Promovierende für den Historiker_innenverband?“, *Portal Lisa*, 27.9.2018. Zugriff am 27.7.2021. https://lisa.gerda-henkel-stiftung.de/offener_brief_historikertag18.

³⁹ Das bestätigten uns einige der Beteiligten bei unserer Nachfrage im Rahmen dieser Studie.

⁴⁰ Vgl. Smith Risser und Waddell, „Beyond the Backchannel“, 203.

- Grandjean, Martin. „A Social Network Analysis of Twitter: Mapping the Digital Humanities Community.“ *Cogent Arts & Humanities* 3/1 (2016). doi: 10.1080/23311983.2016.1171458.
- Holmberg, Kim und Mike Thelwall. „Disciplinary Differences in Twitter Scholarly Communication.“ *Scientometrics* 2 (2014): 1027–1042. doi: 10.1007/s11192-014-1229-3.
- Kimmons, Royce und George Veletsianos. „Education Scholars’ Evolving Uses of Twitter as a Conference Backchannel and Social Commentary Platform.“ *British Journal of Educational Technology* 47/3 (2016): 445–464. doi: 10.1111/bjet.12428.
- König, Mareike. „Scholarly Communication in Social Media.“ In *Science Communication. Handbooks of Communication Science*, Bd. 17, hg. v. Annette Leßmöllmann, Marcelo und Thomas Gloning, 639–656. München: De Gruyter, 2020.
- König, Mareike, und Paul Ramisch. *Historikertage auf Twitter (2012–2018). Datenreport und Dataset*. Zenodo 2022. doi: 10.5281/zenodo.6362301.
- Mahrt, Merja, Isabella Peters und Katrin Weller. „Twitter in Scholarly Communication.“ In *Twitter and Society*, hg. v. Katrin Weller, Axel Bruns, Jean Burgess und Merja Mahrt, 399–410. New York: Peter Lang, 2014.
- McCay-Peet, Lori und Anabel Quan-Haase. „What is Social Media and What Questions Can Social Media Research Help Us Answer?“ In *The SAGE Handbook of Social Media Research Methods*, hg. v. Luke Sloan und Anabel Quan-Haase, 13–26. London: Sage, 2017.
- Meiler, Matthias. „Wissenschaftliches Twittern. Linguistische Bestandsaufnahme und method(olog)ische Auslotung, insbesondere mit Fokus auf Konferenztweets.“ *Linguistik Online* (2021): 87–113. doi: 10.13092/lo.106.7511.
- Paßmann, Johannes. *Die soziale Logik des Likes. Eine Twitter-Ethnografie*. Frankfurt a.M.: Campus, 2018.
- Ross, Claire, Melissa Terras, Claire Warwick, und Anne Welsh. „Enabled Backchannel: Conference Twitter Use by Digital Humanists.“ *Journal of Documentation* 67 (2011): 214–237. doi: 10.1108/00220411111109449.
- Smith Risser, Hilary, und Glenn Waddell. „Beyond the Backchannel: Tweeting Patterns After Two Educational Conferences.“ *Educational Media International* 55/3 (2018): 199–212. doi: 10.1080/09523987.2018.1512449.
- Williams, Matthew L., Pete Burnap und Luke Sloan. „Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users’ Views, Online Context and Algorithmic Estimation.“ *Sociology* 51/6 (2017): 1149–1168. doi: 10.1177/0038038517708140.
- Zimmer, Michael, und Nicholas John Proferes. „A Topology of Twitter Research: Disciplines, Methods, and Ethics.“ *Aslib. Journal of Information Management* 66/3 (2014): 250–261. doi: 10.1108/AJIM-09-2013-0083.

Oliver Kiechle

Noisy digitalia – Kommunikationslärm in Born Digital-Quellen am Beispiel von Spam im Usenet

Abstract: Historical work that is based on source material from the field of digitalia, thus meaning Born-digital sources, often meets a phenomenon, that is so far unknown to historians in such an extent: Communication noise. It can be found as unwanted advertisement, off-topic comments and even blunt offences. Taking the example of Spam messages in the Usenet of the 1980s and 1990s the article will show how communication nuisance developed during early internet communication and what is its part in historical tradition. With Flaming and Social bots two other manifestations of Noisy digitalia will be considered. Based on that it will be discussed, how historical work can deal with noise in digitalia both in a conceptional and a technical way as well.

Keywords: digitalia, communication noise, social media, Usenet, spam, web history

Zusammenfassung: Bei der geschichtswissenschaftlichen Arbeit mit Quellenbeständen aus dem Bereich der *digitalia*, also genuin digital entstandenem Material, macht sich vor allem im Bereich der Social Media-Quellen ein für die historische Zunft in diesem Ausmaß ungewohntes Phänomen bemerkbar: Kommunikationslärm. Dieser *Noise* kann in Form von unerwünschter Werbung, unpassenden Kommentaren bis hin zu offenen Beleidigungen auftreten und vor allem die automatisierte Analyse größerer Datenbestände beeinflussen.

Im Artikel wird am Beispiel von Spam-Nachrichten im Usenet der 1980er und 1990er Jahre beschrieben, wie sich Lärm in der frühen Internet-Kommunikation entwickelt hat und welchen Anteil er in der Überlieferung einnimmt. Ergänzend werden mit *Flaming* und *Social Bots* zwei weitere Spielarten der *Noisy digitalia* beschrieben. Davon ausgehend wird erörtert, wie mit diesen Störgeräuschen im Rahmen einer historischen Beschäftigung mit *digitalia* sowohl konzeptionell als auch technisch umgegangen werden kann.

Schlagwörter: Digitalia, Communication Noise, Social Media, Usenet, Spam, Web history

Der digitale Austausch ist oft mit inhaltlich unpassenden Einstreuungen oder von externen Akteur:innen platziertem Content durchsetzt, der bestenfalls lästig, schlimmstenfalls sogar destruktiv sein kann. Dieser Kommunikationslärm ist zu einem selbstverständlichen Teil der digitalen Welt avanciert und nimmt dort einen unübersehbaren Stellenwert ein. So kann beispielsweise mehr als die Hälfte der verschickten E-Mails im Jahr 2019 als Spam-Nachrichten klassifiziert werden.¹ Diese Störgeräusche – seien es Werbeanzeigen auf Webseiten, invektive Kommentare in Foren und Chats oder eben die allgegenwärtigen Spam-Nachrichten in Mails oder Newsgroups – beeinflussen nicht nur unsere alltägliche Kommunikation, sondern machen sich ebenfalls bei der historischen Analyse entsprechender archivierter Quellen bemerkbar. Im Folgenden soll deshalb zunächst eine Einordnung von genuin digitalen Materialien im Hinblick auf ihren historischen Kontext gegeben werden. Davon ausgehend werden anhand des Beispiels von Spam-Nachrichten im Usenet der 1990er Jahre verschiedene Aspekte des Auftretens von *Noisy digitalia* erörtert. Danach soll der Blick kurz auf weitere Störfeuer in digitalem Quellenmaterial gerichtet und abschließend Auswirkungen auf die historische Analyse thematisiert werden.

1 Digitalia

In der deutschsprachigen Geschichtswissenschaft wird das hier behandelte Material etwas sperrig gerne als genuin digital entstandene Quellen bezeichnet.² Ein wenig griffiger hören wir auf Englisch hingegen von den *Born Digital Sources*, so beispielsweise bei Niels Brügger,³ der neben *Digitized* (also digitalisierten analogen Objekten) und *Born-Digital* mit *Reborn* noch eine dritte Kategorie digitalen Materials beschreibt: „Reborn digital material is born-digital material that has been collected and preserved, and that has been changed in this process to such an extent that it is not identical to the born-digital material

1 Vgl. Statista, „Spam-Mails – Anteil am gesamten E-Mail-Verkehr weltweit bis 2019,“ Zugriff am 19.08.2021, <https://de.statista.com/statistik/daten/studie/872986/umfrage/anteil-der-spam-mails-am-gesamten-e-mail-verkehr-weltweit>.

2 So zum Beispiel Catherina Schreiber, „Genuine Internetdaten als historische Quellen – Entwurf einer korrealistischen Quellentheorie,“ *Zeitschrift für digitale Geschichtswissenschaften* 0/1 (2012), <http://hdl.handle.net/10993/7981>.

3 Vgl. Niels Brügger, *The Archived Web: Doing History in the Digital Age* (Cambridge, Massachusetts/London: The MIT Press, 2018)

from which it was made.“⁴ Die Veränderung von digitalen Daten durch Archivierung beziehungsweise Wiederherstellung wird uns im Zusammenhang mit der Überlieferungsgeschichte des Usenet zu einem späteren Zeitpunkt noch einmal begegnen. In Hinsicht auf eine allgemein praktikable Begrifflichkeit möchte ich jedoch für eine noch stärkere Vereinfachung plädieren und deshalb für Objekte, die auf der rein digitalen Ebene entstanden, gespeichert und transportiert worden sind, den Terminus der *digitalia* etablieren. Dazu zählen unter anderem E-Mails, Webseiten, digitale Textdokumente, Social Media-Inhalte, digitale Bilder und Videos, aber eben auch solche Datenbündel wie Logfiles, Programmcode, Bewegungsdaten oder Computerspiele. All diese *digitalia* können bereits für heutige und perspektivisch noch stärker für künftige Historiker:innen von entscheidender Bedeutung sein. Zeitgeschichtliche Forschung ist – zumindest für Themenbereiche, die ab den späten 1980er Jahren angesiedelt sind – ohne die Berücksichtigung dieser Quellen kaum mehr vorstellbar. Für aktuellere Zeiträume stellen sie bereits jetzt manchmal nur das einzig verfügbare Quellenmaterial dar.⁵

Die in Binärcode gespeicherten digitalen Daten repräsentieren nicht direkt die eigentlichen Inhalte, sondern letztlich nur die Codierung dieser Inhalte in maschinenlesbaren Befehlen, die den Computer beispielsweise dahingehend steuern, die hinterlegten Werte als Pixel auf dem Schirm zu zeigen, sie als Buchstaben auf Papier zu drucken oder sie anhand eines programmierten Algorithmus zu verändern. Das bedeutet, dass jeglicher Zugang zu den Daten und ihrer Bedeutung eine maschinelle Decodierung benötigt: „[D]er Umgang mit Information [wird][...] nur durch die Vermittlung von Maschinen rezipierbar.“⁶ Sowohl die Darstellung als auch die Analyse dieses Materials muss also durch eine Übersetzungsleistung des Computers stattfinden. Wir können hierbei in Anlehnung an den etablierten Begriff der *Computer Mediated Communication* fast schon von einer *Computer Mediated History* sprechen.

Oftmals wissen oder verstehen wir dabei nicht genau, wie die Daten durch die angewandten Prozesse verändert oder gefiltert wurden oder noch werden. Umso wichtiger ist es, zum einen die technischen Kontexte der Entstehung, Codierung, Archivierung und Überlieferung dieser Daten zu berücksichtigen und zum anderen die Zusammenhänge bei ihrer Darstellung und Analyse im Sinne

⁴ Ibidem, 22.

⁵ Dazu ausführlich Ian Milligan, *History in the Age of Abundance? How the Web Is Transforming Historical Research* (Montreal: McGill Queen's University Press, 2019).

⁶ Peter Haber, *Digital Past – Geschichtswissenschaft im Digitalen Zeitalter* (München: Oldenbourg Verlag, 2011), 102.

einer Toolkritik im Blick zu behalten.⁷ Diese teilweise komplexen Prozesse der Überlieferung digitaler Daten führen unter anderem zu der paradoxen Situation, dass zwar einerseits ein Überfluss an Informationen vorhanden zu sein scheint, andererseits jedoch größere Bestände aufgrund des Verlustes von passender Software und Hardware nicht mehr benutzbar sind.⁸

Dabei ergibt sich vor allem bei der Beschäftigung mit großen Datenbeständen aus dem Bereich der digitalen Kommunikation – beispielsweise E-Mail, Social Media oder Chatforen – ein für die historische Zunft in diesem Ausmaß ungewohntes Phänomen, das wohl am besten mit dem Begriff *Noise* bezeichnet werden kann. Dieser Kommunikationslärm tritt beispielsweise in Form von unerwünschten Werbenachrichten auf, kann aber auch aus Off-Topic-Diskussionen bestehen oder gar aus gezielten Verbalattacken in Form von so genanntem *Flaming* oder *Trolling*. Diese Störgeräusche können in unterschiedlicher medialer Art aufscheinen. Im Folgenden werde ich mich jedoch auf die textuellen Erscheinungsformen konzentrieren.

Bei einer direkten Beschäftigung mit dem Material im Sinne eines *Close Reading* werden solche Störungen zwar meist relativ problemlos erkannt und können entsprechend kontextualisiert und gegebenenfalls ausgefiltert werden. Gehen wir jedoch ins *Distant Reading* über, lassen also größere Textmengen computergestützt analysieren, kommt es zu einer deutlichen Beeinflussung der Ergebnisse durch den Kommunikationslärm, sei es bei einer simplen Wortzählung, einer Sentimentanalyse oder beim Topic Modeling.⁹ Das ist nicht per se negativ. Für manche Fragestellung kann gerade der Lärm nützlich sein und auf Entscheidendes hinweisen, meist verfälscht er jedoch die erhoffte inhaltliche Erfassung des Materials. Entsprechend muss dieser Effekt, wenn er nicht durch vorgeschaltetes *Data Cleaning* minimiert werden soll (und kann), zumindest einkalkuliert werden.

⁷ Vgl. Andreas Fickers, „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?“, *Zeithistorische Forschungen/Studies in Contemporary History* 17/1 (2020): 157–168, doi: 10.14765/ZZF.DOK-1765.

⁸ Vgl. den Beitrag von Moritz Feichtinger in diesem Band.

⁹ Zum grundsätzlichen Erkenntniswert von Topic Modeling in der historischen Quellenanalyse vgl. Martin Fechner und Andreas Weiß, „Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts“, *Zeitschrift für digitale Geisteswissenschaften* 2/2 (2017), doi: 10.17175/2017_005. Siehe auch den Beitrag von Melanie Althage in diesem Band.

2 Spam im Usenet

Um sowohl die (forschungs-)konzeptionelle als auch die praktische Problematik besser zu veranschaulichen, sollen nun verschiedene Aspekte der Arbeit mit Usenet-Nachrichten der 1990er Jahre aufgegriffen werden, die ab einem gewissen Zeitpunkt immer stärker von Spam betroffen waren.

Es handelt sich beim Usenet um ein Netzwerk von Newsservern, durch das eine textbasierte Kommunikation der Nutzer:innen ermöglicht wird. Nachrichten werden im Usenet in eine hierarchische Struktur unterschiedlicher Newsgroups eingestellt und dezentral gespeichert. Im Usenet entwickelten sich grundlegende Erscheinungsformen der Online-Kommunikation wie die Netiquette oder auch die Zusammenstellung von Frequently Asked Questions (FAQ) zur Regelung des Informationsaustausches, aber auch negative Aspekte wie Flaming oder der Versand von Spam lassen sich dort früh identifizieren. Entwickelt wurde das grundlegende Konzept im Herbst 1979. Im Sommer 1980 war das Projekt ausgereift genug, um auf einer Entwicklerkonferenz in Delaware vorgestellt zu werden. Das UNIX User Network etablierte sich schnell als eine Art „ARPANET für arme Leute“¹⁰. Es war zu Beginn ganz bewusst nach kooperativen Grundprinzipien organisiert, bald gab es jedoch zunehmende Auseinandersetzungen über Meinungsfreiheit und Zensurversuche im Usenet, die 1986 zur Neuorganisation und 1988 letztlich zur Abspaltung der (inhaltlich unabhängigen) Alternative-Hierarchie führten.¹¹

Spam-Nachrichten sind uns aus unseren Mailpostfächern wohlbekannt. Die massenhaft und normalerweise ungerichtet versendeten Nachrichten oder E-Mails haben oft einen kommerziellen Hintergrund, lassen sich also meist als unerwünschte Werbung klassifizieren. Solche Phänomene treten bereits in der Computerkommunikation der 1970er Jahre erstmals auf, und auch wenn die Bezeichnung Spam noch nicht benutzt wird, ist die Motivation von „Spammern“ in dieser Zeit nicht vorwiegend auf kommerzielle Ziele ausgerichtet: „[W]hatever you were posting, be it duplicated, way too long, saturated with quotes,

10 Das ARPANET war die erste systematische Vernetzung von Computern in den USA, die jedoch nur eine enge Auswahl von Computern aus Regierungsinstitutionen und Eliteuniversitäten einschloss. Eine faktisch solide Darstellung der Frühgeschichte des Usenet bieten Michael Hauben und Ronda Hauben, *Netizens: On the History and Impact of Usenet and the Internet* (Los Alamitos, California: IEEE Computer Society Press, 1997). In ihren Bewertungen bleiben die Haubens jedoch oftmals auffällig unkritisch, was eventuell an ihrer persönlichen Nähe zur Usenet Community gelegen haben könnte.

11 Vgl. dazu ausführlicher Oliver Kiechle, „Ein Gespaltenes Netz? – Das Usenet der 1980er-Jahre zwischen Regulierung und Anarchie,“ in *Zur Geschichte des digitalen Zeitalters*, hg. v. Ricky Wichum und Daniela Zetti (Wiesbaden: VS Verlag für Sozialwissenschaften 2022), 125–142.

contextually inappropriate, had broken the implicature of network conversation that held that you should be in some way relevant.“¹² Spam dient also zu Beginn vor allem dazu, die Möglichkeiten technischer Computernetze auszunutzen, um Aufmerksamkeit zu erregen: „Spammers take the infrastructure of the ‚good things‘ and push them to extremes.“¹³

Diese Motivationslage änderte sich mit der Zeit zunehmend in eine kommerzielle Richtung, was unter anderem durch Vorkommnisse im Usenet vorangetrieben wurde. Auch dort, wo man sich zwar grundsätzlich anti-kommerziell gab, können wir schon früh eine Vorstufe des Spammings im Extensive Cross-Posting feststellen. Hier zirkulierten beispielsweise ab und an Angebote von Computerfirmen durch das verpönte gleichzeitige Posten in mehreren Newsgroups. Für Aufsehen und Diskussionen sorgte dann jedoch vor allem ein Aufruf des Nutzers JJ im Jahr 1988, der an Dutzende Newsgroups die Aufforderung richtete, ihn als ziemlich abgebrannten Studenten doch finanziell zu unterstützen.¹⁴ Wurde der JJ-Vorfall innerhalb des Usenet 1988 zwar kurz und heftig in Hinblick auf die inhaltliche Stabilität des Netzwerks und vor allem auf mögliche Sanktionen gegen JJ diskutiert, kehrte schon kurz danach wieder relative Ruhe ein: „Usenet returned to trading anecdotes, sharing knowledge, joking, and bickering, the moment of self-reflexive panic slipped into memory.“¹⁵

Die Bezeichnung Spam wurde in Anlehnung an einen alten Monty Python-Sketch dann Anfang der 1990er ebenfalls im Usenet geprägt. Einhergehend mit dem *Eternal September* des Jahres 1994,¹⁶ als der Internet-Anbieter AOL seinen Kund:innen neben dem einfachen Zugang zum gerade aufstrebenden World Wide Web auch Eingang zum Usenet verschaffte, nahm nicht nur die Zahl der

12 Finn Brunton, *Spam: A Shadow History of the Internet* (Cambridge, Massachusetts: The MIT Press, 2013), 48.

13 Ibidem, 11.

14 Brad Templeton, „Origin of the Term ‚Spam‘ to Mean Net Abuse“ (o.D.), in *Brad Templeton's Home Page, Essays on Junk E-mail (Spam)*. Zugriff am 22.08. 2021, <https://www.templetons.com/brad/spamterm.html>.

15 Brunton, *Spam*, 48.

16 Dieser begann gar nicht im Herbst, sondern bereits im März 1994. Die Bezeichnung rührte daher, dass es im September durch den Beginn des amerikanischen Studienjahres und entsprechend vieler neuer Studierender alljährlich auch eine im Vergleich zum Rest des Jahres höhere Zahl an neuen Usenet-Nutzer:innen gegeben hatte. Nun stieg die Zahl durch das AOL-Angebot aber permanent und deutlich und vor allem über die ursprünglichen universitären (und meist eher liberal eingestellten) Kreise hinaus an, was viele „alte“ Nutzer:innen zunächst irritierte und später sogar verprellte (vgl. Bryan Pfaffenberger, „A Standing Wave in the Web of Our Communication: Usenet and the Socio-Technical Construction of Cyberspace Values,“ in *From Usenet to CoWebs: Interacting with Social Information Spaces*, hg. v. Christopher Lueg und Danyel Fisher (London: Springer, 2003), 40–41).

Usenet-Nutzer:innen rapide zu, sondern die unerwünschte Kommunikation geriet ab Mitte der 1990er Jahre zu einer Massenerscheinung: Große aufsehenerregende Aktionen gab es 1994 zunächst mit der Nachricht „Global Alert for All: Jesus is Coming Soon“, die von einem Studenten an über 200 Newsgroups versendet wurde, und schließlich vor allem durch die Werbung der Anwaltskanzlei Canter und Siegel, die ihr Angebot für Unterstützung bei der Green Card Lotterie automatisiert an fast alle damaligen Newsgroups verschickte.¹⁷ In der Folge wuchs das Spam-Aufkommen im Usenet immens, so dass manche Newsgroups zu Beginn der 2000er Jahre regelrecht geflutet wurden und die eigentlichen Diskussionsthemen komplett untergingen.

Anders als beim JJ-Vorfall reagierte die Community auf Canter/Siegel umgehend, vor allem der automatisierte Versand der Spam-Nachricht hatte viele Systemadministrator:innen aufgeschreckt. Eine erste direkte Reaktion war eine Flut von E-Mails an den Internetanbieter der Kanzlei, der die Anwälte daraufhin prompt vom Netz nahm (eine eher symbolische Aktion, da diese einen Tag später wieder online waren).¹⁸ Weitreichender war jedoch die Reaktion eines bis heute nicht genauer identifizierten Users namens *Cancelmoose*, der einen *Cancelbot* entwickelte. Dieses Programm nutzte im Prinzip den gleichen technischen Ansatz wie die ursprüngliche Spam-Nachricht, wanderte automatisiert durch die Usenet-Newsgroups und löschte die Green Card-Nachricht, wo immer sie zu finden war. Auch wenn diese Vorgehensweise bei den meisten Nutzer:innen Beifall fand, flammten nun erneute Diskussionen über Zensur und Meinungsfreiheit im Usenet auf, zumal Canter und Siegel im Nachgang versuchten, sich als „icons of free speech on Usenet and the Internet generally“ zu gerieren.¹⁹ Um eine Nachricht als Spam klassifizieren und dann gegebenenfalls löschen lassen zu können, akzeptierte die Usenet Community schließlich den nach seinem Erfinder benannten Bredibart-Index als maßgebliche Kennzahl für exzessives Cross-Posting.²⁰

3 Überlieferung von Spam im Usenet

Das Usenet definierte sich von Beginn an als flüchtiges Medium, da vor allem aufgrund der damaligen Limitierungen bei Übertragungsbandbreite und Spei-

¹⁷ Brunton, *Spam*, 48.

¹⁸ Pfaffenberger, „Usenet and the Socio-Technical Construction of Cyberspace Values,“ 36.

¹⁹ Brunton, *Spam*, 57.

²⁰ Pfaffenberger, „Usenet and the Socio-Technical Construction of Cyberspace Values,“ 39.

cherplatzkosten eine längere Speicherung der Nachrichten nicht vorgesehen und kaum möglich war. Entsprechend gibt es erhebliche Lücken in der historischen Überlieferung des Materials, insbesondere bis in die Mitte der 1990er Jahre. Mehrere kleinere Sammlungen und ein größerer Korpus (die so genannten UTZOO-Tapes) alter Usenet-Nachrichten stellen die zentrale Überlieferung von Usenet-Inhalten aus den 1980er und frühen 1990er Jahren dar. Diese wurden im Jahr 2001 zusammen mit den von Google aufgekauften Beständen der Firma *Deja News*, die ab Mitte der 1990er Jahre versucht hatte, das Usenet systematisch zu archivieren und ihren Nutzer:innen zur Verfügung zu stellen, in die neu geschaffenen *Google Groups* integriert.²¹ Theoretisch lassen sich über *Google Groups* nach wie vor eine enorme Anzahl von historischen Usenet-Nachrichten finden, jedoch wurden in den letzten Jahren nicht nur die Suchmöglichkeiten stark eingegrenzt, sondern auch automatisierte Crawl-Prozesse zur individuellen Zusammenstellung eines Korpus weitestgehend geblockt. Anhand der *Historical Usenet Collection* des *Internet Archive* zeigt sich jedoch, dass diese Blockade offenbar umgangen werden kann.²²

Analog zur gängigen Archivpraxis bedeutet Webarchivierung, dass Quellen selektiert, gesammelt und geordnet werden müssen. Bei *digitalia* finden meist zusätzlich technische Transformationsprozesse statt: Für die Archivierung müssen Daten eventuell bereinigt und gefiltert oder aber auch mit zusätzlichen Metadaten versehen werden, bevor sie in ein Archivformat wie zum Beispiel WebARCHive (WARC) für Internetinhalte überführt werden. Eine bei Usenet-Inhalten zusätzliche Problematik stellt der seit der Veröffentlichung von archivierten Usenet-Nachrichten durch *Deja News* und einer darauf folgenden massiven Gegenwehr eines Teils der Community aufgrund von Datenschutzbedenken eingeführten X-No-Archive-Header dar. Diese Einstellung im Header der Nachricht verhindert im Normalfall eine Archivierung durch automatisierte Prozesse. Zusätzlich besteht auch nach wie vor die Möglichkeit, eigene Postings auf Antrag aus der Archivierung entfernen zu lassen.

Für Spam-Nachrichten aus dem Usenet bedeutet das vor allem, dass diese zum Teil in die archivierten Usenet-Bestände übernommen wurden, zum Teil aber eben auch nicht: Entweder wurden sie bereits vor dem Archivieren systematisch gelöscht (durch Cancelbots) beziehungsweise durch die jeweiligen Nutzer:innen lokal ausgefiltert oder dann während des Archivierungsprozesses ent-

21 Camille Paloque-Bergès, „Usenet as a Web Archive: Multi-Layered Archives of Computer-Mediated-Communication,“ in *Web 25: Histories from the First 25 Years of the World Wide Web*, hg. v. Niels Brügger (New York et al., Peter Lang Publishing, 2017), HAL:halshs-01843627, 236.

22 Zur Usenet-Archivierung vgl. Kiechle, „Das Usenet der 1980er-Jahre zwischen Regulierung und Anarchie“, 127–130.

fernt. Nichtsdestotrotz finden sich in einigen archivierten Newsgroups nach wie vor eine große Anzahl an Spam-Nachrichten, was unter anderem daraus resultiert, dass Spam in *Google Groups* grundsätzlich nicht pauschal ausgefiltert wird.

Ein weiteres Problem von öffentlich zugänglich gemachten Usenet-Nachrichten besteht darin, dass Nutzer:innen im Vertrauen auf die angenommene Flüchtigkeit der Kommunikation dort oftmals ihre E-Mail-Adressen angegeben haben. Diese können nun automatisiert eingesammelt und zum Versand weiterer Spam-Nachrichten benutzt werden. Diesem Umstand trägt beispielsweise das erst seit 2021 existierende Portal *usenetarchives.com* dadurch Rechnung, dass persönliche Informationen auf den zur Verfügung gestellten historischen Usenet-Nachrichten weitgehend unkenntlich gemacht wurden.²³

4 Flaming und Bots

Neben Spam lassen sich noch weitere Arten von Lärm in *digitalia* finden, die an dieser Stelle kurz beschrieben werden sollen: Das *Flaming* beziehungsweise *Trolling* und so genannte *Social Bots*.

Das Phänomen des Flaming, also abwertende oder gar beleidigende Einlassungen, lässt sich schon bei den ersten Gehversuchen der Online-Kommunikation in den 1970er Jahren erkennen, massiv tritt es dann mit der Etablierung und dem raschen Anwachsen des Usenet auf. Beim Flaming handelt es sich um das bewusste Versenden von provozierenden oder beleidigenden Nachrichten, beispielsweise in Online-Foren, Bulletin Boards oder Newsgroups. Die wortnahe Übersetzung von Flaming mit „zündeln“ passt zu diesem Vorgang ganz gut. Es geht meist darum, sachliche Diskussionen zu stören oder gar zu sprengen. Nicht selten entzündet sich an gleichgelagerten Reaktionen auf einen Flame ein regelrechter *Flame War*, der die ursprüngliche Intention einer Online-Unterhaltung komplett in den Hintergrund rückt.²⁴

²³ Usenet Archives, <https://www.usenetarchives.com/>, online seit Mai 2021. Hier lassen sich archivierte Usenet-Nachrichten relativ komfortabel nach inhaltlichen Aspekten durchsuchen. Jedoch ist zum einen nicht ganz klar, welche historischen Usenet-Bestände dort überhaupt eingebunden sind und zum anderen erschwert das Fehlen persönlicher Informationen – so nachvollziehbar diese Herangehensweise in Bezug auf Datenschutzaspekte auch ist – teilweise eine tiefer gehende Analyse.

²⁴ Michael Nycyk, „Trolls and Trolling History From Subculture to Mainstream Practices,“ in *The SAGE Handbook of Web History*, hg. v. Niels Brügger und Ian Milligan (Los Angeles et al.: SAGE, 2019), 580.

Im Usenet wurde früh versucht, dieses Verhalten einzudämmen oder zumindest zu kanalisieren. So wurde bereits in den ersten Versionen der als Netiquette bekannt gewordenen Verhaltensregeln für das Posten im Usenet das Flaming mehr oder weniger geächtet.²⁵ Da es aber zum einen kaum Sanktionsmöglichkeiten gegen diejenigen gab, die Flames posteten und zum anderen viele Nutzer:innen Gefallen an gelungenen Beleidigungen fanden, führte dieser Schritt meist ins Leere. Bei manchen Usern, die nur gelegentlich und oft eher sarkastisch gemeint ausfällig werden wollten, bürgerte es sich ein, den Flame mit den Tags *Flame On/Flame Off* zu kennzeichnen und dadurch als unsachlichen Einwurf sichtbar zu machen.²⁶ Eine Möglichkeit, Flames zu kanalisieren, war die Einrichtung der Newsgroup *net.flames*, auf die aus dem Ruder laufende Diskussionen abgeleitet werden sollten, was jedoch nur zum Teil gelang.

Im Usenet der frühen 1980er war es üblich, seinen echten Namen zu benutzen und meist sogar noch die postalische Adresse anzugeben, was justiziable Beschimpfungen eher selten machte. Die zunehmende Anonymisierung und vor allem eine immens steigende Nutzer:innenzahl führte zu immer schlimmeren Flames. Ab den 1990ern findet sich das Phänomen dann in Form des Trollings auf beinahe allen Kommunikationsebenen des World Wide Web wieder.²⁷ In den heutigen Social Media sind solche Ausfälle weit verbreitet und häufig mit Begriffen wie Hate-Speech, Shitstorm oder Fake News verknüpft.

Ab den 2000er Jahren machen sich zudem vermehrt Social Bots in der Online-Kommunikation bemerkbar, obwohl auch diese beileibe keine neue Erscheinung waren. Bei Bots handelt es sich um algorithmisch gesteuerte und oft mit Hilfe von Machine-Learning programmierte Tools, die versuchen, die menschliche Kommunikation im Internet im Sinne ihrer Programmierer:innen

25 Vgl. eine der ersten und vermutlich am meisten rezipierte Kodifizierung der Usenet-Netiquette: Chuq v. Rospach, „A Primer on How to Work with the USENET Community“, Zugriff am 05.07.2019, <http://www.mit.edu/afs.new/athena/activity/i/isig/Primer/usenet-primer>.

26 Denis Howe, „Flame from FOLDOC“, *Free Online Dictionary of Computing*, Zugriff am 30.08.2021, <https://foldoc.org/flame>.

27 Die Bezeichnung „Troll“ wurde offenbar ebenfalls im Usenet geprägt, vgl. Nycyk, „Trolls and Trolling History“, 580.

zu simulieren und zu beeinflussen.²⁸ Wir finden diese Programme schon im Usenet der 1990er Jahre beispielsweise in Form der vorher erwähnten Cancelbots. Dabei handelt es sich aber meist um Einzelaktionen. Mit Auftreten der großen Social Media Plattformen in den 2000er Jahren halten Social Bots relativ schnell massenhaften Einzug in die Online-Kommunikation. Diese agieren meist zweckgerichtet: Es geht zum Beispiel darum, bestimmte Kanäle mit auf den ersten Blick regulären Messages zu blockieren, personalisierte Werbung für bestimmte Produkte zu platzieren, Trends zu kreieren oder schlicht um Trolling. In diesen Zusammenhang sind aktuelle Diskussionen um eine mögliche Beeinflussung politischer Wahlen durch Bots in sozialen Netzwerken einzuordnen.

5 *Noisy digitalia* in der historischen Analyse

Wie sollten wir nun mit diesen Stördaten in unseren digitalen Quellen, seien es Flame Wars oder Hate Speech, Spam-Fluten oder nichtmenschliche Kommunikation, in der historischen Arbeit umgehen? Zwei grundsätzliche Möglichkeiten scheinen sich mir hierbei anzubieten: Entweder versuchen, sie einzubinden mit all den Weiterungen, die dies nach sich ziehen könnte oder den Lärm weitestgehend auszufiltern, was aber Folgen für die Validität der Daten bzw. der Inhalte haben kann.

Wie eingangs erwähnt, erscheint es wenig problematisch, wenn wir das verwendete Material selbst ansehen und den *Noise* direkt erkennen und einordnen können; nichts anderes beinhaltet die Arbeit mit nicht-digitalen schriftlichen Quellen. Arbeiten wir jedoch mit Hilfe des Computers und analysieren unsere *digitalia* in größerem Maßstab, können umfangreiche Vorarbeiten im Sinne eines Data Cleaning und Preprocessing nötig werden. Bezogen auf die vorher ausgeführten Störkategorien kann sich dies unterschiedlich problematisch gestalten.

In Bezug auf Spam-Nachrichten funktioniert eine Filterung verhältnismäßig gut. Die gewohnten Spam-Filter orientieren sich an bestimmten Schlüsselwörtern, um entsprechende Nachrichten auszufiltern und lassen sich zum Teil im Nachhinein anwenden. Dies geht sowohl mit E-Mails als auch mit Usenet-Nachrichten, da beide Formate zum einen ursprünglich auf demselben RFC-Standard 822 („Standard for the Format of ARPA Internet Text Messages“) beruhen und

²⁸ Vgl. Joseph Seering et al., „The Social Roles of Bots. Evaluating Impact of Bots on Discussions in Online Communities,“ *Proceedings of the ACM on Human-Computer Interaction* 2, Computer Supported Cooperative Work (CSCW) (2018): 1, doi: 10.1145/3274426.

zum anderen manche Mailprogramme zugleich als Newsreader fungieren und entsprechend die gleichen Filter-Plugins angewandt werden können. Trotzdem muss jeweils nachgehalten werden, ob die Filteroptionen nicht auch normale Nachrichten aussortieren. Bei aktuellen historischen Arbeiten zum Usenet wurde das Spam-Problem adressiert: Avery Dame-Griff beschreibt bei seiner Untersuchung von Transgender-Newsgroups das zunehmende Auftreten von Spam-Nachrichten, die nicht nur Werbebotschaften, sondern auch automatisierte Anti-Transgender-Postings transportierten.²⁹ Aus diesem Grund entschied er sich, diese Nachrichten nicht auszufiltern, „as they formed an inextricable part of the Usenet experience. Nevertheless, this choice did consequently limit the applicability of more linguistically focused methodology, such as topic modelling.“³⁰ In ihrer Doktorarbeit zur LGBTQ-Identität im frühen Internet beschäftigt sich Sarah McTavish ebenfalls mit entsprechenden Newsgroups und stellte in Hinblick auf die Gruppe *alt.sex.motss* fest: „[T]he group quickly became consumed with spam messages, both explicit and not.“³¹ Inwiefern diese Nachrichten Teil ihrer Analyse sind, wird jedoch leider offen gelassen.

Wenn wir den Blick auf das Flaming wenden, gibt es bereits Ende der 1990er Jahre, als die Sozialwissenschaften speziell das Usenet als Untersuchungsgegenstand für sich entdecken, erste Versuche, Tools zu entwickeln, die Flames erkennen und ausfiltern können.³² Grundsätzlich arbeiten diese Anwendungen wie moderne Hate-Speech-Detektoren, mit Machine-Learning-Methoden. Das Programm erlernt also bestimmte Schlüsselbegriffe und Kommunikationsmuster und versucht damit, entsprechende Inhalte zu erkennen.³³ Abgesehen davon, dass der Erfolg dieser Methoden wie immer von den Trainingsdaten abhängt, findet Flaming – das sich oft ironisch oder sarkastisch darstellt – manchmal so subtil statt, dass diese Tools an ihre Grenzen stoßen.

Die technischen Möglichkeiten, Bots *ex post* zu identifizieren, sind ebenfalls begrenzt, da die entsprechenden Programme meist versuchen, den ursprüngli-

²⁹ Vgl. Avery Dame-Griff, „Herding the ‚Performing Elephants:‘ Using Computational Methods to Study Usenet,“ *Internet Histories* 11/10 (2019): 12, doi: 10.1080/24701475.2019.1652456.

³⁰ Ibidem.

³¹ Sarah McTavish, „The Queer Eternal September: LGBTQ Identity on the Early Internet and Web“ (Dissertation University of Waterloo, Ontario, 2020), URI: <http://hdl.handle.net/10012/16336>, 105.

³² Melanie Martin, Annotating Flames in Usenet Newsgroups: A Corpus Study (NSF Minority Institution Infrastructure Grant Site Visit to NMSU CS Department, 2002). Zugriff am 30.08.2021, https://www.cs.nmsu.edu/~mmartin/pubs/martin_poster.pdf.

³³ Vgl. Sasha Sax, Flame Wars: Automatic Insult Detection (Stanford University, 2016), Department of Computer Science Report, Stanford University, Zugriff am 30.08.2021, <http://cs224d.stanford.edu/reports/Sax.pdf>.

chen Absender per IP-Adresse zurückzuverfolgen, um einen Bot erkennen zu können. Dies ist bei archivierten *digitalia* jedoch nicht mehr ohne weiteres möglich. Auch imitieren moderne Bots das menschliche Kommunikationsverhalten oft so echt, dass es kaum mehr von dem echter Nutzer:innen zu unterscheiden ist.³⁴ Das führt zu der fast schon paradoxen Situation, dass künstliche Intelligenz – wenn wir fortgeschrittene Machine-Learning-Technik so nennen wollen – versucht zu erkennen, ob eine andere künstliche Intelligenz die Kommunikation von Menschen imitiert hat.

Grundsätzlich würde ich die Frage nach Einbinden oder Ausfiltern von der jeweiligen Forschungsperspektive abhängig machen wollen. Wenn durch Distant Reading alle Aspekte eines digitalen Kommunikationsclusters erfasst werden sollen, verbietet sich eine Filterung fast von selbst. Auch können solche negative Erscheinungen zu entscheidenden Hinweisen auf Bereiche werden, die Close Reading erfordern. Ebenfalls sollte nicht unterschätzt werden, dass Spam und noch mehr das Flaming ein natürlicher Teil der historischen Internetkommunikation ist, der auch nonkonforme Akteur:innen – Rebell:innen, Brigant:innen, Anarchist:innen – repräsentiert.

Geht es jedoch um klar eingegrenzte Inhalte, die vom Noise offensichtlich massiv gestört werden, ist eine starke Filterung sicherlich das Mittel der Wahl, nicht zuletzt um eine sinnvolle linguistische Analyse zu ermöglichen. Grundsätzlich besteht jedoch bei allen Verfahren des Data Cleanings die Gefahr einer determinierenden Vorgehensweise. Die Parameter der Filter werden dann, überspitzt gesagt, einfach so lange angepasst, bis die Daten zur vorgegebenen Fragestellung passen. Dadurch besteht, wie durch die unkritische Verwendung von Tools, deren genaue Funktionsweise nicht nachvollziehbar ist, letztlich die Gefahr, das vielbeschworene „Vetorecht“ der Quellen zu unterminieren, also falsche, beziehungsweise nicht mehr durch das Quellenmaterial gedeckte Deutungen zu provozieren.

In diesem Zusammenhang soll noch einmal betont werden, dass die Kenntnis davon, unter welchen Prämissen bestimmte *digitalia* archiviert worden sind, entscheidend bei der Einordnung der gespeicherten Daten ist. Bei der Webarchivierung wird beispielsweise nicht alles grundsätzlich gespeichert, sondern sowohl aufgrund technischer Kapazitäten, als auch aus Relevanzgründen im Vorfeld die speicherwürdigen Webseiten ausgewählt, mit der Gefahr, bestimmte Randaspekte auszuklammern und nicht zu überliefern. In den letzten Jahren hat eine engagierte Diskussion über die ethischen Aspekte der Webarchivierung

³⁴ Vgl. Adrian Rauchfleisch und Jonas Kaiser, „The False Positive Problem of Automatic Bot Detection in Social Science Research,“ *PLOS ONE* 15/10 (2020), doi: 10.1371/journal.pone.0241045.

zudem dazu geführt, dass manche Inhalte ganz bewusst nicht archiviert oder massiv gefiltert werden, um beispielsweise Opfer rassistischer oder homophober Attacken im Netz zu schützen.³⁵ Dies führt ganz allgemein zu dem Problem, dass hierbei viele rechtliche Fragen, namentlich bezüglich der Urheber- und Persönlichkeitsrechte, noch nicht ausreichend geklärt sind. Diese Probleme betreffen zum einen den Archivierungsprozess, zum anderen aber auch massiv die Forschungsarbeit mit rechtlich heiklen Beständen. Zwar kann die Analyse und Bewertung des Materials eventuell in einem rechtlich sicheren Rahmen stattfinden, soll jedoch im Sinne eines guten Forschungsdatenmanagements die Datengrundlage nachverfügbar gemacht werden, könnte dies problematisch werden. Die Arbeit mit *digitalia* hält jedenfalls offensichtlich noch einige Unwägbarkeiten bereit.

Bibliographie

- Brügger, Niels. *The Archived Web: Doing History in the Digital Age*. Cambridge, Massachusetts, London, England: The MIT Press, 2018.
- Brunton, Finn. *Spam: A Shadow History of the Internet*. Cambridge, Massachusetts: The MIT Press, 2013.
- Dame-Griff, Avery. „Herding the ‚Performing Elephants:‘ Using Computational Methods to Study Usenet.“ *Internet Histories* 11/10 (2019): 1–22, doi: 10.1080/24701475.2019.1652456.
- Fechner, Martin und Andreas Weiß. „Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts.“ *Zeitschrift für digitale Geisteswissenschaften* 2/2 (2017), doi: 10.17175/2017_005.
- Fickers, Andreas. „Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?“ *Zeithistorische Forschungen/Studies in Contemporary History* 17/1 (2020): 157–168, doi: 10.14765/ZZF.DOK-1765.
- Graham, Pamela M. „Guest Editorial: Reflections on the Ethics of Web Archiving.“ *Journal of Archival Organization* 14/3–4 (2017): 103–110. doi: 10.1080/15332748.2018.1517589.
- Haber, Peter. *Digital Past – Geschichtswissenschaft im Digitalen Zeitalter*, München: Oldenbourg Verlag, 2011.
- Hauben, Michael und Ronda Hauben. *Netizens: On the History and Impact of Usenet and the Internet*. Los Alamitos, California: IEEE Computer Society Press, 1997. Zugriff am 30.08.2021, <http://www.loc.gov/catdir/bios/wiley046/97005920.html>.
- Howe, Denis. „Flame from FOLDOC.“ *Free Online Dictionary of Computing*, Zugriff am 30.08.2021, <https://foldoc.org/flame>.

35 Pamela M. Graham, „Guest Editorial: Reflections on the Ethics of Web Archiving,“ *Journal of Archival Organization* 14, 3–4 (2017): 103–110, doi: 10.1080/15332748.2018.1517589.

- Kiechle, Oliver. „Ein gespaltenes Netz? – Das Usenet der 1980er-Jahre zwischen Regulierung und Anarchie.“ In *Zur Geschichte des Digitalen Zeitalters*, hg. v. Ricky Wichum und Daniela Zetti, 125–142. Wiesbaden: VS Verlag für Sozialwissenschaften, 2022.
- Martin, Melanie. „Annotating Flames in Usenet Newsgroups: A Corpus Study.“ NSF Minority Institution Infrastructure Grant Site Visit to NMSU CS Department, 2002, Zugriff am 30.08.2021, https://www.cs.nmsu.edu/~mmartin/pubs/martin_poster.pdf.
- McTavish, Sarah. *The Queer Eternal September: LGBTQ Identity on the Early Internet and Web*, Dissertation University of Waterloo, Ontario, 2020, URI: <http://hdl.handle.net/10012/16336>.
- Milligan, Ian. *History in the Age of Abundance? How the Web Is Transforming Historical Research*. Montreal: McGill Queen's University Press, 2019.
- Nycyk, Michael. „Trolls and Trolling History From Subculture to Mainstream Practices.“ In *The SAGE Handbook of Web History*, hg. v. Niels Brügger und Ian Milligan, 577–89. Los Angeles, London, New Delhi, Singapore, Washington, DC, Melbourne: SAGE reference, 2019.
- Paloque-Bergès, Camille. „Usenet as a Web Archive: Multi-Layered Archives of Computer-Mediated-Communication.“ In *Web 25: Histories from the First 25 Years of the World Wide Web*. Hg. v. Niels Brügger. Peter Lang Publishing, 2017. HAL: halshs-01843627.
- Pfaffenberger, Bryan. „A Standing Wave in the Web of Our Communication': Usenet and the Socio-Technical Construction of Cyberspace Values.“ In *From Usenet to CoWebs: Interacting with Social Information Spaces*, hg. v. Christopher Lueg und Danyel Fisher. Computer Supported Cooperative Work. London: Springer, 2003.
- Rauchfleisch, Adrian, und Jonas Kaiser. „The False Positive Problem of Automatic Bot Detection in Social Science Research.“ *PLOS ONE* 15/10 (2020): e0241045. doi: 10.1371/journal.pone.0241045.
- Rospach, Chuq V. „A Primer on How to Work with the USENET Community.“ Zugriff am 05.07.2021, <http://www.mit.edu/afs.new/athena/activity/i/isig/Primer/usenet-primer>.
- Sax, Sasha. „Flame Wars: Automatic Insult Detection.“ Department of Computer Science Report, Stanford University, Zugriff am 30.08.2021, <http://cs224d.stanford.edu/reports/Sax.pdf>.
- Schreiber, Catherina. „Genuine Internetdaten als historische Quellen – Entwurf einer korrealistischen Quellentheorie.“ *Zeitschrift für digitale Geschichtswissenschaften* 0/1 (2012): 1–15. <http://hdl.handle.net/10993/7981>.
- Seering, Joseph, Juan P. Flores, Saiph Savage und Jessica Hammer. „The Social Roles of Bots.“ *Proceedings of the ACM on Human-Computer Interaction* 2, Computer Supported Cooperative Work (CSCW) (2018): 1–29. doi: 10.1145/3274426.
- Statista. „Spam-Mails – Anteil am gesamten E-Mail-Verkehr weltweit bis 2019 | Statista.“ Zugriff am 19.08.2021. <https://de.statista.com/statistik/daten/studie/872986/umfrage/anteil-der-spam-mails-am-gesamten-e-mail-verkehr-weltweit/>.
- Templeton, Brad. „Origin of the Term ‚Spam‘ to Mean Net Abuse.“ (o.D.), in Brad Templeton's Home Page, Essays on Junk E-mail (Spam). Zugriff am 22.08.2021, <https://www.templetons.com/brad/spamterm.html>.

