

Forschungsdesign 4.0. Datengenerierung und Wissenstransfer in interdisziplinärer Perspektive

Veranstalter: Institut für Sächsische Geschichte und Volkskunde e. V. (ISGV); Sächsische Landesbibliothek – Staats- und Universitätsbibliothek (SLUB)

Datum, Ort: 19.04.2018 - 21.04.2018, Dresden

Bericht von: Cornelia Eisler, Bundesinstitut für Kultur und Geschichte der Deutschen im östlichen Europa, Universität Oldenburg

Neue Technologien, durch die Informationen und große Datenmengen digital schneller und effizienter verarbeitet, analysiert sowie verbreitet werden können, bieten Forschern und Forscherinnen scheinbar unbegrenzte Möglichkeiten. Um Quellenmaterial zu erhalten und für ihre Untersuchungen nutzen zu können, müssten sie lediglich ‚schürfen‘ (*text mining, digging*), ‚zu Tage fördern‘ (*„data is the new oil“*) oder ‚ernten‘ (*data harvesting*). Die Verwendung des Ressourcenvokabulars ruft auf der einen Seite „Goldgräber“- und Entdeckerstimmung hervor, auf der anderen werden damit Ausbeutungsmechanismen und koloniale Tendenzen assoziiert, die bewusst oder unbewusst in die Thematik eingeschrieben werden. Neben den Stimmen, die sich für die digitalen Möglichkeiten und ihr Potenzial begeistern, finden sich jene, die vor zu hohen Erwartungen, vor Risiken und ‚naivem Pioniergeist‘ warnen. Ein in dieser Hinsicht breites Meinungsspektrum offenbarte sich auf der Tagung in Dresden, die das so genannte Forschungsdesign 4.0 zur Diskussion stellte. Im Zentrum standen die Erfahrungen, Reflektionen, Deutungen und Schlussfolgerungen der referierenden Europäischen Ethnologen und Ethnologinnen, (programmierenden) Historiker/innen, Sprachwissenschaftler/innen, Archivare und Archivarinnen in Zeiten des fortgeschrittenen digitalen Wandels. Ungeachtet der verschiedenen Ausgangspositionen drehen sich die Diskussionen immer wieder um zentrale und zugleich vieldeutige Begriffe wie Digitalisierung, Vernetzung (als Wert per se?), Transparenz, Zugänglichkeit, Standardisierung, Effektivität, Sicherung, Nachnutzung und Nachhaltigkeit von Daten sowie Wissensstrukturen.

Das Kollegium des ISGV (Dresden) als Veranstalter bot disziplinenübergreifend ein vielseitiges Tagungsprogramm, das in drei thematische Sektionen aufgliedert war. Diese entsprachen in etwa dem Umgang mit Daten und Ergebnissen im Verlauf des Forschungsprozesses. Die Datengenerierung sowie der Einfluss sozialer Medien und der Digitalisierung standen im Mittelpunkt der ersten Sektion. Wie mit Forschungsdaten umgegangen werden kann, an welche Grenzen Forscher und Forscherinnen stoßen und wie die Ergebnisse präsentiert und vermittelt werden können, innerhalb wie außerhalb des akademischen Bereiches, erörterten die Vortragenden in der zweiten und dritten Sektion. Ebenso wenig wie es sich in der Praxis dabei um einen linearen Prozess handelte, kam dieses Ordnungsprinzip nicht ohne sektionsübergreifende Querverweise aus.

Mit einem historisch-kritischen Rück- und Ausblick eröffnete FRIEDEMANN SCHMOLL (Jena) die Tagung. Er reflektierte in seinem Einführungsvortrag zum Nachlass des „Atlas der Deutschen Volkskunde“ den Umgang mit bereits bestehenden Datenmengen, und zwar den 4,5 Millionen handschriftlichen Fragebögen dieses „ethnographischen Mammutprojektes“, und lotete das Digitalisierungspotential dieses „schwierigen Erbes“ aus. Bereits zu Beginn des 20. Jahrhunderts setzte eine Modernisierungsphase in der Forschung ein, in der die Bewältigung großer Datenmengen neue und vielversprechende Ergebnisse erwarten ließ. Bis in die 1980er-Jahre wurden die Antworten auf teils voreingenommene Fragen zum Alltag der Menschen vorwiegend in den ländlichen Gebieten des Deutschen Reiches und im Ausland, die von circa 20.000 Gewährsleuten um 1935 erhoben wurden, ausgewertet und in Karten verzeichnet, um den prädisponierten Zusammenhang von „Raum – Volk – Kultur“ zu visualisieren. Die „digitale Wiederbelebung“ dieses politischen Projektes, dessen Herangehensweise höchst kritisch zu betrachten ist, wird folglich kontrovers diskutiert. Schmoll fragte nach den Möglichkeiten der Verfügbarmachung und wie mit dem Bestand trotz seiner Geschichte innovativ geforscht werden könnte. Eine einfache Antwort scheint es nicht zu geben, wie

den anregenden Diskussionen im Anschluss zu entnehmen war.

Der Fragestellung, wie Textkorpora aufbereitet werden können, so dass sie über ihr sprachwissenschaftliches Potential hinaus auch für kulturwissenschaftliche Studien verwendbar sind, widmete sich HAUKE BARTELS (Bautzen). In seinem aufschlussreichen Beitrag führte er zunächst in die Korpuslinguistik im Allgemeinen ein und diskutierte die Chancen und Herausforderungen eines solchen „Globalkorpus“ am Beispiel nieder- und obersorbischer Textkorpora als Grundlage für die kulturgeschichtliche Forschung. Zu den Voraussetzungen zählt unter anderem die Lemmatisierung und Normalisierung der Schreibvariation (vor allem bei stark flektierenden Sprachen). Ein Grundfrage ist die Verortung zwischen quantitativen und qualitativen Zugängen, also das Finden einer Balance zwischen der automatischen Sprachverarbeitung unter Nutzung von *big data*, *distant reading*, mit einer hohen Fehlerakzeptanz und der philologischen Herangehensweise, die durch *‘less data’*, *close reading* sowie geringe Fehlertoleranz charakterisiert werden kann. Die Erstellung eines „Globalkorpus“ verspricht neue Perspektiven für die Forschung, doch ist die Aufbereitung der Texte mit einem hohen Aufwand an Personal und Kosten verbunden und setzt eine differenzierte Ablauforganisation voraus. Damit waren zentrale Herausforderungen benannt, auf die insbesondere die Referenten und Referentinnen, die im weiteren Verlauf der Tagung beispielhaft Digitalisierungsprojekte präsentierten, ebenfalls hinwiesen.

Dass im Rahmen solcher Projekte auch ein Mentalitätswechsel der Bearbeiter/innen erwartet wird und die lange Bearbeitungsdauer der Vorhaben nicht immer mit der Entwicklungsgeschwindigkeit technologischer Neuerungen mithalten kann, machte CLEMENS RADL (München) in seinem Überblick über (fast) gescheiterte Editionsprojekte im Kontext der *Monumenta Germaniae Historica* deutlich. Mit seinem Plädoyer für Vorbedingungen und die Nutzbarkeit von Forschungsdaten durch offene Lizenzen stellt sich allerdings die Frage, inwieweit zukünftige Forschungsvorhaben vorhersagbar sind und ob sie es sein sollten.

Eine begriffskritische Auseinandersetzung mit Daten beziehungsweise Datenmaterial und eine stärkere Reflektion dessen, was die Digitalisierung im Bereich der Wissensproduktion impliziert, forderte GERTRAUD KOCH (Hamburg) in ihrem öffentlichen Abendvortrag zur „Datafication qualitativer Forschung“. Sie fragte danach, wie sich die epistemologischen Grundlagen aufgrund der digitalen Durchdringung des Forschungsprozesses verändern, welche Neuerungen damit verbunden sind, aber auch was im Zuge dessen aufgegeben würde. Die „neue politische Ökonomie des Forschens“, die Koch hervorhebt, lasse sich aus den Forderungen nach der Wiederverwendung von Forschungsdaten, nach Transparenz und Effizienz sowie dauerhafter Verfügbarkeit erkennen. Die damit verbundene Strukturierung und Standardisierung bedeute zwar Effizienz, doch zugleich die Reduktion der Informationen. Kontextualität und Referentialität sind zentrale Parameter der qualitativen Forschung, die sich etwa in den Metadaten nur in begrenzter Form wiederfinden. Grundsätzlich aber stelle sich – unter Verweis auf Annette N. Markhams Diskussion des Datenbegriffes – die Frage, wie die Digitalisierung für die qualitative Forschung produktiv gemacht werden könne.

An die Problematik der Kontextualisierung anknüpfend, stellte STEFAN GROTH (Zürich) Überlegungen zur Beurteilung von Dokumenten aus politischen Prozessen als Quellen vor, die in den letzten Jahren – den Forderungen nach mehr Transparenz entsprechend – in zunehmendem Maße in Online-Archiven der Öffentlichkeit zur Verfügung gestellt werden, und zwar über Beschluss- und Protokolltexte hinaus auch Verhandlungsdokumentationen und vorläufige Ergebnisse. Weniger in der statistischen Auswertung dieser Quellen sieht er Potential als vielmehr in der Erkundung der Entstehungszusammenhänge. Es sollte erörtert werden, welche Rolle diese Online-Archive tatsächlich innehaben und welche Dokumente gerade nicht online zur Verfügung gestellt werden. Darüber hinaus sei zu berücksichtigen, dass sich die zeitnahe Publikation und öffentliche Dokumentation auf die Handlungen und Verhandlungen der Beteiligten und somit auch auf die Produktion der politischen Dokumente auswirkt,

was in der Konsequenz zu einer Transformation der Quellen führt.

Während zunächst also die Wirkungen der Digitalisierung und Online-Verfügbarkeit auf die Forschung im Mittelpunkt standen, beschäftigte sich JULIA FLEISCHHACK (Göttingen) umgekehrt mit der Übertragung von Feldforschungsmethoden auf die „Welten“ des Internets. Unter Berufung auf die Arbeiten von Danah Boyd und insbesondere Tom Boellstorffs Studie zu Second Life erörterte sie die Charakteristika der Online-Feldforschung und betonte, dass es eben nicht allein um Technologie in diesem Zusammenhang geht, sondern um die Menschen und ihre kulturellen Praktiken. Zur Diskussion stand anschließend, inwieweit Off- und Online-Welten als Forschungsfelder überhaupt getrennt voneinander aufgefasst werden können.

Wie eng verwoben beide Welten bereits im Forschungsprozess selbst sein können, zeigte BARBARA FRISCHLING (Graz) am Beispiel eines Selbstversuches, indem sie ihr Agieren als Wissenschaftlerin bei gleichzeitiger intensiver Facebook-Nutzung im Nachhinein zu rekonstruieren suchte. Dabei stellte sie fest, dass die Online-Unterstützung bereits so selbstverständlich geworden ist, dass sie sich nicht im Feldforschungstagebuch niederschlug. Rückblickend fragte sich Frischling jedoch, inwieweit der Prozess durch automatisierte und vorstrukturierte Informationsvermittlung, die als so genannte *filter bubbles* oder *echo chambers* deklariert werden, beeinflusst wurde, und im Rahmen der Forschung entsprechend zu reflektieren sei.

Passend dazu stellte TORSTEN SCHRADER (Mainz) die Frage, welche technischen Fertigkeiten zum Repertoire zukünftiger Geisteswissenschaftler/innen, insbesondere Historiker/innen, gehören sollten.¹ Die Beherrschung von Programmiersprachen sei nicht zwangsläufig als Voraussetzung zu verstehen, sondern vielmehr wird das Verständnis für *computational thinking*, für die Architekturmodelle von Software und für die Internetstrukturen als grundlegende Kompetenz gesehen. Die Anpassung methodischer Fähigkeiten an die neuen Formen von Quellen, unter anderem die Befähigung, source code lesen zu können, sollte als Ergänzung der klassischen Quellenkritik verstanden wer-

den. Und schließlich plädiert Schrader dafür, Software und *source code*, die für geschichtswissenschaftliche Analysen genutzt werden, offenzulegen, um die Ergebnisse reproduzierbar und zitierbar zu machen.

Die Forderung nach Offenheit, nicht der Software, sondern der Daten aus qualitativer Forschung diskutierte SABINE IMERI (Berlin) in ihrem Beitrag „Open data“ in den ethnologischen Fächern. Möglichkeiten und Grenzen eines Konzeptes“. Sie legte den Schwerpunkt auf die Charakteristika ethnologischen Forschens, insbesondere auf die Feldforschung. Hier spielen Persönlichkeitsrechte, Vertrauensfragen, langfristige Beziehungen und Zugänge zum Feld eine zentrale Rolle. Das Datenmaterial kann ein Risikopotential für die Beteiligten aufweisen, etwa in Krisenregionen, die untersucht werden, und entstammen komplexen Situationen, die als solche kaum reproduzierbar sind. Die Art der erhobenen Informationen ist von einer anderen Qualität und daher müssten hier spezifische Herangehensweise für die betroffenen Fächer entwickelt werden, um den fächerübergreifenden Forderungen nach Transparenz und Nachvollziehbarkeit der Forschung entgegenzukommen.

Es stellt sich also grundsätzlich die Frage, inwieweit die ökonomisch beeinflussten politischen Forderungen nach einer 4.0-Anpassung Folge geleistet werden kann und sollte. Ebenso wenig wie es ein Forschungsdesign 1.0 bis 3.0 gegeben hat, wird ein Forschungsdesign 4.0 entwickelt werden. Es ist vielmehr anzunehmen, dass sich Forschende in Zukunft bewusst positionieren und eine individuelle Balance zwischen den unterschiedlichen Anforderungen an Forschungsdaten und -prozesse finden müssen, um Anforderungen wie Offenheit, Transparenz, Nachhaltigkeit und Persönlichkeitsrechte, Vertrauensbildung innerhalb des Forschungsfeldes jeweils abzuwägen. Die vielfältigen Herangehensweise allein in den Geisteswissenschaften wird man im Tagungsband im Detail nachvollziehen können.

¹Die Präsentation „Programming Historians. Unverzichtbare Kompetenz, methodisches Plus oder nicht zwingende Eigenschaft zukünftiger Historikerinnen?“ ist online zugänglich: <https://metacontext.github.io/2018-programming-historians/> (Stand: 27.04.2018).

Konferenzübersicht:

Enno Bünz / Thomas Bürger / Ira Spieker (Dresden): Eröffnung und Einführung

Friedemann Schmoll (Jena): Erbe, Altpapier, Archiv? Fragen an den Nachlass „Atlas der deutschen Volkskunde“

Sektion: Datengenerierung und Dokumentation im Forschungsprozess

Sarah Kleinmann (Dresden): Integration digitaler Räume. Die Projekte „Kontaktzonen“ und „Lebensgeschichtliches Archiv“ (Impulsreferat)

Hauke Bartels (Bautzen): Von sprachwissenschaftlicher zu kulturwissenschaftlicher Nutzung: Der Aufbau eines interdisziplinär nutzbaren sorbischen Textkorpus

Clemens Radl (München): Forschungsdaten im Editionsprozess am Beispiel der Monumenta Germaniae Historica

Gertraud Koch (Hamburg): Wissensproduktion 4.0 – Zur Datafication qualitativer Forschung

Ruth Dorothea Eggel (Bonn): Mapping Cyber-Spaces. Potentiale und Herausforderungen in der Anwendung von Mental Maps als Methode in digitalen Online-Räumen

Stefan Groth (Zürich): Narratologisches Doppel: Text und Kontext in Online-Archiven und Interaktionssituationen

Julia Fleischhack (Göttingen): Online-Spuren. Methodische Überlegungen zum Ethnografieren im ‚Digitalen‘

Barbara Frischling (Graz): Feldforschung in der Filter-Bubble. Reflexionen zum (digitalen) Forschungsalltag

Sektion: Präsentation von und Umgang mit Forschungsdaten

Christian Schuffels (Dresden): Forschungsdaten = Open Data? Die „Sächsische Biografie“ und der „Codex diplomaticus Saxoniae“ (Impulsreferat)

Torsten Schrade (Mainz): Programming Historians – unverzichtbare Kompetenz, methodisches Plus oder nicht zwingende Eigenschaft zukünftiger Historiker_innen?

Yannick Weber (Mainz): RI OPAC und Regestendatenbank. Verschränkte Speicher unterschiedlich strukturierter Wissens

Klaus Graf (Aachen): Forschungsdaten und Nachschlagewerke – Wie Gedächtnisinstitutionen sich vernetzen müssen

Walter Fanta (Wien): Die Grenzen von XML/TEI am Beispiel von Robert Musils Nachlass in MUSILONLINE

Sektion: Vermittlung und Formate von Forschungsergebnissen

Nadine Kulbe (Dresden): Kontextualisierungen. Der Nachlass von Adolf Spamer (Impulsreferat)

Sabine Imeri (Berlin): „Open Data“ in den ethnologischen Fächern. Möglichkeiten und Grenzen eines Konzepts

Martin Munke (Dresden): Citizen Science/Bürgerwissenschaften: Projekte, Probleme, Perspektiven am Beispiel Sachsen

Eberhard Wolff (Zürich): Auf der Suche nach dem Grenznutzen von Digitalisierung und Online-Verfügbarkeit in der Kulturwissenschaft

Isabelle Berens (Marburg): Hessische Zeitungen im Ersten Weltkrieg: Ein Kooperationsprojekt zwischen den wissenschaftlichen Bibliotheken und der Landesgeschichtsforschung

Annalena Schmidt / Marek Slodička (Bautzen): „Sorbenwissen“ – ein Portal für Wissenschaft und Öffentlichkeit

Tagungsbericht *Forschungsdesign 4.0. Datengenerierung und Wissenstransfer in interdisziplinärer Perspektive*. 19.04.2018 - 21.04.2018, Dresden, in: H-Soz-Kult 18.06.2018.