

Positionspapier zu Infrastrukturen für historische Sozialdaten in der Zeitgeschichte

Autorinnen und Autoren (in alphabetischer Reihenfolge):

Prof. Dr. Kerstin Brückweh	Berliner Hochschule für Technik, BHT
Prof. Dr. Christina von Hodenberg	Deutsches Historisches Institut London, DHIL
Prof. Dr. Lutz Raphael	Universität Trier
Prof. Dr. Sabine Reh	Bibliothek für Bildungsgeschichtliche Forschung (BBF) des DIPF Leibniz-Institut für Bildungsforschung und Bildungsinformation und Humboldt-Universität zu Berlin
Dr. Pascal Siegers	GESIS – Leibniz-Institut für Sozialwissenschaften
Dr. Clemens Villinger	Friedrich-Schiller-Universität Jena
Kathrin Zöller	Bibliothek für Bildungsgeschichtliche Forschung (BBF) des DIPF Leibniz-Institut für Bildungsforschung und Bildungsinformation

Stand: März 2023

Inhalt

Einleitung.....	2
Zentrale Ergebnisse der Bedarfsanalyse	3
Dateninfrastrukturen für die Zeitgeschichte.....	4
Zusammenarbeit mit Dateninfrastrukturen für die Geschichtswissenschaften	12
Finanzierung und Governance	12
Nächste Schritte	13

Die Erstellung dieser Publikation wurde gefördert durch die Deutsche Forschungsgemeinschaft (DFG)-
Projektnummer 418958624.

Einleitung¹

Unter „historischen Sozialdaten“ werden qualitative und quantitative Daten verstanden, die von privaten Organisationen oder öffentlichen Einrichtungen unter Anwendung sozialwissenschaftlicher oder statistischer Forschungsmethoden zur Beschreibung von Gesellschaften erhoben und gesammelt wurden und als Quellen durch die historisch arbeitende Wissenschaft, insbesondere Zeithistoriker:innen wiederverwendet werden können.² Zu dieser Quellengattung gehören somit die aus zeitgenössischer, im weitesten Sinne sozialwissenschaftlich vorgehender Forschung generierten Materialien sowie die Dokumente audio-visueller Medienarchive und die Datenbestände amtlicher statistischer Erhebungen. Das im Januar 2020 gestartete DFG-Projekt „Sozialdaten als Quellen der Zeitgeschichte“ hat sich mit den Problemen und Potentialen auseinandergesetzt, die die Erschließung dieser in der Geschichtswissenschaft bislang nur relativ selten genutzten Quellen für die zeitgeschichtliche Forschung mit sich bringt. Die hinter dem Projekt stehende Arbeitsgruppe „Sozialdaten als Quellen der Zeitgeschichte“³ hat einerseits die konzeptionellen und methodischen Debatten sowie die in den letzten zehn Jahren neu entstandenen Forschungsthemen vertieft.⁴ Andererseits hat sie den Bedarf für neue Infrastrukturen, die speziell auf einen historiographischen Umgang mit historischen Sozialdaten ausgerichtet sind, erhoben und analysiert⁵ und stellt hiermit ihre Empfehlungen vor.

Der wissenschaftliche Wert von Sozialdaten für die historische Forschung ist unbestritten und geht über die Zeitgeschichte hinaus. Zeithistoriker:innen nutzen in den letzten zehn Jahren vermehrt historische Sozialdaten, um zeitgenössische Forschungsergebnisse der Sozialwissenschaften in geschichtswissenschaftlicher Perspektive kritisch zu prüfen, neue Fragen an die Daten zu stellen und neue Themenstellungen zu entwickeln.

Für die Sozialwissenschaften existiert seit der Gründung des Zentralarchivs für Empirische Sozialforschung an der Universität zu Köln (heute: Datenarchiv der GESIS Leibniz-Institut für

¹ Die Autoren danken dem Kollegium des Landesarchivs Baden-Württemberg (Grundsatzabteilung) für die wertvollen Hinweise zur Verbesserung des Papiers. Der Dank geht insbesondere an Clemens Rehm, Kai Naumann und Andreas Neuburger.

² Zur Definition vgl. Villinger, Clemens/ Zöllner, Kathrin: Tagungsbericht: Sozialdaten und die Geschichtsschreibung von Wohneigentum und Sozialräumen, in: H-Soz-Kult, 08.06.2021, <https://www.hsozkult.de/conferencereport/id/fdkn-127526> (Zugriff am 15.2.2022).

³ Sie ist Teil des bei der Werner Reimers Stiftung angesiedelten Arbeitskreises „Sozialdaten und Zeitgeschichte“, der zunächst von sog. Mikrodaten ausging und damit vor allem Daten der amtlichen Statistik im Blick hatte; mittlerweile werden gleichermaßen qualitative und quantitative Daten in den Blick genommen.

⁴ Dazu erscheint in der Zeitschrift *Geschichte und Gesellschaft* ein Themenheft „Sozialdaten als Quellen der Zeitgeschichte“ (2022, 1). Siehe zudem Villinger, Clemens/ Zöllner, Kathrin: Tagungsbericht: Sozialdaten und die Geschichtsschreibung von Wohneigentum und Sozialräumen, in: *Moderne Stadtgeschichte* (2021) 1, S. 158-163; Villinger, Clemens: Conference Report: Contemporary Historians and the Reuse of Social Science-Generated Datasets: An International Dialogue on the Challenges Presented by Social Science Data, in: *German Historical Institute London Bulletin XLIV* (2022) 1, S. 150-156.

⁵ Villinger, Clemens/ Zöllner, Kathrin: Tagungsbericht: Sind das Daten und muss das (deshalb) weg?: Rechtliche und ethische Grundlagen zeithistorischer Forschung mit Sozialdaten, in: H-Soz-Kult, 15.12.2021, <https://www.hsozkult.de/conferencereport/id/fdkn-127403> (Zugriff am 4.7.2022); dies.: Tagungsbericht: HT 2021: Sozialdaten für die zeitgeschichtliche Forschung: Wie bauen wir die digitale Infrastruktur aus?, in: H-Soz-Kult, 15.1.2022, <https://www.hsozkult.de/conferencereport/id/fdkn-127816> (Zugriff am 4.7.2022).

Sozialwissenschaften) im Jahr 1960 und der Forschungsdatenzentren (FDZ) beim Rat für Sozial- und Wirtschaftsdaten (RatSWD) seit 2004 eine leistungsfähige, aber fragmentierte Infrastruktur für den Zugang zu einem umfangreichen Bestand von Sozialdaten für die wissenschaftliche Nutzung. Es handelt sich dabei vor allem um Daten aus großen Umfrageprogrammen seit den 1980er Jahren und amtliche Statistiken.⁶ Spätestens seit den 1950er-Jahren wurden historische Sozialdaten detailliert dokumentiert und langfristig gesichert (wobei auch ältere Datenbestände verfügbar, aber weniger gesichert sind). Der Zugang zu den Daten ist möglich, wenn auch nicht auf die Bedarfe von Historiker:innen zugeschnitten, weil die (digitale) Verknüpfung von Daten aus verschiedenen Quellen (zum Beispiel unterschiedlichen FDZ) meist nicht möglich ist und Historiker:innen vielfach (im Interesse der üblichen Nutzung unterschiedlicher Quellenarten) die in den sozialwissenschaftlichen Fächern verbreitete Trennung zwischen qualitativen und quantitativen Zugängen zu überwinden versuchen. Stärker als bisher muss zukünftig bei der Diskussion über den Erhalt sozialwissenschaftlich erhobener Forschungsdaten die Perspektive der Historiker:innen mitgedacht werden. Die Erforschung historischer Fragestellungen insbesondere über die Zeit seit den 1970er Jahren wird ohne Rückgriff auf historische Sozialdaten als Quellen vielfach nicht mehr zufriedenstellend möglich sein. Diese Daten sollten zudem langfristig so erhalten werden, dass auch neue, digitale Forschungsmethoden für deren Auswertung angewandt werden können.

Zentrale Ergebnisse der Bedarfsanalyse

Auf der Basis der im Projekt durchgeführten Fachgespräche, Workshops und der Auswertung von Curricula zeigt unsere Bedarfsanalyse, dass die Nutzung historischer Sozialdaten in der deutschen zeithistorischen Forschung noch immer eine Ausnahme darstellt.⁷ Das hat mehrere Ursachen, wobei der hohe Mehraufwand einer Analyse von Sozialdaten im Vergleich zur Auswertung „klassischer“ Archivquellen den am häufigsten genannten Grund darstellt.

Der Mehraufwand bei der Auswertung von historischen Sozialdaten ergibt sich aus zwei Gründen: Erstens müssen diese Daten durch Zeithistoriker:innen zunächst kritisch in ihren Entstehungskontext eingebettet werden. Dafür benötigen die Forschenden nicht nur Zugang zu den quantitativen oder qualitativen Daten und deren Dokumentation, sondern sie müssen auch Informationen über die Motivation und Interessen der Akteur:innen sammeln, die Konzeption und Umsetzung jeweiliger Studien (z.B. durch Projektanträge, Gremienprotokolle) rekonstruieren und unter Umständen – vor allem, wenn keine Unterlagen mehr vorliegen – mit „Erzeuger:innen-Interviews“ fehlende Kontextinformation selbst generieren. Diese Vorarbeiten sind wesentlich umfangreicher als bei

⁶ Der langfristige Erhalt von Daten aus weniger institutionalisierten Forschungskontexten an Universitäten und Forschungsinstituten – gerade auch im Bereich der qualitativen Sozialforschung – erfolge deutlich weniger systematisch, so dass weite Teile der sogenannten „long tail“ Daten bis heute nicht systematisch erschlossen wurden.

⁷ Für die Bedarfsanalyse siehe Zöller, Kathrin/ Villinger, Clemens/ Siegers, Pascal/ Reh, Sabine/ Raphael, Lutz/ von Hodenberg, Christina/ Brückweh, Kerstin: Sozialwissenschaftliche Forschungsdaten als historische Quellen: Welche Infrastrukturbedarfe hat die zeitgeschichtliche Forschung?, in: RatSWD Working Paper 277 (2022), <https://doi.org/10.17620/02671.66> (Zugriff am 4.7.2022).

herkömmlichen, in Archiven oder Bibliotheken vorgehaltenen Quellen. Zweitens benötigen Zeithistoriker:innen Kompetenzen in der Datenanalyse, die sie sich in der Praxis entweder fachfremd oder autodidaktisch aneignen müssen. Lehrangebote zur (sozialwissenschaftlichen) Analyse qualitativer und quantitativer Sozialdaten sind in den geschichtswissenschaftlichen Curricula deutscher Universitäten derzeit kaum verankert. Selbst wenn der Zugang zu Sozialdaten in den dezentralen sozialwissenschaftlichen Archiven bzw. FDZ transparent geregelt ist, gestaltet sich die Suche nach passenden Beständen aufgrund der fragmentierten Infrastruktur aufwendig. Die Qualität historischer Sozialdaten für die zeithistorische Nachnutzung lässt sich oft erst nach detaillierter Sichtung der Studiendokumente einschätzen.

Aus den Arbeitsergebnissen des DFG-Projektes „Sozialdaten als Quellen der Zeitgeschichte“ lassen sich drei kurzfristige Bedarfe für Infrastrukturen ableiten (Punkte 1 bis 3), mit denen Methodenkompetenzen entwickelt, die Auffindbarkeit, Qualität und Zugänglichkeit historischer Sozialdaten gewährleistet sowie rechtliche und forschungsethische Fragen der Nutzung adressiert werden können:

- 1) Die Gründung eines Lehr- und Weiterbildungszentrums für zeithistorische Sozialdatenanalyse,
- 2) Die Schaffung eines Informationsportals historische Sozialdaten,
- 3) Die Gründung eines Forums Ethik und Recht in der Zeitgeschichte.

Abschließend wird die künftige Sicherung historischer Sozialdaten thematisiert (Punkt 4), und zwar als Anbieterspflicht durch Forschende an öffentliche Archive. Damit nehmen wir Fragen des Forschungsdatenmanagements auf und zielen darauf ab, mit dem Archivrecht einen in der DSGVO gegebenen, aber unzureichend wahrgenommenen Rechtsrahmen für die langfristige Sicherung von Sozialdaten zu nutzen, damit diese als historische Quellen für die Geschichtsschreibung moderner Industriegesellschaften nicht verloren gehen.

Dateninfrastrukturen für die Zeitgeschichte

Eine zeithistorische Auswertung von Sozialdaten erfordert eine quellenkritische Kontextualisierung der Datengenerierung und -verwendung. Zudem können – gerade bei der Nutzung qualitativer Sozialdaten – theoretisch Verknüpfungen mit anderen Informationsquellen über Personen, zu Betrieben oder anderen Organisationen hergestellt werden. Solche Verknüpfungen werden von sozialwissenschaftlichen Datenhaltern jedoch regelmäßig untersagt, weil dafür die informierte Einwilligung der Studienteilnehmenden fehlt. Eine zeithistorische Forschungsdateninfrastruktur muss diese Nutzungshürden überwinden. Dazu sind Unterstützungsstrukturen auf drei Handlungsfeldern notwendig.

(1) Lehr- und Weiterbildungszentrum für zeithistorische Sozialdatenanalyse

Weil für die Auswertung von historischen Sozialdaten spezielle Kompetenzen benötigt werden, deren

Vermittlung derzeit nicht in der zeitgeschichtlichen Lehre verankert ist, schlagen wir die Gründung eines **Lehr- und Weiterbildungszentrums für zeithistorische Sozialdatenanalyse** in enger Abstimmung mit dem NFDI4Memory-Konsortium vor. Das Zentrum soll Lehrangebote für Methoden zur Auswertung historischer Sozialdaten entwickeln. Die Angebote des Zentrums sollen in die von NFDI4Memory geplante Plattform der Task Area 4 (*Data Literacy*) eingebettet werden, um die Sichtbarkeit der NFDI für die Kompetenzentwicklung in der Zeitgeschichte zu nutzen.⁸

Die von Zeithistoriker:innen angewendeten Methoden bei der Auswertung qualitativer und quantitativer Daten hängen vom jeweiligen Erkenntnisinteresse ab. Deshalb werden Lehrangebote benötigt, die spezifisch auf die Bedarfe und Erkenntnisinteressen der zeithistorischen Forschung zugeschnitten sind. So sind für Historiker:innen mehrdimensional beschreibende Verfahren wichtig wie zum Beispiel Korrespondenzanalysen, multidimensionale Skalierungen oder die multimethodischen Vorgehensweisen der qualitativen Forschung. Geschichtswissenschaftliche Forschung zielt meist nicht auf die Entdeckung von Gesetzmäßigkeiten in der sozialen Welt ab, sondern darauf, historische Ereignisse und Prozesse zu rekonstruieren und Strukturen in ihrem zeitlichen Wandel in unterschiedlicher Form nachzuvollziehen. Deshalb sind fortgeschrittene Verfahren der Kausalanalyse, die sich in den Sozialwissenschaften als Standards etabliert haben, tendenziell weniger bedeutsam als Verfahren, die Stabilität und Wandel in den historischen Sozialdaten aufzeigen und visualisieren. Zudem benötigt die Zeitgeschichte wissens- und wissenschaftsgeschichtliche Ansätze, um diese Bestände als historische Quellen kritisch zu kontextualisieren.

Vermittelt werden sollte darüber hinaus ein allgemeines Verständnis von Datafizierungsprozessen, von Datenkompetenz und von Programmierkompetenz in Open-Source Tools, die sich in jüngerer Vergangenheit fachübergreifend etabliert haben (wie etwa R oder Python). Auch Kenntnisse von Verfahren der *Digital Humanities*, die Texte und Bilder automatisierten Auswertungen unterziehen (etwa das Topic Modeling), sind wichtig. Insgesamt sollte es nicht nur um die Vermittlung von Anwenderwissen über Auswertungssoftware gehen (etwa SPSS/Stata, MAXQDA oder Text Mining-Tools wie Voyant u.a.), sondern um themengeleitete *Data Literacy*.

Auch wenn die Besonderheiten der zeithistorischen Sozialdatenanalyse einen Schwerpunkt in diesem Lehrangebot einnehmen sollten, dürfen Grundkenntnisse der Methoden der empirischen (quantitativen und qualitativen) Sozialforschung und der von Sozialforscher:innen häufig verwendeten Verfahren im Programm nicht fehlen. Denn Zeithistoriker:innen sollten in der Lage sein, die älteren Erhebungstechniken aus den Sozialwissenschaften zu verstehen und die Qualität der Ergebnisse zu beurteilen. Auch diejenigen, die selbst nicht in eine Reanalyse der Daten einsteigen, müssen die zeitgenössischen Forschungsmethoden kritisch reflektieren können. Das betrifft in der quantitativen Methodenlehre neben Verfahren der deskriptiven Statistik (z.B. Kreuztabellen oder Maße der zentralen Verteilung) auch regressionsanalytische Anwendungen. Hinsichtlich qualitativer Methoden sind grundlegende Einsichten in codierende und interpretativ- rekonstruierende Verfahren zentral. Die Grundlage für die Auswertung historischer Sozialdaten sollte eine solide *Data Literacy* in den

⁸ Zu geplanten measures in Task Area 4 vgl. <https://4memory.de/task-areas/task-area-4-data-literacy/> (Zugriff am 30.6.2022).

Geschichtswissenschaften bilden. Der Bedarf, diese *Data Literacy* zu stärken, geht weit über das Feld der Zeitgeschichte hinaus. In der zur Förderung empfohlenen Initiative NFDI4Memory ist die Entwicklung von Datenkompetenz in den historisch arbeitenden Disziplinen eines von fünf Handlungsfeldern, weil die Digitalisierung der Forschungspraxis die Vermittlung von Grundlagen der Datenorganisation, -bearbeitung und -auswertung erfordert. Lehrangebote zur Entwicklung der *Data Literacy* sollten zügig in die Curricula der historischen Studiengänge integriert werden, um alle Studierenden frühzeitig an den Umgang mit Forschungs- und Sozialdaten sowie digitalen Auswertungsverfahren heranzuführen. Grundkenntnisse in Datenstrukturen, Datenbankmanagement und Datenauswertung sind in Zukunft auch für Historiker:innen unverzichtbar. Sie sind in vielen Berufsfeldern jenseits der Hochschullaufbahn stark nachgefragt, wie im Archivwesen, im Journalismus, in Museen und im Verlagsgeschäft. Zudem müssen Anreize geschaffen werden, die das Engagement in diesem Feld fachintern belohnen, etwa durch Netzwerkbildung, Publikationsmöglichkeiten, attraktive Fortbildungsangebote und Preise.

(2) Informationsportal historische Sozialdaten

Die Entwicklung fachspezifischer Dateninformationsportale hat derzeit Konjunktur. Es existieren bereits Vorbilder aus der empirischen Bildungsforschung (www.forschungsdaten-bildung.de), der Migrationsforschung (<https://ethmigsurveydatahub.eu/>) oder der qualitativen Sozialforschung (Qualidata Network). Für die Forschenden stellen solche Portale gegenüber generischen Suchportalen einen Mehrwert dar, weil die Studieninformationen um Merkmale erweitert werden, die eine konkrete Einschätzung der Nutzbarkeit und Qualität der Daten für den eigenen Forschungszweck ermöglichen. Darüber hinaus informieren fachspezifische Portale die Forschenden über besonders relevante methodische oder rechtliche Grundsatzfragen bei der Nutzung von Daten. Deshalb ist ein fachspezifisches Dateninformationsportal für die Zeitgeschichte ein wichtiger Baustein, um Hürden bei der Nutzung historischer Sozialdaten als Quellen zu verringern. Wir schlagen deshalb ein **Informationsportal historische Sozialdaten** vor. Es sollte mindestens zwei Komponenten enthalten:

(a) Unterstützung bei der Datensuche

Daten aufzufinden ist trotz der Bestrebungen, Dateninfrastrukturen nach den FAIR-Prinzipien⁹ zu gestalten und Suchportale zu vernetzen, für Zeithistoriker:innen eine Hürde. Hauptgrund dafür ist, dass es – anders als zum Beispiel in Großbritannien – in Deutschland noch immer kein zentrales Suchportal für historische Sozialdaten gibt. Zudem sind die existierenden Kataloge so organisiert, dass eine Suche nach Inhalten neuere Daten in den Trefferlisten priorisiert und die Verschlagwortung historische Themen und Begriffe nicht ausreichend berücksichtigt. Beispielsweise ist eine Suche nach Zeitpunkten oder -abschnitten derzeit nicht ideal implementiert. Zeithistoriker:innen brauchen aber eine

⁹ Zu den FAIR-Prinzipien bei Forschungsdaten vgl. z.B. die Informationsplattform <https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/> (Zugriff am 15.2.2022)

kombinierte Suchoption nach Zeitpunkten und -räumen sowie Themen. Die Informationen aus existierenden Suchportalen könnten für ein zeithistorisches Suchportal unter diesem Gesichtspunkt optimiert werden. Das **Informationsportal historische Sozialdaten** kann eine Suchunterstützung entsprechend dieser Bedarfe anbieten. Unter Umständen ist es dazu notwendig, die Metadaten zu den Studien um fehlende Attribute zu ergänzen (z.B. können Informationen zu den Geburtsjahrgängen von Studienteilnehmenden interessant sein oder Angaben zu Nutzungsbedingungen, die in bestehenden Suchportalen häufig fehlen). Das bedeutet allerdings viel Arbeit für die zusätzliche Indexierung der Studien. Beim Aufbau des Portals sollte proaktiv eine enge Abstimmung mit den Metadatenstandards erfolgen, die sich derzeit im Konsortium NFDI4Memory (Task Area *Data Connectivity*) in wechselseitigen Diskussionen entwickeln.

Im Zuge der wissenschafts- und wissenschaftsgeschichtlichen Analyse von historischen Sozialdaten erzeugen Historiker:innen auch Material, das von Archiven und Forschungsdatenzentren in der Regel nicht in die Archivierung übernommen wird, weil es nicht deren Sammelrichtlinien entspricht. Dazu gehören beispielsweise Erzeuger:inneninterviews, Anonymisierungslisten, Aufbereitungsskripte oder Feldnotizen. Solche Kontextinformationen können selbst den Status von Forschungsdaten haben und müssen zur Sicherstellung der Transparenz und intersubjektiven Überprüfbarkeit ebenfalls aufbewahrt werden, idealerweise in direkter Nachbarschaft zum Originalbestand oder durch gemeinsame Verlinkung im Informationsportal historische Sozialdaten. Optimal wäre es, künftige Repositorien des Konsortiums NFDI4Memory oder das Archivportal D für diese Zwecke zu nutzen.¹⁰

(b) Nutzenden-Forum historische Sozialdaten

Ergänzt wird die Suchfunktion um ein „Nutzenden-Forum historische Sozialdaten“, das detaillierte Datenbeschreibungen (im Sinne einer Quellenkritik) und Praxisberichte aus dem Forschungsalltag bei der Nutzung von Sozialdaten (Datapapers) publiziert. Damit wird das Informationsangebot um Hinweise für die wissenschaftliche Verwendung historischer Sozialdaten erweitert und ein möglichst niedrigschwelliges Angebot für Diskussionsbeiträge bereitgestellt. Ziel ist einerseits, die Suche nach Daten um praktische Anleitung zur Datennutzung zu ergänzen und Forschenden so Orientierung zu geben. Andererseits schafft das Forum wissenschaftliche Anreize für die Publikation von Datapapers. Damit soll dauerhaft die Erschließung historischer Sozialdaten als eigenständige wissenschaftliche Leistung sichtbar und zitierbar werden.¹¹

Das **Informationsportal historische Sozialdaten** senkt die Hürden bei der Nutzung, indem die Suche

¹⁰ <https://www.archivportal-d.de/>

¹¹ Für die Einrichtung eines solchen Forums können bestehende technische Lösungen in das Portal eingebunden werden oder eigene Rubriken in einschlägigen Zeitschriften eingerichtet werden: als Beispiel siehe die neue Rubrik in der Zeitschrift *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte VSWG*, <https://journaldata.zbw.eu/journals/vswg> (Zugriff am 20.7.2022). Außerdem bieten die sozialwissenschaftlichen Infrastrukturen mittlerweile an einigen Stellen Lösungen für das Teilen von Auswertungsskripten an, wobei diese Funktion über die Repositorien der Fachzeitschriften erfolgt. Die Skripte sind in der Regel ausreichend, um die Ergebnisse der Datenanalysen in den Publikationen nachzuvollziehen.

nach passenden Daten erleichtert wird und Antworten auf forschungsmethodische Fragen beantwortet werden. Der Erfolg des Portals ist allerdings von einer langfristigen Sicherung des Betriebs im Anschluss an eine Aufbauphase abhängig. Es benötigt deswegen eine nachhaltige institutionelle Anbindung, eine dauerhafte Finanzierung und Personalmittel (vgl. Abschnitt „Finanzierung und Governance“). Technische Plattformen sollten nicht neu entwickelt, sondern aus dem Portfolio der NFDI oder der European Open Science Cloud (EOSC) gewählt werden.

(3) Forum für Ethik und Recht in der Zeitgeschichte: Ein Beratungsangebot

Die Nutzung historischer Sozialdaten wirft rechtliche und forschungsethische Fragen auf, zum Beispiel wenn für ältere Bestände die nach aktuellem Recht geforderte informierte Einwilligung zur Speicherung und Verarbeitung (personenbezogener) Daten in der Regel nicht oder nur unvollständig vorliegt. Diese Frage ist vor allem bei qualitativen Sozialdaten von Bedeutung, die sich teils nur unter großem Informationsverlust anonymisieren lassen.

Rechtliche und ethische Fragen stellen sich in vielfältigen Nutzungszusammenhängen. Insbesondere wenn Historiker:innen noch unveröffentlichte Daten für ihre Forschung aufbereiten und nutzen (also kein Archiv oder FDZ die Daten bereits geprüft, erschlossen und bereitgestellt hat), müssen urheber- und datenschutzrechtliche Fragen beantwortet werden (z.B. wer die Daten verwahrt und den Zugang zu den Daten kontrolliert, ob personenbezogene Informationen an Historiker:innen übergeben und ob die Daten im Zuge der Nutzung publiziert werden dürfen). In der Praxis lassen sich informierte Einwilligungen oft nicht mehr mit vertretbarem Aufwand nachträglich einholen. In solchen Fällen gibt es andere rechtliche Erwägungsgründe, die für die Beurteilung legitimer und illegitimer Nutzung herangezogen werden – etwa das Vorliegen eines öffentlichen Interesses an der Durchführung des Forschungsprojektes. Nur handelt es sich dabei gerade nicht um Standardfragen zum Datenschutz für Forschungsdaten, weil die juristische Diskussion auf neue Datenerhebungen konzentriert ist.

In vielen Fällen bauen Handlungsentscheidungen im Umgang mit historischen Daten auch auf Überlegungen und Abwägungen auf, beispielsweise, wenn es um den Schutz der Befragten oder die Interessen der Sozialwissenschaftler:innen geht, die die Daten erhoben haben. Die dafür nötigen rechtlichen und forschungsethischen Überlegungen wurden in der Zeitgeschichte noch nicht systematisch entwickelt, ein Standardverfahren fehlt. Ein vielversprechender Ansatz für ältere Daten ist das Rechtsinstitut des Lösungsderivats und der Schutzfristen im Archivrecht.¹² In der jüngeren Vergangenheit hat sich in den Sozialwissenschaften gezeigt, dass ein gemeinsames Austauschforum für ethische Fragen (initiiert und getragen vom RatSWD) die Diskussion von Beurteilungsstandards erleichtert.¹³

Deshalb schlagen wir die Gründung eines **Forums für Ethik und Recht in der Zeitgeschichte** vor. Das Forum soll rechtliche Kompetenz in Fragen der Nutzung historischer Sozialdaten bündeln, einen *Code*

¹² Becker, Irmgard Ch. (Hrsg.): Schutzfristen - Festlegung und Verkürzung. Beiträge zum Workshop der Archivschule Marburg am 3. Mai 2011, 2012, https://www.archivschule.de/uploads/Publikation/VOE54/VOe54_Internet.pdf (Zugriff am 3.11.2022).

¹³ Vgl. Forum Forschungsethik beim RatSWD, <https://www.konsortswd.de/ratswd/themen/forschungsethik/> (Zugriff am 15.2.2022).

of Conduct für die Fachcommunity entwickeln (auch in Abstimmung mit Jurist:innen und der Task Area *Data Culture* des Konsortiums NFDI4Memory¹⁴) sowie als Anlaufstelle für die ethische und juristische Beratung von Forschungsvorhaben dienen. Solche Verfahren sind in der Medizin, der Psychologie oder den Sozialwissenschaften bereits flächendeckend in Form von Ethikkommissionen etabliert. Die Maßstäbe und Standards des Forums sollten in enger Abstimmung mit Archiven, Forschungsdatenzentren und Spezialbibliotheken in Europa und weltweit entwickelt werden.

(4) Daten als Quellen für die historische Forschung sichern: Anbietungspflicht von Sozialdaten durch Forschende an öffentliche Archive

Die Bedarfsanalyse und die Diskussionen mit Wissenschaftler:innen in den Projektworkshops haben gezeigt, dass die existierende Forschungsdateninfrastruktur für die Sozialwissenschaften keinen ausreichenden Zugang zu historischen Sozialdaten ermöglicht. Das liegt daran, dass in den Sozialwissenschaften zwar eine lange Tradition gemeinsamer Datenerhebungsprogramme existiert, aber die Bereitschaft von Primärforschenden, ihre Forschungsdaten zu erhalten und zu teilen, in den vergangenen Jahrzehnten zu gering ausgeprägt war. Aus der zeitgenössischen Literatur ist bekannt, dass zahlreiche für die zeithistorische Forschung wichtige Datenbestände – insbesondere aus dem Bereich der qualitativen Sozialforschung – von Historiker:innen nur mit großem Aufwand gefunden und erschlossen werden können.¹⁵ Besonders problematisch ist, dass zu Quellen werdende Forschungsdaten in vielen Fällen im individuellen Besitz ihrer Erzeuger:innen verbleiben. Die spätere Nutzung dieser Daten hängt von der Bereitschaft und Fähigkeit der Erzeuger:innen ab, interessierten Historiker:innen Zugang zu gewähren. Gerade für ältere Forschungsprojekte fehlen Vorgaben für die langfristige Sicherung der mit sozialwissenschaftlichen Methoden erhobenen Daten in Archiven oder Datenzentren. Das schränkt die (zeit-)historische Forschung ein, weil Sozialdaten entweder nicht zugänglich sind oder nicht erhalten wurden, obwohl sie relevante Quellenbestände darstellen. Auch ist die digitale Bereitstellung massenhafter Datenquellen (z.B. ein Scan von Millionen von Volkszählungsbögen) mit Investitionen verbunden, die die Budgets der Forschung übersteigen.

Das Schriftgut aus Universitäten und öffentlich finanzierten Forschungsinstituten (insbesondere der Leibniz- und Helmholtz-Gemeinschaft sowie der Max-Planck-Gesellschaft) unterliegt entweder den Vorgaben des Bundesarchivgesetzes oder den Bestimmungen der jeweiligen Landesarchivgesetze. Die Hochschulen sind aufgrund Artikel 5 DSGVO gehalten, Regeln darüber zu erlassen, welche Daten wie lang für die primären Aufgaben gebraucht werden (Aufbewahrungsfristen). Nach Ablauf des Gebrauchs sind diese entweder zu löschen oder im Einklang mit Artikel 89 DSGVO zu Daten von bleibendem Wert zu erklären. Die Aufbewahrungsfrist von Forschungsdaten ist in diesem Kontext bislang unbestimmt. Nur in Ausnahmefällen regeln Verordnungen der Universitäten, was mit Forschungsdaten geschehen

¹⁴ Siehe die NFDI4Memory Task Area 5: *Data Culture*, Measure 3: Ethical and Legal Implications, <https://4memory.de/task-areas/task-area-5-data-culture/> (Zugriff am 30.6.2022).

¹⁵ Vgl. Böick, Marcus: *Die Treuhand. Idee – Praxis – Erfahrung 1990-1994*, 4. Aufl., Göttingen 2020; zu diesem Problem siehe auch: Brückweh, Kerstin/ von Hodenberg, Christina/ Raphael, Lutz/ Reh, Sabine/ Siegers, Pascal/ Villinger, Clemens/ Zöller, Kathrin: *Sozialdaten als Quellen der Zeitgeschichte. Zur Einführung*, in: *Geschichte und Gesellschaft* (2022), 48, 1, 5-27.

soll.¹⁶ Historische Sozialdaten werden noch nicht ausreichend als Unterlagen von bleibendem Wert wahrgenommen und spielen deshalb in der Archivierungspraxis der Universitäten und in anderen öffentlichen Archiven oft noch keine Rolle.

Entgegen dieser Fehlwahrnehmung gewinnen solche aus den Sozialwissenschaften überlieferten Forschungsdaten in der Forschungspraxis von (Zeit-)Historiker:innen zunehmend an Bedeutung. Das Bundesarchivgesetz nennt als Kriterium für die Archivwürdigkeit die "Bedeutung [...] für die Erforschung und das Verständnis von Geschichte und Gegenwart, auch im Hinblick auf künftige Entwicklungen"¹⁷, was auf historische Sozialdaten in Gänze zutrifft. Auch nach anderen Archivgesetzen und Archivsatzungen der Länder und Kommunen ist eine Archivwürdigkeit sehr oft gegeben. Demnach entspricht die verbreitete Gewohnheit, Daten aus öffentlich geförderter Forschung in der Verfügungsgewalt der Erzeuger:innen zu belassen, nicht der Rechtslage.

Die Leitlinien für die gute wissenschaftliche Praxis der Deutschen Forschungsgemeinschaft (DFG) schreiben bereits vor, dass Forschungsdaten zur Sicherstellung der Replizierbarkeit von Forschungsergebnissen mindestens zehn Jahre nach Abschluss eines Forschungsprojektes gesichert werden müssen.¹⁸ Jenseits dieser Frist fehlen in der Bundesrepublik verbindliche Regelungen für den Umgang mit Forschungsdaten. Obwohl Universitäten, Fachhochschulen und außeruniversitäre Forschungseinrichtungen, an denen ein Großteil der öffentlich geförderten Forschungsprojekte durchgeführt wird, den Archivgesetzen des Bundes oder der Länder unterliegen, wird die Pflicht der Träger, Sozialdaten ihren zuständigen Archiven anzubieten, nicht hinreichend wahrgenommen. Weil Sozialdaten als Quellen für den historischen Forschungsprozess unverzichtbar sind, muss an diese Anbietungspflicht für Forschungsdaten nachhaltig erinnert werden. Die Verantwortung für die Kuratierung der Daten obliegt den zuständigen Archiven der Organisationen, insbesondere den Universitäts- und Landesarchiven. Sie müssen nach Wichtigkeits- und Qualitätsgesichtspunkten auswählen und die Metadatenaufbereitung übernehmen, sofern diese fehlt.

Wir wollen einen Diskussionsprozess zwischen den Sozialwissenschaften (z.B. vertreten durch den RatSWD), den Trägern öffentlicher Archive und der (Zeit-)Geschichte anstoßen, um einen rechtskonformen Umgang mit historischen Sozialdaten sicherzustellen. Es sollen Leitlinien für den Umgang mit historischen Sozialdaten formuliert werden, die deren langfristigen Erhalt für die

¹⁶ Zum Beispiel sieht die Universität Bielefeld vor, dass Daten aus Forschungsprojekten nach einer Frist von zehn Jahren dem Universitätsarchiv angeboten werden müssen. Vgl. die Richtlinien über Aufbewahrung, Aussonderung, Archivierung und Vernichtung von Unterlagen für die Universität Bielefeld vom 4. Juni 2018, in: Verkündungsblatt Universität Bielefeld, Amtliche Bekanntmachungen, Nr. 12/47, S. 112-125, https://www.uni-bielefeld.de/uni/universitaetsarchiv/aktuelles/AufbewRichtl_VerkBlatt_Jg47-12_2018-06-04_09.pdf (Zugriff am 15.2.2022). Dies entspricht den Empfehlungen der Arbeitsgemeinschaft Hochschularchive NRW für Aufbewahrungsfristen an nordrhein-westfälischen Hochschulen vom Januar 2022, die für Forschungsdaten eine Einzelfallprüfung durch das zuständige Hochschularchiv vorsehen, <https://www.archive.nrw.de/arbeitsgemeinschaft-hochschularchive-nrw/information-aufbewahrungsfristen> (Zugriff am 10.02.2023).

¹⁷ Vgl. § 1 Abs. 11 Gesetz über die Nutzung und Sicherung von Archivgut des Bundes (Bundesarchivgesetz – BArchG), <https://www.bundesarchiv.de/DE/Content/Downloads/Rechtliches/bundesarchivgesetz.pdf?blob=publicationFile> (Zugriff am 15.2.2022).

¹⁸ Vgl. die DFG-Leitlinien zur Sicherung guter wissenschaftlicher Praxis, insb. Leitlinie 17 zur „Archivierung“, <https://www.bundesarchiv.de/DE/Content/Downloads/Rechtliches/bundesarchivgesetz.pdf?blob=publicationFile> (Zugriff am 15.2.2022).

Nachnutzung sicherstellen.¹⁹ In Zuge dieser Diskussion müssen sowohl die Forschungseinrichtungen als auch die öffentlichen Archive ihren Umgang mit solchen Daten neu justieren. Es wäre hier auch zu fragen, ob und unter welchen Bedingungen Erzeuger:innen ihre Sozialdaten ersatzweise auch an ein Repositorium oder FDZ übergeben können, soweit dies als vertrauenswürdige digitales Langzeitarchiv zertifiziert ist.²⁰

Der Zugang zu historischen Sozialdaten in den öffentlichen Archiven kann über die Bestimmungen der Archivgesetze organisiert werden, indem die geltenden Schutzfristen eingehalten werden oder Verfahren für Schutzfristverkürzungen für Forschungsprojekte Anwendung finden. Diese Verkürzungen stellen eine Alternative zu den Zugangskategorien nach Anonymisierungsgraden dar, wie sie in den sozialwissenschaftlichen FDZ praktiziert werden. Öffentliche Archive haben gegenüber anderen Datenrepositorien auch den Vorteil, dass sie in der DSGVO privilegiert sind. Sie haben den Status „sicherer Datenhäfen“, in denen auch personenbezogene Daten langfristig gesichert werden können.²¹ Die ähnlichen Privilegien sonstiger wissenschaftlicher Einrichtungen aus Artikel 89 Abs. 2 DSGVO unterscheiden sich insoweit, als das Recht Betroffener auf Mitteilung (Art. 19) und auf Datenübertragbarkeit (Art. 20) bei den Archiven zurücktritt, weil diese Rechte voraussichtlich die Verwirklichung der Archivzwecke ernsthaft beeinträchtigen würden.

Die rechtlichen Rahmenbedingungen für den Zugang zu historischen Sozialdaten werden sich in der laufenden Legislaturperiode unter Umständen deutlich ändern. Die Regierungskoalition plant ein Gesetz zur Verbesserung des Zugangs zu Forschungsdaten, das nationale Spielräume in der DSGVO nutzen möchte, um den Zugang und die Verknüpfung von Daten zu vereinfachen.²² Davon kann auch die zeitgeschichtliche Forschung profitieren, wenn die Bedarfe frühzeitig in den legislativen Prozess eingebracht werden, wie dies andere wissenschaftspolitische Akteure bereits tun.²³

¹⁹ Für die Archivierung von Geodaten ist so ein Prozess kürzlich erfolgreich abgeschlossen worden und könnte als Vorbild dienen: <https://www.gda.bayern.de/aktuelles/leitlinien-archivierung-von-geobasisdaten/> (Zugriff am 30.11.2022)

²⁰ Beispiele für Zertifikate sind das *CoreTrustSeal* (<https://www.coretrustseal.org/>, Zugriff am 20.7.2020) oder das nestor-Siegel auf Basis der DIN-31664.

²¹ Vgl. Villinger/ Zöller: Sind das Daten und muss das (deshalb) weg? (s. Fußnote 4).

²² Vgl. den Abschnitt zum Thema Forschungsdaten in dem im November 2021 verabschiedeten Koalitionsvertrag zwischen SPD, Bündnis 90/ Die Grünen und FDP (S. 21), <https://www.wiwo.de/downloads/27830022/8/koalitionsvertrag-2021-2025.pdf> (Zugriff am 15.2.2022).

²³ https://www.konsortswd.de/wp-content/uploads/220727_RatSWD-U15-Joint_Statement_FD-Gesetz-final.pdf oder <https://www.nfdi.de/wp-content/uploads/2022/10/NFDI-Positionspapier-zum-Datenzugangsrecht-der-Forschung.pdf>

Zusammenarbeit mit Dateninfrastrukturen für die Geschichtswissenschaften

Die hier vorgeschlagenen Maßnahmen zur Unterstützung der zeitgeschichtlichen Forschung vernetzen Zeithistoriker:innen mit den existierenden und geplanten Dateninfrastrukturen in den Sozial- und Geisteswissenschaften. Eine effiziente Realisierung der vorgeschlagenen Vorhaben setzt eine Einbettung in die neu zu schaffenden Dienste der NFDI4Memory voraus.

Die Vorschläge für eine zeithistorische Dateninfrastruktur enthalten kein eigenes Repository für historische Sozialdaten, die von Zeithistoriker:innen wiederentdeckt und erschlossen werden. Da solche Daten in der Regel nicht unter die Sammelrichtlinie der sozialwissenschaftlichen Forschungsdatenzentren fallen (insbesondere, wenn die Bestände verschiedene digitale Objekttypen umfassen), werden Repositorien aus dem Angebotsportfolio der NFDI-Konsortien eine zentrale Rolle für die langfristige Sicherung solcher Daten spielen.²⁴

Das anvisierte **Informationsportal historische Sozialdaten** widerspricht nicht dem Gedanken einer stärkeren Standardisierung des Forschungsdatenmanagements in den Geschichtswissenschaften. Vielmehr geht es darum, eine Schnittstelle zwischen den besonderen Bedarfen der zeitgeschichtlichen Forschung und den Angeboten der NFDI zu schaffen, d.h. die spezifischen Informationsbedarfe der Historiker:innen bei der Darstellung von Informationen über Daten zu berücksichtigen. Damit werden die Angebote der NFDI für Zeithistoriker:innen attraktiv. Gleichzeitig profitiert eine zeitgeschichtliche Dateninfrastruktur von der Sichtbarkeit, Reichweite und Stabilität der NFDI. Innovationen in der digitalen Quellenkritik können der gemeinsame Rahmen für eine IT-gestützte Geschichtswissenschaft sein, in die sich die hier vorgeschlagenen Maßnahmen nahtlos einfügen. Mit dem Fokus auf historische Sozialdaten wird auch eine Verbindung zu sozialwissenschaftlichen Infrastrukturen in KonsortSWD geschaffen, weil die Metadaten (d.h. die Daten über Sozialdaten) für den Datenkatalog zum überwiegenden Teil aus den Archiven und Forschungsdatenzentren der Sozialwissenschaften stammen werden.

Finanzierung und Governance²⁵

Während in der Bundesrepublik verschiedene Förderinstrumente für den projektförmigen Aufbau von Forschungsdateninfrastrukturen existieren (z.B. das DFG-Programm Informationsinfrastrukturen für Forschungsdaten oder die ebenfalls DFG-finanzierten Fachinformationsdienste), sind die Optionen für eine langfristige Finanzierung begrenzt. Einerseits können Universitätsbibliotheken und Hochschulrechenzentren langfristige Dienste zur Unterstützung der Forschung leisten. Deren Hauptaugenmerk liegt allerdings auf der Versorgung der Angehörigen der eigenen Hochschule, weshalb die Erbringung überregionaler Angebote für einzelne wissenschaftliche Communities in der

²⁴ Bisher existieren Archive für Datenkollektionen von Interviews aus der *Oral History*, siehe <https://www.oral-history.digital/> (Zugriff 30.6.2022), jedoch nicht für quantitative oder mixed-methods- Datenbestände.

²⁵ Details zum Finanzbedarf und potentiellen Trägerstrukturen der hier vorgeschlagenen Infrastrukturkomponenten finden sich in den Anlagen zum Abschlussbericht des DFG-Projektes.

Regel für die Hochschulen nicht in Frage kommt. Allerdings braucht der Aufbau und Betrieb von Forschungsdateninfrastrukturen eine langfristige Perspektive. Der Rat für Informationsinfrastrukturen (RfiI) hat in seinen Analysen deutlich gemacht, dass der projektförmig finanzierte Aufbau von Dateninfrastrukturen in den meisten Fällen nicht in ein tragfähiges Betriebsmodell mündet. Projektergebnisse gehen verloren, obwohl sie für die Forschung nützlich wären. Das derzeit einzig verfügbare Finanzierungsinstrument für langfristig konzipierte Dateninfrastrukturen ist deren Integration in die Institute der Leibniz-Gemeinschaft. Gerade die außeruniversitären Forschungsinstitute können über ihr Engagement im Bereich der Dateninfrastrukturen einen überregional bedeutsamen Beitrag zur Förderung datengestützter Forschung leisten und damit ihre Position in den Fächern konsolidieren.

Finanziert werden könnte dies über eine strategische Erweiterung eines oder mehrerer Leibniz-Institute über einen sogenannten strategischen Sondertatbestand. Dann könnte eine zentrale Struktur den Aufbau der technischen Komponenten und deren wissenschaftliche Einbettung sicherstellen.²⁶

Unabhängig von der Trägerstruktur müssen kontinuierlich die Bedarfe der Forschenden im Aufbau der Infrastruktur berücksichtigt werden. Daher sollte ein gemeinsamer Beirat von Nutzenden als integrierendes Steuerungsgremium eingerichtet werden. Das gilt auch dann, wenn das Modell einer zentralisierten Infrastruktur gewählt würde.

Nächste Schritte

In der Vergangenheit wurden Dateninfrastrukturen in der Regel aus einer projektförmigen Finanzierung aufgebaut und bei ausreichender Relevanz für die Communities verstetigt. Zu Beginn kann ein Netzwerk aus interessierten Partner:innen eine projektförmig finanzierte Startphase absichern (siehe das Modell von *eLabour*). In einem Kooperationsmodell könnten die hier vorgeschlagenen Komponenten von verschiedenen Partnern beantragt und betrieben werden. Die Konsultation mit Forschenden der Zeitgeschichte hat gezeigt, dass die Kompetenzbildung derzeit höchste Priorität haben sollte. Der Aufbau der drei Komponenten muss daher nicht unbedingt synchron erfolgen. Wünschenswert ist aber auch im verteilten Modell eine gemeinsame Governance-Struktur, die einerseits die Forschenden systematisch in die Entwicklung der Infrastruktur einbezieht und andererseits die Konsistenz der Initiativen sicherstellt.

²⁶ Vorbilder für solche Initiativen sind z.B. das DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation und die Bibliothek für Bildungsgeschichtliche Forschung (BBF) des DIPF, das über eine strategische Erweiterung den Verbund Forschungsdaten Bildung verstetigt hat. Das Zentrum für Marine Tropenökologie hat über einen kleinen strategischen Sondertatbestand den Aufbau eines eigenen Datenportals finanziert (das sog. digitale ZMT).