

Symposion

Forschungsdateninfrastruktur

Zur Einführung

Hubert Knoblauch

Im Zuge der verschiedenen Initiativen zur Speicherung, Archivierung und Nachnutzung sozialwissenschaftlicher Forschungsdaten ist es insbesondere im Zusammenhang mit den unterschiedlichen Datensorten der qualitativen Forschung zu durchaus kontroversen Debatten gekommen, an denen viele Disziplinen beteiligt sind. Die Entwicklungen haben auch in der Soziologie (zum Beispiel in dieser Zeitschrift) einen ebenso kontroversen Niederschlag gefunden. Mit der Realisierung des Programms zur Schaffung einer Nationalen Forschungsdateninfrastruktur (NFDI) und vor allem dem Erfolg des vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) koordinierten KonsortSWD hat sich die Situation zwischenzeitlich noch einmal verändert. Denn jetzt geht es nicht mehr nur um die Frage, *ob*, sondern vielmehr *wie* (auch qualitative) sozialwissenschaftliche Daten gespeichert, archiviert und nachgenutzt werden sollen und welche das sind. Dies ist die Frage, die hier aufgenommen und anhand von Beiträgen aus Forschungsdatenzentren (FDZ) diskutiert werden soll, die auf qualitative Daten spezialisiert sind. Neben der Vorstellung und Darstellung dieser Einrichtungen sollen auch die mit der Archivierung einhergehenden Probleme und deren Lösungen angesprochen werden. Und schließlich sollen damit auch die Forschenden in der Soziologie auf diese Entwicklungen aufmerksam gemacht werden, um damit die Möglichkeit zu schaffen, sich am Aufbau der FDZ und der Gestaltung von Speicherung, Archivierung und Nachnutzung aktiv zu beteiligen.

Faire Prinzipien – für die Nachnutzung von Forschungsdaten UND die Primärforschung

Nicht zuletzt durch technologische Entwicklungen in Gang gesetzt, besteht in Deutschland seit ungefähr 15 Jahren ein forschungspolitisches Interesse am Aufbau wissenschaftlicher Informationsinfrastrukturen für die Geistes- und Sozialwissenschaften (DFG 2006; DFG 2019; RatSWD 2010; Wissenschaftsrat 2011; Wissenschaftsrat 2012). Zu den Einrichtungen der Informationsinfrastrukturen, deren gemeinsames Ziel die Dissemination wissenschaftlichen Wissens ist, zählen – neben Bibliotheken, Archiven und Fachinformationszentren – die Forschungsdatenzentren. Sie sind Einrichtungen des Datenteilens, und zunehmend finden sich auch solche FDZ, die auf Datensorten der qualitativen empirischen Sozialforschung spezialisiert sind (Huschka et al. 2013; Knoblauch, Wilke 2018; Wilke, Pröbrock, Pach 2019). Vor diesem Hintergrund möchten wir diesen Beitrag einerseits dazu nutzen, um die Möglichkeiten eines fortgeschrittenen Forschungsdatenmanagements (FDM) digitaler Daten zu skizzieren. Als Primärforscher möchten wir andererseits auf Vorbehalte seitens der Forschenden gegenüber dem Datenteilen eingehen, um neben Möglichkeiten auch Herausforderungen zu beleuchten, die aus der forschungspolitisch beschleunigten Entwicklung des Datenteilens hervorgehen.

Digitalität und Forschungsdatenmanagement

Etwas später als in der quantitativen Forschung setzt sich die Digitalisierung seit den 2000er Jahren auch in der qualitativen Forschung durch. Während zunächst vor allem die digitale Erhebung und Auswertung von Materialien sowie die digitale Langzeitarchivierung (LZA) in die qualitative Forschungspraxis Einzug hielten, gelangte seit den 2010er Jahren auch das Nachnutzungspotenzial genuin digitaler (*born digitals*) und digitalisierter Daten (*Digitalisate*) in den Fokus, das durch Informations- und Kommunikationstechnologien (IuK) ermöglicht wird. Digitale Daten können Dank der Intraaktivität der Digitaltechnik auf vielfältige Weise organisiert und weiterverarbeitet werden. Qualitative Methodensoftware erleichtert ihre Auswertung, außerdem kann sie mit geringem Aufwand für die Nachnutzung aufbereitet werden (Wilke, Knoblauch, Pröbrock 2021). Die *genuine* Digitalität vieler in der gegenwärtigen qualitativen Forschung erhobener bzw. gesammelter Materialien ist ein Vorteil für

die systematische Nachnutzbarkeit von Forschungsdatenkorpora. Die Retro-digitalisierung analog erhobener Daten hingegen ist mit einem sehr hohen Ressourcenaufwand verbunden. Zu Buche schlagen hierbei nicht allein die Arbeiten, die in die Transferierung in ein digitales Speicherformat investiert werden müssen, sondern auch die Rekonstruktion einer Korpusstruktur, das heißt die Verknüpfung der unterschiedlichen digitalen Dateien miteinander, mit Teil- und Gesamtstudien sowie mit Kontextmaterialien wie Transkripten, Annotationen, Berichten etc. und schließlich die nachträgliche Erstellung von Metadaten. Wo diese aufwändige Arbeit jedoch möglich ist, kann sie zu neuen Perspektiven der Sekundäranalyse führen, wie sie derzeit unter anderem auch mit Blick auf audiovisuelle Daten erkundet werden.¹

Im *Datenlebenszyklus* (Dierkes 2021: 305 ff.) von *genuin* digitalen Daten ist die Nachnutzbarmachung erheblich erleichtert. Einige Arbeitsschritte werden in digitalen Prozessen sogar automatisch vollzogen. Digitale Aufzeichnungsgeräte produzieren parallel zur technologischen Registrierung zugleich technische Metadaten. Als Computerdateien lassen sich Forschungsdaten daher leichter organisieren und zum Beispiel über das Datum ihrer Erstellung, die Dateinomenklatur oder das spezifische Speicherformat auch kontextualisieren. Außerdem lassen sich digitale Forschungsdaten unabhängig von ihrer ursprünglichen Materialität einheitlich repräsentieren und bearbeiten. Auf diese Weise werden die Forschungsressourcen mittelfristig geschont, wodurch Lebenszyklen, die die Nachnutzbarkeit umfassen und Innovationen ermöglichen (Demmer, Engel, Fuchs 2020), überhaupt erst breiter in Betracht kommen.

Forschungsgenerierte FDZ für qualitative Forschungsdaten

Die Entwicklung von FDZ der qualitativen Sozialforschung wird bereits seit rund einem Jahrzehnt systematisch ausgebaut, sodass einige Infrastrukturen besonders im Bereich einzelner Datensorten oder Disziplinen (zum Beispiel Sprach- und Erziehungswissenschaften) heute in der Lage sind, ein umfangreiches Dienstleistungsportfolio anzubieten (Hollstein, Strübing 2018; Wilke,

¹ Wie etwa in der Kooperation des FDZ-aviDa mit dem SFB 1265 »Re-Figuration von Räumen«, wo die Nachnutzung von analogen Videoaufzeichnungen aus Kontrollräumen der 1990er Jahre zum Vergleich mit der Analyse gegenwärtiger Kontrollräume genutzt wird (vgl. dazu Knoblauch, Janz, Schröder 2021). Gemeinsam mit dem SFB 1265 ist auch eine Sekundärnutzung anderer raumbezogener Daten geplant.

Pröbrock, Pach 2019; Steinhardt et al. 2020). Dieses umfasst Beratung und Hilfestellungen hinsichtlich sämtlicher Aspekte des FDM, das heißt Services sowohl für Primärforscher*innen, die ihre Forschungsdaten teilen möchten, als auch für Forscher*innen, die fremde Forschungsdaten in Forschung oder Lehre nachnutzen möchten. Zahlreiche FDZ für qualitative Forschungsdaten haben sich aus der Primärforschung selbst heraus entwickelt. So ist zum Beispiel *Qualiservice* aus dem *Archiv für Lebenslaufforschung* des SFB 186 »Statuspassagen und Risikolagen im Lebenslauf« hervorgegangen, das ursprünglich eingerichtet wurde, »um die umfangreichen qualitativen Interviewdaten« des SFB mit anderen Forscher*innen zu teilen (Witzel 2000).² *eLabour* wurde seit 2015 im Rahmen eines Verbunds von vier soziologischen Forschungseinrichtungen (SOFI Göttingen, ISF München, IfS Jena, sfs Dortmund) mit drei IT-Partnereinrichtungen (L3S Hannover, SUB und GWDG Göttingen) für die Nachnutzung umfangreicher arbeits- und industriesoziologischer Interviewdatenkorpora entwickelt.³

Auch das *FDZ-aviDa* (Leitung: Hubert Knoblauch) ist aus der Primärforschung hervorgegangen. Das DFG-Forschungsprojekt wurde in seiner Pilotphase (2018–2021) in Kooperation des Fachgebiets Allgemeine Soziologie mit der Universitätsbibliothek sowie der Zentraleinrichtung Campusmanagement (ZECM) der Technischen Universität Berlin (TUB) und dem Lehrstuhl für Kultur- und Religionssoziologie der Universität Bayreuth aufgebaut. Dabei wurde zunächst, ausgehend von dem auf *DSpace* basierenden TUB-Repositorium *DepositOnce*, der Grundstein für eine nachhaltige webbasierte Plattform mit hoher Usability gelegt, die dem (wissenschaftsinternen) Austausch von *videographischen* Forschungsdaten (Knoblauch, Tuma, Wilke 2019) dient. In der Pilotphase wurde ein in drei Teilstudien gegliederter Pilotkorpus (ca. 900 Videodateien) ausgewählt, aufbereitet und publiziert, der Metadatenschemata von *DepositOnce* (basierend auf Dublin Core Extended) um *aviDa*-spezifische Felder erweitert, ein gestaffeltes Zugriffsrechte-management entwickelt und implementiert sowie Tools für die automatische Transkodierung (Matroska-Containerformat) und Video-Streaming in *DSpace* entwickelt. Seit Januar 2021 werden weitere audiovisuelle Daten aufgenommen und, abhängig vom jeweiligen Schutzniveau, teilweise Open Access publiziert. Die Metadaten der auf *aviDa* veröffentlichten Forschungskorpora sind stets frei recherchierbar. In der gegenwärtig zweiten Förderphase (2021–2024) wird *aviDa* über die Bedarfe der *Videographie* (Tuma,

² Siehe den Beitrag von Jan-Ocko Heuer, Betina Hollstein und Kati Mozygemba auf S. 459 ff.

³ Dazu Heidemarie Hanekop auf S. 447 ff.

Schnettler, Knoblauch 2013) hinaus auf die Anforderungen der gesamten Bandbreite von in natürlichen Situationen erhobenen audiovisuellen Daten der Geistes- und Sozialwissenschaften ausgeweitet werden.

Zahlreiche weitere forschungsgenerierte FDZ ließen sich an dieser Stelle aufzählen. Ohne Anspruch auf Vollständigkeit sollen wenigstens die folgenden genannt werden: das *FDZ-Bildung* am Deutschen Institut für Internationale Pädagogische Forschung (DIPF), das Forschungsdatenzentrum *Archiv für Gesprochenes Deutsch* am Leibniz-Institut für Deutsche Sprache (FDZ-AGD), das *Forschungsdatenzentrum Betriebs- und Organisationsdaten* (FDZ-BO) am Deutschen Institut für Wirtschaftsforschung (DIW) sowie das Forschungsdatenzentrum des Leibniz-Instituts für Wirtschaftsforschung Halle (FDZ-IWH).⁴ Für qualitative Forschung, in der Daten erhoben werden, die eventuell geteilt werden sollen und für deren Ansatz bereits FDZ etabliert sind, lohnt es sich, mit den entsprechenden Einrichtungen in Verbindung zu treten, um von deren breitem Serviceangebot im Bereich des FDM (LZA und Datenteilen) zu profitieren.⁵ Die Entwicklung von neuen Forschungsdateninfrastrukturen aus den jeweiligen Feldern der Primärforschung heraus empfiehlt sich hingegen insbesondere für Forschungsfelder, Methoden und/oder Datensorten, für die noch keine befriedigenden Lösungen für die Herausforderungen durch die Nachnutz(barmach)ung entwickelt wurden. Explorative Pilotprojekte und/oder Kooperationen mit bestehenden FDZ, die von der jeweiligen Forschungscommunity getragen werden, können einen wesentlichen Beitrag dazu leisten, Nachnutzungspotenziale in diesen Domänen auszuloten und gegebenenfalls geeignete Infrastrukturen für die Nachnutzung zu entwickeln und bereitzustellen, die den eigenen Herangehensweisen entsprechen.

Vorbehalte gegen die Nachnutz(barmach)ung qualitativer Forschungsdaten

Das Öffnen und Teilen von Daten ist allerdings mehr als eine technische Frage. Vor allem methodologische und forschungsstrategische Gesichtspunkte spielen eine Rolle bei der Beurteilung, die den FDZ als Institutionen des Datenteilens in den verschiedenen Communities der qualitativen Forschung

4 Für eine vollständige Liste aller vom RatSWD akkreditierten FDZ siehe www.konsortswd.de/datenzentren/alle-datenzentren/.

5 Das FDZ-aviDa steht für die Beratung zum Management audiovisueller Daten insbesondere bei geplanten Forschungsprojekten bereit. Interessierte wenden sich bitte an die beiden Autoren. Siehe auch www.konsortswd.de/datenzentren/alle-datenzentren/avida/.

zukommt. Prägend erweist sich vor allem die Rolle ethnographischer Grundsätze, die für die qualitativen Methoden und ihr Datenverständnis von konstitutiver Bedeutung sind. Die aus der quantitativen Forschung stammende Idee des Datenteilens, die traditionell eng mit der (mitunter positivistisch gefassten) Vorstellung von Replizierbarkeit verknüpft ist, stößt gerade im Kontext jener ethnographischen Forschungsansätze häufig auf Skepsis oder gar Widerspruch, deren »Datenerhebungsprozesse« nicht auch registrierende Technik wie audiovisuelle Aufzeichnungsgeräte involvieren (Knoblauch 2005), sondern weitgehend an die subjektiven Erfahrungen, Handlungen und Praktiken der Ethnograph*innen im Feld gebunden sind. Insbesondere Feldnotizen gelten häufig als sehr persönliche Aufzeichnungen, die von idiosynkratischen Codes, Semantiken und Verweisstrukturen gekennzeichnet sind, die der Nachnutzung unzugänglich sind. Feldzugänge und Feldaufenthalte sind als Vertrauensbekundungen gerahmt, wodurch das Teilen von nicht aufbereiteten Aufzeichnungen von den Primärforscher*innen nicht selten als Vertrauensbruch gegenüber ihren Informant*innen verstanden würde. Auch werden in der Ethnographie Daten oft in solchen Situationen erhoben, die aus rechtlichen und/oder ethischen Gründen Dritten nicht zugänglich gemacht werden können (von Unger, Narimani, M'Bayo 2014). Entsprechend positioniert sich eine Reihe von Publikationen und Beiträgen auf Workshops kritisch zum Öffnen und Teilen von Daten der qualitativen Sozialforschung (Hirschauer 2014; Strübing et al. 2018; Eisewicht, Grenz 2018; von Unger 2018; Laudel, Bielick 2019; Wazlawik, Christmann 2020).

Neben diesen grundsätzlichen Vorbehalten werden zahlreiche weitere Argumente ins Feld geführt, wenn die Initiative des Datenteilens in der qualitativen Forschung thematisiert wird. Hierbei geht es insbesondere um Fragen des Datenschutzes bzw. der Anonymisierung. Dort wo systematisch schriftlich fixierte Einverständniserklärungen der Forschungsteilnehmer*innen eingeholt werden, ist die Nachnutzung außerhalb des unmittelbaren Projektkontexts häufig noch nicht explizit adressiert. Da die Datenschutzgrundverordnung zum Zweck der Nachnutzung explizite Vereinbarungen vorschreibt, können Daten, die ohne entsprechenden Passus in den korrespondierenden Erklärungen erhoben wurden, aus formalen Gründen in der Regel nicht geteilt werden. Ausnahmen stellen vor allem solche Daten dar, die so anonymisiert werden können, dass alle personenbezogenen Informationen anschließend vollständig getilgt sind. Allerdings schließt die Frage nach der Nachnutzbarkeit hier unmittelbar an die technische Machbarkeit an: Zwar *können* die meisten Daten vollständig anonymisiert werden, doch

sind sie dann nurmehr begrenzt nachnutzbar. Es gilt daher, bei der Öffnung qualitativer Daten zwischen dem zu unterscheiden, was machbar erscheint, und dem, was sinnvoll ist. Zugleich wird deutlich, dass es für unterschiedliche Datensorten differenzierte Lösungen geben muss.

Ein weiterer ernstzunehmender Vorbehalt bezieht sich auf befürchtete Veränderungen der Forschung und der Forschungsförderung. Auf einschlägigen Workshops lassen sich häufig Stimmen vernehmen, die darauf hinweisen, dass durch die Institutionalisierung des Nachnutzbarkeitsziels in den Forschungs- und Förderinstitutionen die Freiheit der Wissenschaft eingeschränkt werden könnte. Eine mögliche Nebenfolge der gegenwärtigen Initiative wird darin gesehen, dass sich zukünftige Forschungsvorhaben auf Felder beschränken könnten, die die Vereinbarung entsprechender Einverständnisse erleichtern würden, oder auf Datensorten, deren Anonymisierung vergleichsweise unproblematisch ist. Während diese Befürchtung eines vorauseilenden Gehorsams einen stark hypothetischen Charakter aufweist (und ihr zudem das erkenntnisorientierte Beharrungsvermögen der Wissenschaftler*innen entgegensteht), ist bei verschiedenen Förderlinien bereits eine den Prinzipien qualitativer Forschung unangemessene Festlegung auf oder Priorisierung von teilbaren Daten zu beobachten. Noch ist die pauschale Forderung von Open Data auf wenige Programme beschränkt, doch ist nicht zuletzt bei dem Auf- bzw. Ausbau einer Nationalen Forschungsdateninfrastruktur, gegebenenfalls im Rahmen einer Begleitforschung, darauf zu achten, eine Ausweitung dieser Entwicklung zu vermeiden und die Institutionalisierung des Datenteilens an den Forderungen der (häufig disziplinär organisierten) Fachgesellschaften und den mittlerweile vielfach hochgradig inter- und transdisziplinären Forschungsverbänden und -communities zu orientieren.

Eine offene oder indirekte Verpflichtung zur Öffnung von Forschungsdaten ist mit den Prinzipien der qualitativen Forschung unvereinbar. Anstelle einer simplen Open Data-Strategie müssen Daten vielmehr, wie es bereits der Praxis in den FDZ entspricht, in Schutzklassen differenziert werden, wobei nur ausgewählte Korpora überhaupt geteilt werden können. Unbedingt gilt, dass die Entscheidung zum Öffnen sich nicht an forschungspolitischen Vorgaben orientieren kann, sondern allein vor dem Hintergrund von forschungsethischen, methodologischen und epistemologischen Kriterien getroffen werden muss, die die Forschung selbst leiten. Ein eindeutiges Bekenntnis zu dieser Selbstverständlichkeit der freien wissenschaftlichen Forschung seitens der öffentlichen Forschungsförderung wäre wünschens-

wert und müsste, gegebenenfalls in Abhängigkeit von zukünftigen Entwicklungen, mit Nachdruck eingefordert werden. Eine wichtige Rolle spielen hier sicher – neben den institutionell häufig weniger klar umschriebenen Forschungscommunities – die Fachgesellschaften und Fachinformationsdienste, die sich zum Teil schon mit eigenen Stellungnahmen zur Nachnutzung von Forschungsdaten zu Wort gemeldet haben (zum Beispiel dgv 2018; DGS 2019; Akademie für Soziologie 2019; DGfE, GEBF, GFD 2020).

Fazit

Von der (auch zukünftigen) Freiwilligkeit der Nachnutz(barmach)ung von Forschungsdaten hängt die Akzeptanz des Datenteilens schon gegenwärtig maßgeblich ab. Für die heterogene qualitative Forschungslandschaft ist aber auch die Förderung von Ansätzen, Projekten und Initiativen des Datenteilens aus den jeweiligen Communities selbst heraus entscheidend. Dabei zeigen sich sowohl die nutzungsorientierten Bedarfe als auch die besonderen Anforderungen, die unterschiedliche Datensorten stellen. Da eine Reihe von Datensorten bisher noch gar nicht erfasst wurde, können entsprechende Lösungen so aus der feldspezifischen bzw. methodologischen Expertise der Forschungscommunities im Dialog mit den FDZs entwickelt werden. Und auch die Debatten darüber, welche Materialien »Daten« sind und geteilt werden sollten bzw. sinnvoll nachgenutzt werden können, müssen in den Communities der Forschenden stattfinden und schließlich auf Projektebene aufgegriffen und vor dem Hintergrund eines jeweils sehr spezifischen Forschungsdesigns von Fall zu Fall neu geführt werden. Zukünftig muss es weiterhin Ziel bleiben, insbesondere vor dem Hintergrund der Entwicklung einer NFDI und der gegenwärtigen Vernetzung großer Konsortialverbände, lokale Stärken zu festigen, das heißt, die Communities, die im Umgang mit ihren Datensorten besondere Expertise aufweisen, dabei zu unterstützen, eigene Lösungen zu entwickeln, zu erproben und gegebenenfalls geeignete Infrastrukturen aufzubauen (RFII 2018: 4).

Vor diesem Hintergrund wird die weitere Ausgestaltung des Konzepts der »verteilten Archivierung« von großer Bedeutung sein, das der Rat für Sozial- und Wirtschaftsdaten (RatSWD) bereits seit einigen Jahren umsetzt und nun im Rahmen der NFDI bzw. des seit Oktober 2020 DFG-geförderten Konsortiums für Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften (KonsortSWD) forciert. Das Konzept basiert im Kontext einer

föderierten FDZ-Struktur darauf, lokal vorhandene Expertise zu nutzen, um den Herausforderungen durch die methodologische Heterogenität von Forschungsdaten in der Archivierung und Nachnutzung angemessen zu begegnen. Die besondere Anforderung bei der »verteilten Archivierung« besteht darin, die in verschiedenen Kompetenzzentren archivierten Forschungsdaten in einer gemeinsamen Infrastruktur so miteinander zu vernetzen, dass sie so FAIR⁶ (Wilkinson et al. 2016) wie möglich sind. Allerdings sollten nicht nur Forschungsdaten FAIR für Nachnutzer*innen sein. Zugleich und insbesondere gilt es, auch darauf zu achten, die Kultur des Datenteilens für die Primärforschung nach *fairen* Prinzipien zu gestalten.

René Wilke, Hubert Knoblauch

Literatur

- Akademie für Soziologie 2019: Wissenschaftliche Daten sind kein Privateigentum einzelner Forschender, sondern ein kollektives Gut. <https://akademie-soziologie.de/wp-content/uploads/2019/11/AS-Stellungnahme-Nachnutzung-Forschungsdaten-01-2019.pdf>, letzter Aufruf am 19. Juli 2021.
- Demmer, Christine / Engel, Juliane / Fuchs, Thorsten 2020: Erkenntnis, Reflexion und Bildung – zur Frage neuer Formen der Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten. *Erziehungswissenschaft*, 31. Jg., Heft 61, 39–49.
- DFG 2006: DFG-Positionspapier: Wissenschaftliche Literaturversorgungs- und Informationssysteme. www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier.pdf, letzter Aufruf am 19. Juli 2021.
- DFG 2019: Information für die Wissenschaft Nr. 9. Ankündigung: Nationale Forschungsdateninfrastruktur. www.dfg.de/foerderung/info_wissenschaft/2019/info_wissenschaft_19_09, letzter Aufruf am 19. Juli 2021.
- DGfE, GEBF, GFD 2020: Empfehlungen zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten im Kontext erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung. www.forschungsdaten-bildung.de/files/Stellungnahme_zum_FDM_DGfE-GEBF-GFD.pdf, letzter Aufruf am 19. Juli 2021.
- DGS 2019: Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie. Stellungnahme des Vorstands und Konzils der DGS. *SOZIOLOGIE*, 48. Jg., Heft 2, 191–199.

⁶ FAIR steht für findable, accessible, interoperable und reusable, wobei die FAIR-Prinzipien auch schon hinsichtlich der Metadatenebene wertvoll sind.

- dgv 2018: Positionspapier zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten. www.d-g-v.org/sites/default/files/dgv-positionspapier_fdm.pdf, letzter Aufruf am 19. Juli 2021.
- Dierkes, Jens 2021: Planung, Beschreibung und Dokumentation von Forschungsdaten. In Markus Putnigs / Heike Neuroth / Janna Neumann (Hg.), *Praxishandbuch Forschungsdatenmanagement*. Berlin, Boston: De Gruyter, 303–326.
- Eisewicht, Paul / Grenz, Tilo 2018: Die (Un-)Möglichkeit allgemeiner Gütekriterien in der Qualitativen Forschung – Replik auf den Diskussionsanstoß zu »Gütekriterien Qualitativer Forschung« von Jörg Strübing, Stefan Hirschauer, Ruth Ayaß, Uwe Krähnke und Thomas Scheffer. *Zeitschrift für Soziologie*, 47. Jg., Heft 5, 364–373.
- Hirschauer, Stefan 2014: Sinn im Archiv? Zum Verhältnis von Nutzen, Kosten und Risiken der Datenarchivierung. *SOZIOLOGIE*, 43. Jg., Heft 3, 300–312.
- Huschka, Denis / Knoblauch, Hubert / Oellers, Claudia / Solga, Heike (Hg.) 2013: *Forschungsinfrastrukturen für die Qualitative Sozialforschung*. Berlin: SCIVERO.
- Knoblauch, Hubert 2005: Focused Ethnography. *FQS*, 6. Jg., Ausgabe 3, Artikel 44, doi.org/10.17169/fqs-6.3.20.
- Knoblauch, Hubert / Janz, Arne / Schröder, Joshua 2021: Kontrollzentralen und die Polykontextualisierung von Räumen. In Martina Löw / Volkan Sayman / Jona Schwerer / Hannah Wolf (Hg.), *Am Ende der Globalisierung? Über die Refiguration von Räumen*. Bielefeld: Transcript, 206–230.
- Knoblauch, Hubert / Tuma, René / Wilke, René 2019: Videovignette I: Videographie und videographische Daten. doi.org/10.14279/depositonce-9138.
- Knoblauch, Hubert / Wilke, René 2018: Forschungsdateninfrastrukturen für audiovisuelle Daten der Qualitativen Sozialforschung – Bedarfe und Anforderungen. *RatSWD Working Paper 267/2018*, 47–58.
- Laudel, Grit / Bielick, Jana 2019: Forschungspraktische Probleme bei der Archivierung von leitfadengestützten Interviews. *FQS*, 20. Jg., Ausgabe 2, Artikel 10, doi.org/10.17169/fqs-20.2.3077.
- RatSWD 2010: Recommendations for Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences. *RatSWD Working Paper 150/2010*.
- RFII 2018: Zusammenarbeit als Chance. Zweiter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für die Wissenschaft in Deutschland. www.rfii.de/download/rfi-diskussionspapier-maerz-2018/, letzter Aufruf am 19. Juli 2021.
- Steinhardt, Isabel et al. 2020: *Das Öffnen und Teilen von Daten qualitativer Forschung: eine Handreichung*. (Weizenbaum Series, 6). Berlin: Weizenbaum-Institut für die vernetzte Gesellschaft – Das Deutsche Internet-Institut. doi.org/10.34669/wi.ws/6.
- Strübing, Jörg / Hirschauer, Stefan / Ayaß, Ruth / Krähnke, Uwe / Scheffler, Thomas 2018: Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. *Zeitschrift für Soziologie*, 47. Jg., Heft 2, 83–100.
- Tuma, René / Schnettler, Bernt / Knoblauch, Hubert 2013: *Videographie. Einführung in die interpretative Videoanalyse sozialer Situationen*. Wiesbaden: Springer VS.

- von Unger, Hella / Narimani, Petra / M'Bayo, Rosaline (Hg.) 2014: Forschungsethik in der qualitativen Forschung. Wiesbaden: Springer VS.
- von Unger, Hella 2018: Archivierung und Nachnutzung qualitativer Daten aus forschungsethischer Perspektive. RatSWD Working Paper 267/2018, 91–100.
- Wazlawik, Martin / Christmann, Bernd (Hg.) 2020: Forschungsdatenmanagement und Sekundärnutzung qualitativer Forschungsdaten. Perspektiven für die Forschung zu sexualisierter Gewalt. Wiesbaden: Springer VS.
- Wilke, René / Knoblauch, Hubert / Pröbrock, Willi 2021: Videovignette II: Forschungsdatenmanagement. Fit für das Teilen digitaler Forschungsdaten. [dx.doi.org/10.14279/depositonce-12165](https://doi.org/10.14279/depositonce-12165).
- Wilke, René / Pröbrock, Willi / Pach, Helen 2019: Infrastrukturen für Forschungsdaten der qualitativen Sozialforschung. SOZIOLOGIE, 48. Jg., Heft 4, 467–486.
- Wilkinson, Mark D. et al. 2016: The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018.
- Wissenschaftsrat 2011: Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. www.wissenschaftsrat.de/download/archiv/10465-11.html, letzter Aufruf am 19. Juli 2021.
- Wissenschaftsrat 2012: Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. www.wissenschaftsrat.de/download/archiv/2359-12.html, letzter Aufruf am 19. Juli 2021.
- Witzel, Andreas 2000: Archive for Life Course Research, University of Bremen, Germany. *FQS*, 1. Jg., Ausgabe 3. doi.org/10.17169/fqs-1.3.1045.

Unterstützung bei der Analyse digitaler Gesellschaften

Die Digitalisierung gesellschaftlicher Prozesse bringt eine Fülle an neuen Verhaltensdaten mit sich. Dieser Artikel will einen knappen Überblick über neue Angebote zur Unterstützung der Forschung mit diesen Daten geben, die derzeit bei GESIS aufgebaut werden. Der Beitrag soll damit gleichzeitig eine Einladung an diejenigen Kolleginnen und Kollegen sein, die Forschungsfragen mit solchen Daten bearbeiten.

Digitale Verhaltensdaten sind alle Spuren sozialen Verhaltens, die durch die Nutzung digitaler Technologien entstehen oder nutzbar gemacht werden können. Digitale Verhaltensdaten sind damit eine Teilmenge dessen, was oft als *Big Data* bezeichnet wird. Insbesondere sind dies Daten aus Online-Umgebungen wie Facebook, Twitter oder Wikipedia. Außerdem zählen dazu Sensordaten aus digitalen Endgeräten wie Smartphones oder Wearables (zum Beispiel Fitnesstracker). Hierzu gehören beispielsweise auch Mobilitätsdaten, wie sie durch Ridesharing-Dienste erhoben werden.

Diese Daten haben sowohl das Potential, Informationen zum neuen (Kommunikations-)Verhalten zu liefern, als auch neue Perspektiven auf klassische Fragestellungen der Soziologie zu eröffnen (Edelmann et al. 2020; Pötzschke, Braun 2017). Als Begleiterscheinungen von Digitalisierung sind so einerseits oft Filterblasen, politische Polarisierung, digitale Ungleichheit, *Fake News*, psychologische Auswirkungen sozialer Medien oder *Hate Speech* von soziologischem Interesse. Gleichzeitig findet, nicht erst seit der Corona-Pandemie, gesellschaftliche sowie politische Teilhabe in immer stärkerem Maße online und durch die Nutzung digitaler Endgeräte statt. Aber auch in klassischen sozialen Kontexten erzeugt die Nutzung digitaler Endgeräte wie beispielsweise Smartphones oder Tablets ein breites Spektrum an neuen Daten (zum Beispiel Nutzungsdauer des Smartphones, Nutzung von Apps, Kontaktdaten, Browserverläufe, Positionsdaten), aber auch Multimediadaten (Text, Audio, Video, Bild), Angaben zur triaxialen Beschleunigung oder physiologische Messwerte. Als Beobachtungsdaten scheinen diese Datenquellen eine neue Objektivität zu versprechen.

Diese Potentiale spiegeln sich in einem stark gestiegenen Interesse an der Nutzung digitaler Verhaltensdaten in den Sozialwissenschaften wider (Edelmann et al. 2020), beispielsweise in der steigenden Anzahl an Publikationen in hochrangigen Journalen, in der Schaffung neuer Professuren wie zum Beispiel in München oder Aachen, neuer Einrichtungen für Digitalisierungsforschung (zum Beispiel das Weizenbaum-Institut oder das *Center for Advanced Internet Studies* CAIS), neuer Studiengänge (zum Beispiel Aachen, Bamberg, Mannheim, München) und neuer Konferenzen (zum Beispiel die *International Conference on Computational Social Science* IC2S2, *BigSurv* oder *digital societies*).

Die besondere Herausforderung digitaler Verhaltensdaten ist, dass nicht nur – wie in verschiedenen Konsortien der Nationalen Forschungsdateninfrastruktur – Fragen des Forschungsdatenmanagements (bestehender Daten) beantwortet werden müssen. Vielmehr gilt es auch, Datenerhebung und Analyse – und die dafür nötigen Tools und Methoden für diese Datentypen – neu zu denken. Dies gilt, obwohl (und gerade weil) viele digitale Verhaltensdaten ohne unser dezidiertes Zutun als Forscher:innen entstehen. Um sie gewinnbringend einsetzen zu können, sind etliche Schritte nötig, bei denen spezielles Wissen Voraussetzung und Unterstützung ist.

Herausforderungen

Den genannten, vielfältigen Vorteilen digitaler Verhaltensdaten für soziologische Fragestellungen stehen große Herausforderungen gegenüber: Für diese Daten bedarf es neuer Methoden der Erhebung, Aufbereitung und Analyse sowie der Archivierung. Die allermeisten Sozialwissenschaftler:innen haben entsprechende Kenntnisse in ihrem Studium noch nicht erwerben können und müssen sich diese selbst aneignen. Einzelne Forscher:innen oder Gruppen müssen gegenwärtig außerdem beispielsweise technische Tools zur Erhebung und Methoden zur Aufbereitung und Analyse digitaler Verhaltensdaten aufwändig selbst entwickeln und können erhobene Datensätze nicht effektiv bereitstellen und teilen.

Für die Bewertung und Verbesserung der Datenqualität digitaler Verhaltensdaten fehlen weitestgehend etablierte Konzepte. Zudem stellt die Arbeit mit diesen Daten oft noch ethisches und rechtliches Neuland dar. Dezentral entwickelte Tools, Methoden und Daten sind oft nicht langfristig nutzbar und teils schwer zugänglich. Dies beeinflusst die Quantität und Qualität von Forschung mit digitalen Verhaltensdaten. Darüber hinaus gibt es eine Reihe spezifischer Herausforderungen bei der Analyse von Daten, die in digitalisierten Gesellschaften generiert werden (Lazer et al. 2020; Wagner et al. 2021).

Für die systematische Erschließung und Qualitätsbewertung digitaler Verhaltensdaten ebenso wie vor allem für die Vermittlung der nötigen Methodenkenntnisse ist eine Infrastruktur erforderlich, eine Anlaufstelle, die best-practices identifiziert und zentrale Angebote bündelt. GESIS hat sich daher dazu entschlossen, neben seinem bestehenden Angebot in der Unterstützung der Umfrageforschung, sein Portfolio für die Forschung mit digitalen Verhaltensdaten auszubauen.

In »der wegweisenden Verbindung von Umfragedaten mit digitalen Verhaltensdaten, die wechselseitige Stärken und Schwächen der jeweiligen Datentypen nutzbringend ausgleicht,« (Wissenschaftsrat 2021: 13)¹ sah der Wissenschaftsrat bei der Begutachtung des Antrages zum Aufbau einer Infrastruktur für digitale Verhaltensdaten ein »europaweites Alleinstellungsmerkmal« (ebd.). GESIS wird daher ab dem Jahr 2022 den Aus- und Aufbau von Angeboten für digitale Verhaltensdaten forcieren.

¹ Exemplarisch dafür Quinlan et al. (2018).

GESIS Angebote

Forscherinnen und Forscher brauchen Unterstützung, um möglichst großen wissenschaftlichen Mehrwert aus diesen neuen Daten ziehen zu können. Auch für digitale Verhaltensdaten ist diese Unterstützung in allen Phasen des Forschungsprozesses nötig. Entsprechend wird GESIS bei der Erhebung, der Aufbereitung, der Analyse, der Archivierung und Bereitstellung unterstützen. Außerdem bildet der Bereich der Kompetenzvermittlung einen Schwerpunkt der Arbeit.

Eine wesentliche Herausforderung in der Nutzung digitaler Verhaltensdaten ist das Thema Datenqualität, das sowohl den Bereich der Messung als auch die Möglichkeit von Verallgemeinerungen der Ergebnisse betrifft.² Für die *Erhebung digitaler Verhaltensdaten* erweitert GESIS daher seine etablierte Dateninfrastruktur sozialwissenschaftlicher Erhebungen (zum Beispiel ALLBUS oder die Wahlstudie GLES) um ein *Access Panel* für digitale Verhaltensdaten. Ein Teil der Stichprobe soll auf Grundlage vorhandener Umfragen rekrutiert werden, so dass bereits Vorwissen über die Befragten existiert, was Methodenstudien zur Datenqualität erlaubt. Neben einer Umfragekomponente wird das geplante *Access Panel* von Beginn an auch die Erhebung von Web-Tracking-Daten ermöglichen, das heißt, es wird aufgezeichnet, welche Webseiten die Nutzenden besuchen, und es werden die Inhalte dieser Webseiten gespeichert. Nicht zuletzt können auf diese Weise auch neue Trends im Bereich Social Media verfolgt werden, wenn zum Beispiel eine Plattform an Attraktivität verliert.

Digitale Verhaltensdaten zeichnen sich im Vergleich zu Umfrage- aber auch zu Textdaten vor allem durch ihre Masse, Dynamik, Heterogenität und komplexere Datenstruktur aus. Um solche Daten einer effizienten Nutzung zuzuführen, ist es notwendig, sie *aufzubereiten*. Dies bedeutet, Daten in geeignete Formate zu exportieren und Inhalte automatisiert aufzubereiten, bei Bedarf mit Informationen anzureichern und mit weiteren Datenquellen verknüpfbar zu machen. GESIS plant daher für diese Schritte, Methoden und Skripte bereitzustellen. Dies erleichtert auch die Verknüpfung zum Beispiel mit Umfragedaten.

² Die Gründe sind oft verzerrte Demographien der Nutzenden, intransparente Schnittstellen für den Zugriff auf die Daten (APIs), die nicht zwingend zufällige Stichproben liefern, sowie durch den Einsatz von Algorithmen verursachte Verzerrungen (Wagner et al. 2021; Sen et al. 2019; Pfeffer et al. 2018).

Die *Analyse* digitaler Verhaltensdaten erfordert Methoden, die in der Soziologie derzeit noch nicht breit genutzt werden. Hinzu kommt oft der Bedarf an technischer Expertise zum verteilten Rechnen und die dafür nötige Soft- und Hardwareinfrastruktur. Dies alles vor dem Hintergrund, dass Forschungsprozesse transparent und Ergebnisse replizierbar sein sollen. GESIS plant daher eine Analyseinfrastruktur mit Fernzugriff anzubieten, die es ermöglicht, auch große und komplexe Datensätze digitaler Verhaltensdaten effizient und reproduzierbar auszuwerten. Ebenso wird diese Plattform aber auch die Analyse klassischer sozialwissenschaftlicher Makro- und Mikrodatensätze ermöglichen und damit einen Beitrag zur Replizierbarkeit auch bestehender Forschungsarbeiten leisten. Über diese Analyseinfrastruktur wird GESIS einerseits selbst Methoden zu digitalen Verhaltensdaten zielgruppengerecht aufbereiten und in Form sogenannter »Notebooks«³ zur Verfügung stellen und andererseits auch die Informatik-Community durch die Veröffentlichungen von Aufgaben und Testdatensätzen einbeziehen. Soziolog:innen können so einfach auf neue Methoden zugreifen, sie Schritt für Schritt im Browser nachvollziehen, ihre Performanz anhand von einheitlichen Testdatensätzen vergleichen und sie auf eigene Datensätze anwenden. Ebenso können die Methoden für eigene Fragestellungen angepasst werden. Zur Aufbereitung und Analyse entwickelte Methoden (zum Beispiel Wang et al. 2019) können folglich gut sichtbar angeboten und effizient nachgenutzt werden. Die webbasierte Analyseplattform (GESIS-Notebooks) wird dabei auch Codes in sozialwissenschaftlich häufig genutzten Programmiersprachen (R sowie Stata oder SPSS) ausführen können.

Für nachhaltige sozialwissenschaftliche Forschung ist schließlich eine effektive *Archivierung* von Forschungsdaten unerlässlich. Sie garantiert die Transparenz und Replizierbarkeit wissenschaftlicher Forschung, ermöglicht die Durchführung von Meta-Analysen, verhindert redundante Datenerhebungen und ermöglicht ein langfristiges Monitoring der Gesellschaft. Für digitale Verhaltensdaten besteht derzeit keine geeignete Infrastruktur (Kinder-Kurlanda et al. 2017). GESIS plant daher, Instrumente zu schaffen, die es sowohl technisch und methodisch als auch rechtlich abgesichert ermöglichen, digitale Verhaltensdaten effektiv zu archivieren und zugänglich zu machen, sodass sowohl von GESIS als auch extern erhobene Datensätze mit digitalen Verhaltensdaten nachhaltig archiviert, sichtbar vertrieben und effizient nachgenutzt werden können.

3 <https://notebooks.gesis.org>

methodischer und technischer *Kompetenzen* ein Hindernis für einen breiten Einsatz digitaler Verhaltensdaten zur Beantwortung wichtiger soziologischer Fragen. Zur Erhebung, Aufbereitung und Analyse von digitalen Verhaltensdaten bedarf es oft neuer methodischer Ansätze, die bisher nicht im Mittelpunkt der sozialwissenschaftlichen Methodik standen. Zudem lassen sich digitale Verhaltensdaten aufgrund ihrer fehlenden Strukturiertheit und ihrer zum Teil qualitativen Natur nicht ausreichend mit den in der empirischen Sozialforschung sonst üblicherweise genutzten statistischen Methoden analysieren. GESIS wird daher sein stark nachgefragtes Weiterbildungsprogramm in den Bereichen Programmiersprachen, Erhebung sowie Aufbereitung und Analyse von digitalen Verhaltensdaten durch spezifische Formate ausweiten. Darüber hinaus existiert, ähnlich wie im Umfragebereich, ein zunehmender Beratungsbedarf zu allen Fragen rund um digitale Verhaltensdaten. Auch hier wird GESIS Forschenden seine Expertise anbieten.

Ausblick

Auch wenn dieser Beitrag nur einen Überblick geben kann und wichtige Details der Angebote erst mit deren Ausbau Form annehmen werden, versteht es sich von selbst, dass eine Forschungsinfrastruktur entlang gesicherter Erkenntnisse aufgebaut werden muss. So wird begleitende Forschung zu Methoden und Datenqualität für die Ausgestaltung dieser Dienste eine wichtige Rolle spielen, zum Beispiel zur internationalen Vergleichbarkeit von Ergebnissen. Ein wichtiges Ziel dabei ist die systematische Identifikation und Quantifizierung von Fehlerquellen in digitalen Verhaltensdaten, um damit einen Beitrag zu einem *Total Error Framework* für digitale Verhaltensdaten zu leisten (Sen et al. 2019.). Darauf können dann Empfehlungen zur Identifikation, Reduktion und Vermeidung von Fehlern aufbauen.

Angesichts vielfältiger inhaltlicher Aspekte ist GESIS bei diesem Vorhaben nicht nur auf das direkte Feedback der Forschenden als Nutzer:innen angewiesen, sondern auch auf externe Expertise. So werden *Seed-Projekte*⁴ dabei helfen, zum Beispiel die gesellschaftlichen, ethischen und rechtlichen Rahmenbedingungen von digitalen Verhaltensdaten genauer zu betrachten, neue Trends aufzugreifen und Potentiale und Herausforderungen für die so-

⁴ Projekte zur Bewertung und Erschließung relevanter und innovativer Forschungstrends im Bereich digitaler Verhaltensdaten.

zialwissenschaftliche Forschung zu untersuchen. Aktuell ist bereits ein Forschungsvorhaben in Kooperation mit dem Leibniz-Institut für Medienforschung (Hans-Bredow-Institut) zur Forschung im Bereich Medienrecht vereinbart.

Die Potentiale digitaler Verhaltensdaten zu erschließen erfordert einen dialogischen Aufbau der hier skizzierten Infrastruktur. GESIS freut sich daher auf Hinweise und Vorschläge zu den umrissenen Themenbereichen und wird auch weiterhin den Austausch mit Soziolog:innen suchen.

Julian Kohne, Bernhard Miller, Markus Strohmaier,
Claudia Wagner und Christof Wolf

Literatur

- Edelmann, Achim / Wolff, Tom / Montagne, Danielle / Bail, Christopher A. 2020: Computational Social Science and Sociology. *Annual Review of Sociology*, vol. 46, 1–21, doi: 10.1146/annurev-soc-121919-054621.
- Kinder-Kurlanda, Katharina E. / Weller, Katrin / Zenk-Möltgen, Wolfgang / Pfeffer, Jürgen / Morstatter, Fred 2017: Archiving Information from Geotagged Tweets to Promote Reproducibility and Comparability in Social Media Research. *Big Data & Society*, vol. 4, no. 2, 1–14, doi: 10.1177/2053951717736336.
- Lazer, David et. al. 2020: Computational Social Science: Obstacles and opportunities. *Science*, vol. 369, no. 6507, 1060–1062, doi:10.1126/science.aaz8170.
- Pfeffer, Jürgen / Mayer, Katja / Morstatter, Fred. 2018: Tampering with Twitter's Sample API. *EPJ Data Science*, vol. 7, no. 50. doi: 10.1140/epjds/s13688-018-0178-0.
- Pötzsche, Steffen / Braun, Michael 2017: Migrant sampling using Facebook advertisements: A case study of Polish migrants in four European countries. *Social Science Computer Review*, vol. 35, no. 5, 633–653, doi: 10.1177/0894439316666262.
- Quinlan, Stephen / Gummer, Tobias / Roßmann, Joss / Wolf, Christof 2018: Show me the money and the party! – variation in Facebook and Twitter adoption by politicians. *Information, Communication & Society*, vol. 21, no. 8, 1031–1049, doi: 10.1080/1369118X.2017.1301521.
- Sen, Indira / Flöck, Fabian / Weller, Katrin / Weiß, Bernd / Wagner, Claudia 2019: A Total Error Framework for Digital Traces of Online Human Behavior. *Public Opinion Quarterly*. Preprint: <https://arxiv.org/pdf/1907.08228.pdf>.
- Wagner, Claudia / Strohmaier, Markus / Olteanu, Alexandra / Kıcıman, Emre / Contractor, Noshir / Eliassi-Rad, Tina 2021: Measuring algorithmically infused societies. *Nature*, vol. 595, no. 7866, 197–204, doi: 10.1038/s41586-021-03666-1.

- Wang, Zijang / Hale, Scott / Adelani, David Ifeoluwa / Grabowicz, Przemyslaw / Hartmann, Timo / Flöck, Fabian / Jurgens, David 2019: Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. In Ling Liu, Ryan White (eds.), WWW 19 The World Wide Web Conference. New York: ACM, 2056–2067, doi: 10.1145/3308558.3313684.
- Wissenschaftsrat 2021: Stellungnahme zum Antrag auf strategische Erweiterung von GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim und Köln, großer strategischer Sondertatbestand im Rahmen der Ausführungsvereinbarung WGL. Drs. 8820-21, www.wissenschaftsrat.de/download/2021/8820-21.pdf?__blob=publicationFile&v=8, letzter Aufruf am 20. August 2021.

Archivierung und Sekundäranalyse qualitativer Forschungsdaten in der Arbeits- und Industriesoziologie

Der Aufbau der Forschungsdateninfrastruktur für qualitative soziologische Forschungsdaten tritt in eine wichtige Phase, in der Weichen für die Zukunft gestellt werden. Neben dem Konsortium für die Sozial- und Wirtschaftswissenschaften (KonsortSWD) sind seit einigen Jahren auch qualitative Forschungsdatenzentren involviert, die von Forschungseinrichtungen mit dem Fokus auf bestimmte Themenfelder und Datentypen entwickelt werden (Wilke, Pröbrock, Pach 2019). Im folgenden Beitrag wird über den Aufbau eines Forschungsdatenzentrums für qualitative Forschungsdaten in der Arbeits- und Industriesoziologie (AIS) berichtet. Der Text basiert auf Erfahrungen mit Sekundäranalysen in Projektverbänden, aus denen das FDZ eLabour hervorgegangen ist.¹

Sekundäranalysen sind unter qualitativ ausgerichteten Soziolog:innen durchaus umstritten. In der AIS wächst das Interesse an qualitativen Sekundäranalysen und am Angebot des FDZ eLabour, aber nur wenige Wissenschaftler:innen führen selbst Sekundäranalysen durch. Die verbreitete Zurückhaltung gegenüber qualitativen Sekundäranalysen hat Gründe, die im ersten Teil des Beitrags anhand von Erfahrungen aus dem Projektverbund eLabour diskutiert werden (Dunkel, Hanekop, Mayer-Ahuja 2019). Im zweiten Abschnitt

1 Re-sozIT 2013–2015 (<http://sofi.uni-goettingen.de/projekte/re-sozit/projektinhalt/>) und eLabour 2015–2020 (<http://elabour.de/>) wurden vom BMBF gefördert. Die Leitung lag jeweils bei Nicole Mayer-Ahuja. In beiden Forschungsverbänden war ich wissenschaftliche Koordinatorin.

berichte ich über Herausforderungen und Erfahrungen beim Aufbau der qualitativen Forschungsdateninfrastruktur in der Arbeits- und Industriesoziologie sowie Lösungsmöglichkeiten für die diskutierten Probleme qualitativer Sekundäranalysen. Abschließend sollen neben den Chancen auch mögliche Rückwirkungen des Aufbaus und der Verbreitung einer solchen Infrastruktur auf die qualitative Forschung und die Forschungspraxis angesprochen werden.

Probleme und Chancen qualitativer Sekundärforschung

Im Unterschied zur weit verbreiteten Sekundäranalyse quantitativer Forschungsdaten führen qualitative Methoden und Forschungspraktiken zu einer spezifischen Widerständigkeit gegenüber der Nachnutzung qualitativer Forschungsdaten. Typische Merkmale in diesem Zusammenhang sind die Gegenstandsangemessenheit und das Primat der Forschungsfrage bei der Datenerhebung (Strübing et al. 2018) sowie die enge Beziehung der Forschenden zum empirischen Feld und den untersuchten Personen im Erhebungsprozess (Hanekop, Dunkel 2019). Qualitative Forschungsdaten werden in einem offenen Prozess generiert, in dessen Verlauf Erhebungsmethode und Untersuchungssample angepasst und mit dem Erkenntnisfortschritt verfeinert und justiert werden. Die fortschreitende Präzisierung von Forschungsfrage und Forschungsgegenstand steuert diesen Prozess, insofern sind Erhebung und Interpretation nicht strikt voneinander zu trennen. Diese inhaltliche und persönliche Trennung von Erhebung und Auswertung erfolgt unvermeidlich bei der Weitergabe qualitativer Forschungsdaten an »fremde« Wissenschaftler:innen. Daraus ergeben sich typische Probleme soziologischer Sekundäranalysen von qualitativen Forschungsdaten (vgl. dazu auch Birke, Mayer-Ahuja 2017):

- Die Analyse qualitativer Forschungsdaten durch Wissenschaftler:innen, die nicht an der Primärforschung beteiligt waren, trennt diese von dem expliziten und insbesondere von dem impliziten Wissen der Primärforscher:innen über Forschungsgegenstand und Erhebungsprozess. Auch das Erkenntnisinteresse der Primärforschenden, das die Generierung der Forschungsdaten geleitet hat, erschließt sich nicht ohne weiteres aus den Forschungsdaten und auch nicht immer vollständig aus der Publikation der Forschungsergebnisse. Die in der AIS typischen Betriebsfallstudien

bieten zahlreiche Beispiele dafür, wie das Erkenntnisinteresse von Forschergruppen Form und Inhalt der zeitgenössischen Forschungsdaten prägt (Pongratz, Trinczek 2010), die im Forschungsverbund eLabour nachgenutzt wurden.

- Die Forschungsfrage der Sekundärforschung weicht meist von der Primärforschungsfrage ab, da diese selten als Replikation² angelegt ist, sondern ausgehend vom aktuellen Forschungsstand neue Forschungsziele formuliert. Qualitative Sekundäranalysen müssen also mit Forschungsdaten arbeiten, deren Generierung durch abweichende oder gänzlich andere Forschungsfragen und Erkenntnisinteressen geprägt ist. Dies könnte dazu führen, dass der (abweichende) sekundäranalytische Forschungsgegenstand im Primärmaterial mehr oder weniger unscharf oder sogar verzerrt erfasst ist. Je größer die inhaltliche Distanz zwischen Primär- und Sekundärforschungsfrage, umso geringer die Chancen, Daten zu finden, die Aussagen zur Sekundärfrage ermöglichen. Soziologische Sekundäranalysen wählen häufig einen selektiven Zugriff auf qualitative Primärdaten, der aber zwingend die primäre Fragestellung reflexiv in Rechnung stellen muss. Das heißt, Sekundärforscher:innen müssen diese nachvollziehen und verstehen können.
- Datenverändernde Datenschutzmaßnahmen beeinflussen ebenfalls die Datenqualität für die Sekundärforschung. In Betriebsfallstudien der AIS betreffen solche Datenveränderungen oft wichtige Kontextinformationen (Betriebe, Ereignisse etc.), die entfernt oder verfremdet werden müssen, um das Risiko der Re-Identifikation der Befragten zu minimieren. Dies betrifft nicht nur Informationen in der Studie, sondern auch das Herstellen von Feldzusammenhängen und die Verknüpfung von Betriebsfallstudien in der Sekundäranalyse. Eine weitere, ebenso gravierende und oft nicht sinnvoll durchzuführende Datenveränderung betrifft die Pseudonymisierung von Personen, über die in Interviews gesprochen wird.

Aus diesen Risiken für die Durchführung guter Sekundärforschung erwachsen gerade unter erfahrenen, qualitativ arbeitenden Soziolog:innen Unsicherheit und Zweifel am wissenschaftlichen Ertrag qualitativer Sekundäranalysen,

² Eine sekundäranalytische Replikation der Primärforschungsfrage verspricht nur begrenzt neue Erkenntnisse. Sie kann aber in Lehre, Qualifizierung und Methodenentwicklung nützlich sein. Hierfür sind archivierte qualitative Forschungsdaten gut geeignet und häufig alternativlos. Oft sind qualitative soziologische Daten als Quellen für Zeithistoriker:innen interessant (ca. die Hälfte der Nutzungsverträge des FDZ eLabour).

zumal der Aufwand für eine sinnvolle Archivierung und Kuratierung erheblich ist. Warum also sollen qualitative Forschungsdaten in der Soziologie archiviert und nachgenutzt werden, wenn die Qualität der mit großem Aufwand kuratierten Daten für Sekundärforschung eingeschränkt ist? Die Antworten sind so vielfältig wie die möglichen sekundäranalytischen Forschungsansätze; im folgenden einige Erfahrungen aus dem Verbund eLabour:

- Daten aus Forschungsprojekten, die mit guten zeitlichen und personellen Ressourcen ausgestattet waren oder in denen die Gelegenheit bestand, zu historisch einmaligen Zeitpunkten Befragungen durchzuführen, können trotz der oben genannten Probleme für Sekundäranalysen sehr ertragreich sein.
- Der Rückgriff auf Forschungsdaten, die zu einem früheren Zeitpunkt erhoben wurden, ermöglicht eine empirisch fundierte Analyse von neuen Forschungsfragen in einer Längsschnittperspektive.³ Im Verbund eLabour wurden Sekundärforschungsprojekte zum Thema »Wandel der Arbeit nach dem Fordismus« durchgeführt (Dunkel, Hanekop, Mayer-Ahujja 2019). In der Längsschnittperspektive konnte auf der Basis von Primärstudien aus unterschiedlichen Jahrzehnten und Entwicklungsstadien für verschiedene Phänomene gezeigt werden, wie und durch welche gesellschaftlichen Prozesse sie ihre heutige Gestalt erhalten haben.
- Die Kombination bzw. der Vergleich von Forschungsdaten aus mehreren Studien zu einem Forschungsgegenstand in unterschiedlichen Feldern, zum Beispiel Branchen, Regionen, Ländern (Querschnittanalysen), ermöglichen vergleichende Sekundäranalysen mit einer breiteren empirischen Basis, als dies in Einzelprojekten üblicherweise möglich ist.
- Aktuelle empirische Erhebungen können sehr ertragreich mit Sekundäranalysen kombiniert werden. Insbesondere dann, wenn es möglich ist, in der aktuellen Erhebung die Fragestellung und Methoden der Primärerhebung fortzuführen (»echter« Längsschnitt).
- Weitere Ansätze qualitativer Sekundäranalyse sind vorstellbar. Die Erfahrung zeigt, wie wichtig eine gezielte Methodenentwicklung für qualitative Sekundäranalysen ist. Hierfür wäre – flankierend zur Förderung der Forschungsdateninfrastruktur – gezielte Forschungsförderung wünschenswert.

³ Diese Längsschnittperspektiven sind natürlich keine echten Längsschnitte, weil die Primärstudien unterschiedliche Fragestellungen und Methoden verfolgen. Allerdings wurden im SOFI und dem ISF München bestimmte Forschungslinien über viele Jahre mit zahlreichen Projekten verfolgt, die aufeinander aufbauen.

Eine zentrale Herausforderung für Sekundäranalysen mit neuer Forschungsfrage besteht darin, geeignete Primärstudien mit solchen Daten zu finden, die den sekundäranalytischen Gegenstand beinhalten, obwohl er nicht ihr primärer Gegenstand ist. Das Problem kumuliert bei der Kombination unterschiedlicher Studien, die nicht auf eine gemeinsame Analyse angelegt waren. Gleichzeitig bietet gerade die typische Offenheit qualitativer Methoden Chancen, solche sekundäranalytisch passenden Inhalte in den Erzählungen der Befragten, in geschilderten oder beobachteten Zusammenhängen und Ereignissen zu finden. Qualitative Methoden generieren häufig einen inhaltlichen »Überschuss« über das primäre Forschungsinteresse hinaus, der für Sekundäranalysen genutzt werden kann. Die Herausforderung bei der Suche nach passenden Stellen für die Sekundärforschungsfrage ist, dass sie in der Regel zufällig und überraschend auftauchen. Es ist kaum möglich, sie mit Hilfe der primären Erhebungsinstrumente zu suchen, denn nicht selten wurden sie unbeabsichtigt eingefangen. In diesem Prozess des Aufspürens von passenden Inhalten hat die Sekundärforschung Ähnlichkeiten mit qualitativen Erhebungsmethoden im Feld. Allerdings können Sekundärforscher:innen nachträglich keinen Einfluss auf die Daten nehmen, sondern müssen im Zweifel die Suche auf andere Studien ausweiten. Die sekundäranalytische Methode des Aufspürens geeigneter Daten ist eine wissenschaftliche Kompetenz, die Kreativität und Erfahrung erfordert. Bei der Kombination unterschiedlicher Studien sind außerdem Feldkenntnis und Überblick über den Diskurs zum Forschungsstand in der untersuchten Zeitspanne nützlich.

Für die Sekundäranalysen im Verbund eLabour konnten Rahmenbedingungen genutzt werden, die sich als sehr hilfreich erwiesen haben, um die Folgen der Trennung der Daten aus dem Primärforschungskontext abzuschwächen. Das Problem differierender Fragestellungen zwischen unterschiedlichen Projekten der Primär- und Sekundärforschung ist leichter lösbar, wenn die Projekte in einer Forschungslinie oder in einem Diskurszusammenhang entstanden sind. Die Trennung vom Wissen und der Erfahrung der Primärforscher:innen ist weniger einschneidend, wenn es personelle Überschneidungen zwischen Primär- und Sekundärforscher:innen gibt oder ein Austausch möglich ist. Das Problem, keine passenden Daten in den Primärstudien zu finden, kann gelöst werden, wenn weitere Studien verfügbar sind. In einer peer-to-peer Konstellation wie eLabour ist das leichter zu organisieren als in einer großen, thematisch übergreifenden Infrastruktur.

Um die hohen Anforderungen ertragreicher qualitativer Sekundärforschung an die Archivierung und Kuratierung geeigneter Forschungsdaten zu

erfüllen und mit vertretbaren Ressourcen zu bewältigen, ist der Aufbau einer geeigneten Infrastruktur und der Einsatz von IT-Methoden unerlässlich.

Doch wie sieht eine geeignete Infrastruktur aus, welche Anforderungen soll und kann sie erfüllen? Wie können die Beziehungen zwischen den Wissenschaftler:innen, ihren Fachgremien und Organisationen und den Infrastrukturbetreibern gestaltet werden? Im folgenden Abschnitt werden diese Fragen auf der Grundlage der Erfahrungen beim Aufbau des Forschungsdatenzentrums eLabour diskutiert. eLabour steht für einen themenzentrierten bottom-up Ansatz der Infrastrukturentwicklung, der auf peer-to-peer Netzwerken und Organisationen in der AIS-Community aufbaut.

Erfahrungen beim Aufbau der qualitativen Forschungsdateninfrastruktur in der Arbeits- und Industriesoziologie

Das FDZ eLabour ist aus dem Zusammenschluss von Forschungsinstituten aus der AIS⁴ in dem BMBF-Verbundprojekt eLabour hervorgegangen, in dem die aktuelle Forschung zum Wandel von Arbeit in ihrer zeitlichen Genese analysiert und dazu qualitative Studien seit den 1960er Jahren sekundäranalytisch ausgewertet wurden (Dunkel, Hanekop, Mayer-Ahuja 2019). Diese Fokussierung des FDZ auf ein Themenfeld und bestimmte qualitative Methoden erleichtert die Bewältigung der oben diskutierten inhaltlichen Herausforderungen qualitativer Sekundäranalysen. Durch die verteilte Trägerschaft sind umfangreiche Forschungsdaten verfügbar. Gleichzeitig fördert die Verankerung in der AIS-Community den direkten Austausch zwischen Sekundär- und Primärforscher:innen. Im Folgenden werden Erfahrungen beim Aufbau des FDZ eLabour geschildert, die für die Gestaltung der qualitativen Forschungsdateninfrastruktur insgesamt von Interesse sein können, auch wenn sie nur teilweise übertragbar sind.

Das FDZ als Intermediär zwischen Primär- und Sekundärforschung

Das FDZ eLabour versteht sich heute als Intermediär zwischen den Primärforscher:innen in der Rolle als Datenhalter und den Sekundärforscher:innen als Datennutzer. Dieses Grundverständnis hat eine inhaltliche, organisatorische und vertragliche Dimension.

⁴ SOFI Göttingen, ISF München, Sozialforschungsstelle Dortmund, IfS Jena mit Unterstützung der Sektion Arbeits- und Industriesoziologie.

Inhaltlich sind die Primärforscher:innen in die Bereitstellung »ihrer« Forschungsdaten einbezogen. Die Aufbereitung der Forschungsdaten kann nicht ohne die Übergabe, Dokumentation und Erläuterung der Primärforscher:innen erfolgen. Um möglichst viel Erfahrungswissen aufzunehmen, muss dieser Prozess nicht mit der Aufnahme einer Studie in das FDZ abgeschlossen sein. Denn viele Fragen treten erst bei der Aufbereitung einer Studie durch Nutzer:innen auf und lassen sich im direkten Kontakt am besten klären. Erfahrungsgemäß sind viele Primärforscher:innen bereit, solche inhaltlichen Rückfragen zu ihren Studien zu beantworten. Oft sind sie mit den an eLabour beteiligten Instituten verbunden.

Organisatorisch ist eLabour ein Forschungsdatenzentrum mit Sitz in Göttingen und Vertretungen in den beteiligten soziologischen Forschungsinstituten. Die meisten Forschungsdaten im FDZ kommen aus diesen Instituten oder von Wissenschaftler:innen, die über Kooperationsbeziehungen bzw. den Austausch im Rahmen der Sektion AIS mit dem FDZ verbunden sind. In der Regel sind sie in einer Doppelrolle – als Datenhalter und Datennutzer – an eLabour interessiert.

Das FDZ eLabour schließt für jede eingestellte Studie einen Vertrag mit dem/der Datenhalter:in ab, in dem geregelt ist, welche Daten und Materialien der Studie eingebracht werden, welche datenschutzbezogenen Risiken sie enthalten und wie sie für die Nachnutzung bereitgestellt werden sollen. In den meisten Verträgen bleiben die datengebenden Einrichtungen selbst als Datenhalter verantwortlich. Sie können die Studie aber auch in die Verantwortung des FDZ eLabour übergeben. Das FDZ eLabour ist in einer intermediären Rolle beteiligt, wobei es berät und die Rahmenbedingungen für die Archivierung und Nutzung gewährleistet.

Die Beziehungen zu den Datennutzer:innen sind in Nutzungsverträgen geregelt, in denen die Bedingungen für den Zugang zu einer oder mehreren Studien vereinbart werden. Gleichzeitig sichern die Nutzer:innen rechtsverbindlich den datenschutzkonformen Umgang mit den Daten zu, der für bestimmte Daten auch zusätzliche Nutzungsaufgaben umfassen kann.

Die technische Plattform des FDZ

Um die Anforderungen der Sekundäranalyse zu erfüllen und den damit verbundenen hohen Aufwand in den Griff zu bekommen, wurde für das FDZ eLabour eine technische Plattform entwickelt, die alle Aufgaben vom Einlesen der Forschungsdaten, der Bearbeitung und Kuratierung bis hin zum

Freigabeprozess und Zugang zu den Forschungsdaten sowie die Suche und Bearbeitung durch Nutzer:innen unterstützt. Die Entwicklung war ein komplizierter, interdisziplinärer Prozess, in dessen Verlauf wir gemeinsam mit unseren fachlich erfahrenen und aus anderen Projekten vertrauten IT-Partnern GWDG⁵ und L3S⁶ einige unerwartete Lernprozesse zu bewältigen hatten. Besonders wichtig für den Erfolg war die mutige Entscheidung, den Prototypen nach drei Jahren zu verwerfen und die Entwicklung der Plattform mit den gewonnenen Erkenntnissen neu aufzusetzen.

eLabour erleichtert durch teil-automatisierte Prozesse das Einlesen, Archivieren und Bereitstellen der Forschungsdaten. Diese Prozesse bieten zugleich ein hohes Maß an Flexibilität und Kontrolle. Hier war der immanente Widerspruch zwischen der Offenheit qualitativer Methoden einerseits und dem für die Infrastrukturentwicklung unerlässlichen Mindestmaß an Regeln, Standards und Formalisierung andererseits zu bearbeiten. Beispielsweise ist es möglich, das Einlesen von heterogenen Datenstrukturen durch Übergabe einer komplexen Ordnerstruktur teil-automatisch zu steuern. Außerdem sind alle Prozesse so angelegt, dass sie eine Ergänzung oder Verbesserung der archivierten Studien (Kuratierung, Kontextualisierung) erlauben. Alle Aufgaben sind modular ausführbar. Die erforderlichen Berechtigungen werden auf Grundlage eines IT-basierten Rollenmodells zugewiesen, kontrolliert und dokumentiert. Die IT-gestützte Kontrolle umfasst auch den datenschutzbezogenen Freigabeprozess der Daten und die Steuerung der Zugangsmöglichkeiten für Nutzer:innen. Durch die Definition einer Freigabeklasse auf der Ebene von Dokumenten können differenzierte Zugangs- und Nutzungsbedingungen realisiert und kontrolliert werden, um Datenschutzmaßnahmen und vertragliche Auflagen durchzusetzen. Für Nutzer:innen bietet die Plattform jederzeit Online-Zugang zu den vertraglich vereinbarten Forschungsdaten. Darüber hinaus umfasst sie ein avanciertes Suchmodul und die Möglichkeit, eigene Datensätze für die Sekundärforschung zusammenzustellen und im Rahmen der Nutzungsverträge mit anderen Nutzer:innen zu teilen.

5 Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) und Forschungsabteilung der Staats- und Universitätsbibliothek Göttingen (SUB). Die Plattform eLabour wurde maßgeblich von der GWDG entwickelt, wird von der GWDG bereitgestellt und weiterentwickelt. Es besteht eine dauerhafte Kooperationsbeziehung zum FDZ eLabour.

6 Forschungszentrum L3S an der Gottfried Wilhelm Leibniz Universität Hannover. Das L3S hat die Suchplattform als Teil der Plattform eLabour entwickelt und kooperiert auch weiterhin eng mit eLabour. Alle Partner sind Mitglieder im Trägerverein eLabour e.V.

Alle an der Plattform eLabour beteiligten Einrichtungen sind offen für Kooperationen im Rahmen des Infrastrukturaufbaus, auch mit der Möglichkeit, Teile der technischen Plattform für die Nutzung durch andere Einrichtungen zu öffnen.

Aufbereitung qualitativer Forschungsdaten: der Primärforschungskontext

Das FDZ eLabour kann auf einen umfassenden Bestand an empirisch gehaltvollen qualitativen Studien aus den beteiligten Instituten seit den 1970er Jahren zurückgreifen, teilweise bereits in digitalisierter Form. Thematisch sind die Daten geprägt durch die Forschungslinien der Institute und die Diskurse in der AIS. Methodisch handelt es sich in der Regel um Betriebsfallstudien (Pongratz, Trinczek 2010).

Oft enthält eine Studie mehr als 500 Interviews, Arbeitsplatzbeobachtungen und Felddokumente. Entgegen der ursprünglichen Erwartung, diesen Datenbestand rasch über die Plattform eLabour für Sekundärforschung nutzen zu können, hat sich gezeigt, dass gute Sekundärforschung mehr benötigt als empirische Dokumente und deren Dokumentation. Gerade bei älteren Studien ist es wichtig, den Erhebungsprozess nachvollziehbar und das Kontextmaterial über das zeitgenössische empirische Feld verfügbar zu machen; am besten ergänzt durch nicht verschriftlichtes Erfahrungswissen der Primärforscher:innen. Diese Art der Aufbereitung und Anreicherung von Interviewprotokollen ist aufwändig und nur schrittweise zu bewältigen. Daher werden weitere Primärstudien aus den Beständen der Partnerinstitute in der Reihenfolge aufbereitet, in der sie von Sekundärforscher:innen angefragt werden. Bisher stehen zehn umfangreiche Studien aus dem Zeitfenster von 1977 bis 2016 zur Verfügung. Laufend kommen weitere Studien auch aus aktuell abgeschlossenen Projekten hinzu. Wenn eine Studie nachgefragt wird, ermitteln wir den Aufwand und suchen gemeinsam mit den Nachfragenden (gegebenenfalls mit Forschungsförderung) nach einer Lösung. Voraussetzung für dieses Verfahren ist die Sicherung der Forschungsdaten und Materialien aus abgeschlossenen Projekten in einem geeigneten Repository. Das FDZ eLabour bietet hierfür ein Originaldatenarchiv, in dem Datenhalter ihre Daten auf dem höchsten Datenschutzniveau und ohne Zugang für andere ablegen können.

Auch bei der Kontextualisierung gehen wir schrittweise vor. Die Struktur der Primärstudie und die zentralen Kontextinformationen werden als Metadaten IT-gestützt in der Plattform erfasst, zusätzlich werden gegebenenfalls

unbearbeitete Kontextdokumente (oder Verweise) aufgenommen, sodass Sekundärforscher:innen sie auswerten können. Nutzer:innen sind aufgefordert, sich an der Verbesserung zu beteiligen. Für alle weiteren Fragen bietet das FDZ den direkten Austausch mit Primärforscher:innen.

Aufbereitung qualitativer Forschungsdaten: Datenschutz und Freigabeprozess

Qualitative, soziologische Forschungsdaten können trotz Pseudonymisierung stets überraschende personenbeziehbare Informationen enthalten, die zusätzliche, datenverändernde Schutzmaßnahmen erforderlich machen. Diese können die Sekundäranalyse der Daten erheblich beeinträchtigen, insbesondere dann, wenn sie feldbezogene Kontexte verändern oder entfernen.

Eine zentrale Herausforderung bei der Durchführung datenverändernder Maßnahmen besteht darin, dieses Spannungsverhältnis zwischen den Anforderungen des Datenschutzes und der Sekundäranalyse stets zu reflektieren. Diese Entscheidungen werden in eLabour im Rahmen des Freigabeprozesses getroffen. Ziel des Freigabeprozesses ist, das gesamte Spektrum an Datenschutzmaßnahmen auszuschöpfen, das die DSGVO mit den Ausnahmeregelungen für die Wissenschaft bietet, um eine möglichst hohe Qualität der Forschungsdaten zu erhalten.

Im Freigabeprozess werden alle empirischen Dokumente geprüft und erhalten eine Freigabeklasse. Gleichzeitig werden Datenschutzmaßnahmen durchgeführt, um die Risiken so weit zu minimieren, wie dies für eine sinnvolle wissenschaftliche Verwendung möglich ist. Wenn der Datenschutz nur durch weitreichende und sehr aufwändige Datenveränderungen zu erfüllen ist, wird im Datenschutzkonzept von eLabour empfohlen, die Veränderungen zu unterlassen und die Daten nicht zugänglich zu machen. Anderenfalls stünde ein hoher Aufwand einer schlechten Qualität der Daten für die Sekundäranalyse gegenüber. Besonders sensible Daten werden nicht zugänglich gemacht, sondern verbleiben im Originaldatenarchiv. Dies wird für Sekundärforscher:innen dokumentiert. In der Plattform eLabour werden die Dokumente entsprechend ihrer Freigabeklasse für die wissenschaftliche Nutzung bereitgestellt. Im Nutzungsvertrag werden für Freigabeklassen unterschiedliche Nutzungsbedingungen festgelegt. Dazu zählen Datenschutzmaßnahmen bei den Nutzer:innen und Auflagen bei der Weitergabe und Veröffentlichung von Informationen und Zitaten aus den Daten. Die Durchführung des Freigabeprozesses und die Kontrolle der Zugangsbedingungen in der Plattform eLabour werden technisch unterstützt und kontrolliert.

Für alle Aufbereitungsschritte qualitativer Forschungsdaten gilt das schwierige Spannungsverhältnis zwischen Zugang zu den Forschungsdaten, guter Datenqualität für die Sekundärforschung und vertretbarem Aufwand der Datenaufbereitung. Bei der Beurteilung des vertretbaren Aufwands muss auch reflektiert werden, welche Aufgaben und Belastungen für Primärforscher:innen zumutbar sind, welche Wirkung sie auf qualitative Forschungspraktiken haben und wie die zusätzlich benötigten Ressourcen (Fördermittel) dafür bereitgestellt werden.

Ausblick

Der Weg des Infrastrukturaufbaus und das Konzept von eLabour sind aus dem Kontext der AIS und der Institute gewachsen und haben sich hier als sinnvoll und praktikabel erwiesen. In anderen Konstellationen, insbesondere für größere Infrastrukturen, bieten sich wahrscheinlich andere Konzepte an. Zwei grundlegende Erfahrungen möchte ich jedoch abschließend hervorheben, da sie den Aufbau qualitativer Forschungsinfrastrukturen insgesamt betreffen.

Erstens sind die Verbesserung der Forschungsmöglichkeiten und die Qualität der Forschung für uns nicht nur allgemeine, übergeordnete Ziele, sondern gleichzeitig das wichtigste Entscheidungskriterium bei konzeptionellen wie auch bei sehr konkreten Gestaltungsfragen. Das mag banal klingen. Jedoch fordert ein gängiges Argument in der gegenwärtigen Infrastrukturdebatte, dass die Bereitstellung von Daten zukünftig ein Qualitätsmerkmal von Forschung sein soll. Damit wird die Beziehung zwischen Forschungsqualität und Forschungsdateninfrastruktur nach meinem Verständnis auf den Kopf gestellt. Stattdessen sollten Entscheidungen insbesondere Weichenstellungen in Bezug auf die Forschungsdateninfrastruktur darauf geprüft werden, ob sie die Möglichkeiten und die Qualität der Forschung fördern und von der Forschung umsetzbar sind. Notfalls auch mit dem Ergebnis, dass bestimmte Forschungsdaten nicht bereitgestellt werden können. Zumal eine Forschungsdateninfrastruktur nicht nur Hindernisse beseitigt, sondern gelegentlich aus Sicht der Forschenden auch neue schafft. Für die qualitative Soziologie können die Risiken einer Verselbständigung der Infrastrukturentwicklung gravierend sein, denn sie tangieren nicht nur die Frage, ob und wie Forschungsdaten archiviert und nachgenutzt werden, sondern auch die Bedingungen unter denen qualitative Forschung zukünftig

stattfindet. Unter anderem sollte der Zusatzaufwand nicht zu Lasten der Forschung gehen. Ich plädiere dafür, solche Risiken ernst zu nehmen, sie aufmerksam zu reflektieren und sich aktiv in die Planung und den Aufbau der Forschungsinfrastruktur einzumischen.

Der zweite, unmittelbar anschließende Punkt bezieht sich auf das Verhältnis zwischen Wissenschaftler:innen und Infrastruktur(betreibern) sowie zwischen Primär- und Sekundärforscher:innen. eLabour steht hier – wie andere qualitative FDZ – für eine tragende Rolle der Forschungsinstitute und die weitreichende Einbeziehung der Wissenschaftler:innen einer Community. Damit bietet es im Kleinen ein institutionelles Setting, das der Verselbständigung der Infrastruktur entgegenwirkt. In großen Infrastrukturprojekten ist die enge Einbindung der Forschenden und ihrer Organisationen schwerer umzusetzen. Dem könnte die Soziologie entgegenwirken, zum Beispiel indem Vertreter:innen der Forschenden nicht nur im übergreifenden Beirat mitwirken, sondern auch in den Arbeitsbereichen, in denen konkrete Entscheidungen getroffen werden. Gleichzeitig würde die intensive Einbeziehung der Forschenden die Akzeptanz und Verbreitung der Forschungsinfrastruktur fördern. Soziolog:innen und ihre Fachgemeinschaften sollten sich daher intensiv in die Gestaltung der Forschungsinfrastruktur einmischen.

Heidemarie Hanekop

Literatur

- Birke, Peter / Mayer-Ahuja, Nicole 2017: Sekundäranalyse qualitativer Organisationsdaten. In Stefan Liebig / Wenzel Matiaske / Sophie Rosenbohm (Hg.), *Handbuch empirische Organisationsforschung*. Berlin: Springer, 105–126.
- Dunkel, Wolfgang / Hanekop, Heidemarie / Mayer-Ahuja, Nicole 2019: Blick zurück nach vorn, Sekundäranalysen zum Wandel von Arbeit nach dem Fordismus. *International Labour Studies*. Frankfurt am Main, New York: Campus.
- Hanekop, Heidemarie / Dunkel, Wolfgang 2019: Aufbereitung, Archivierung und Nachnutzung qualitativer Datensätze – das Forschungsdatenzentrum eLabour. In Wolfgang Dunkel / Heidemarie Hanekop / Nicole Mayer-Ahuja (Hg.): *Blick zurück nach vorn. Sekundäranalysen zum Wandel von Arbeit nach dem Fordismus*. *International Labour Studies*, Frankfurt am Main, New York: Campus, 25–58.
- Pongratz, Hans J. / Trinczek, Rainer 2010: *Industriesoziologische Fallstudien. Entwicklungspotenziale einer Forschungsstrategie*. Berlin: Springer.

- Strübing, Jörg / Hirschauer, Stefan / Ayaß, Ruth / Krähnke, Uwe / Scheffer, Thomas 2018: Gütekriterien qualitativer Sozialforschung. *Zeitschrift für Soziologie*, 47. Jg., Heft 2, 83–100.
- Wilke, René / Pröbrock, Willi / Pach, Helen 2019: Infrastrukturen für Forschungsdaten der qualitativen Sozialforschung. *SOZIOLOGIE*, 48. Jg., Heft 4, 467–486.

»Data Sharing« mit Qualiservice

Data Sharing in der qualitativen Sozialforschung:
Potentiale und Herausforderungen

Im Unterschied zur quantitativen Forschung ist das Archivieren und Teilen (Data Sharing) von qualitativen Forschungsdaten in den Sozialwissenschaften noch nicht weit verbreitet. Zugleich äußern viele qualitativ arbeitende Forscher:innen eine große Bereitschaft zum Teilen ihrer Forschungsdaten, aber auch großen Bedarf an Informationen und Beratungsangeboten zum Thema Management und Teilen von Forschungsmaterialien (Medjedović, Witzel 2010; Medjedović 2014; FID Soziologie 2018).

Zu den Vorteilen des Data Sharings zählen erhöhte Transparenz und Chancen für verbesserte Qualität der Forschung sowie forschungsökonomische Aspekte (Knoblauch 2013; RatSWD 2018). Häufig enthalten die in qualitativen Forschungsprojekten aufwändig generierten Daten weit mehr Informationen als zur Beantwortung der Forschungsfrage erforderlich und könnten unter neuen Aspekten, mit anderen Methoden oder in Kombination mit anderen Daten weiter ausgewertet werden. Dies generiert neue Möglichkeiten, beispielsweise für Zeitvergleiche und Follow-up-Studien, Cross-disziplinäre Datennutzungen, Methodenforschung und als Inspiration für neue Forschungsfragen. Aber das Archivieren und Teilen von qualitativen Forschungsmaterialien bringt nicht allein Vorteile für die Scientific Community, sondern auch für die datengebenden Forschenden selbst: Sicherheit vor Datenverlust, Erfüllung datenschutzrechtlicher und forschungsethischer Standards, bessere Datenaufbereitung und Datenzugänglichkeit – und mit letzterer auch erhöhte Sichtbarkeit der eigenen Forschungsarbeit (Wilke, Pröbrock, Pach 2019: 473).

Auf der anderen Seite schafft das Teilen von qualitativen Forschungsdaten besondere Herausforderungen sowohl für die datengebenden Primärforscher:innen als auch für die Sekundärnutzenden – und damit auch für die sie unterstützenden Forschungsdateninfrastrukturen. Herausforderungen resultieren vor allem aus der Vielfalt und Komplexität, der Kontextabhängigkeit und der Personengebundenheit qualitativer Forschungsdaten.¹ Im Folgenden skizzieren wir die besonderen Anforderungen an das Forschungsdatenmanagement qualitativer Forschungsmaterialien und stellen die Lösungsmöglichkeiten vor, die das Forschungsdatenzentrum Qualiservice an der Universität Bremen hierzu in den letzten Jahren entwickelt hat. Der Beitrag schließt mit einem Ausblick auf aktuelle Herausforderungen und Perspektiven im Rahmen der Nationalen Forschungsdateninfrastruktur.

Das Forschungsdatenzentrum Qualiservice

Das FDZ Qualiservice an der Universität Bremen ist ein vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditiertes Forschungsdatenzentrum, das qualitative sozialwissenschaftliche Forschungsdaten archiviert und für wissenschaftliche Nachnutzungen zur Verfügung stellt. Im Unterschied zu Archiven und Forschungsdatenzentren, die sich auf bestimmte Themenfelder konzentrieren, nimmt Qualiservice qualitative Forschungsdaten themenunabhängig auf. Dabei lag der Schwerpunkt lange auf Interviewdaten. Mittlerweile kuratiert Qualiservice die gesamte Bandbreite qualitativer sozialwissenschaftlicher Daten, insbesondere auch ethnographische Daten wie Beobachtungsprotokolle, Feldnotizen, Text-, Audio- und Videodaten.

Bezogen auf die Archivierung von qualitativen Forschungsdaten ist Qualiservice neben dem ESDS Qualidata des UK Data Archive in Essex (Corti 2018) in Europa einer der Vorreiter und hat auch in Deutschland die Diskussion um das Data Sharing in den Sozialwissenschaften maßgeblich vorangetrieben und begleitet. Qualiservice wurde Ende der 1990er Jahre als

¹ Die Grenzen und Möglichkeiten der Archivierung und Nachnutzung qualitativer Daten wurden in der letzten Dekade in der Soziologie intensiv diskutiert. Meilensteine der Diskussion, in der die Anforderungen an das Forschungsdatenmanagement qualitativer sozialwissenschaftlicher Forschungsdaten herausgearbeitet wurden, bilden die Beiträge von Knoblauch und Solga (2011), die Beiträge in Huschka et al. (2013), die Resolution der DGS-Sektionen Biographieforschung und Qualitative Methoden der empirischen Sozialforschung (2014), die Stellungnahme des RatSWD (2015) sowie die Beiträge der RatSWD-Tagung zu Archivierung und Zugang zu qualitativen Daten in Bremen 2018 (RatSWD 2018).

»Archiv für Lebenslaufforschung« von Forschenden des Sonderforschungsbereiches 186 »Statuspassagen und Risikolagen im Lebensverlauf« der Universität Bremen eingerichtet. Ziel war es zunächst, die rund 850 qualitativen Interviews, die in der über zehnjährigen Laufzeit des SFBs erhoben worden waren, zu archivieren und anderen Forscher:innen für Sekundäranalysen zur Verfügung zu stellen (Witzel 2000).

Bald entstand der Gedanke, das Angebot auch auf Forscher:innen außerhalb von Bremen zu erweitern. Um den Bedarf der Community zu eruieren, führte Qualiservice zusammen mit GESIS eine DFG-geförderte Machbarkeitsstudie durch. Befragt wurden alle Sozialwissenschaftler:innen, die in den letzten zehn Jahren qualitatives Interviewmaterial erhoben hatten, insgesamt rund 1.750 Projekte (Opitz, Mauer 2005; Medjedović, Witzel 2010; Medjedović 2014). Mit den Ergebnissen dieser Befragung wurde das Konzept für ein Forschungsdatenzentrum entworfen. Im Rahmen eines weiteren DFG-Projekts wurden dafür zwischen 2011 und 2014 die Basiskomponenten entwickelt. Hierzu gehörten unter anderem ein Metadatenschema für qualitative Interviewdaten (Betancort, Haake 2014), Workflows für deren Archivierung und Sekundärnutzung und ein Anonymisierungskonzept (Kretzer 2013). Als ein FDZ, das von Forscher:innen initiiert und getragen wird, suchte Qualiservice von Beginn an den engen Austausch mit der Community. Workshops mit Forschenden, Nutzer:innen und Kolleg:innen anderer Datenzentren im In- und Ausland sind zentraler Bestandteil der Entwicklung des FDZ. Ein Beirat aus Fachwissenschaftler:innen und Expert:innen für Datenschutz und Forschungsethik berät das FDZ und begleitet seine Entwicklung.²

Seit 2015 ist Qualiservice am SOCIUM, dem sozialwissenschaftlichen Forschungszentrum der Universität Bremen, angesiedelt. Die technisch-infrastrukturelle Entwicklung von Qualiservice erfolgt in Kooperation mit PANGAEA, einem international zertifizierten Weltdatenarchiv.³ Die Staats- und Universitätsbibliothek Bremen unterstützt die Entwicklung von Metadaten für qualitative Forschungsdaten und ist für die Publikation der Stu-

2 Aktuell gehören dem Beirat an: Günter Mey (Vorsitz), Hansjörg Dilger, Grit Laudel, Herwig Reiter, Margrit Schreier, Jörg Strübing, Hella von Unger und die Datenschutzbeauftragte der Max-Planck-Gesellschaft Heidi Schuster. Frühere Mitglieder waren Nigel Fielding, Andreas Witzel sowie Louise Corti vom UK Data Archive.

3 PANGAEA ist Mitglied des »International Council for Science World Data System (ICSU-WDS)« und ist als »World Data Center (WDC)« zertifiziert und akkreditiert. www.pangaea.de.

dienreports verantwortlich. GESIS ist Partner bei der verteilten Archivierung von Mixed-Methods-Studien: Quantitative Daten werden von GESIS, qualitative Daten bei Qualiservice archiviert und bereitgestellt.

Zusammen mit der Bibliothek der Humboldt-Universität verantwortet Qualiservice seit 2019 den Fachinformationsdienst Sozial- und Kulturanthropologie (FID SKA). Im Rahmen des FID SKA wird bei Qualiservice ein Bereich für ethnographische Forschungsdaten aufgebaut, sodass neben dem bisherigen Fokus auf Interviewdaten auch Bild-, Ton- und Videomaterial, Feldnotizen und Beobachtungsprotokolle archiviert und für die geschützte Weiternutzung in Forschung und ausgewählte Materialien auch für die Lehre angefordert werden können.

Im Juni 2020 wurde die Leiterin von Qualiservice, Betina Hollstein, vom Bundesministerium für Bildung und Forschung in den RatSWD berufen. Damit ist jetzt das erste Mal auch eine qualitative Dateninfrastruktur im Rat vertreten. Aktuell bereitet Qualiservice die Implementation des Regelbetriebs vor. Der Übergang in den Vollbetrieb ist für das Frühjahr 2022 geplant. Daneben koordiniert Qualiservice im Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten (KonsortSWD) seit Beginn des Jahres den Aufbau eines Verbunds für Qualitative Forschungsdaten (Qualidata-Net) als Teil der Nationalen Forschungsdateninfrastruktur.

Lösungen für die Herausforderungen des Data Sharings in der qualitativen Sozialforschung

In diesem Abschnitt stellen wir die Lösungsmöglichkeiten vor, die Qualiservice in den letzten Jahren entwickelt hat, um den besonderen Anforderungen an das Forschungsdatenmanagement qualitativer Forschungsmaterialien gerecht zu werden. Im Mittelpunkt stehen zunächst die Anforderungen an eine sichere Archivierung und Bereitstellung von Forschungsdaten, und anschließend dann die Möglichkeiten zur Förderung und Verbesserung der Nachnutzung. Zum Schluss wenden wir uns der grundlegenden Bedeutung zu, die der Zusammenarbeit von Forschenden und Fachcommunities mit den Forschungsdateninfrastrukturen bei der Entwicklung von Lösungen für das Teilen qualitativer Forschungsdaten zukommt.

Forschungsdaten sicher archivieren und bereitstellen

Qualitative Forschung zeichnet sich durch eine prinzipielle Offenheit des Forschungsprozesses, gegenstandsbezogene Methodenvielfalt und oftmals multimethodische Erhebungsstrategien sowie daraus resultierende heterogene Datenarten und Datenformate aus, die unstrukturierte, reiche Informationen enthalten und unter Umständen auf spezifische Weise miteinander verknüpft bzw. aufeinander bezogen sind. Diese Heterogenität und Komplexität qualitativer Forschungsmaterialien stellt hohe Anforderungen an die Archivierung und Aufbereitung für die Sekundärnutzung. Hinzu kommt, dass qualitative Forschung auf der Interaktion der Forschenden mit Menschen beruht, die oft (auch über längere Zeiträume) tiefe Einblicke in ihr persönliches Leben gewähren, sodass hochgradig sensible bzw. schutzwürdige und stark personenbezogene Forschungsmaterialien entstehen.

Qualitative Forschungsdaten erfordern somit ein hohes Maß an Kontrolle und Vorkehrungen, um den Persönlichkeits- und Datenschutz der beteiligten Personen zu gewährleisten. Deshalb haben Maßnahmen zum Datenschutz und zur Datensicherheit bei Qualiservice oberste Priorität. Unser umfangreiches Datenschutzkonzept enthält zahlreiche technische und organisatorische Maßnahmen zum Schutz der Forschungsdaten und der zugehörigen Kontextmaterialien. Es ist allerdings auch klar, dass der Schutz und die Sicherheit der Beteiligten und ihrer Daten über die Behandlung rechtlicher Fragen hinausgeht und fachliche sowie forschungsethische Aspekte einschließt (vgl. RatSWD 2015; RatSWD 2018; von Unger 2021). Um der besonderen Verantwortung der Forschenden als auch der sie unterstützenden Forschungsdatenzentren gerecht zu werden und Aspekte wie Vertraulichkeit, Partizipation und Schutz der Forschungsteilnehmenden zu berücksichtigen, hat Qualiservice von Beginn an einen engen Austausch mit den Forschenden und den Fachcommunities gesucht und auch große forschungsethische Expertise in seinem wissenschaftlichen Beirat versammelt (vgl. Fußnote 2).

Generell werden bei Qualiservice die rechtlichen und forschungsethischen Aspekte bei der Archivierung und dem Teilen qualitativer Daten ausführlich mit den Forschenden besprochen, um zu Lösungen zu kommen, die alle beteiligten Personen und Personengruppen vor möglichen Beeinträchtigungen schützen. Zudem werden alle Forschungsmaterialien nach dem Vier-Augen-Prinzip sowohl von den Forschenden als auch bei Qualiservice auf die Erfüllung datenschutzrechtlicher und forschungsethischer

Anforderungen geprüft. Entscheidungen über die erforderlichen Schutzmaßnahmen werden stets gemeinsam mit den Forscher:innen getroffen.

Die Archivierung und Bereitstellung von Forschungsdaten mit Qualiservice erfolgt obligatorisch auf Grundlage einer »Informierten Einwilligung« der Forschungsteilnehmenden. In Zusammenarbeit mit Jurist:innen hat Qualiservice für das Einholen einer schriftlichen Einwilligung Vorlagen erstellt, die die Anforderungen der Europäischen Datenschutz-Grundverordnung (DSGVO) erfüllen und von den Forschenden auf der Qualiservice-Website heruntergeladen und an die Bedarfe des eigenen Forschungsprojekts angepasst werden können. Qualiservice unterstützt Forschende dabei sowohl mit individueller Beratung und in themenbezogenen Workshops als auch mittels einer Handreichung, die den rechtlichen und ethischen Hintergrund der »Informierten Einwilligung« erläutert und Hinweise zum Umgang mit den Vorlagen gibt (Kretzer et al. 2020). Darüber hinaus steht Qualiservice im Austausch mit Datenschutz-Expert:innen und Forscher:innen, um Lösungen für Forschungssituationen zu finden, in denen das Einholen einer (schriftlichen) Einwilligung aus ethischen oder forschungspraktischen Gründen nicht möglich ist (vgl. Huber, Imeri 2021).

Die Übergabe von Forschungsmaterialien an Qualiservice erfolgt stets erst nach Unterzeichnung einer Datenübergabvereinbarung. Dieser Vertrag bietet den Datengebenden die Möglichkeit, spezifische Bedingungen für die Archivierung und Sekundärnutzung festzulegen. Dies kann beispielsweise ein Embargo sein, das die Nutzung der Daten erst nach einem bestimmten Zeitraum erlaubt, oder ein Ausschluss von bestimmten Nutzungen, wie eine Verwendung der Daten in der Lehre.

Nach dem Upload der Forschungsdaten durch die Forschenden über eine gesicherte Verbindung beginnt ein umfangreicher Prozess der Datenkuratation bei Qualiservice, der je nach Datentyp, Sensibilität der Inhalte und Festlegungen durch die Datengebenden neben rechtlichen, technischen und Qualitätsprüfungen auch spezifische inhaltliche und formale Prüfungen und Bearbeitungsschritte umfasst. Dies betrifft beispielsweise die Anonymisierung bzw. Pseudonymisierung der Forschungsmaterialien, die in der Regel durch die Forschenden erfolgt, aber von Qualiservice vollständig geprüft und ergänzt wird. So sind Interviewtranskripte und andere Textdaten oftmals so anonymisierbar, dass sie sinnvoll nachgenutzt werden können, während zum Beispiel bei Bild- und Tondaten die Anonymisierung ein Problem darstellen kann, das bestimmte Zugangswege und Nachnutzungsszenarien ausschließt. Somit müssen je nach Datentyp und Inhalten unterschiedliche

Kurationsschritte erfolgen und Sicherheitsmaßnahmen ergriffen werden, um adäquate Nachnutzungsmöglichkeiten zu schaffen. Die für diese unterschiedlichen Datentypen entwickelten Workflows für die Datenkuration sind in enger Zusammenarbeit mit Forschenden, unter anderem in der Form von »Use Studies«, erprobt und optimiert worden.

Das Kurationspersonal besteht aus ausgebildeten Sozialwissenschaftler:innen unterschiedlicher disziplinärer Herkunft mit Erfahrungen in der qualitativen Sozialforschung. Diese arbeiten in einem besonders gesicherten und nur für das Kurationssteam zugänglichen »Safe Center«, das wie die übrige IT-Infrastruktur in enger Zusammenarbeit mit dem CTS-zertifizierten World Data Center PANGAEA eingerichtet wurde und höchste Sicherheitsanforderungen erfüllt. Die Kuration beginnt stets damit, dass direkte Identifikationsdaten wie beispielsweise Kontaktdaten der Forschungsteilnehmenden aus den Forschungsmaterialien entfernt und auf separaten Rechnern sowie örtlich getrennt gespeichert werden. Anschließend werden die Forschungsdaten und Kontextmaterialien geprüft und abgabefähige Datensätze für die Sekundärnutzung einerseits und für die Langzeitarchivierung andererseits erstellt.

Zum Abschluss der Kuration können die Datengebenden den vollständig aufbereiteten Datensatz mit einem »Proofread« begutachten. Alle Datensätze bekommen eine permanente Identifikation in Form eines Digital Object Identifiers (DOI) zur sicheren Auffindbarkeit und sind damit zitierbar. Für die Archivierung und Bereitstellung von Mixed-Methods-Studien wurde ein gemeinsamer Workflow mit GESIS entwickelt, bei dem die qualitativen Datensätze bei Qualiservice sowie die quantitativen Datensätze bei GESIS unmittelbar aufeinander verweisen.

Die Zugangsformen bzw. Nutzungsbedingungen für die Daten werden von Qualiservice mit den Forschenden gemeinsam besprochen und festgelegt. Grundsätzlich sind aus den oben genannten datenschutzrechtlichen und forschungsethischen Gründen weder die Forschungsdaten selbst noch die oftmals ebenfalls sensiblen Kontextmaterialien⁴ unmittelbar verfügbar bzw. abrufbar. Angeboten wird vielmehr ein kontrollierter Zugriff – das heißt, die Daten stehen nicht öffentlich für jeden bereit, sind aber verfügbar für wissenschaftliche Nachnutzungen. Dabei müssen die nachnutzenden Forscher:innen den Zweck der Nutzung angeben und eine Nutzungsvereinbarung unterzeichnen. Die Daten können dann je nach Einstufung entweder

⁴ Mit Ausnahme des Studienreports und der Metadaten; siehe nächster Abschnitt.

als »Scientific Use File« extern heruntergeladen werden oder an einem Gastwissenschaftler:innen-Arbeitsplatz vor Ort in einem besonders gesicherten »Safe Room« genutzt werden. Es ist möglich, dass unterschiedliche Versionen und Formate der Daten eines Forschungsprojekts, etwa Transkripte und audiovisuelle Daten, unterschiedliche Sicherheitskategorien erhalten, sodass jeweils angemessene Nachnutzungen möglich sind.

Neben der umfangreichen Kuration ist ein zweites besonderes Merkmal der Arbeit von Qualiservice die Begleitung und Unterstützung der Datengebenden bereits während des Forschungsprozesses. Dies geschieht nicht nur über Informationsmaterialien und Workshops zu spezifischen Themen, sondern auch im persönlichen Kontakt während einer individuellen Beratung, in der für die jeweiligen Bedarfe des Forschungsprojekts passgenaue Absprachen und Lösungen entwickelt werden sollen. Dies bedeutet konkret, dass den Forschenden jederzeit ausgebildete sozialwissenschaftliche Domain-Expert:innen als Berater:innen zur Verfügung stehen, die beispielsweise Informationen, Erfahrungen und »Best Practices« vermitteln können, um so den Bedarf von Forschenden nach Unterstützung beim Forschungsdatenmanagement zu bedienen. Neben dem persönlichen Kontakt kann die Kommunikation mit Qualiservice auch per Ticketsystem über eine sichere und dialogische Verbindung erfolgen. Sie ermöglicht zugleich die fortlaufende Dokumentation des Beratungsprozesses.

Zudem berät Qualiservice Forschende bereits vor Projektantragstellung bei der Planung der Ressourcen, die sowohl bei Qualiservice als auch im Forschungsprojekt für die Dokumentation und die Aufbereitung der Daten berücksichtigt werden müssen und bei den Forschungsförderern beantragt werden können. Das Resultat kann ein gemeinsam entwickelter Datenmanagementplan sein, der eine Orientierung im Forschungsverlauf bietet, aber selbstverständlich an die Entwicklung des qualitativen Forschungsprojekts angepasst werden kann und sollte. Im Idealfall beginnt somit die Zusammenarbeit von Qualiservice und Forschenden bereits mit bzw. vor der Antragstellung und schließt den gesamten weiteren Datenlebenszyklus ein.

Forschungsdaten auffindbar und nachnutzbar machen

In der qualitativen Sozialforschung ist der Kulturwandel zum Data Sharing erst in seinen Anfängen und bedarf der Unterstützung auch von Seiten der Forschungsdateninfrastrukturen. Besonders wichtig ist dabei, dass archivierte Daten gut nachgewiesen werden und möglichst breit sichtbar sind, sodass

an einer Nachnutzung Interessierte schnell und einfach die Informationen erhalten, die sie benötigen, um die Eignung der Daten für ihr Forschungsanliegen einschätzen zu können. Diesem Zweck dienen einerseits die Metadaten zu den Datensätzen, andererseits der von den Forschenden verfasste Studienreport. Beide präsentieren öffentlich zugänglich zentrale (Kontext-) Informationen über die Forschungsdaten. Die Daten selbst können dann für wissenschaftliche Nutzungen beantragt werden (s. oben).

Eine breite Such- und Auffindbarkeit der Datensätze wird über Metadaten sichergestellt, die sich an internationalen Metadatenstandards orientieren und die Daten möglichst detailliert beschreiben. Das von Qualiservice verwendete Metadatenschema wurde von Betancort Cabrera und Haake (2014) entworfen und wird fortlaufend weiterentwickelt. Es basiert auf dem in den Sozialwissenschaften international gängigen Metadatenstandard der »Data Documentation Initiative« (DDI), der derzeit in Version DDI 3.2 vorliegt. Zugleich werden die Arbeit der DDI Alliance Qualitative Data Model Working Group und weitere Metadatenschemata wie Dublin Core, DCAT und der REFI-QDA Standard berücksichtigt, um die Austauschbarkeit (Interoperabilität) der Metadaten zu optimieren. Über die beständige Fortentwicklung dieser Metadaten sind alle bei Qualiservice archivierten Datensätze international auf unterschiedlichen Plattformen les- und auffindbar.

Da Metadaten vorrangig der Auffindbarkeit von Datenkollektionen dienen und mittels eines möglichst standardisierten/kontrollierten Vokabulars Daten beschreiben, sind in der Regel weitere Informationen erforderlich, um die Eignung der Forschungsdaten für die jeweilige Nachnutzung beurteilen zu können. Hierzu dient insbesondere der Studienreport, in dem die Datengebenden das Forschungsprojekt, die Datengenerierung und die weitere Bearbeitung der Daten sowie Nachnutzungsmöglichkeiten erläutern. Dieser Studienreport wird über den Dokumentenserver E-LIB der Staats- und Universitätsbibliothek Bremen (SuUB) veröffentlicht und mit einem persistenten Identifikator (DOI) versehen. Er stellt somit eine wichtige Methodenpublikation des Forschungsprojekts dar.

Für Datengebende bietet der Studienreport die Möglichkeit, den institutionellen, konzeptionellen, methodologischen und situativen Kontext der Datenproduktion so umfassend darzustellen, wie es ihnen sinnvoll erscheint. Auf diese Publikation lässt sich dann verweisen, wenn an anderer Stelle der Projektverlauf und die methodologischen Entscheidungen nicht so umfangreich dargestellt werden können, wie dies angesichts der Offenheit und Flexibilität qualitativer Forschung oftmals wünschenswert bzw. erforderlich wäre.

Qualiservice unterstützt Forschende bei der Erstellung des Studienreports mit individueller Beratung, mit Workshops zur Forschungsdokumentation und Kontextualisierung sowie mit einer Handreichung (Heuer et al. 2020). Darüber hinaus können Forschende alle weiteren Kontextmaterialien, die sie für das Verständnis der Forschungsdaten als relevant erachten,⁵ bei Qualiservice archivieren und für Nachnutzungen bereitstellen. Qualiservice behandelt diese Kontextmaterialien (mit Ausnahme des Studienreports) wie die Forschungsdaten selbst und führt entsprechende datenschutzrechtliche und technische Prüfungen und gegebenenfalls Ergänzungen durch.

Generell empfiehlt Qualiservice, die Dokumentation des Forschungskontextes als Bestandteil des Forschungsprozesses aufzufassen und so weit wie möglich in diesen zu integrieren. Dies hat nicht nur den Vorteil, dass der Aufwand für die Kontextualisierung geringer ist als bei einer Rekonstruktion des gesamten Forschungsprozesses im Nachhinein, sondern ermöglicht auch eine weitaus genauere Darstellung des Forschungsverlaufs inklusive der in der qualitativen Forschung häufigen Modifikationen von Teilen des Forschungsdesigns. Zudem profitiert nach unserer Erfahrung auch das Primärforschungsprojekt selbst von einer solchen fortlaufenden Dokumentation, da Informationen besser zwischen Personen und über längere Zeiträume hinweg festgehalten werden können; im Idealfall unterstützt die Dokumentation somit auch die Datenauswertung und die (Selbst-)Reflexion im Forschungsprojekt.

Um die Nachnutzbarkeit der Forschungsdaten zu erhöhen, hat Qualiservice das Konzept der flexiblen Anonymisierung entwickelt (Kretzer 2013) und arbeitet fortlaufend an seiner Umsetzung für textgebundene Forschungsdaten. Kern des Konzepts ist der Vorschlag, anstelle des Löschens Schwärzens oder gar Verfälschens von sensiblen personenbezogenen Daten diese durch (oftmals abstraktere) sozialwissenschaftlich relevante Informationen zu ersetzen. So wird beispielsweise ein Herzinfarkt zu einer koronaren Herzkrankheit abstrahiert. Mit dem von Qualiservice entwickelten und kürzlich publizierten Anonymisierungstool »QualiAnon« lässt sich dieses Konzept für alle Arten von Texten umsetzen, indem eine flexible Anonymisierung/Pseudonymisierung ermöglicht wird. So können für die Sekundärnutzung erforderliche Informationen »aufgeklappt« werden, während für den jeweiligen Forschungszweck weniger wichtige Informationen abstrahiert bzw. »zugeklappt« werden können und eine Identifikation der

⁵ Dies können zum Beispiel Projektanträge, Berichte, Leitfäden, Transkriptionsregeln, Codierschemata, Memos, Fallbeschreibungen etc. sein.

betroffenen Personen vermieden wird. Das Tool ist nicht nur für die Kuration in Forschungsdatenzentren, sondern auch für die Datenvorbereitung durch die Forschenden konzipiert und ist über Qualiservice verfügbar.

Data Sharing im Austausch mit der Community entwickeln

Qualiservice hat das Teilen von qualitativen Forschungsdaten von Anfang an als eine kollaborative Aufgabe verstanden, die nur in enger Anbindung an und in ständigem Austausch mit den Forschenden und den Fachcommunities möglich ist. Mehr noch: Qualiservice versteht sich als ein Forschungsdatenzentrum von Forschenden, mit Forschenden und für Forschende. So wurde Qualiservice von Forschenden initiiert und wird bis heute von forschungsaktiven Sozialwissenschaftler:innen geleitet und repräsentiert. Die Bedarfe von Forschenden und Fachcommunities spielten von Beginn an eine zentrale Rolle und wurden wie oben skizziert schon 2003 bis 2005 mit der ersten »Machbarkeitsstudie« für ein qualitatives Forschungsdatenzentrum erfasst. Auf Basis der Ergebnisse dieser Befragung wurden dann die Basiskomponenten für ein FDZ konzipiert und in ständigem Austausch mit den Forschenden weiterentwickelt.

Bis heute werden nicht nur Informationsmaterialien und Beratungsangebote, sondern auch die Workflows zur Datenarchivierung und Bereitstellung in »Use Studies«, das heißt in Zusammenarbeit mit Forschenden entwickelt, erprobt und optimiert. Auch bei der Archivierung und Bereitstellung von Daten aus einzelnen Projekten, sind es stets die Forschenden, die – in Abstimmung mit Qualiservice – die zentralen Entscheidungen treffen: von der Auswahl der Forschungsdaten und Kontextmaterialien über die Datenzugänge bis hin zu den Voraussetzungen für das Teilen bzw. die Nachnutzung. Es ist in unseren Augen gerade diese Flexibilität und Forschungsorientierung, die zur Akzeptanz von Qualiservice und zur Bereitschaft zum Teilen von qualitativen Forschungsdaten beiträgt. Im Gegenzug sind es oftmals Impulse aus der Kooperation mit Forschenden, die zur Weiterentwicklung des Angebots von Qualiservice führen.

Die zentrale Rolle des Austauschs mit den Forschenden und Fachcommunities bedeutet aber auch, dass Qualiservice bei vielen das Data Sharing von qualitativen Forschungsdaten betreffenden Fragen auf Rückmeldungen aus und Diskussionen in den Fachcommunities angewiesen ist, um die qualitative Forschung bestmöglich zu unterstützen und befördern. So sind in vielen wichtigen Fragen neben den Forschungsdateninfrastrukturen auch

die Forscher:innen selbst und weitere wissenschaftspolitische Akteure gefordert, sich an der Diskussion über Standards, Best Practices und Empfehlungen zu beteiligen. Dies betrifft etwa die wichtige Frage der Eignung von Daten für das Data Sharing und die damit zusammenhängende Frage einer möglichen Auswahl von (geeigneten) Datensätzen. Während Einigkeit darüber besteht, dass sich bestimmte Forschungsdaten nicht oder nur sehr eingeschränkt für das Teilen eignen, wäre darüber hinaus innerhalb der Fachcommunities zu diskutieren, ob eine Auswahl unter den für das Data Sharing angebotenen Daten getroffen werden sollte und welche Kriterien gegebenenfalls in Bezug auf die Daten bzw. ihre Nachnutzbarkeit angelegt werden sollten.

Perspektiven im Rahmen der Nationalen Forschungsdateninfrastruktur

Forschungsdateninfrastrukturen können sich nur langfristig etablieren, wenn sie in enger Anbindung zu den Forschenden und den wissenschaftlichen Communities arbeiten. Dies ist auch eine zentrale Anforderung an die derzeit im Entstehen begriffene Nationale Forschungsdateninfrastruktur. Im Rahmen des NFDI-Konsortiums für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftsdaten KonsortSWD (vgl. Hollstein et al. 2021) koordiniert Qualiservice seit Anfang des Jahres den Aufbau eines Verbunds für Qualitative Forschungsdaten (QualidataNet). Beteiligt sind neben Qualiservice das FDZ Bildung am DIPF, das FDZ Betriebs- und Organisationsdaten am DIW Berlin, das FDZ-DZHW für die Hochschul- und Wissenschaftsforschung und das Archiv für Gesprochenes Deutsch.

Etabliert werden soll eine auf Nutzer:innen und Services ausgerichtete, nachhaltige vernetzte Infrastruktur für qualitative Forschungsdaten. Diese soll die bisherige heterogene, fragmentierte und für Forschende wenig transparente Archivlandschaft ersetzen und den Forschenden eine zentrale Anlaufstelle bieten, die sie beim Management und Teilen ihrer Forschungsdaten bestmöglich unterstützt. Der *single point of entry* soll einen deutlich vereinfachten Überblick über die Datenbestände und Informations- und Beratungsangebote bieten und die Forschenden – Primärforscher:innen wie Sekundärnutzer:innen – mit dem jeweils für sie am besten geeigneten FDZ und dessen fachlicher bzw. methodischer Expertise zusammenbringen. Diese Struktur soll sukzessive ausgebaut werden, indem neben den vom RatSWD akkreditierten FDZ weitere Archive und Bestände qualitativer Daten inkludiert werden.

Für den Erfolg dieser interdisziplinär ausgerichteten Vernetzungsstruktur wird ein wichtiger Faktor sein, wie gut es gelingt, die Bedarfe der Forschungscommunities aufzunehmen und die Angebote daran auszurichten. Nicht zuletzt sind hierbei auch Wissenschaftspolitik und Forschungsförderorganisationen gefordert, neue Wege bei der – ideellen wie finanziellen – Förderung des Datenteilens zu beschreiten.

Jan-Ocko Heuer, Betina Hollstein und Kati Mozygemba

Literatur

- Betancort Cabrera, Noemi / Haake, Elmar 2014: Das Qualiservice Metadatenchema: Version 1.1. Qualiservice Technical Reports, 2014/01. <http://nbn-resolving.de/urn:nbn:de:gbv:46-00103643-13>.
- Corti, Louise 2018: 20 Years of Archiving and Sharing Qualitative Data in the UK. In Rat für Sozial- und Wirtschaftsdaten (Hg.), Archivierung und Zugang zu qualitativen Daten. RatSWD Working Paper 267/2018. Berlin: RatSWD, 14–25.
- DGS-Sektionen Biographieforschung und Qualitative Methoden der empirischen Sozialforschung 2014: Resolution zur Archivierung und Sekundäranalyse von Daten der Sektionen für Biographieforschung und für Methoden der Qualitativen Sozialforschung der DGS. https://soziologie.de/fileadmin/sektionen/biographieforschung/datenarchivierung/Resolution_Datenarchivierung.pdf, letzter Aufruf am 31. Juli 2021.
- FID Soziologie 2018: Evaluationsbericht – Bedarfserhebung 2018, https://sociohubfid.de/p/Bedarfserhebung_2018, letzter Aufruf am 31. Juli 2021.
- Heuer, Jan-Ocko / Kretzer, Susanne / Mozygemba, Kati / Huber, Elisabeth / Hollstein, Betina 2020: Kontextualisierung qualitativer Forschungsdaten für die Nachnutzung: eine Handreichung für Forschende zur Erstellung eines Studienreports. Qualiservice Working Paper 1-2020. <https://doi.org/10.26092/elib/166>.
- Hollstein, Betina / Miller, Bernhard / Siegers, Pascal / Wolf, Christof 2021: KonsortSWD: Vom Netzwerk zur integrierten Dateninfrastruktur der Gesellschaftsforschung. Bausteine Forschungsdatenmanagement, 2. Jg., Heft 2, 10–22. <https://doi.org/10.17192/bfdm.2021.2.8330>.
- Huber, Elisabeth / Imeri, Sabine 2021: Informed Consent in Ethnographic Research: A Common Practice Facing New Challenges. Qualiservice Working Paper (in Vorbereitung).
- Huschka, Denis / Knoblauch, Hubert / Oellers, Claudia / Solga, Heike (Hg.) 2013: Forschungsinfrastrukturen für die qualitative Sozialforschung. Berlin: Scivero Verlag.

- Knoblauch, Hubert 2013: Einige Anforderungen an Forschungsinfrastrukturen aus der Sicht der qualitativen Forschung. In Denis Huschka / Hubert Knoblauch / Claudia Oellers / Heike Solga (Hg.), *Forschungsinfrastrukturen für die qualitative Sozialforschung*. Berlin: Scivero Verlag, 27–32.
- Knoblauch, Hubert / Solga, Heike 2011: Thesen zur Handhabung quantitativer und qualitativer Daten in Forschungsinfrastrukturen der Sozialwissenschaften: Ist eine Integration möglich? RatSWD Working Paper 190/2011. www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_190.pdf, Berlin: RatSWD.
- Kretzer, Susanne 2013: Arbeitspapier zur Konzeptentwicklung der Anonymisierung/Pseudonymisierung in Qualiservice. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-47605-2>.
- Kretzer, Suanne / Mozygamba, Kati / Heuer, Jan-Ocko / Huber, Elisabeth 2020: Erläuterungen zur Verwendung der von Qualiservice bereitgestellten Vorlagen für die informierte Einwilligung. Qualiservice Working Paper 2-2020. Bremen. <https://doi.org/10.26092/elib/192>.
- Medjedović, Irena 2014: *Qualitative Sekundäranalyse: Zum Potenzial einer neuen Forschungsstrategie in der empirischen Sozialforschung*. Wiesbaden: Springer VS.
- Medjedović, Irena / Witzel, Andreas 2010: Wiederverwendung qualitativer Daten: Archivierung und Sekundärnutzung qualitativer Interviewtranskripte. Wiesbaden: VS.
- Opitz, Diane / Mauer, Rainer 2005: Erfahrungen mit der Sekundärnutzung von qualitativem Datenmaterial – Erste Ergebnisse einer schriftlichen Befragung im Rahmen der Machbarkeitsstudie zur Archivierung und Sekundärnutzung qualitativer Interviewdaten. *Forum Qualitative Sozialforschung*, 6. Jg., Ausgabe 1, Art. 43. doi.org/10.17169/fqs-6.1.510.
- RatSWD 2015: Stellungnahme des RatSWD zur Archivierung und Sekundärnutzung von Daten der qualitativen Sozialforschung. Berlin: RatSWD. www.ratswd.de/dl/RatSWD_Stellungnahme_QualiDaten.pdf, letzter Aufruf am 31. Juli 2021.
- RatSWD (Hg.) 2018: Archivierung und Zugang zu qualitativen Daten. RatSWD Working Paper 267/2018. Berlin: RatSWD. <https://doi.org/10.17620/02671.35>.
- von Unger, Hella 2021: Ethical Reflexivity as Research Practice. *Historical Social Research*, 46. Jg., Heft 2, Art. 9.
- Wilke, René / Proßbrock, Willi / Pach, Helen 2019: Infrastrukturen für Forschungsdaten in der qualitativen Sozialforschung: Überblick und aktuelle Herausforderungen. *SOZIOLOGIE*, 48. Jg., Heft 4, 467–486.
- Witzel, Andreas 2000: »Archiv für Lebenslaufforschung« des Sfb 186 »Statuspassagen und Risikolagen im Lebensverlauf« an der Universität Bremen. *Forum Qualitative Sozialforschung*, 1. Jg., Ausgabe 3, Art. 28. doi.org/10.17169/fqs-1.3.1045.