



DGS / KWI / GOETHESTRASSE 31 / 45128 ESSEN

**DGS DEUTSCHE
GESELLSCHAFT
FÜR SOZIOLOGIE**

**VORSITZENDE · BEAUFTRAGTE FÜR
ÖFFENTLICHKEITSARBEIT
PROF. DR. NICOLE BURZAN
KULTURWISSENSCHAFTLICHES INSTITUT
GOETHESTRASSE 31 · D-45128 ESSEN
NICOLE.BURZAN@KWI-NRW.DE**

**GESCHÄFTSSTELLENLEITERIN
DR. SONJA SCHNITZLER
TEL +49 (0) 201 / 72 04 208
FAX +49 (0) 201 / 72 04 111
SONJA.SCHNITZLER@KWI-NRW.DE**

Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie

Stellungnahme des Vorstands und Konzils der DGS¹, Stand: 08.01.2019

1. Einleitung

Seit längerem wird die Bereitstellung und Nachnutzung von Forschungsdaten durch deutsche Wissenschaftsorganisationen, darunter die Deutsche Forschungsgemeinschaft (DFG 2015) und der Rat für Sozial- und Wirtschaftsdaten (RatSWD 2015), gefordert, gefördert und diskutiert. Es wurden Empfehlungen und Richtlinien zum Umgang mit Forschungsdaten verabschiedet², und Drittmittelgebende legen Grundsätze für die Archivierung und Bereitstellung von Forschungsdaten zugrunde, die von Forscherinnen und Forschern, die Primärdaten generieren, schon zum Zeitpunkt der Antragstellung zu konkretisieren sind. Teilweise wird bei Anträgen zur Finanzierung der Nachweis einer Daten- und nicht nur einer Forschungslücke verlangt.

¹ Wir danken den Sektionsvorständen und beteiligten DGS-Mitgliedern (u.a. Betina Hollstein, Wolfgang Ludwig-Mayerhofer, Jörg Strübing) für ihre Mitarbeit und Anregungen.

² S. http://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/. Weiterhin wurden u.a. folgende Texte einbezogen: Corti et al. 2005; Graduate School of Social Sciences der Universität Bremen/Zentralarchiv für Umfrageforschung der Universität Köln 2005; Stanat 2014; DGS-Sektionen Biographieforschung und Methoden der Qualitativen Sozialforschung 2014; Schönbrodt et al. 2017; DGfE 2017; von Unger 2018; RatSWD 2018a/b.

Zugleich werden in der Wissenschaft Nutzen und Probleme des Forschungsdatenmanagements diskutiert. Die Vielzahl von Forschungsweisen und Zielkonflikte in der Umsetzung wissenschaftlicher Arbeit geben Anlass für intensive Diskussionen und eine Verständigung über sachkundige Standards. Zum jetzigen Stand der Diskussion lassen sich vorab vier Punkte formulieren:

1. Nicht alle wissenschaftlich erhobenen Forschungsdaten können einer Nachnutzung zugänglich gemacht werden.
2. Die Auswahl der Forschungsgegenstände und der Forschungsfragen darf sich nicht an einer Nachnutzung von Daten ausrichten.
3. Rechtliche, forschungsethische, den Aufwand und den gegenstandsadäquaten Ertrag abwägende sowie organisatorisch-technische Überlegungen sind zu berücksichtigen bei der Frage, welche Daten und weiteren Informationen wann, für wen, zu welchem Zweck und in welcher Form bereitgestellt werden können und sollen.
4. Dort, wo umfängliche Bereitstellungen und Nachnutzungen von Forschungsdaten sinnvoll sind, sind die personellen und technischen Rahmenbedingungen zu verbessern und weitere Fachkompetenz aufzubauen.

Angesichts der Vielgestaltigkeit von Forschungskonzepten, Method(ologi)en und Daten in der soziologischen Forschung dienen die folgenden Ausführungen als ein Zwischenschritt, um fachspezifische Grundsätze für den Umgang mit Forschungsdaten weiterzuentwickeln. Die DGS macht damit auf den relevanten Punkt aufmerksam, dass heterogene Forschungsansätze (und damit Typen von Forschungsdaten) zu differenzierten Beurteilungen von Möglichkeiten der Bereitstellung und Nachnutzung von Forschungsdaten führen müssen.

Im Folgenden werden Vorteile und Herausforderungen der Bereitstellung und Nachnutzung von Forschungsdaten thematisiert, wobei auch grundsätzlich bestehende Zielkonflikte, Grenzen der Nutzung sowie Förder- und Diskussionsbedarfe deutlich werden.

2. Forschungsdatenmanagement: Nutzen, Herausforderungen, Aufgaben

Die Bereitstellung und Nutzung von Forschungsdaten über den Primärkontext der Erhebung hinaus ist in vielen Fällen mit Vorteilen verbunden. So fördern Sekundäranalysen die Transparenz des Forschungsprozesses, indem über Auswertungsverfahren und Interpretationen auf einer gemeinsamen Datengrundlage diskutiert und indem teilweise Metaanalysen oder Replikationen vorgenommen werden können, was über neue inhaltliche Erkenntnisse hinaus nicht zuletzt der Qualitätssicherung empirischer Forschung dienen kann. Nachnutzungen ermöglichen weiterhin die Analyse von (sozialen) Veränderungen bis hin zur Bearbeitung von Problemstellungen, die zum Erhebungszeitpunkt noch nicht absehbar waren bzw. nicht im Vordergrund standen. Auf diese Weise können ebenfalls Kooperationen zwischen Fächern bzw. Forschungsdisziplinen unterstützt werden. Zusätzliche Vorteile bestehen darin, dass eine intensive Nutzung bereits erhobener Daten nicht allein zeit- und kostensparend für die Forschenden ist, sondern auch auf der Seite der Forschungsteilnehmenden den Aufwand und die Belastungen reduziert. Schließlich können Sekundärdaten auch in der Lehre mit Gewinn eingesetzt werden.

Zugleich ist die wissenschaftliche Nachnutzung von Daten – wie die Forschungspraxis zeigt – mit erheblichen Herausforderungen in Form von Restriktionen sowie Ziel- und Interessenkonflikten verbunden. Dies kann bedeuten, dass eine Nachnutzung von Forschungsdaten nicht möglich ist.

Ein wichtiger Punkt betrifft die *Durchführbarkeit der Forschung in Verbindung mit Schutzbedürfnissen und -rechten* der an der Forschung Beteiligten. Gesetzlich festgeschriebene Datenschutzrechte (z.B. von Befragten) etwa führen dazu, dass entweder eine Einwilligung der betreffenden Personen vorliegen muss, damit personenbezogene Daten weitergegeben werden können, oder dass eine umfassende Anonymisierung der Daten vorzunehmen ist (vgl. RatSWD 2017). Je nach der Art der von Forschenden generierten Daten kann dies mehr oder weniger einfach oder aber

unmöglich sein. Vergleichsweise unproblematisch sind etwa Zusammenstellungen öffentlich zugänglicher Daten und in vielen Fällen auch standardisierte Surveydaten. Insbesondere bei Projekten, deren Hauptzweck in der Bereitstellung von Daten für die Wissenschaftsgemeinschaft besteht (in Deutschland z.B. SOEP, NEPS oder pairfam) bestehen aufwändige, aber routinierte Verfahren der Anonymisierung und Bereitstellung. Für Daten im Zusammenhang z.B. mit biographischen Interviews, ethnographischen Feldprotokollen oder audiovisuellen Aufzeichnungen im privaten oder semiöffentlichen Raum sind die Möglichkeiten, Grenzen und ein potenzieller Schaden differenzierter zu beurteilen und gegeneinander abzuwägen.

Ausschlaggebend ist, dass der Feldzugang durch Kriterien der Nachnutzung nicht beeinträchtigt werden darf und – genereller – Forschungen und Forschungsfragen nicht mit Blick auf eine etwaige Nachnutzung einem Druck zur Modifizierung unterliegen dürfen. So kann bei bestimmten Forschungssettings die Einwilligung von Forschungspersonen davon abhängen, dass ihnen Vertraulichkeit innerhalb des Projektkontexts und die Nicht-Weitergabe personenbezogener Daten garantiert werden (wenn die Betroffenen z.B. Sanktionen fürchten, etwa als hierarchisch untergeordnete Arbeitnehmer/innen, bei abweichendem Verhalten, in Bezug auf Nachteile gegenüber Konkurrent/innen etc.). Bereits ein Unbehagen der Beteiligten kann hier zu einer deutlich sinkenden Teilnahmebereitschaft führen. Die Garantie von Vertraulichkeit ist also in zahlreichen – gerade als ‚heikel‘ geltenden oder sogenannte vulnerable Gruppierungen einbeziehenden – Forschungskontexten essentiell. Eine informierte Einwilligung entbindet Forschende zudem nicht von ihrer forschungsethischen Verantwortung, so dass sie nur innerhalb der gegebenen Rahmenbedingungen Möglichkeiten der Nachnutzung von Daten bestmöglich ausloten können. Entsprechend darf die Sekundärnutzung von Daten nicht als ‚Normalmodell‘ von Forschung gelten und darf das Ausmaß der Eignung zu erhebender Daten für eine solche Sekundärnutzung die Entscheidung über die Bewilligung von Forschungsmitteln *nicht* beeinflussen.

Um bei einer Bereitstellung von Daten einen Schutz vor missbräuchlichem Zugriff zu gewährleisten, unterscheidet etwa der RatSWD (2018b: 8) drei Varianten der Archivierung: die Sicherung von Daten in der Forschungseinrichtung der Primärforschenden oder in einer Einrichtung der Forschungsinfrastruktur a) ohne Zugriffs- oder Nutzungsrechte durch Dritte, b) mit Datenzugang für Dritte vor Ort und c) mit ortsunabhängigem Datenzugang (z.B. in Form von scientific oder public use files). Im Falle der Varianten b) und c) ist dabei der hohe Aufwand für die Anonymisierung, Aufbereitung und Dokumentation (nicht zuletzt vor dem Hintergrund unbekannter späterer Nutzungskontexte) abzuwägen. Überdies gilt es zu reflektieren, inwiefern Veränderungen von rechtlichen Grundlagen, politischen Haltungen und generell Machtverhältnissen zu Einschränkungen der langfristigen Rechteabsicherung an der Forschung Beteiligter führen können.

Damit besteht eine weitere Herausforderung darin, den *Aufwand*, die Kosten und die Möglichkeiten der Erfassung von weitergabefähigen Daten sowie von weiteren Informationen (z.B. Metadaten) gegenüber dem erwarteten *Nutzen* abzuwägen. Dies betrifft nicht allein eher organisatorisch-technische, personelle und finanzielle Aspekte (nicht zuletzt z.B. bei Qualifikationsarbeiten Einzelner oder bei der Aufteilung von Zeitbudgets in Projekten: der Aufwand für ein auf die Weitergabe von Daten ausgerichtetes Forschungsdatenmanagement bindet Zeit, die nicht mehr z.B. für die Primärforschung und für das Verfassen entsprechender Publikationen verfügbar ist), für die die Mittel für den Aufbereitungsaufwand i.w.S. von Forschungsförderinstitutionen zur Verfügung gestellt werden müssten. Es betrifft auch grundsätzlich die Frage, inwiefern die Daten (z.B. anonymisiert oder nicht, mit oder ohne spezifische Kontextdaten) überhaupt eine *Basis für sinnvolle Sekundäranalysen* sein können (vgl. z.B. Witzel et al. 2008; Huschka et al. 2013; Hirschauer 2014). Zugespitzt: Je gründlicher die Anonymisierung bestimmter Datensorten erfolgt, desto unmöglicher wird ihre angemessene rekonstruktive Interpretation. Anonymisierte Daten etwa könnten ohne entsprechende Kontextinformationen (die dem Datenschutz

unterliegen, s.o.) u.U. nicht sinnvoll außerhalb ihres ursprünglichen Forschungszusammenhangs auswertbar sein. Einer Verpflichtung, umfängliche Kontextinformationen zu liefern, steht über Aufwands- und Datenschutzaspekte hinaus die Problematik gegenüber, dass fließende, eine saubere Trennung häufig verhindernde Übergänge bestehen zwischen von den Forschenden konstruierten ‚Daten‘ und ‚Kontext‘ bzw. zwischen ‚Kontextinformationen‘ und zentralen forschersischen (Interpretations-)Leistungen (z.B. Memos, Konstruktion neuer Variablen). Auch hier gilt, dass die Frage danach, welche Kontextinformationen in welcher Form sinnvoll weitergegeben werden können, differenziert je nach Forschungskonzept und Datentypen zu beantworten ist. Auch für die sekundäranalytische quantitative Forschung wird dabei im Übrigen diskutiert (RatSWD 2018b: 9-10), inwiefern die Forschenden ihre datenmodifizierenden Arbeiten und abgeleiteten Daten wiederum Dritten zur Verfügung stellen sollen (und wer dies ggf. entscheidet bzw. kontrolliert).

Nicht zu vergessen ist zudem, dass Publikationen zum Forschungsprojekt ein wesentliches Instrument der Nachnutzung von Daten darstellen, die durch Infrastrukturen der Datennachnutzung nicht von basalen Zeige- und Nachweispflichten entbunden werden dürfen.

Im Zusammenhang mit unscharfen Grenzen zwischen Datenweitergabe und der Weitergabe forschersischer Leistungen unterschiedlicher Art sind auch *Rechte der Primärforschenden* (z.B. Urheber-, Leistungsschutz- oder Nutzungsrechte) zu klären. So ist zu diskutieren, wer über Nutzungsrechte zu welchem Zeitpunkt entscheiden kann und wie entsprechende Zitationen³ zur Wahrung der Rechte von Primärforschenden aussehen müssten (vgl. Kuschel 2018). Diese Frage stellt sich z.B. bei ausführlichen Feldnotizen und Beobachtungsprotokollen – d.h. auch in Forschungsprozessen, in denen Datenerhebungen und -analysen nicht strikt

³ s.a. RatSWD 2015: 6; demzufolge wird die genaue Art der Kompensation für die Überlassung von Nutzungsrechten, etwa durch Zitation des Eigentümers, weltweit diskutiert.

voneinander getrennt sind – anders dar als z.B. bei standardisierten Befragungsdaten. Von einer Motivation dazu, innovative Primärerhebungen durchzuführen, kann z.B. eher dann ausgegangen werden, wenn Forschende hier Entscheidungsrechte haben und einschlägige Publikationen vorlegen können, bevor sie Daten ggf. an die Wissenschaftsgemeinschaft weitergeben. Angesichts beispielweise z.T. längerer Zeiträume zwischen einer ersten Manuskripterstellung und dem Zeitpunkt der Veröffentlichung in begutachteten Fachjournalen kann hier nicht von einem standardisierbaren Zeitrahmen ausgegangen werden. Je nach Anlage des Forschungsprojekts kann eine adäquate Bereitstellungsfrist also sehr unterschiedlich ausfallen. Die Primärforschenden haben jedenfalls das Recht, die Bedingungen für die Nachnutzung einzuschränken (vgl. gemeinsame Resolution der Sektionen Biographieforschung und Methoden der Qualitativen Sozialforschung zur Archivierung und Sekundärnutzung von Daten, 2014). Forderungen z.B. von forschungsfinanzierenden Organisationen oder Publikationsorganen dürfen dieses Recht nicht relativieren.

Schließlich bestehen auch ungeklärte *technische und organisatorische Herausforderungen und Unterstützungsbedarfe*. So müssen Forschungsdatenzentren in Forschungseinrichtungen, Hochschulen und Universitäten verlässliche Infrastrukturen anbieten, um während des Forschungsprozesses eine sichere Erhebung, Haltung (z.B. Dateiformate, Server, Auffindbarkeit durch DOI etc. betreffend) und gegebenenfalls Übermittlung von Daten sowie im Anschluss nachhaltig sichere Archivierung (u.a. auch im Sinne des Schutzes vor Diebstahl) zu gewährleisten. Hierfür ist nicht nur eine geschützte Hardware erforderlich. Zugleich ist personell und organisatorisch sicherzustellen, dass eine langfristige Haltung, gegebenenfalls eine Weitergabe sowie die Löschung von Daten in verantwortlicher und fachlich kompetenter Weise gewährleistet sind, etwa auch, wenn mit der Erhebung und Kodierung befasste Personen die Organisation verlassen haben oder wenn sich Lizenzbedingungen verändern. An dieser Stelle sind damit erneut rechtliche Fragen zu

klären. Die Einrichtungen der Datenschutzbeauftragten der jeweiligen Organisationen sind in die Lage zu versetzen, den Forschenden in allen Schritten des Forschungsprozesses als zuverlässige, kompetente und rasch verfügbare Partner zur Seite stehen, von der Beratung hinsichtlich des Persönlichkeitsschutzes für die Informantinnen und Informanten oder des Urheber-/Leistungsschutzrechts (etwa bei der Erhebung medialer Daten) über die Formulierung von Einwilligungserklärungen bis hin zu Regeln für sicheren Datentransfer bzw. der Datenspeicherung.

Wichtig ist im Zusammenhang mit einer organisatorisch-technischen Infrastruktur auch die entsprechende Ausstattung von Forschungsdatenzentren als kompetente Partner in Fragen der Speicherung, Archivierung und Bereitstellung, die beispielsweise Recherchertools weiterentwickeln oder die die Verknüpfung zwischen Datensätzen/Methodenberichten und auf dieser Basis veröffentlichter Literatur optimieren. Forschende sollten frühzeitig Informationsangebote über die Möglichkeiten und Grenzen rund um die Bereitstellung und Nachnutzung von Forschungsdaten (z.B. sichere Archivierung, Umgang mit sensiblen Informationen) erhalten. Dies schließt auch Informationen über Datenschutzbestimmungen, Anonymisierungsanforderungen, Praxisbeispiele des Einholens der informierten Einwilligung etc. ein, die allerdings die Verantwortung der Forschenden für ein forschungsethisch angemessenes Handeln über alle Forschungsphasen hinweg nicht einschränken. Auch hierbei können Kompetenzen in Forschungsdatenzentren genutzt werden, deren Austausch mit interdisziplinären und internationalen Partnern zu unterstützen ist. Wichtig ist die wissenschaftliche Professionalität der bereitstellenden Organisationen sowie deren wirtschaftliche und ideologische Unabhängigkeit. Der Datenzugang in Einrichtungen der Forschungsdateninfrastruktur muss dabei – bei entsprechenden Möglichkeiten der Zugangsbeschränkung – kostenlos und langfristig gesichert sein.

3. Schluss

Die Soziologie als Reflexionswissenschaft nutzt ihre Stärken, um Möglichkeiten und Herausforderungen des Forschungsdatenmanagements nicht zuletzt fachspezifisch zu

hinterfragen. Angesichts der Vielfalt empirischer Zugänge kann in der Soziologie nicht von einem ‚Normalfall‘ des Forschungsdatenmanagements gesprochen werden, sondern es sind stets je nach Forschungsrichtung und Datentypen differenzierte Beurteilungen vorzunehmen. Eine grundsätzliche Aufgabe des wissenschaftlichen Diskurses besteht dabei bereits darin zu reflektieren, was unter den Begriff der von Forschenden konstruierten „Forschungsdaten“ subsumiert wird – mit den daraus folgenden angesprochenen Herausforderungen.⁴

Dort, wo sinnvolle Möglichkeiten einer Nachnutzung von Primärdaten bestehen, unterstützt die DGS ausdrücklich den Ausbau notwendiger Infrastrukturen, die Transparenz guter Praxisbeispiele (u.a. ‚Leuchtturmprojekte‘) und weitere Maßnahmen, um die genannten Vorteile von Nachnutzungen bestmöglich zu optimieren und auszubauen. Dort, wo eine begründete Skepsis gegenüber der Bereitstellung und Weitergabe von Daten besteht, setzt die DGS sich dafür ein, dass Primärforschenden, die bestimmte Daten nicht öffentlich zur Verfügung stellen, keine Nachteile (z.B. hinsichtlich der Bewilligung von Forschungsgeldern) entstehen und dass die Durchführbarkeit von forschungsethisch verantwortlichen Forschungen erhalten bleibt.

Literatur

Corti, Louise; Witzel, Andreas; Bishop, Libby (Hrsg.) (2005): Schwerpunkttausgabe „Sekundäranalyse qualitativer Daten“. Forum Qualitative Sozialforschung 6 (1).

Deutsche Forschungsgemeinschaft (DFG) (2015): Leitlinien zum Umgang mit Forschungsdaten. Bonn.

DGfE (2017): Stellungnahme der DGfE zur Archivierung, Bereitstellung und Nachnutzung qualitativer Forschungsdaten in der Erziehungswissenschaft. o.O.

⁴ Beispielsweise könnten dazugehören: sogenannte „Rohdaten“, etwa in Form von Audio-/Videoaufzeichnungen oder Screenshots, „aufbereitete Daten“, z.B. in Form von Transkriptionen oder Datenmatrizen aus standardisierten Befragungen, oder „Auswertungsdaten“, z.B. Regressionsberechnungen, Fallinterpretationen, Memos etc., wobei sich diese Formen von ‚Daten‘ mehr oder weniger gut voneinander unterscheiden lassen.

DGS-Sektionen Biographieforschung und Methoden der Qualitativen Sozialforschung (2014): Resolution zur Archivierung und Sekundärnutzung von Daten der Sektionen für Biographieforschung und für Methoden der Qualitativen Sozialforschung der DGS. o.O.

Graduate School of Social Sciences an der Universität Bremen; Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln (2005): Archivierung und Sekundärnutzung qualitativer Interviewdaten – eine Machbarkeitsstudie. Abschlussbericht. DFG-Geschäftszeichen SCHU 348/8-1 und -2.

Hirschauer, Stefan (2014): Sinn im Archiv? Zum Verhältnis von Nutzen, Kosten und Risiken der Datenarchivierung. In: *Soziologie* 43 (3), 300-312.

Huschka, Denis; Knoblauch, Hubert; Oellers, Claudia; Solga, Heike (Hrsg.) (2013): Forschungsinfrastrukturen für die qualitative Sozialforschung. Standards und disziplinspezifische Lösungen. Berlin.

Kuschel, Linda (2018): Wem „gehören“ Forschungsdaten? In: *Forschung & Lehre* 9/2018, 764-766.

Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2015): Stellungnahme des RatSWD zur Archivierung und Sekundärnutzung von Daten der qualitativen Sozialforschung. Berlin.

Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2017): Handreichung Datenschutz. Berlin.

Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2018a): Archivierung und Zugang zu qualitativen Daten, Working Paper Series 267. Berlin.

Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2018b): Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte. Output 3, 2. Auflage. Berlin.

Schönbrodt, Felix; Gollwitzer, Mario; Abele-Brehm, Andrea (2017): Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien. Im Auftrag des DGPs Vorstands (17.09.2016). In: *Psychologische Rundschau* 68 (1), 20-35.

Stanat, Petra (2014): Bereitstellung und Nutzung quantitativer Forschungsdaten in der Bildungsforschung: Memorandum des Fachkollegiums „Erziehungswissenschaft“ der DFG. o.O.

von Unger, Hella (2018): Forschungsethik, digitale Archivierung und biographische Interviews. In: Lutz, Helma; Schiebel, Martina; Tuidier, Elisabeth (Hrsg.): *Handbuch Biographieforschung*. Wiesbaden, 681-693.

Witzel, Andreas; Medjedovic, Irena; Kretzer, Susanne (2008): Sekundäranalyse qualitativer Daten: zum gegenwärtigen Stand einer neuen Forschungsstrategie. In: *Historical Social Research* 33 (3), 10-32.