

267

Archivierung und Zugang zu qualitativen Daten

Doris Bambey, Louise Corti, Michael Diepenbroek,
Wolfgang Dunkel, Heidemarie Hanekop,
Betina Hollstein, Sabine Imeri, Hubert Knoblauch,
Susanne Kretzer, Christian Meier zu Verl,
Christian Meyer, Alexia Meyermann, Maike Porzelt,
Marc Rittberger, Jörg Strübing, Hella von Unger
und René Wilke

Oktober 2018

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienenen empirischen Forschungsarbeiten sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD. Das Bundesministerium für Bildung und Forschung hat die Publikationen nicht beeinflusst.

Die RatSWD Working Paper Series wird herausgegeben von:

seit 2014 Regina T. Riphahn (Vorsitzende des RatSWD)

2009–2014 Gert G. Wagner

2007–2008 Heike Solga

Inhaltsverzeichnis

Archivierung und Zugang zu qualitativen Daten	1
<i>Betina Hollstein und Jörg Strübing</i>	
20 years of archiving and sharing qualitative data in the UK	14
<i>Louise Corti</i>	
Flexible Strategien für eine forschungsfreundliche Archivierung und Nachnutzung qualitativer Forschungsdaten.....	26
<i>Susanne Kretzer und Michael Diepenbroek</i>	
FDZ für qualitative Forschungsdaten in der Arbeits- und Industriesoziologie: das interdisziplinäre Zentrum eLabour.....	36
<i>Wolfgang Dunkel und Heidemarie Hanekop</i>	
Forschungsdateninfrastrukturen für audio-visuelle Daten der Qualitativen Sozialforschung – Bedarf und Anforderungen.....	47
<i>Hubert Knoblauch und René Wilke</i>	
Bereitstellung und Nachnutzung qualitativer Daten in der Bildungsforschung. Das Forschungsdatenzentrum (FDZ) Bildung am DIPF	59
<i>Doris Bambey, Alexia Meyermann, Maike Porzelt und Marc Rittberger</i>	
Archivierung und Verantwortung. Zum Stand der Debatte über den Umgang mit Forschungsdaten in den ethnologischen Fächern.....	69
<i>Sabine Imeri</i>	
Probleme der Archivierung und sekundären Nutzung ethnografischer Daten	80
<i>Christian Meier zu Verl und Christian Meyer</i>	
Archivierung und Nachnutzung qualitativer Daten aus forschungsethischer Perspektive	91
<i>Hella von Unger</i>	
Zentrale Ergebnisse und Empfehlungen aus dem Workshop für Archivierung und Sekundäranalyse qualitativer Forschungsdaten.....	101
<i>Betina Hollstein und Jörg Strübing</i>	

Archivierung und Zugang zu qualitativen Daten

Betina Hollstein und Jörg Strübing

1. Einleitung

Der vorliegende Band dokumentiert die Beiträge und Ergebnisse eines vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) und dem Datenservicezentrum Qualiservice am 27. und 28. April 2018 an der Universität Bremen veranstalteten Workshops zum Thema „Archivierung und Zugang zu qualitativen Daten“.¹ Ziel des Workshops mit mehr als 50 Forschenden verschiedener sozial- und geisteswissenschaftlicher Disziplinen sowie Vertreterinnen und Vertretern von Forschungsdatenzentren war es, bisherige Erfahrungen zu evaluieren und Anforderungen für eine verbesserte Archivierungsinfrastruktur in der qualitativen Forschung zu benennen. Es ist bislang nicht selbstverständlich, dass Daten aus der qualitativen Sozialforschung außerhalb der Entstehungskontexte archiviert und in kuratierter Form für externe Sekundärnutzungen in Lehre und Forschung bereitgestellt werden. Während Archivierung und Sekundäranalysen quantitativer Daten in den Sozialwissenschaften seit vielen Jahrzehnten fest etabliert sind, ist die Diskussion um die Archivierung und Bereitstellung qualitativer Daten deutlich jünger. In Großbritannien gibt es mit *Qualidata* als Teil des UK Data Archives in Essex seit 1994 ein zentrales Archiv für Forschungsdaten, das ganz dezidiert auch qualitatives Material kuratiert, archiviert und für Sekundärnutzungen in Forschung und Lehre bereitstellt (Corti, in diesem Band). In Deutschland sind entsprechende institutionalisierte Archivierungsformen für qualitative Daten jedoch noch im Aufbau. Die Aktivitäten des Bremer Archivs für Lebenslaufforschung (ALLF, später Qualiservice; Kretzer und Diepenbroek, in diesem Band) und des RatSWD haben hierbei wichtige Impulse gesetzt. So untersuchte das Archiv für Lebenslaufforschung zusammen mit der GESIS bereits Anfang der 2000er Jahre die Bedarfe für ein Servicezentrum für qualitative Interviewdaten (Opitz und Mauer 2005, Medjedović und Witzel 2010, Medjedović 2014), schob gemeinsam mit *Qualidata* erste Studien zu Sekundäranalysen von qualitativen Daten an (Corti/Witzel/Bishop 2005) und koordinierte die ersten Vernetzungsaktivitäten qualitativer Archive in Europa (Rasmussen 2010/2011). Daneben trugen insbesondere die Aktivitäten des RatSWD und Initiativen des Wissenschaftsrates (WR) und der Deutschen Forschungsgemeinschaft (DFG) für eine verbesserte Forschungsdateninfrastruktur² dazu bei, dass die Diskussion um Möglichkeiten und Grenzen einer sekundäranalytischen Verwendung qualitativer Forschungsdaten und die dafür erforderliche Archivinfrastruktur auf breiterer Basis geführt wurde (Kluge und Opitz

¹ Unser Dank gilt den Mitarbeiterinnen und Mitarbeitern der Geschäftsstelle des RatSWD sowie -für die lokale Organisation des Workshops- dem Team des Datenservicezentrums QualiService an der Universität Bremen.

² Etwa die von WR und DFG veranstaltete Tagung zu „Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. Stellenwert – Förderung – Zukunftsperspektiven“ in Bonn, 9. Dezember 2011. Vgl. auch die dazu entwickelten Empfehlungen (Wissenschaftsrat 2011).

1999; Mruck et al. 2000; Corti/Witzel/Bishop 2005; Knoblauch und Solga 2011; Hirschauer 2014; RatSWD 2015).³

Insbesondere Wissenschaftsorganisationen und Forschungsförderer, aber auch Fachzeitschriften formulieren zunehmend die Erwartung, auch qualitative Daten zu archivieren und wenn möglich für Nachnutzungen, z. B. im Rahmen von Sekundäranalysen zur Verfügung zu stellen. Diese Forderungen sind zum einen forschungsökonomisch begründet, wenn es darum geht, die meist aufwändig erhobenen Daten für weitere Auswertungen verfügbar zu machen, teure Doppelerhebungen zu vermeiden und einer Überforschung des Feldes zu begegnen. Hinzu kommt, dass im Rahmen der üblichen Projektlaufzeiten Daten häufig unterausgewertet bleiben. Daneben sind auch Überlegungen zu Transparenz, Qualitätssicherung und -kontrolle Motive für die Archivierungs- und Datenkurationsvorhaben. Schließlich besteht die Hoffnung, dass Datenarchive und Sekundäranalysen neue Forschungsperspektiven eröffnen können, etwa in Hinblick auf Themen und Felder, übergreifende Vergleiche, historische Untersuchungen, Metaanalysen oder langfristige Längsschnittstudien.

In weiten Teilen der standardisierten Sozialforschung, insbesondere in der Survey- und Panelforschung, ist all dies bereits lang etablierte Praxis. Standardisierte Forschungsdaten werden nicht nur häufig extern in Forschungsdatenzentren kuratiert, es hat sich über die Zeit auch eine typische Arbeitsteilung ergeben: Einige zentrale Institutionen wie das Zentrum für Umfragen, Methoden und Analysen (ZUMA), das Deutsche Institut für Wirtschaftsforschung (DIW Berlin) oder National Educational Panel Study (NEPS) erheben selbst regelmäßig sozialwissenschaftliche Daten, etwa das sozio-ökonomische Panel (SOEP am DIW Berlin) oder die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS). Das jedoch nicht, um sie vorrangig selbst auszuwerten, sondern um Forschenden unterschiedlicher thematischer und institutioneller Herkunft diese Daten für eigene Projekte zur Verfügung zu stellen. Umgekehrt gibt es eine wachsende Zahl von Forschenden, z. B. an den Universitäten, die selbst keine eigenen Daten mehr erheben oder dies nur noch ergänzend tun und sich stattdessen auf zunehmend verfeinerte und komplexere Auswertungen zentral verfügbarer Survey- und Paneldaten spezialisieren.

Während in der standardisierten Forschung Sekundäranalysen und die Kultur des *data sharing* weit verbreitet sind und die eigenständige Generierung projektspezifischer Daten nicht die Regel, sondern eher die Ausnahme ist, sind Erhebung und Auswertung im qualitativen Forschungsprozess häufig eng miteinander verzahnt. Qualitativ Forschende gehen meist mit großer Selbstverständlichkeit davon aus, dass sie ihr Datenmaterial selbst generieren und sie konstitutiver Bestandteil im funktionalen Zusammenhang dieses Geneseprozesses sind. Die Orientierung an Gegenstandsangemessenheit, empirischer Sättigung und theoretischer Durchdringung (Hollstein 2011; Strübing et al. 2018) bedeutet für qualitativ Forschende in der Regel, Daten zum konkreten Forschungsgegenstand in der jeweils aktuellen Forschungssituation neu und an die konkrete Fragestellung angepasst zu generieren.

Archivierung und Sekundärnutzung sind in der qualitativen Sozialforschung also keine verbreitete Praxis; vielmehr orientiert sich letztere forschungspraktisch wie institutionell an der Eigenlogik ihrer Methoden und Methodologien. Da diese im heterogenen Feld qualitativer Forschung sehr unterschiedlich sind, stellen sich hier auch die Archivierungs- und Nachnutzungsperspektiven sehr unterschiedlich dar. So sind etwa im Bereich der Lebenslauf- und Bildungsforschung Längsschnittstudien auf der Basis qualitativer Interviews eine wichtige Erkenntnisquelle und die Vernetzung

³ S.a. Resolution zur Archivierung und Sekundäranalyse von Daten der Sektionen für Biographieforschung und für Methoden der Qualitativen Sozialforschung der DGS, o.J. (ca. 2013; nachfolgend als „Resolution Biografieforschung“ zitiert; https://www.sozioologie.de/fileadmin/user_upload/Sektionen/Biographieforschung/Resolution_Datenarchivierung.pdf; Zugriff am 11.09.2018).

entsprechender Archive ist vergleichsweise weit fortgeschritten (Rasmussen 2010/2011). Anders sieht es aus, wenn es um ethnographisches Material (Beobachtungsnotizen, Feldtagebücher etc.) geht, dem die Perspektive der Primärforschenden sehr viel stärker „eingeschrieben“ ist und dessen sekundäranalytische Nachnutzung ohne die nicht vollumfänglich explizierbaren Erfahrungen der Forschenden im Feld nicht gut gelingen kann (Meier zu Verl und Meyer, in diesem Band). Es gibt hier also kein ‚one size fits all‘. Qualitative Daten stellen an die Archivierung, Datenkuration und mögliche Sekundärnutzung große Anforderungen. Auf diese möchten wir im Folgenden kurz eingehen.

2. Was sind qualitative Daten?

Jede Entwicklung einer Dateninfrastruktur muss am jeweils relevanten Datenbegriff ansetzen. Die Resolution der DGS-Sektionen für Biografieforschung sowie für Methoden der Qualitativen Sozialforschung bemerkt dazu: „Die Forderung nach einer generalisierten Datenarchivierung auch für qualitative Daten setzt einen Begriff des Datums (als dekontextualisierbares Sinnquantum vor jeder Analyse) voraus, der für das Gros der qualitativen Sozialforschung unzutreffend ist“ (Resolution Biografieforschung, S. 1). Die als qualitative Daten bezeichneten Forschungsmaterialien zeichnen sich zunächst einmal durch eine mehrdimensionale Heterogenität aus: Sie sind *in ihrer Materialität* heterogen insofern sie z. B. als digitale Textdateien, digitale oder analoge Audio- oder Videoaufnahmen, materiale Sammlungen von Artefakten oder auch – vor allem in der Ethnographie – als körperlich-sensorische Erfahrungen und deren z. B. handschriftliche Aufzeichnung in Forschungstagebüchern auftreten (Meier zu Verl und Meyer, in diesem Band). Damit ist schon eine zweite Dimension angesprochen, der *Datentypus*: Transkribierte Leitfadeninterviews, narrative Interviews, Expertinnen- und Expertengespräche oder Gruppendiskussionen werden schon in der Primärforschung unterschiedlich bearbeitet. Sie unterscheiden sich aber noch einmal deutlich von auditiven und bildlichen Materialien oder den diversen im Prozess der Feldforschung wahrgenommenen und gesammelten Eindrücken und Informationen. Unterschiedlich ist dabei auch die Art und Intensität der *Involviertheit der Forschenden* in die Daten: Aufgezeichnete Tischgespräche, im Feld gesammelte Fotografien oder Videos sind wesentlich weniger abhängig vom Zutun der Forschenden als moderierte Gruppendiskussionen oder qualitative Interviews. Entsprechend variiert damit auch das Ausmaß, in dem Forschende und die Forschungssituation ein Teil jenes *Kontextes* sind, der für kompetente Anschlussverwendungen von Forschungsdaten vorhanden, also dokumentiert und archiviert sein muss. Für Archivierungsfragen ebenfalls relevant ist die Frage der *prozessualen Einbindung* von Forschungsdaten: Welche Fälle, welche Daten und Datenarten für ein jeweiliges Projekt wichtig sind, entscheidet sich, streng am Gütekriterium der „Gegenstandsangemessenheit“ orientiert (Hollstein 2011; Strübing et al. 2018), häufig erst im Projektverlauf. Datengewinnung und -analyse sind in vielen qualitativen Projekten nicht distinkte Phasen, sondern im Rahmen iterativ-zyklischer Forschungsdesigns ineinander verschränkt. Und schließlich entstehen diese Daten in unterschiedlich sensiblen und vulnerablen Feldern (von Unger, in diesem Band): Bei Expertinnen- und Expertengesprächen mit öffentlichen Personen ist ein *informed consent* auch für Sekundärnutzungen häufig unproblematisch, während ethnographische Feldnotizen oder Videodaten in kleinkriminellen Milieus hochgradig sensible Materialien sind, die außerhalb der von wechselseitigem Vertrauen geprägten Beziehung zwischen Forschenden und Teilnehmenden ein hohes Schädigungspotential haben.

3. Archivierung und Sekundäranalyse: Chancen und Risiken

3.1. Nachfrage und Bedarf

Zweifelloos kann die qualitative Forschung von einer projektunabhängigen, auf Dauer gestellten und professionell betriebenen Archivierungsstruktur profitieren. Allein schon zur Gewährleistung der Aufbewahrungsfristen ist es sinnvoll, die von Dynamik und Fluktuation geprägten lokalen Organisationsstrukturen projektförmiger Forschung zu verlassen und die Daten in ein stabiles, den Anforderungen des modernen Datenschutzes genügendes Umfeld einzubringen, in dem auch technische Medienwechsel über längere Zeiträume mitvollzogen werden. Vorliegende Befragungen zeigen, dass auch viele Sozialforscherinnen und -forscher die Archivierung und Nachnutzung ihrer Daten befürworten: In einer DFG-finanzierten Machbarkeitsstudie untersuchten das Archiv für Lebenslaufforschung (ALLF) und die GESIS die Bedarfe für ein Servicezentrum für qualitative Interviewdaten (Opitz und Mauer 2005, Medjedović und Witzel 2010, Medjedović 2014). Hierbei sprachen sich 80 Prozent der mit Interviewmaterial arbeitenden Sozialforscherinnen -und forscher für den Aufbau einer entsprechenden Infrastruktur aus. 60 Prozent der Untersuchungsteilnehmerinnen und Untersuchungsteilnehmer⁴ erklärten sich zur Übergabe ihrer Studien an ein Datenzentrum bereit. Über ein Drittel der Befragten hat selbst schon Sekundäranalysen durchgeführt, dabei mehrheitlich aber auf eigene Daten zurückgegriffen (Opitz und Mauer 2005). Zwei Drittel der Befragten können sich vorstellen, bei einem entsprechenden Angebot qualitative Daten sekundär zu nutzen (ebenda). Auch eine kürzlich vom Fachinformationsdienst Kultur- und Sozialanthropologie durchgeführte Umfrage konstatiert unter den befragten Ethnologinnen und Ethnologen „eine grundsätzliche Zustimmung zu Formen der Recherchierbarkeit und Nachnutzbarkeit von Forschungsdaten“ (Imeri 2017: 171; Imeri, in diesem Band). Festzuhalten ist allerdings, dass das Spektrum der Einschätzungen zu Fragen von Forschungsdatenmanagement hier besonders breit ist und auch stark zwischen Datentypen variiert. Im *UK Data Archive*, das in den letzten 15 Jahren etwa 1.000 qualitative Studien archiviert hat, wurde über die Hälfte der Studien für Sekundäranalysen abgerufen. Dabei dominieren Lehr- und Lernzwecke und Nachnutzungen im Rahmen von Qualifikationsarbeiten (Bishop und Kuula-Luomi 2017). Insofern kann erwartet werden, dass die Nachfrage nach Daten aus qualitativ-empirischer Forschung – variierend nach Disziplinen, Forschungsfeldern, Datentypen und Methodentraditionen – zunehmen wird, insbesondere dann, wenn ein entsprechendes Angebot vorhanden und gut erschlossen ist.

3.2. Datenkuration

Für die Datenkuration stellen qualitative Daten allerdings eine besondere Herausforderung dar. Dabei berührt die Archivierung und Sekundärnutzung kaum oder nur wenig standardisierter Daten nicht nur datenschutzrechtliche Fragen (RatSWD 2015, Gebel et al. 2015). Da in qualitativen Erhebungen zahlreiche Informationen generiert werden, die eine Re-Identifizierung von Personen oder Organisationen ermöglichen, reicht eine formale Anonymisierung, bei der nur direkte Identifikationsmerkmale wie Namen und Adressen entfernt werden, in aller Regel nicht aus. Zugleich sind bestimmte Maßnahmen, wie Schwärzen und Maskieren problematisch. Ersteres ist mit Datenverlust verbunden; letzteres verfälscht die Daten. Aus diesem Grund wird zumeist pseudonymisiert. Eine solche (reversible) Pseudonymisierung ist allerdings aufwändig, nur begrenzt automatisierbar (Kretzer 2013a, Kretzer und Diepenbroek, in diesem Band) und bei bestimmten Datentypen (Bild, Video) eventuell gar nicht möglich. Dabei besteht bei allen diesen Maßnahmen eine grundsätzliche Spannung zwischen

⁴ Angeschrieben wurden alle Sozialwissenschaftlerinnen und Sozialwissenschaftler, die zwischen 1994 und 2003 qualitatives Interviewmaterial erhoben haben (insgesamt 1750 Projekte). Von den 1.104 Projektleitenden nahmen knapp 40 Prozent (430) an der Befragung teil.

der Anonymisierung der Daten und dem Forschungsinteresse, das sich auf gerade diese Daten bezieht. Eine besondere Herausforderung bezogen auf die Bereitstellung und Nachnutzbarkeit qualitativer Daten ergibt sich aus der Kontextsensitivität: Qualitative Methoden zielen – auf unterschiedliche Weise – auf methodisch kontrolliertes Sinnverstehen (Hollstein und Ullrich 2003). Es geht um das Verstehen von Handlungen bzw. Äußerungen in ihrem Bedeutungszusammenhang, das heißt in ihrem Kontext. Sekundärnutzerinnen und Sekundärnutzer, die an der Datenerhebung nicht beteiligt waren, fehlt dieses Kontextwissen aus „erster Hand“ (Medjedović 2014: 217). Zu unterscheiden ist dabei u. a. zwischen dem kommunikativen oder situativen Kontext des Interviews, dem Projektkontext, dem lokalen Kontext und dem Makro-Kontext (ebenda). Zur Erhaltung des Nachnutzungspotentials müssen die Kontexte der Studien und der erhobenen Daten ausführlich dokumentiert werden (Kretzer 2013b). Für die Nachnutzungspotentiale sind ausführliche Kontextdaten essentiell, wobei zu beachten ist, dass Kontexte in Abhängigkeit vom Forschungsansatz und Methodologie unterschiedlich definiert werden und je nach Auswertungsinteresse unterschiedlich relevant sind (Medjedović 2014).

In der sozialwissenschaftlichen Fachöffentlichkeit wurden Möglichkeit und Nutzen der Wiederverwendung qualitativer Daten in Forschung und Lehre in den letzten Jahren verstärkt diskutiert. Die Sektionen Methoden der Qualitativen Sozialforschung und Biografieforschung der Deutschen Gesellschaft für Soziologie (DGS)⁵ sowie die Neue Gesellschaft für Psychologie (NGfP)⁶ haben hierzu eine Reihe von Bedingungen für die Archivierung und Nachnutzung qualitativer Daten formuliert: Hierzu gehören neben dem informierten Einverständnis der Studienteilnehmenden und der Anonymisierung der Daten auch die Forderung nach der Freiwilligkeit der Datenüberlassung, die Notwendigkeit der Kontextualisierung, d. h. der Bereitstellung von Kontextdaten für die Sekundärnutzung, sowie Kontrollmöglichkeiten der Primärforschenden bzgl. der Bedingungen der Nachnutzung. Die in der Machbarkeitsstudie befragten Forschenden nennen als wichtigste Voraussetzungen für die Sekundärnutzung von Interviewdaten die umfassende Dokumentation des Forschungsprozesses und eine gute Datentranskription (Medjedović 2014). Als wichtigste Bedingungen für die Datenüberlassung werden die Wahrung der Anonymität der interviewten Personen und die datenschutzrechtliche Zulässigkeit der Weitergabe genannt (ebenda). Da zur Wahrung der Anonymität die Entfernung direkter Identifikatoren nicht ausreicht, handelt es sich nicht nur um eine rechtliche, sondern auch um eine forschungsethische Frage. Auf die Diskussion in der Community reagierend, hat der RatSWD eine Empfehlung verabschiedet, die darauf verzichtet, qualitativ Forschende auf die Freigabe ihrer Daten zu verpflichten (RatSWD 2015). Der RatSWD empfiehlt die sichere Archivierung qualitativer Daten sowie die Etablierung einer Kultur des *data sharings* und der Sekundärnutzung (ebenda). Festzuhalten ist, dass bei der Archivierung und der Bereitstellung qualitativer Daten Fragen des Datenschutzes und der Datensicherheit neben der Sicherung der Primärforschung einen herausragenden Stellenwert haben. Gleichzeitig können entsprechende Maßnahmen zur Folge haben, dass die Nutzbarkeit für Sekundäranalysen eingeschränkt wird (Medjedović und Witzel 2010, Medjedović 2014, Hirschauer 2014). Wie stark diese Einschränkungen sind, hängt nicht zuletzt von der Art der Daten und der Art und Fragestellung der Sekundärnutzung ab.

⁵ http://www.sozioologie.de/fileadmin/user_upload/Sektionen/Biographieforschung/Rundbrief-66-1.pdf (Zugriff am 10.10.2018).

⁶ <http://www.ngfp.de/2009/12/ngfp-befurwortet-die-einrichtung-eines-servicezentrums-fur-qualitative-daten/> (Zugriff am 10.10.2018).

3.3. Sekundäranalysen

Ein Vorteil der Archivierung qualitativer Daten, das betonen Knoblauch und Wilke in ihrem Beitrag, liegt zweifelslos in der Möglichkeit, Datensätze weitgehend vollständig zugänglich zu machen (in diesem Band: 47–58). Dies ist gerade bei qualitativen Studien ein Plus, weil die *Belegfunktion* von empirischen Daten für theoretische Argumente an qualitativen Daten nur im Rahmen umfangreicher Publikationen erfüllt werden kann. Bei den üblichen Formaten für Zeitschriftenaufsätze lässt sich die empirische Fundierung der vorgetragenen Thesen meist nur exemplarisch zeigen. In Datenarchiven hingegen ließe sich diese Fundierung in ihrer Breite und in ihren Differenzierungen nachverfolgen. Ob allerdings *Replikationen*, wie sie in der standardisierten Forschung üblich sind, in der qualitativen Forschung möglich bzw. überhaupt wünschenswert sind, ist fraglich (Strübing 2018). Denn was die Daten bedeuten, was sie über die zu untersuchenden Gegenstandsbereiche zu sagen haben, erfordert im Grundsatz den Nachvollzug des iterativ-zirkulären Forschungsprozesses und der in der Primärforschung entwickelten Theorie-Empirie-Verhältnisse. Umgekehrt besteht bei weniger reflektiertem Umgang mit den archivierten Daten die Gefahr einer Entwertung dieser Qualität qualitativer Forschung und die Daten würden reduziert auf die Funktion eines Fensters zur Realität.

Wenn es Sekundäranalysen aber nicht primär um den kritischen Nachvollzug der Primärforschung geht, sondern um neue Projektzuschnitte, die neue Fragen an das Datenmaterial stellen und den archivierten Daten neue Funktionen zuweisen, kann erwartet werden, dass mit Daten, die sorgfältig kuratiert und durch hochwertige Metadaten und Kontextbeschreibungen erschlossen werden, einiges gewonnen werden kann. So lassen sich je nach Material etwa Vorher-Nachher-Vergleiche durchführen, bei denen die Archivdaten kontrastiv auf aktuelles Datenmaterial bezogen werden können. Oder es lassen sich in eine vergleichende Querschnittsperspektive nicht mehr nur Studien, sondern auch deren Daten einbeziehen, die ähnliches in anderen Kontexten untersucht haben. Mittlerweile gibt es für Interviewdaten bereits einige exemplarische Sekundäranalysen, auch im deutschsprachigen Raum (u. a. Heinz/Wachtveitl/Witzel et al. 1986/87, Kühn und Witzel 2004, Gläser und Laudel 2008, Fischer und Witzel 2008, Behrmann und Hollstein 2012, Beckmann et al. 2013), sowie insbesondere auch Studien, die die Möglichkeiten, Schwierigkeiten und Probleme von Sekundäranalysen qualitativer Daten systematisch untersuchen. Zu nennen sind etwa die Beiträge in Special Issues des Forums Qualitative Sozialforschung (2005, Vol. 6, No. 1)⁷ und des Journals Historical Social Research (2008, Vol. 33, No. 3)⁸. Unstrittig besteht hier aber nach wie vor ein großer Bedarf, insbesondere auch an reflektierten Praxisbeispielen und beispielhaften Studien („best practice“), die zeigen, wie der methodologisch angemessene Gebrauch dieses Materials, auch in Kombination mit Primärdaten einer gemischt operierenden Studie, aussehen kann.

3.4. Primärforschung

Eine Frage, die bei allen Bemühungen um die Etablierung von Infrastrukturen für Archivierung und Sekundäranalyse qualitativer Daten beantwortet werden muss, ist die nach den Konsequenzen für die Primärforschung. Anders als in vielen Forschungsdesigns der standardisierten Forschung gewinnt die qualitative Forschung ihre Ergebnisse in den meisten Fällen aus intensiven Interaktionen mit dem zu untersuchenden Feld (Ausnahmen sind hier teilweise die konversations- und die diskursanalytische Forschung, die auch mit existierenden Materialkorpora arbeiten). Viele Interviewgespräche und Gruppendiskussionen, teilnehmende Beobachtungen oder ethnographische Feldaufenthalte stehen und fallen mit der Kompetenz der Forschenden, eine Vertrauensbeziehung zu den Teilnehmenden zu

⁷ <http://www.qualitative-research.net/index.php/fqs/issue/view/13> (Zugriff am 11.10.2018).

⁸ <https://www.gesis.org/hsr/volltext-archiv/2008/333-qualitative-data/> (Zugriff am 11.10.2018).

etablieren. Dies gilt umso mehr, wenn prekäre, deviante oder politisch aufgeladene Milieus erforscht werden. Dort Vertrauen zu gewinnen ist oft ein mühsamer und langwieriger Prozess. Vertrauensbeziehungen zum Feld sind das Kapital qualitativer Sozialforschung und die Aushandlung des Feldzugangs nimmt in dieser Art von Forschung nicht nur breiten Raum ein, sie ist auch ein nicht zu unterschätzender Teil des Datenmaterials.

Aus Gründen der Forschungsethik und des Datenschutzes werden die Teilnehmenden wie in jeder empirischen Sozialforschung mit dem *informed consent* über die Vorgehensweise, die Forschungsziele und die geplante Verwendung des Datenmaterials aufgeklärt und ihr ausdrückliches und freiwilliges Einverständnis eingeholt und dokumentiert. Wenn Archivierung und Nachnutzung Teile der Datenverwendung sein sollen, müssen diese ebenfalls gegenüber den Teilnehmenden benannt werden. Gerade in hochgradig vertrauensbedürftigen Settings, in denen es z. B. um Intimität, Devianz, Scham oder innerorganisatorische Machtverhältnisse geht, besteht aber ein großer Unterschied darin, ob die Daten allein zwischen Teilnehmenden und Primärforschenden zirkulieren oder ob in Aussicht gestellt wird, sie Dritten zur Verfügung zu stellen, die den Teilnehmenden nicht persönlich bekannt sind und die daher auch nicht Teil der Vertrauensbeziehung sein können. Vorbehalte qualitativ forschender Wissenschaftlerinnen und Wissenschaftler gegenüber der Archivierung und Nachnutzung gründen sich zu einem erheblichen Teil auf der Befürchtung, dass mit einem entsprechend erweiterten *informed consent* der Zugang zum Feld erschwert und in manchen Bereichen unmöglich wird (RatSWD 2015, von Unger, in diesem Band).⁹ Mögliche Vorbehalte der Teilnehmenden sind durchaus berechtigt, denn wenn z. B. Informationen über nicht gesetzeskonformes Handeln gut dokumentiert und durch Metadaten erschlossen in institutionalisierten Archiven lagern, sind z. B. staatliche Zugriffe im Rahmen der Strafverfolgung, aber auch andere Formen einer die Teilnehmenden gefährdenden Nachnutzung möglich – und allein die potentielle Möglichkeit kann den Feldzugang ruinieren. Auch Anonymisierung ist hier kein in allen Fällen zuverlässiger Schutz, weil sie bei hinreichender, z. B. kriminalistischer oder auch krimineller Energie gerade bei den sehr kontextreichen qualitativen Daten rückgängig gemacht werden kann. Aus diesem Grund sollten Regeln für die Archivierung und Sekundäranalyse etabliert werden, die die Qualität der Primärforschung und ihr Primat gegenüber nachgelagerten Verwendungsformen des dort produzierten Materials sichern, denn ohne gute Primärforschung kann es auch keine qualitativ hochwertigen Archivbestände und damit auch keine gute Sekundärforschung geben. Ebenso bedarf es eines Primats der Gegenstandangemessenheit der Datengenerierung. Forschungsdaten der Primärforschung sollten nicht mit Blick auf ihre spätere Archivierbarkeit ausgewählt und formatiert werden, sondern sich strikt am Qualitätskriterium der Gegenstandangemessenheit orientieren. Erst aus den so gewonnenen Daten lassen sich Materialien für zu bestimmende Arten von Nachnutzungen auswählen. Primärforschende müssen also in die Entscheidungsprozesse über Datenaufbereitung und -weitergabe eingebunden werden und weiterhin ihre Verantwortung gegenüber dem von ihnen erforschten Feld wahrnehmen.

3.5. Geistiges Eigentum

Schließlich ist festzuhalten, dass in die Generierung von Daten in der Primärforschung erhebliche Eigenleistungen der Forschenden eingehen. Sie gestalten situativ den Verlauf von Interviewgesprächen und sichern sowohl Spezifik und Tiefgründigkeit des Materials als auch dessen Vergleichbarkeit. Sie kontextuieren das aufgezeichnete Material mit weiteren Beobachtungen. Sie treffen

⁹ Die unseres Wissens nach einzige empirische Studie, die (indirekt) Hinweise zu dieser Frage gibt, ist die Studie von Kuula (2010). Von 169 Interviewpartnerinnen und -partnern, die nachträglich um die Einwilligung zur Sekundärnutzung gebeten wurden, gaben fast alle (165) ihr Einverständnis. Es ist allerdings davon auszugehen, dass die Zustimmung bei sehr sensiblen Themen oder gefährdeten Gruppen geringer ausfällt.

permanent Auswahlentscheidungen und sind, insbesondere in der Ethnografie im Wortsinne, Autorinnen und Autoren der Daten. Auch die Leistungen der Teilnehmenden selbst bei der Datenproduktion sind erheblich. Beides wirft die Frage auf, wie mit diesen Formen geistigen Eigentums umgegangen wird, wenn das Material für Nachnutzungen verfügbar gemacht wird. Ein wichtiger Schritt ist hier die Vergabe von DOIs, zum einen für die Daten(sätze), zum anderen auch für die Studienreports, in denen die Primärforschung dokumentiert wird (Kretzer 2013b). Dies ermöglicht die Zitation der Datensätze und der Erhebungskontexte der Primärforschung. Es macht zugleich die Leistung der Datenproduzierenden sichtbar und soll in der Zuerkennung wissenschaftlicher Reputation münden. Noch ist nicht absehbar, wie gut das funktionieren kann, denn Reputation lässt sich nicht verordnen, sie wird – auf Widerruf – von der *scientific community* gewährt. Hier bedarf es also neben der Etablierung konkreter Verfahren auch einer genaueren Erforschung der Reputationsbildungsprozesse.

4. Die Beiträge in diesem Band

Die Beiträge auf dem Workshop lassen sich grob nach ihrer Zielrichtung unterscheiden: Einige stellen vor allem Erfahrungen mit unterschiedlich weit gediehenen Archivierungsprojekten vor (so etwa Qualidata, Qualiservice, elabour oder das FDZ Bildung am DIPF), während andere vor allem bestimmte Probleme der Archivierung und Sekundärnutzung (z. B. ethische Erwägungen) bzw. Probleme der Archivierung und Sekundärnutzung bestimmter Datentypen (insbesondere ethnografisches Datenmaterial) thematisieren. In ihrer Summe machen sie deutlich, dass Archivierung und Sekundärnutzung sich hier nicht einfach analog zu den Praktiken und Institutionalisierungen der standardisierten Sozialforschung gestalten lassen, sondern nach differenzierten, Datenarten und Forschungsfelder berücksichtigenden, Vorgehensweisen verlangen.

Louise Corti (Essex) berichtet in ihrem Beitrag von Erfahrungen mit über zwanzig Jahren Archivierung und *data sharing* qualitativer Daten im Rahmen des *UK Data Archives*. Sie betont, dass eine gemeinsame formale Archivierungsstruktur für verschiedene Disziplinen möglich und sinnvoll ist, wobei die Erfordernisse unterschiedlicher Datentypen und methodischer Zugänge über die Art der Dokumentation und jeweils adäquate Zugangswege geregelt werden können. Sie erläutert Möglichkeiten der Datenkuration und -dokumentation und gibt einen Überblick über die Nachnutzung qualitativer Daten in Großbritannien und Finnland. Der Beitrag zeigt die Schlüsselrolle von Archiven, nicht nur bei der Beurteilung ethischer und rechtlicher Aspekte von Forschung, sondern auch für die Einschätzung des wissenschaftlichen Wertes und der Nutzbarkeit. Mit Blick auf Infrastrukturen für qualitative Daten betont er die Notwendigkeit der Kooperation aller Beteiligten, gerade auch über disziplinäre und methodologische Grenzen hinweg.

Susanne Kretzer und Michael Diepenbroek (Bremen) stellen das Datenservicezentrum *Qualiservice* an der Universität Bremen vor, das in Deutschland zu den Vorreitern der Archivierung von qualitativen Forschungsdaten gehört. Seit Ende der 1990er Jahre setzt sich *Qualiservice* (damals noch Archiv für Lebenslaufforschung) für die Kultur des *data sharings* ein und entwickelt im Austausch mit den verschiedenen Fachcommunities praktische Lösungen für die Archivierung, Datenkuration und -bereitstellung. Der Schwerpunkt der Arbeit von *Qualiservice* lag bislang auf qualitativen Interviewdaten; in Kooperation mit dem Fachinformationsdienst für Kultur- und Sozialanthropologie werden künftig auch ethnographische Daten mit einbezogen. In ihrem Beitrag erläutern die Autorinnen und Autoren die Arbeitsweise von *Qualiservice* und die flexiblen Strategien der Archivierung und Nachnutzung qualitativer Forschungsdaten.

Wolfgang Dunkel (München) und Heidemarie Hanekop (Göttingen) stellen in ihrem Beitrag mit *eLabour* ein Forschungsdatenzentrum für vorwiegend qualitative Daten vor, das seinen thematischen Schwerpunkt in der Arbeits- und Industriosozilogie hat. Mehrere Forschungszentren, u. a. das SOFI in Göttingen und das ISF in München, an denen über Jahrzehnte in einer Vielzahl empirischer Projekte große Mengen an Daten (vorwiegend Interviewdaten und Protokolle von Betriebsbegehungen und Arbeitsplatzbeobachtungen) produziert wurden, haben *eLabour* mit der Intention gegründet, diese umfangreichen, aber kaum kuratierten und oft auch nicht digitalisierten Daten, in systematischer Weise zu archivieren und für mögliche Nachnutzung zu erschließen. Derzeit ist *eLabour* im Evaluationsprozess, um als Forschungsdatenzentrum beim RatSWD akkreditiert zu werden.

Auch Hubert Knoblauch und René Wilke (Berlin) stellen mit dem DFG-geförderten Projekt *aviDa* eine Initiative für ein Forschungsdatenarchiv vor. In diesem Fall ein Archiv, das sich auf bestimmte Datensorten spezialisieren will: Es geht um audio-visuelle Daten in der qualitativen Sozialforschung, insbesondere Videodaten aus den *workplace studies* und ähnlichen videografischen Interaktionsstudien, die sich noch einmal von anderen nicht ethnographisch gewonnenen Video-Datenarten unterscheiden. In ihrem Beitrag verweisen sie auf die Lücke, die sich in der allmählich entstehenden Archiv-Landschaft bei den nicht text- und sprachorientierten Datenarten zeigt und die sie, bezogen speziell auf im Forschungsprozess von den Forschenden selbst aufgezeichnete Videomaterialien, schließen wollen. Der Beitrag plädiert deutlich für einen Ausbau der Archivierungsinfrastruktur, legt aber Wert auf die Feststellung, dass Archivierung und Freigabe zur Sekundärnutzung auf Freiwilligkeit beruhen müssen, da hier geistiges Eigentum berührt ist.

Thema des Beitrags von Doris Bambey, Alexia Meyermann, Maike Porzelt und Marc Rittberger (Frankfurt) sind die Bereitstellung und Nachnutzung qualitativer Daten in der pluridisziplinären empirischen Bildungsforschung. Nach einer Skizze der aktuellen förderpolitischen Regelungen und des Diskurses in der Fachcommunity, gehen die Autorinnen und Autoren auf den Verbund Forschungsdaten Bildung (VFDB) ein, einer wesentlichen Maßnahme zur Entwicklung einer übergreifenden und koordinierten Forschungsdateninfrastruktur, der vor allem vom Forschungsdatenzentrum (FDZ) Bildung am Deutschen Institut für Pädagogische Forschung (DIPF), GESIS und dem IQB getragen wird. Der Beitrag stellt den Verbund-Workflow zur förderierten Archivierung und Bereitstellung qualitativer und quantitativer Forschungsdaten sowie das am FDZ Bildung praktizierte Vorgehen dar, bezogen auf die qualitativen Daten der Bildungs- und der videobasierten Schul- und Unterrichtsforschung. Abschließend benennen die Autorinnen und Autoren infrastrukturelle und nutzungsbezogene Anforderungen an die Datenzentren, die qualitative Forschungsdaten bereitstellen.

Sabine Imeri (Berlin) gibt in ihrem Beitrag „Archivierung und Verantwortung“ einen Überblick über den Stand der Debatte zum Umgang mit Forschungsdaten in den ethnologischen Fächern. Kontrovers diskutiert werden etwa Formen der Kontextualisierung sowie – angesichts großer Heterogenität der Datenformen – der Auswahl von Daten zur Archivierung. Wichtige Themen sind Fragen des Datenschutzes, von Persönlichkeitsrechten und möglicher Auswirkungen auf die Feldforschung selbst. Für eine breite Akzeptanz von Datenarchivierung und Nachnutzung sind die Anforderungen an Professionalität und Vertrauenswürdigkeit künftiger Datenrepositorien entsprechend hoch: Gefordert sind Lösungen mit ausgeprägter Sensibilität für die Eigenheiten der Wissensproduktion in der ethnografischen Forschung sowie für die spezifische Verknüpfung von methodischen und forschungsethischen Fragen. Bedarf besteht hier an fachlich spezialisierten Repositorien, die möglichst flexible Prozesse der Archivierung und kontrollierte Prozesse der Nachnutzung genauso anbieten können, wie fachspezifische Beratung.

Christian *Meier zu Verl* und Christian *Meyer* (Konstanz) nehmen in ihrem Text vor allem einen Typus von Forschungsdesigns in den Blick; eben die Ethnografie und die dabei entstehenden Konstellationen von Feld, Forschenden und Datenmaterial. Dabei adressieren sie nicht allein forschungspraktische, sondern auch methodologische und epistemologische Fragen und konstatieren der Ethnografie einen Methodenpluralismus, der sehr unterschiedliche Arten objektivierte Wissens hervorbringt, die im Forschungsprozess nur zum Teil in Daten überführt, oft aber von den Forschenden als implizites Orientierungswissen inkorporiert werden. Insbesondere macht die ethnografische Forschungsweise prozedurales und episodisches, häufig prä-reflexives Wissen aus, das nur situativ aktivier- und abrufbar ist. Daraus leiten sie als eher ernüchternde These ab, dass die Archivierung und Sekundärnutzung der in ethnographischen Prozessen generierten wissenschaftlichen Wissensformen nicht gelingen kann – zumindest nicht, solange man nicht „den Körper der Ethnografin mit zu den Akten leg[t]“ (S. 81).

Hella von *Unger* (München) argumentiert in ihrem Beitrag „Archivierung aus forschungsethischer Perspektive“, dass angesichts immer nur unvollständiger Anonymisierungen sowohl von Teilnehmenden als auch von den Forschenden selbst die bei Archivierung und Weitergabe qualitativer Forschungsdaten auftretenden Schädigungsrisiken lediglich begrenzt antizipiert werden können. Weil in Anbetracht des erforderlichen Aufwandes zu Schädigungsminimierung der Aufwand in einem problematischen Verhältnis zum Nutzen steht, seien Archivierung und Nachnutzung forschungsethisches nur für bestimmte Datenarten und bei hinreichender Vorbereitung und Begleitung vertretbar.

5. Schluss

Der Workshop, dessen Ergebnisse in diesem Band dargestellt werden, bestand neben diesen, im Folgenden abgedruckten, Impulsreferaten vor allem aus Diskussionsrunden, die in unterschiedlichen Formaten Voraussetzungen, Bedarfe, Probleme und Lösungsideen für Fragen der Archivierung und Nachnutzung qualitativer Daten erörterten. Die in diesem einleitenden Beitrag nur kurz angerissenen Themen wurden in den Diskussionsrunden intensiv bearbeitet und, gesättigt mit Erfahrungen aus der Forschungs- und Archivierungspraxis, spezifiziert. Den Abschluss bildete eine Podiumsdiskussion, in der ein vorläufiges Fazit sowie Forderungen an Wissenschaftsorganisationen und Forschungspolitik formuliert wurden. Die Ergebnisse des Workshops finden sich am Ende dieses Bandes als knappes Fazit versammelt. Sie sollten künftigen Bemühungen um eine im Rahmen der Nationalen Forschungsdateninfrastruktur (NFDI) zu etablierende Archivierungskultur in der qualitativen Sozialforschung Ziel und Richtung geben.

Literaturverzeichnis

- Beckmann, Sabine; Patrick Ehnis; Thomas Kühn und Marina Mohr (2013): Qualitative Sekundäranalyse – Ein Praxisbericht. In: Denis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die qualitative Sozialforschung*. Berlin, SCIVERO, 137–152.
- Behrmann, Laura und Betina Hollstein (2012): Starthilfe oder Hemmschuh? Arbeitsmarkteinstieg und soziale Netzwerke gering qualifizierter junger Erwachsener. In: Jürgen Mansel und Karsten Speck (Hrsg.): *Jugend und Arbeit. Empirische Bestandsaufnahmen und Analysen*. Weinheim, Beltz Juventa Verlag, 79–99.
- Bishop, Libby und Arja Kuula-Luumi (2017): Revisiting Qualitative Research Data Reuse: A Decade On. *SAGE Open* 7(1).
<http://journals.sagepub.com/doi/10.1177/2158244016685136> (Zugriff am 08.10.2018).
- Fischer, Martin und Andreas Witzel (2008): Zum Zusammenhang von berufsbiografischer Gestaltung und beruflichem Arbeitsprozesswissen. Eine Analyse auf Basis archivierter Daten einer Längsschnittanalyse. In: Martin Fischer und Georg Spöttl (Hrsg.): *Forschungsperspektiven in Facharbeit und Berufsbildung. Strategien und Methoden der Berufsbildungsforschung*. Frankfurt am Main (u.a.), Peter Lang, 24–47.
- Gebel, Tobias; Matthis Grenzer; Julia Kreusch; Stefan Liebig; Heidi Schuster; Ralf Tscherwinka; Oliver Watteler und Andreas Witzel (2015): Verboten ist, was nicht ausdrücklich erlaubt ist: Datenschutz in qualitativen Interviews. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research* 16(2), Art. 27. <http://nbn-resolving.de/urn:nbn:de:0114-fqs1502279> (Zugriff:10.02.2016).
- Gläser, Jochen und Grit Laudel (2008): Creating Competing Constructions by Reanalysing Qualitative Data. *Historical Social Research / Historische Sozialforschung* 33(3), 115–147.
- Heinz, Walter R.; Erich Wachtveitl und Andreas Witzel (1986/1987). *Berufsfindung und Berufsberatung. Eine interpretative Sekundäranalyse. Abschlussbericht an die DFG. Teil I und II*. Universität Bremen.
<http://nbn-resolving.de/urn:nbn:de:gbv:46-ep000103261> / <http://nbn-resolving.de/urn:nbn:de:gbv:46-ep000103279> (Zugriff am 24.10.2018).
- Hollstein, Betina (2011): Qualitative Approaches. In: John Scott und Peter J. Carrington (eds.): *Sage Handbook of Social Network Analysis*. London/New Delhi, SAGE Publications Ltd, 404–417.
- Hollstein, Betina und Carsten G. Ullrich (2003): Einheit trotz Vielfalt? Zum konstitutiven Kern qualitativer Sozialforschung. *Soziologie* 32(4), 29–43.
- Hirschauer, Stefan (2014): Sinn im Archiv? Zum Verhältnis von Nutzen, Kosten und Risiken der Datenarchivierung. *Soziologie* 43(3), 300–312.
<http://publikationen.soziologie.de/index.php/soziologie/article/view/795/1052> (Zugriff am 08.10.2018).
- Imeri, Sabine (2017): Open Data? Zum Umgang mit Forschungsdaten in den ethnologischen Fächern. In: Jonas Kratzke und Vincent Heuveline (Hrsg.): *E-Science-Tage 2017: Forschungsdaten managen*. Heidelberg, heiBOOKS. <http://doi.org/10.11588/heibooks.285.377> (Zugriff am 08.10.2018).
- Kluge, Susann und Diane Opitz (1999): Die Archivierung qualitativer Interviewdaten: Forschungsethik und Datenschutz als Barrieren für Sekundäranalysen? *Soziologie, Mitteilungsblatt der DGS*, Heft 4, 48–63.
- Knoblauch, Hubert und Heike Solga (2011): Thesen zur Handhabung quantitativer und qualitativer Forschungsinfrastrukturen. Ist eine Integration möglich? RatSWD Working Paper 190/2011. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD).
https://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_190.pdf (Zugriff am 08.10.2018).
- Kretzer, Susanne (2013a): Arbeitspapier zur Konzeptentwicklung der Anonymisierungs-/Pseudonymisierung in Qualiservice. <http://nbn-resolving.de/urn:nbn:de:0168-ssor-47605-2> (Zugriff am 08.10.2018).
- Kretzer, Susanne (2013b): Infrastruktur für qualitative Forschungsprimärdaten – Zum Stand des Aufbaus eines Datenmanagementsystems von QualiService. In: Denis Huschka, Hubert Knoblauch, Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die qualitative Sozialforschung*. Berlin, SCIVERO Verlag, 93–110.
https://www.ratswd.de/dl/downloads/forschungsinfrastrukturen_qualitative_sozialforschung.pdf (Zugriff am 08.10.2018).

-
- Kühn, Thomas und Andreas Witzel (2004): Die Arbeitskraftunternehmer-These aus berufsbiografischer Perspektive: In: Hans J. Pongratz und G. Günter Voß (Hrsg.): *Typisch Arbeitskraftunternehmer? Befunde der empirischen Arbeitsforschung*. Berlin, edition sigma, 229–254.
- Kuula, Arja (2010): Methodological and Ethical Dilemmas of Archiving Qualitative Data. *IASSIST Quarterly* 34(3–4) und 35 (1–2), 12–17. http://www.iassistdata.org/sites/default/files/iqvol34_35_kuula.pdf (Zugriff am 08.10.2018).
- Medjedović, Irena (2014): *Qualitative Sekundäranalyse. Zum Potenzial einer neuen Forschungsstrategie in der empirischen Sozialforschung*. Wiesbaden, VS Verlag für Sozialwissenschaften.
- Medjedović, Irena und Andreas Witzel (2010): Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte. Wiesbaden, VS Verlag für Sozialwissenschaften.
- Mruck, Katja; Louise Corti, Susann Kluge und Diane Opitz (2000). Zu diesem Band. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research* 1(3), Art. 1. <http://nbn-resolving.de/urn:nbn:de:0114-fqs000311> (Zugriff am 08.10.2018).
- Opitz, Diane und Reiner Mauer (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 / Forum: Qualitative Social Research* 6(1), Art. 43. <http://nbn-resolving.de/urn:nbn:de:0114-fqs0501431> (Zugriff am 08.10.2018).
- Rasmussen, Karsten Boye (ed.) (2010/2011): Special Issue “Bremen Workshop 2009”. *Qualitative and Qualitative Longitudinal Resources in Europe. Mapping the Field and Exploring Strategies for development*. *IASSIST Quarterly* 34 (3–4) und 35 (1–2). http://www.iassistdata.org/sites/default/files/iqvol34_35.pdf (Zugriff am 08.10.2018).
- RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2015): Stellungnahme des RatSWD zur Archivierung und Sekundärnutzung von Daten der qualitativen Sozialforschung. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). http://www.ratswd.de/dl/RatSWD_Stellungnahme_QualiDaten.pdf (Zugriff am 10.02.2016).
- Strübing, Jörg (2018): Problem, Lösung oder Symptom? Die Forderung nach Replizierbarkeit. *Forschung & Lehre* 25(2), 102–105.
- Strübing, Jörg; Stefan Hirschauer, Ruth Ayaß, Uwe Krähnke, Thomas Scheffer (2018): Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. *Zeitschrift für Soziologie* 47(2), 83–100.
- Wissenschaftsrat (2011): Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. Empfehlung des Wissenschaftsrats vom 28.01.2011, Drucksache 10465-11. Berlin, Wissenschaftsrat (WR). <https://www.wissenschaftsrat.de/index.php?area=&ptyp=3&year=2011&keyword=Forschungsinfrastruktur&suchen=suchen&id=836&rpp=&searchdata=1&L=0> (Zugriff am 08.10.2018).

Betina Hollstein ist Professorin für Mikrosoziologie und Qualitative Methoden an der Universität Bremen. Sie ist Ko-Sprecherin des SOCIUM – Forschungszentrum Ungleichheit und Sozialpolitik, leitet das Datenservicezentrum *Qualiservice* und arbeitet im wissenschaftlichen Beirat des Fachinformationsdienstes Soziologie mit. Zwischen 2014 und 2017 war sie im Vorstand der Deutschen Gesellschaft für Soziologie zuständig für die Forschung.

Jörg Strübing ist Professor an der Universität Tübingen, er forscht und lehrt zu Methoden und Methodologien der qualitativen Sozialforschung. Von 2008–2012 war er Sprecher der Sektion Methoden der qualitativen Sozialforschung in Deutschen Gesellschaft für Soziologie (DGS) und von 2014 bis 2017 Mitglied im RatSWD. Dort vertritt er seitdem als ständiger Gast die Interessen der qualitativen Forschung.

20 years of archiving and sharing qualitative data in the UK

Louise Corti

Abstract

In this presentation I will provide an overview of archiving qualitative data in the UK. The UK Data Archive has been collecting and providing free access to high quality data for the social science research base since 1967, and since then has nurtured a close working relationship with data owners and producers across a range of sectors. Our major UK funder, the Economic and Social Research Council (ESRC), established its Data Policy in 1995, including qualitative research, which led to real success for changing the culture of data sharing in the UK. I briefly show how data are being promoted and re-used and compare the state of qualitative data archiving across Europe. I discuss the art of managing and sharing qualitative data, building on the procedures that we have developed, and illustrate an example of sharing emotionally sensitive data.

1. Introduction

This paper provides a short tour through the 20 plus years of archiving and sharing qualitative data in the UK. Disciplinary definitions of qualitative research and data are considered, and what this might mean for archiving. It outlines what users are doing with archived data and how data can be best prepared to reveal context in the data, and improve usability. The practicalities and challenges of sharing qualitative data are highlighted followed by an example of a ‘sensitive’ qualitative data collection that has been shared. The paper notes strategies used by the UK to successfully bring users to the data and encourage the practice of secondary analysis. It concludes by emphasising the power of collaboration and how important it is that data centres in one country work closely together to provide united services to users.

2. Who is sharing qualitative data?

If we consider whether experienced researchers are worried about sharing their qualitative research data, many will likely say yes. Over my 25-year period of working in this field, I see that this is often down to lack of appreciation of what this means and what is involved, and an assumption that it will expose them, their methods, and possibly their participants to some kind of potential damage or harm. Time and resources are also frequently noted as being especially precious and in competition with new research endeavours. However once we begin to unpick concerns and demonstrate the processes of how data can be, and have been successfully prepared and published, the worry often subsides.

If we look across the disciplines that routinely use qualitative methods, what is captured as *data* varies quite a lot. Disciplines like sociology, anthropology and education tend to prefer interviewing and ethnographic methods, generating data such as interview or focus group recordings and transcripts, diaries, other kind of observations, fieldnotes and kinship diagrams. Political science qualitative methods prefer to focus on text analysis using historical and legal documents, newspaper sources, political manifestoes and records of meetings, for example, in the policy environment or in human rights organisations. Humanities scholars such as historians or linguists often focus on a close examination of the source, using methods of text or audio-visual analysis to discover linguistic features of speech and on verbal communication and create transcribed and annotated text and audiovisual data. Figures 1a–d demonstrate the variety of transcriptions that could be created depending on the disciplinary approach taken. They vary from a basic summary of an interview, in Figure 1a, while Figure 1b shows a more detailed transcription of an interview. Figure 1c shows non-verbal communication that has been recorded, and Figure 1d shows a more detailed orthographic transcription, that captures some of the linguistic features of spoken word including pronunciation.

Figure 1a

Short Summary

This is a short passage from an interview conducted by Stef Scagliola on 15 April 2007 at the Imperial War Museum with major general Julian Thompson on the use of oral history in military history. As curator of the Dutch Veterans Interviewproject initiated by the Dutch Veterans Institute in the Netherlands, she was visiting the Imperial War Museum to do research on the requirements and best practices for a large scale oral history collections related to military topics. Julian Thompson agreed to be interviewed about the benefits and risks of using oral sources for research and on how he used the collections of the IWM for his publications. In the interview he stresses four major points. Firstly, that almost all history starts with oral accounts of events and that the trustworthiness of written accounts is often overestimated, secondly that the author of a book needs to know the subject well in order to distinguish facts from fiction, thirdly that despite these merits, oral history is not regarded as a serious discipline in the academic realm. An other important value, is it social character.

Transcript on the basis of light editing

J.T. - The problem which I recognize, and people here will be the first to admit it, with oral history, there are a number of problems, is that if you interview people who are old, their memories sometimes are bad. Quite often they have had a chance to discuss with other people what went on, so they have come to an agreed story.

Figure 1b

Algemene gegevens interview

Titel	Erfgoed van de Oorlog, Bystander Memories, interview 01
Geïnterviewde	[echte naam]
Interviewer	[echte naam]
Plaats	
Tijd	
Trefwoorden	
Samenvatting	
Transcriptie gemaakt door	
Eigenaar van het bestand	

00:00:00

[image: 00:00:00]

IV: En kropen die mannen dan gewoon in de schuur, verstopten ze zich..

GI: Ja, bovenaan was een scheve kap, en daar kropen ze met een ladder naartoe, moesten wij een ladder daar neer zetten en dan kropen ze er naartoe, en dan werd dat dicht gepakt, en dan de ladder weer weg natuurlijk dan een paar uur zaten ze daar stil, der was er zelfs een uit Eindhoven, die was hier gekomen van Eindhoven naar zijn familie, om ja, hier veiliger te zitten, maar die moest hier ook onderduiken, die moest hier ook wegkruipen voor die razzia die er gehouden werd.

IV: Nelly, jullie hadden ook een appelboomgaard. Wat gebeurde daar mee?

GI: Ja, wij hadden een heel grote aard, appelboomgaard daar en peren natuurlijk. En die hadden wij, die moesten geplukt worden natuurlijk, en die hadden wij verkocht aan de zusters Ursuline van Venray, maar aangezien de mannen niet durfden te komen plukken vanwege die razzia allemaal, hebben de nonnen zelf, de zusters zelf, kwamen de appels plukken,

Figure 1c

I: *eh, eh, ik wilde u eens vragen: wat vindt u van deze werkplek?*
 R: *ja, wat vind ik van deze werkplek... dat is een moeilijke vraag, da da da daar daar moet ik even over nadenken. Ik vind het een hele aardige omgeving, ja, een hele aardige omgeving. Leuke collega's, inhoudelijk interessant werk, enzo. Het gebouw is wat minder, maar ja, je kunt niet alles hebben in het leven, hé?*
 I: *kunt u daar wat dieper op in gaan, wat bedoelt u met leuke collega's?*
 R: *wat ik met leuke collega's bedoel?*

Figure 1d

la 'borea e il fa'vornio ||
 una 'volta ala 'borea 'venne 'volta di 'prender mari'bo || a'ndò dal fa'vornio e 'li 'disse ||
 'vwo(i) 'esere(e) [il] 'mir(o) 'spozzo || il fa'vornio 'er(a) un 'ti'fo atak'ato a(i) hwa'trimi ||
 e 'le 'dome non 'a an'darvan(o) a 'd'gemio || le 'disse || 'no | per'ke non 'a'i ne'ank(e) un 'soldo
 di 'd'òe || la 'borea | 'punta sul 'virvo | si 'miz(e) a 'sof'jare hon 'tute le 'sur(e) 'for'se || sof'jo
 per 'tre 'd'gomi | e 'nevi'ho 'f'ito 'f'ito || 'kwand(o) 'ebe 'fini'bo di 'stender(e) il su(o) ar'd'gento
 (i) 'tomo | 'disse || 'ek'òbi la 'mi(a) 'd'ote | 'tu kie di'f'evi he non 't'je l' 'o || e a'ndò
 a 'ri'fo'sarsi delta fa'òia || il fa'vornio skro'l'o he 'spate | e si 'miz(e) a 'sof'jare 'lui ||
 la hamp'ana e (i) 'monti restar'ono 'sot(o) un 'f'ar'bo 'haldo he 'f'ol'se fin l' 'ult'imo 'f'joko
 di 'neve || la 'borea | 'ri'fo'sa'fasi per 'beme | 'vide he 'di'eta 'd'òe non restar'va 'f'ju
 'nula || dov' e a'nd'òa 'òuta la 'òur(a) 'd'òe || la hant'son'o il fa'vornio || in'soma | mi 'vwo(i)
 an'korra per mari'bo || la 'borea 'i'ris'pose || 'no | nom 'v'ore(i) 'ma(i) 'esere [la] 'òur(a)
 'spozza | per'ke (i) un 'd'gomo 'se(i) ha'f'a'f'e di man'darmi m' 'fumo 'òuta la 'd'òe ||

Source: Scagiola and Calamai (2017) from a workshop on tools for oral history, Arezzo

One thing is clear: a generally similar approach to the formal practices of archiving data can be taken, regardless of the flavour of data. This does not mean putting all research methods and their outputs into one box, but it means that we can utilise a common underlying approach. And, we can think about different levels and types of documentation depending on the likely needs of future users. Again, this does not imply we must fit the description of data into one box, but we can consider different ‘layers’ to help reveal the context of raw data. Finally, we can treat and control data in a variety of ways to make sure we are being legally and ethically compliant, and to ensure that ‘data’ are only shared, where they can be shared using appropriate pathways for access.

3. Qualitative data sharing in the UK: The UK Data Archive

My own organization, the UK Data Archive, is a department at the University of Essex and was established in 1967 by the Economic and Social Research Council (ESRC) as a data bank for social science. The Essex Archive brings with it 50 years of experience in curating and providing access to data and currently directs the national flagship service for data in the UK – the UK Data Service (UKDS). The Service supports the curation, long-term preservation, and access to data and supports the re-use of data for research and teaching. It specialises in social surveys, historical databases and qualitative research data and increasingly, bio-medical studies and data from real-time streaming devices.

The ESRC was an early adopter of data archiving and was the first major research funder to develop and implement an overarching research data policy across the spectrum of social sciences (ESRC 2015). Data Management Plans were later formally required as part of funding. The UKDS works as the organisation to operate the policy, often acting as the ‘Data Police’ when it comes to monitoring and liaising with researchers about data sharing on completion of their funded award. For qualitative researchers, discussions can be challenging, for example, when maybe they forgot to read their contracts or they were unaware that the policy also covers qualitative data! Or they did not seek consent to share outputs, despite setting out plans to share data in their submitted data management plans. Each time researchers are reminded by ESRC about their contractual obligations and to set out in advance in their application and DMP any problems foreseen in sharing data.

The data teams at UKDS have built up years of experience and awareness of the range of issues confronted by researchers and their data regarding the reality of sharing their outputs. Great emphasis is placed on advocacy, support and training for managing and sharing data, to ensure it becomes more familiar across all disciplines. The UK has had almost 25 years to embed principles and everyday practices of data sharing, with the methods literature embracing arguments for and against data sharing

and secondary analysis. For Germany, the state of enculturation for sharing and re-using qualitative data is still in its infancy, but it can learn from the UK, and feel relieved to hear that hostility, confusion and protest are not uncommon emotions to hear as research communities adjust to new practices. Indeed, UK anthropologists sought a formal exemption from the Research Council to be exempted from the ESRC Research Data Policy, but it was (thankfully) turned down.

At the UKDS, the 7000 plus data collections are available for research, learning and teaching and are used across many sectors and by many different disciplines across the world. But, first we must rewind back to the origins of UK qualitative data sharing. *Qualidata* was set up in 1994, following a small grant awarded to Paul Thompson, a Professor of Sociology at Essex University who set up the National Life Story Collection (NLSC) at the British Library, who decided that he wanted to find out the fate of the qualitative data created from well-known qualitative investigations in the UK; and how the investigators felt about sharing these data, if, indeed, they still existed! On conducting a survey and establishing that some great material did still exist, but some famous studies having been destroyed, he bid for a Search and Rescue grant, and was awarded 5 years of funding from the ESRC to catalogue the whereabouts of data arising from qualitative research grants, funded by the ESRC, and to try to rescue them and put them in accessible places.

Ironically, the UK Data Archive, also based at Essex since 1967, was housed literally 300 metres cross the square, but was not interested in dealing with qualitative data at that time. The recruitment in *Qualidata* of a traditional archivist, a survey methodologist and a Historical Sociology Professor, enabled them to come together to develop a fused approach across the survey data publishing, oral history and traditional (paper based) archiving domains to create a new systematic approach to preparing, documenting and cataloguing qualitative data. Guinea pig studies were chosen, representing fieldwork from some of the UK Pioneers of Sociology (Thompson 2017). With *Qualidata* emerged a new culture of preserving and re-using qualitative data by defining approaches and methods for preparing and documenting data. In turn this work spawned a new (critical) literature on secondary analysis of qualitative data from the mid-1990s, encouraging those against sharing data to come out and demand that the qualitative research community also share its concerns. The debate slowly made its way into key methods literature, for example, Silverman (2016). The main concerns are briefly addressed in the following section, and have helped shape and refine robust methods for archiving and curating data. The *Qualidata* approach has been sought as a model for many embryonic archives across the world, and while some have flourished then faltered (often because the message has not been delivered in a convincing way, or that it has not been grasped as business-critical by funders), we should congratulate the Bremen QualiService, especially Andreas Witzel (and colleagues of the UKDS since 1998!), which will be applying for accreditation as a research data centre (RDC) by the RatSWD in the near future.

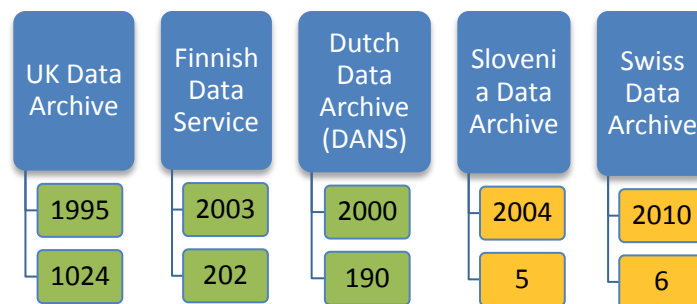
The centre, which reached a maximum of seven staff at its peak in 1998, gained funding until 2000, when money for smaller data initiatives ran out. It did continue its life by joining the UK Data Archive in 2000, when Louise Corti took a new job as Director of User Services, and brought the *Qualidata* unit she has been codirecting with her. It was fully incorporated into business in 2000, along with the separate (funding also running out) History Data Service. A year was spent enculturating these staff in new ways of working and, equally, helping staff in the UK Data Archive to accept this unfamiliar 'new kind' of data, which they did. Where integration is working well at home, then promoting positive messages to the research community is easier. Today, the UKDS has around 60 staff in total, with substantial benefits to sharing infrastructure. While many of the basic archiving processes are shared, such as acquisition and licensing, access pathways and cataloguing and promotion, a small

number of qualitative research specialists deal with evaluating data, ethical issues, preparing and documenting data, and training in re-use of data.

4. The European Landscape

While the UK is very well resourced with 1000+ published qualitative data collections, other European countries lack this resource. The UK funding bodies have continued to support infrastructure for qualitative data, which include the continuing support for the UK Data Service and one-off funding for a longitudinal data repository at the University of Leeds (University of Leeds 2017). In 2018, very few of the main social science data services are actively ingesting qualitative data; while some countries have had smaller pilots as noted above, others have built more lightweight self-deposit repositories that are able to take a greater volume of data to include qualitative data. Figure 2 shows a snapshot of the number of individual qualitative data collections in national data repositories at May 2018. All of these archives are part of a long running European Consortium of Social Science Data Archives that is now a more formally supported European infrastructure (CESSDA ERIC 2018). CESSDA aims to provide large-scale, integrated and sustainable data services to the social sciences and bring together archives to share expert guidance on matters like repository certification, data policy, data management and data discovery and persistent identifiers.

Figure 2: Qualitative data collection in national repositories

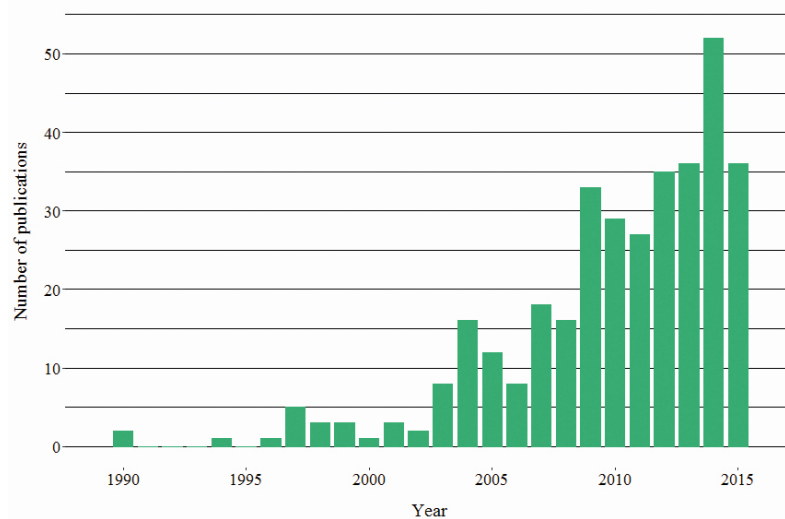


Source: Own illustration based on desk research

Legend: Numbers in the first row under the name of the archive indicate the respective founding year of the archive. The bottom row provides the number of individual qualitative data collections in the archive as of May 2018.

5. Reuse of qualitative data

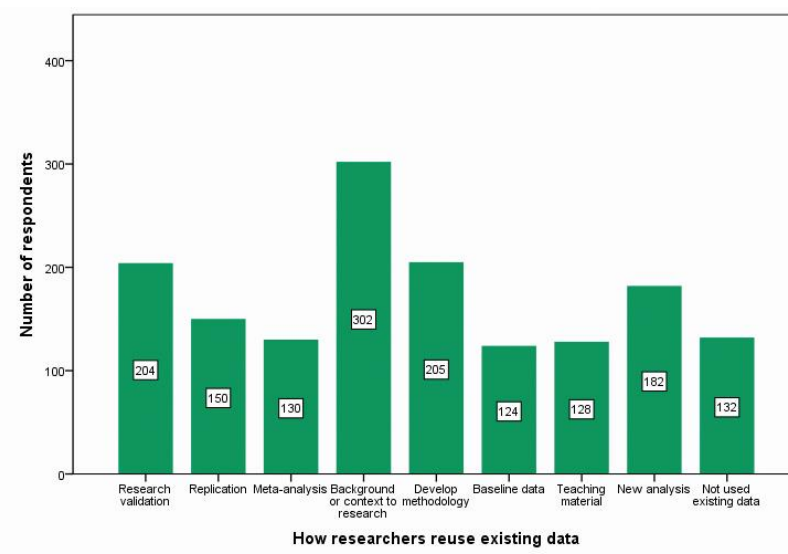
To examine demand for archived data for re-use, Bishop and Kuula-Luumi (2017) looked at measure of popularity of data re-use over a 25-year period (1990–2015) using published articles in Thomson–Reuters Web of Science citation metrics portal (web of knowledge.com), across all disciplines, that had ‘re-used’ qualitative data in their analyses or discussed the technique. The number of publications identified was almost zero from the period 1990 until 1997 when five items were published (Figure 3) and rose to over 50 per year by 2014, and reaching reach a total of 347 over the 25-year period.

Figure 3: Mentions of re-use of qualitative data in the citation literature (1990–2015)

Source: Bishop and Kuula-Luomi (2017)

The case of two archives demonstrates how data are being re-used in practice: The UK Data Service and the Finnish Social Science Data Archive (FSD). The data are mostly text-based, arising from in-depth interviews, focus groups, essays, and observations; and the scale of studies varies hugely. Accompanying audio material is often not offered, due to disclosure risk, but there are sometimes related materials, such as images. Where raw data from a study has not been archived, there are often rich documentation about the conduct of the research itself and fieldwork.

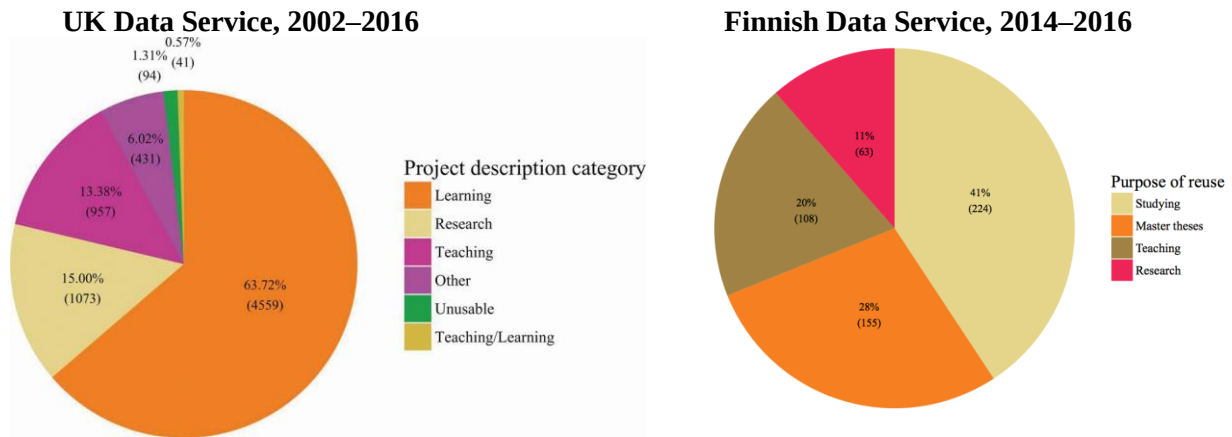
Building on types of use, originally classified by Corti and Thompson (2004), they fall into eight broad categories: providing background description and historical context; comparative research, restudy or follow-up; secondary analysis; text mining and natural language processing; replication of published work; research design and methodological advancement; and teaching and learning. A survey carried out in 2016 of UK holders of Wellcome and ESRC research grants supports this pattern of use (van den Eynden et al. 2016), as shown in Figure 4.

Figure 4: Reason for re-using data

Source: van den Eynden et al. (2016)

20 years of evidence, reflected in Bishop and Kuula-Luumi's (2017) paper on re-use of data at the UKDS and FSD suggests that people are undertaking secondary analysis of qualitative collections rather than replication, and there is a great deal of use for teaching and learning across all instructional levels, from undergraduate to postgraduate; indeed, over two thirds of total use (Figure 5).

Figure 5: Re-use purposes of qualitative data in two social science data archives



Source: Bishop and Kuula-Luumi (2017)

Capturing the methodological perspectives and details under which studies are undertaken, provides great value for new users who may be unfamiliar with the raw data, but also as teaching exemplars of showing various methods and fieldwork approaches. Evidence from Finland suggests that when data are especially rich, users do not demand much documentation about the study itself. Finally some topics seem to be popular, such as health, family practices and work. Crow and Edwards (2012) present a range of researchers' own methodological challenges when using archives. Further, it can be enlightening to read other's experiences of re-using data, for example, those captured in case studies of re-use in research and teaching published by the UK Data Service (2018).

There isn't room here to debate some of the more challenging epistemological and practical problems when re-using qualitative data. The sociology literature in the UK was the first to encounter critical debate around the role of research fit and context. Some of the real and perceived problems of re-using data, are uncovered by Corti and Thompson in their original early 2004 chapter on Secondary Analysis of Qualitative Data, and relate to: concerns about ethical re-use of data; not enough data available to analyse; unfamiliarity with the secondary analysis method; 'enough' contextual information; biased selection and selective sampling in secondary analysis.

6. The practicalities: preparing qualitative data

The role of a data archive includes the acquisition, preservation, documentation, cataloguing, anonymisation, and dissemination of data, to prepare a collection for future use. The current 'Qualidata' systemic approach to combining traditional archiving practice with survey data archiving protocols seeks to anticipate what future users are likely to need. Whatever one presents, a user will still want to review and assess the provenance of the data and identify any limitations or assumptions for their proposed analysis. What ends up in an archive, is in part driven by what the donor has to or is prepared to offer and what the archive wishes to take. Thus it is hard to assess what might be missing as part of a donation, or what an archivist might have weeded out. Archiving can never be a 'scientifically-neutral activity', nor are data passive resources to be mined by researchers. Indeed, archiving can be even viewed as a value-laden political practice, with what and how material gets archived being influenced by power.

One of the key roles then of the data archivist is to appraise data using criteria of research value, usability, formats, and of course, ethical and legal. Where collections contain detailed documentation about the research process such as administrative documents, grant proposals and final reports, this can help appraise the completeness of the material. Once a collection is accepted, and legal issues around ownership and suitable access conditions are agreed, various 'data processing' activities can begin. The UKDS assigns priorities to collections, so that the inflow of data meets the resources available for preparation.

Despite which data formats come in, users typically want data in a user-friendly format, such as a common word-processing package for text; and long-term preservation requires formats that can be accessed by researchers, now and in the future. Data archives undertake various data curation activities which include checking and validation of the collection's contents, ensuring that consent and confidentiality agreements are met; anonymising direct identifiers where needed; and storing data for the longer term. In term of format, clear speech demarcation and the use of speaker tags are crucial when transcribing interviews and focus groups (e.g. transcription template, UK Data Service 2017).

Research should conform to ethical and legal requirements especially with respect to data protection. Informed consent is required to participate in research and to specify what kinds of uses are to be made of the information is also necessary. Many qualitative researchers have been successful in adding clauses for future data sharing into their consent forms. Where personal information might be passed on, data protection issues considerations must be built in – including that identifying information be removed or pseudonymised, or the participant must agree (e.g. oral history testimony). Where (legally protected) personal information needs to be anonymised, pseudonyms or generic descriptors should be used to edit identifying information, rather than blanking-out information, follow any article publication strategy, and be indicated in the text. Pre-planning and agreeing with participants during the consent process, on what may and may not be recorded or transcribed, can be a much more effective way of creating data that accurately represents the research process and the contribution of participants. Consideration should be given to the level of anonymity required to meet the needs agreed during the informed consent process. Obtaining informed consent for data sharing or regulating access to data should also be considered together with anonymization.

At the UKDS, documenting the data collection to enable informed use requires that information is collated about the study, methods, questionnaires and data is compiled into a User Guide. It is useful to think about levels of context that can help us document data. For example, in addition to situational features, context can also include factors resulting from everyday interaction and interpretive processes. 'Data-level' documentation includes information about participants/settings and the data files collected. File-level attributes can be detailed in a number of ways, including providing a summary at the top of a text file (e.g. an interview transcript), attaching descriptive information to a separate file, or setting out file-level information for all data files in a collection in one document. A data list, in an MS Excel spreadsheet can be used to identify: descriptive attributes or biographical characteristics of participants or entities studied, such as: age, gender, occupation or location; and identifying details of the data items, such as file name, description, file format and size; connections between related files e.g. audio, images and so on; and indicate where parts of the data might be missing, such as partial transcripts or those completely missing from a collection. See for example a data list from the study, *Being a Doctor* (Nettleton 2009).

What an archive can receive and store in terms of annotated data depends on the software in which annotations/coding have been created. CAQDAS packages, like NVivo and Atlas-ti, are proprietary and not all the value-added work can be exported. But they can aid with data description tasks as they

contain features to help create attributes about sources used, interview settings and interviewees, pseudonyms used, and the final coding list and analytic memos that can also be exported and archived with the data. Finally, it is essential to provide an audit trail of what was done at each stage to provide full transparency. Many repositories use a read file for users, for example: <http://doc.ukdataservice.ac.uk/doc/2000/read2000.htm>.¹ This should include information on Privacy Impact Assessment, legal agreements on condition of use, and enhancements by the depositor or the archive.

A systematic catalogue or ‘metadata’ record is created for studies, providing an overview of the study, the size and content of the data files, availability and terms and conditions of access. The Data Documentation Initiative (DDI), is seen as the de facto standard for archiving and providing systematic resource discovery for social science collections. The UKDS uses DataCite DOIs (Digital Object Identifiers) for persistent identification of their data collections. Many data archives provide web-based delivery of data via access facilities in a secure and managed environment, where data files can be downloaded, after the appropriate level of user authentication and authorisation that has been agreed for that data collection. Under certain circumstances, sensitive and confidential data can be safeguarded by regulating or restricting access to and use of data.

Data presented digitally via user-friendly tools and interfaces, enables more efficient discovery and manipulations of large data. The UK Data Service developed a fully searchable qualitative data platform, QualiBank (UK Data Service, 2014) where users can search for key terms and explore the entirety of published texts. QualiBank also allows users to select extracts of text, such as a paragraph from an interview transcript, highlight these and cite that extract in an article with a persistent web address. The reader can view and use the quoted extract in context as a highlighted paragraph (e.g. extract: <https://discover.ukdataservice.ac.uk/QualiBank/Document/?cid=q-1dba72b1-d148-40e7-b3dc-a81ae230ca80>).² Fielding and Corti (2016) discuss how being able to persistently point to quote or extracts of qualitative data from publications is a useful feature for the reader to engage themselves ‘within’ the primary data source.

7. Sharing data from an emotionally sensitive topic

Seymour’s 2012 study of *Managing suffering at the end of life: a study of continuous deep sedation until death*³ covers discussion about delicate issues which prompt heightened ethical and legal concerns about how to ‘treat’ such data beyond the immediate needs of the research team. A tension arises between the desire to share valuable data and the need to protect these assets. While the ‘topic’ is undeniably sensitive, the data themselves need not be, at least not by the definition of common data protection legislation. Seymour discussed data sharing with relevant governance committees where a two-stage consent process was agreed: first, for participation in the research and then second, afterwards to allow for data archiving/sharing. Seymour identified the danger of clinicians as qualitative researchers being tempted to switch to their clinical roles in order to help vulnerable participants, and thus leading to a reluctance to add additional “burden” of asking for permission to share data; yet missing an opportunity for participants to feel further empowered by the increased reach of their voices. Preparing the data for the Archive was straightforward; a suitable anonymisation strategy was devised and carried out at the transcription stage, ensuring that the transcriber had signed a confidentiality agreement. A safeguarded access pathway enabled 32 anonymised transcripts from

¹ Accessed on 08.08.2018.

² Accessed on 08.08.2018.

³ See <https://www.researchcatalogue.esrc.ac.uk/grants/RES-062-23-2078/read> (accessed on 08.08.2018).

the interviews to be available for registered users who sign up to a legally-binding end user agreement. The detailed study protocol is also available to freely download (UKDS and Seymour 2017).

8. Getting users to come

Data services can provide a range of materials that encourage data usage and skills. For the UKDS these have included: helping change attitudes to data re-use and sharing through advocacy campaigns; gathering supportive research data champions, such as VIPs with famous data; publishing case studies of deposit and use; promoting data via ‘data papers’; creating ‘teaching packs’ with lecturers; running webinar and courses on secondary analysis of qualitative data; holding a competition for creative use of data; pressurising funders to create funding opportunities for dedicated secondary analysis projects; running an annual dissertation prize, based on re-use of data.

9. End note

If there is one issue I want as a take away message it is that the best way to change culture and practice is through a unified mission. Bringing together scholars, qualitative data curators, traditional archivists and librarians, technologists and preservation specialists, as we have seen at the RatSWD data-oriented meeting, enables a shared understanding of the landscape, and an appreciation of each other’s specialisms and areas for overlap. With research data centres continually at risk with short-term funding, they must work together and share wherever possible.

References

- Bishop, Libby and Arja Kuula-Luumi (2017): Revisiting Qualitative Data Reuse: A Decade On. SAGE Open 7(1). DOI: 10.1177/2158244016685136 (accessed on 08.08.2018).
- CESSDA ERIC (2018): <https://www.cessda.eu/> (accessed on 20.05.2018).
- Corti, Louise und Nigel Fielding (2016): Opportunities From the Digital Revolution. SAGE Open 6(4). DOI: 10.1177/2158244016678912 (accessed on 08.08.2018).
- Corti, Louise and Paul Thompson (2004): Secondary analysis of archived data. In Clive Seale; Giampietro Gobo, Jaber F. Gubrium and David Silverman (eds.): Qualitative Research Practice. London, Sage, 327–343.
- Crow, Graham and Rosalind Edwards (eds.) (2012): Perspectives on working with archived textual and visual material in social research. Special issue of International Journal of Social Research Methods 15(4). <https://doi.org/10.1177/2158244016678912> (accessed on 08.08.2018).
- ESRC (2015): Research Data Policy. Economic and Social Research Council. <https://esrc.ukri.org/funding/guidance-for-grant-holders/research-data-policy/> (accessed on 20.05.2018).
- Nettleton, Sarah (2009): Being a Doctor: a Sociological Analysis, 2005–2006. [data collection]. UK Data Service. SN: 6124. <http://doi.org/10.5255/UKDA-SN-6124-1> (accessed on 08.08.2018).
- Silverman, David (ed.) (2016): Qualitative Research. 4th Edition. London, Sage.
- Thompson, Paul (2017): Pioneers of Social Research, 1996–2012. [data collection]. UK Data Service, SN: 6226, 3rd Edition. <http://doi.org/10.5255/UKDA-SN-6226-4> (accessed on 08.08.2018).
- University of Leeds (2017): Timescapes Repository. University of Leeds. <http://timescapes.researchdata.leeds.ac.uk> (accessed on 20.05.2018).
- van den Eynden, Veerle; Gareth Knight; Anca Vlad; Barry Radler; Carol Tenopir; David Leon; Frank Manista; Jimmy Whitworth and Louise Corti (2016): Survey of Wellcome researchers and their attitudes to open research. Technical Report. Figshare. DOI: <https://doi.org/10.6084/m9.figshare.4055448.v1> (accessed on 08.08.2018).
- UK Data Service and Jayne Seymour (2017): Depositor Story: Consent for data sharing in sensitive qualitative research: burden or benefit? UK Data Service. <https://www.ukdataservice.ac.uk/deposit-data/stories/seymour> (accessed on 20.05.2018).
- UK Data Service (2014): Introducing the UK Data Service QualiBank. UK Data Service. <https://www.ukdataservice.ac.uk/media/428412/qualibankguide.pdf> (accessed on 20.05.2018).
- UK Data Service (2017): UK Data Service model transcript. <http://data-archive.ac.uk/media/136055/ukdamodeltranscript.pdf> (accessed on 20.05.2018).
- UK Data Service (2018): Case studies of reuse. UK Data Service. <https://www.ukdataservice.ac.uk/use-data/data-in-use/case-studies> (accessed on 20.05.2018).

Louise Corti is Associate Director at the UK Data Archive and leads the UK Data Service's Collections Development and Data Publishing teams. Louise actively researches and publishes on key aspects of data management, sharing and reuse of social science data. She was instrumental in helping operationalise the ESRC's Research Data Policy from 1995 and extending this to fully accommodate qualitative data. She regularly acts as a consultant for qualitative data archives setting up round the world, and is currently helping UKDS to scale up for managing big data.

Flexible Strategien für eine forschungsfreundliche Archivierung und Nachnutzung qualitativer Forschungsdaten

Susanne Kretzer und Michael Diepenbroek

Abstract

In den letzten Jahren wurde ein intensiver Diskurs über die Archivierung und Sekundärnutzung qualitativer Daten aus der empirischen Sozialforschung in der Scientific Community geführt. Um das Sekundärnutzungspotenzial von qualitativen Primärstudien auszuschöpfen, bedarf es professioneller Datenservices, die Forschungsmaterialien forschungsorientiert und nachnutzungsfreundlich aufbereiten und bereitstellen. *Qualiservice* legt seinen Fokus auf interviewbasierte Studien, die idealerweise schon während der Primärforschung von *Qualiservice* beratend begleitet werden, um so eine möglichst optimale Datenvorbereitung für die Nachnutzung zu realisieren und die Einhaltung von rechtlichen und forschungsethischen Anforderungen zu gewährleisten. Die Nachnutzung qualitativer Interviewstudien in Forschung und Lehre erfordert ein sicheres Management der Anonymisierung und transparente Nutzungsbedingungen. Voraussetzung für eine Sekundärnutzung sind eine gute Kontextualisierung auf Studienebene mithilfe von Studienreports sowie die Erarbeitung von Mikrokontexten, um Sekundärforschenden die Bewertung und den weitgehenden Nachvollzug des Forschungsgeschehens zu ermöglichen. Um Studien, die archiviert wurden und/oder für die Sekundärnutzung angeboten werden, auch international in entsprechenden Katalogen und Plattformen sichtbar zu machen, unterstützt *Qualiservice* Forschende dabei, ihre Studien mit aussagekräftigen Metadaten zu versehen.

Im vorliegenden Beitrag werden praktische Lösungen für die Archivierung und Bereitstellung qualitativer Daten vorgestellt, die *Qualiservice* im Rahmen eines aktuellen DFG-Infrastrukturprojekts implementiert und auf dem Workshop *Archivierung und Zugang zu qualitativen Daten* präsentierte.

1. Zum aktuellen Stand von *Qualiservice*

Qualiservice ist Nachfolger des *Archivs für Lebenslaufforschung (ALLF)*, das aus dem Sonderforschungsbereich 186 *Statuspassagen und Risikolagen im Lebensverlauf* an der Universität Bremen hervorging. Das ALLF führte gemeinsam mit dem *GESIS Leibniz-Institut für Sozialwissenschaften* von 2003 bis 2005 eine von der DFG finanzierte Machbarkeitsstudie durch, die die Angebots- und Nachfrageseite für ein Servicezentrum für qualitative Interviewdaten untersuchte (Medjedović 2011, Opitz und Mauer 2005)¹ und zur Basis der konzeptionellen Überlegungen (Medjedović und Witzel 2010: 95–143) für *Qualiservice* wurde. In einer ersten DFG geförderten Aufbauphase für die Vorbereitung eines Forschungsdatenzentrums (FDZ) entwickelte *Qualiservice* von 2012–14 in enger Abstimmung mit der Scientific Community einen Workflow für die Archivierung und Aufbereitung qualitativer Forschungsdaten (Kretzer 2013: 93). 2014 schließlich integrierte Betina Hollstein, Professorin für Mikrosoziologie und Methoden, als Projektleitung *Qualiservice* in das *SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik*. *Qualiservice* leistete dann zunächst Beratungen für Forschende hinsichtlich der Archivierung und Datenvorbereitung und koordinierte sich schließlich mit dem neuen Projektpartner für die informationstechnischen Dienstleistungen *PANGAEA*. Aktuell kann in einer dritten DFG-Förderphase die Implementierung von *Qualiservice* als FDZ beginnen und damit eine Verstetigung vorbereitet werden. Einen starken und erfahrenen Partner wie *PANGAEA* (s.u.) für die informationstechnische Infrastruktur gewonnen zu haben, ermöglicht die an die Anforderungen der qualitativ Forschenden flexibel anpassbaren Abläufe.

Die Nachfrage nach einer Archivierungsmöglichkeit für qualitative Daten ist groß und nicht nur auf interviewbasierte Studien und die soziologischen Disziplinen beschränkt (bspw. Imeri 2017). Dies zeigen viele Anfragen potenzieller Datengeber*innen aus unterschiedlichsten Fächern wie bspw. der Ethnologie, der Geographie, den Sport-, Politik-, Religions- und Kulturwissenschaften an *Qualiservice* und damit werden schon jetzt neue Anforderungen offenbar, deren Erfüllung schon bei der Implementierung für eine zukünftige Erweiterung berücksichtigt werden müssen.

Die Projektpartner und ihre Aufgaben sollen nachfolgend in einem kurzen Überblick vorgestellt werden:

► SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik an der Universität Bremen:²

Seit 2014 ist *Qualiservice* im SOCIUM verortet. Die zentrale Geschäftsführung von *Qualiservice* steuert von hier aus alle Services wie die Beratung und Begleitung von Forschungsprozessen, Datenübergaben, Vereinbarungen zur Nutzung, die Kommunikation mit der Scientific Community und die Kooperation mit nationalen und internationalen Partnern. Das Safe Center, in dem übernommene Forschungsmaterialien unter sicheren Bedingungen aufbereitet und bspw. auf eine ausreichende Anonymisierung hin untersucht werden, ist ebenfalls im SOCIUM angesiedelt und Wissenschaftler*innen können an einem Gastarbeitsplatz mit einer gesonderten Nutzungsvereinbarung in sensiblen Daten forschen. In der zentralen Geschäftsstelle werden auch rechtliche Regelungen und forschungsethische Überlegungen mit der Praxis der Datenarchivierung und Sekundärnutzung abgeglichen und umgesetzt.

¹ Eine Publikationsliste zu Datenarchivierung und Sekundärnutzung von ALLF und *Qualiservice* steht unter <http://qualiservice.org/index.php?id=92> zum Download (Zugriff am 15.05.2018).

² <http://www.socium.uni-bremen.de/> (Zugriff am 15.05.2018).

► PANGAEA³ – zertifiziertes „World Data Center“ und „Data Publisher“

Seit 1993 archiviert und publiziert PANGAEA Forschungsdaten und prägt in internationalen Kooperationen den weltweiten Aufbau der dafür erforderlichen Infrastrukturen. Ursprünglich aus der Umwelt- und Klimaforschung kommend, arbeitet PANGAEA mittlerweile interdisziplinär und erweitert fortlaufend seine Workflows. Gerade für die Archivierung und Bereitstellung qualitativer und sensibler Daten entwickelte PANGAEA geeignete flexible Strukturen, die an die Anforderungen von *Qualiservice* angepasst werden können. PANGAEA stellt Metadaten international für unterschiedliche Metadatenschemata zur Verfügung.

► Staats- und Universitätsbibliothek Bremen (SuUB)⁴

Die Studienreports für eine ausführliche Information über archivierte Studien werden über den zertifizierten Dokumentenserver E-LIB von der SuUB publiziert. Für die Metadatenerfassung wurde von der SuUB in Kooperation mit der DDI Working Group on Qualitative Data (Hoyle et al. 2013) ein Metadatenschema für qualitative Daten (Betancort Cabrera und Haake 2014) erarbeitet und fließt nun in das von PANGAEA entwickelte Metadatenschema panFMP⁵ ein.

► GESIS Leibniz-Institut für Sozialwissenschaften⁶

Für die verteilte Archivierung von Mixed-Methods-Studien wird mit der GESIS ein Workflow entwickelt, um quantitative und qualitative Forschungsdaten aus einer Studie anwendungsfreundlich zugänglich zu machen.

2. Antizipierte Datenvorbereitung durch Begleitung der Primärforschung

Voraussetzung für die Archivierung und Datenvorbereitung sind Ressourcen, die im Primärforschungsprozess für die Dokumentation, Kontextualisierung und Primärdatenvorbereitung, wie bspw. die Anonymisierung, verfügbar sind. Diese können bei einigen Förderinstitutionen in den Projektantrag eingestellt werden. Um für jede Studie den Aufwand für die notwendigen Mittel einzuschätzen, empfiehlt sich eine Erörterung mit *Qualiservice* über das zu erwartende Datenvolumen und die voraussichtlichen Datentypen (Text, Bild, Ton). Am besten gelingt dies mit einem gemeinsam entwickelten Datenmanagementplan. Mit dessen Hilfe können einerseits die benötigten Ressourcen für die Datenvorbereitung, die in der Primärforschung zu leisten sind und andererseits der Aufwand im FDZ etwa für die Datenkuration und -publikation im FDZ aufgezeigt werden. Ein Datenmanagementplan und damit auch die Datenvorbereitung für die Archivierung kann jederzeit an die Forschungsrealität angepasst werden.

Qualiservice möchte Forscher*innen schon während der Primärforschung begleiten. Die nach und nach entstehenden Dokumentationen über die Forschung können so zeitnah auf ihren Informationsgehalt für die spätere Sekundärnutzung begutachtet und optimiert werden. Der fortlaufende Kontakt zwischen FDZ und Primärforscher*innen ermöglicht es, forschungsethische Aspekte (von Unger 2018: 681, 2014: 15) sowohl bei den vorbereitenden Arbeiten als auch für die Übergabe- und späteren Nutzungsbedingungen zu berücksichtigen. Die notwendige Kontextualisierung der Forschungsdaten, die Sekundärnutzer*innen brauchen, um das primäre Forschungsgeschehen nachzuvollziehen und die Dateneignung für potenzielle Sekundärforschungsprojekte zu bewerten, wird zu einer fortlaufenden reflexiven Aufgabe während der

³ <https://www.pangaea.de/> (Zugriff am 15.05.2018).

⁴ <https://www.suub.uni-bremen.de/> (Zugriff am 15.05.2018).

⁵ <http://www.panfmp.org/> (Zugriff am 15.05.2018).

⁶ <https://www.gesis.org/home/> (Zugriff am 15.05.2018).

Forschung und nicht zu einer Rekonstruktion gar auf Basis einer Gedächtnisleistung im Nachhinein. Eine gelungene Kontextualisierung bezieht sich sowohl auf die Studienebene als auch auf die Mikroebene des Projektes. *Qualiservice* hat schon vor einigen Jahren einen Leitfaden für die Kontextualisierung in Form eines Studienreports entwickelt (Kretzer 2013: 105–106), der je nach Forschungsdesign flexibel gehandhabt werden kann, und diesen Leitfaden⁷ weiter optimiert. Der Studienreport soll Informationen geben über:

- Konzept der Studie
- Sampling/Stichprobenziehung
- Sampling Population/Beschreibung der Befragtengruppe
- Methodik/Beschreibung der Datenerhebung
- Auswertungsprozesse/Ergebnissicherung
- Datenschutz und Datensicherheit
- Bezugsrahmen der Studie
- Anträge und Berichte, interne Arbeitspapiere

Für die Begleitung der Primärforschung sind unterschiedliche Formate vorgesehen:

- *Qualiservice* und Forscher*innen können in einem Ticketsystem, einer sicheren dialogischen Verbindung miteinander kommunizieren. Das Ticket kann während einer gesamten Projektlaufzeit genutzt werden. Das Ticketsystem dokumentiert alle Kontakte und Arbeitsschritte. Forschende können über das Ticket fortlaufend Forschungsmaterialien und -dokumentationen übergeben. Dadurch kann die Arbeitsbelastung für die Datenübergabe an *Qualiservice* zum Projektabschluss eine erhebliche Entlastung erfahren. Sensible Daten werden für die Kuration zeitnah aus dem Ticket in das Safe Center übernommen. Das Ticketsystem bietet eine hohe Transparenz für die Forscher*innen: Sie können in ihr Ticket Einsicht nehmen und vor der Veröffentlichung von unsensiblen Metadaten und Studienreports, der Bereitstellung der Daten für die Sekundärnutzung, sei es ortsfern oder im Safe Center, die aufbereiteten Daten mit einem proof read begutachten und freigeben.
- *Qualiservice* wird regelmäßig Workshops für Nutzer*innen anbieten, die sich mit den Anforderungen der Archivierung und Bereitstellung für die Sekundärnutzung auseinandersetzen möchten. Forscher*innen können in den Workshops konkret die Archivierung und Datenvorbereitung ihrer eigenen Forschung planen. Workshops können auch als Inhouse-Veranstaltungen organisiert werden.
- Schließlich wird *Qualiservice* bei Bedarf Projekte auch vor Ort besuchen.

3. Begleitung der Datenvorbereitung abgeschlossener Projekte

Eine nachträgliche Vorbereitung der Daten gestaltet sich in der Regel recht aufwendig und es empfiehlt sich für die Archivierung im Nachhinein, also schon beendeter Projekte, Ressourcen bei Förderinstitutionen zu beantragen. Häufig sind die beteiligten Wissenschaftler schon in einem anderen Projekt tätig und so müssen externe Fachkräfte die vorbereitenden Arbeiten für die Archivierung leisten und sich mit der Studie vertraut machen. Insofern ist anzuraten, die Daten hinsichtlich ihrer Potenziale für eine Sekundärnutzung zu analysieren und eine nachträgliche Archivierung von diesem Ergebnis abhängig zu machen.

⁷ Der gesamte Leitfaden steht unter <http://qualiservice.org/index.php?id=48> (Zugriff am 15.05.2018).

4. Archivierungsmöglichkeiten bei Qualiservice

Qualiservice bietet Archivierungsstrategien für unterschiedliche Nutzergruppen an. Es können

- Querschnittstudien und Einzelprojekte
- Längsschnittstudien
- Verbundprojekte und Sonderforschungsbereiche
- Institutsspezifische Forschung

archiviert werden. Alle Studien sind über ihre Metadaten (s. u.) für Datensuchende auf Plattformen und in Katalogen auffindbar. Die Studien sind zitierbar und bekommen einen Digital Object Identifier (DOI). Anteile der Forschungsmaterialien, die keine sensiblen Daten enthalten - dies trifft bspw. auf den Studienreport und ggf. auf ergänzendes Kontextmaterial zu - sind allgemein über das Internet zugänglich. Sensible Daten, wie etwa Primärdaten, Interviewtranskripte und Postscripta, werden prinzipiell nicht in das Internet gestellt. Ebenso prinzipiell werden nur anonymisierte Daten⁸ zur Sekundärnutzung gegeben. Dafür müssen Datensuchende eine Nutzungsvereinbarung mit *Qualiservice* schließen. Angaben in der Vereinbarung werden geprüft. Primärforscher*innen können in ihrer Datenübergabevereinbarung Bedingungen für die Sekundärnutzung setzen. In besonderen Fällen ist eine Einsicht in die Forschungsdaten nur im Safe Center möglich.

Aus diversen Forschungsdesigns und unterschiedlich gerahmter Forschungsorganisation (bspw. Forschungsverbund; Einzelprojekt) resultieren diverse Anforderungen an die Archivierung und Nachnutzung. Für die Archivierung solch vielfältiger Forschungsmaterialien bedarf es eines flexiblen Datenmanagementsystems, das adäquate Workflows bietet, um die Integrität und Authentizität der Forschungsdaten bei der Archivierung und für die Nachnutzung zu erhalten. Im Anschluss werden einige Beispiele für die flexible Vorgehensweise bei *Qualiservice* genannt:

- *Qualiservice* eröffnet Forscher*innen die Möglichkeit, in einer Übergabevereinbarung individuelle Bedingungen für die Archivierung und Sekundärnutzung festzulegen. Dies kann über ein zeitliches Embargo, das die Nutzung der Daten erst ab einem bestimmten Zeitraum erlaubt, bis hin zu Nutzungsausschlüssen, bspw. für die Lehre, reichen.
- Durch die Kombination verschiedener Erhebungsmethoden in einer Studie wird es erforderlich, die unterschiedlichen Datenarten (Interview, Bild- und Audioformate, Beobachtung), die entstehen, entsprechend unterschiedlich für die Archivierung und Sekundärnutzung zu behandeln. Während Interviewtranskripte häufig anonymisierbar sind, stellt die Anonymisierung etwa bei Audio- und Bildformaten ein Problem dar, das bisher die Sekundärnutzung weitgehend einschränkt. Es müssen also für eine Studie, in der mit der Triangulation mehrerer Methoden Daten erhoben wurden, ggf. unterschiedliche Kurations- und Sicherheitsmaßnahmen als auch Nachnutzungsoptionen eingerichtet und ebenfalls für die langfristige Nutzbarkeit der Daten unterschiedliche Erhaltungsmaßnahmen ergriffen werden.
- Bei einem Längsschnittdesign werden bei *Qualiservice* die einzelnen Erhebungswellen unterhalb der Studienebene einzeln beschrieben, um Unterschiede im methodischen Vorgehen oder bei den

⁸ Ein Arbeitspapier zum Anonymisierungskonzept von *Qualiservice* steht unter <https://www.ssoar.info/ssoar/handle/document/47605> zum Download (Zugriff am 15.05.2018). Zurzeit wird das Konzept hinsichtlich der neuen Datenschutzbestimmungen überarbeitet.

Merkmale der Forschungspartner*innen vermerken zu können. Die Wellen können, falls erforderlich, jeweils divergierend archiviert und bereitgestellt werden.

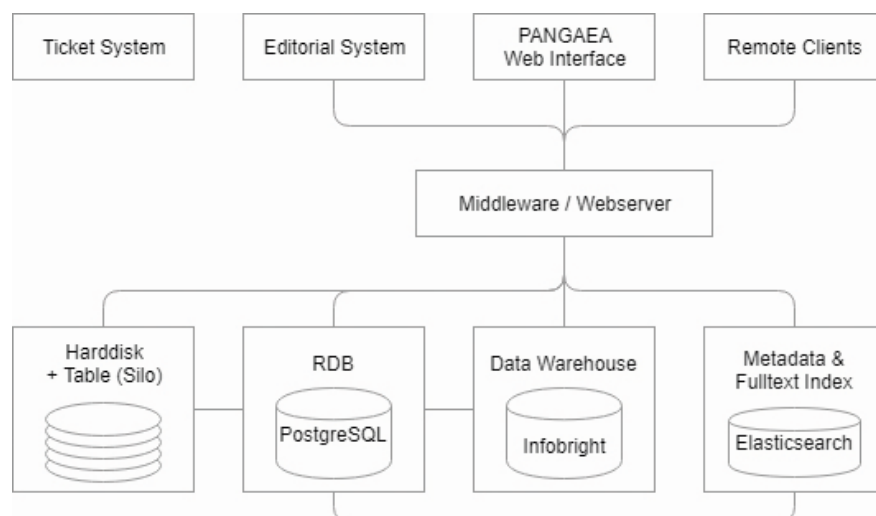
► Verbundprojekte werden als eine Studie organisiert, die sich dann in die unterschiedlichen Projektanteile gliedert. Bei Studien mit internationalen Forschungspartner*innen wird bspw. das gesamte Forschungsvorhaben dargestellt und dann in die jeweiligen nationalen Teilprojekte untergliedert. Dies erlaubt auch ein angemessenes Management der Daten, das durch besondere rechtliche Regelungen (z. B. im Datenschutz, Datentransfer) in den beteiligten Ländern notwendig sein kann. Daraus resultiert auch die Möglichkeit, Teilprojekte einer Studie mit besonderen Nutzungszugängen und Sicherheitsvorkehrungen zu belegen. Bspw. können die deutsch- und englischsprachigen Teilprojekte für die Sekundärnutzung außerhalb des Forschungsdatenzentrums genutzt werden, während andere Teilprojekte in weiteren Sprachen nur im Safe Center als Gastwissenschaftler beforscht werden dürfen. Auch können, falls dies erforderlich ist, Datenübergabevereinbarungen mit allen Teilprojekten geschlossen werden, so dass internationale Forscher*innen die Möglichkeit erhalten, ihre Bedingungen für die Sekundärnutzung zu benennen. Ist ein Datentransfer nach Deutschland nicht zulässig, können mit einem Archiv des betreffenden Landes technische Maßnahmen zur Auffindbarkeit der Daten abgestimmt werden. Verbundprojekte können einen gemeinsamen Studienreport verfassen.

► Ähnlich verhält es sich mit Sonderforschungsbereichen (SFB). Sie werden als ein Projekt organisiert und dargestellt, aber dann in die Teilprojekte untergliedert. Allerdings stehen hier mit den unterschiedlichen Teilprojekten auch unterschiedliche Forschungsthemen zur Archivierung an. Das bedeutet bspw., dass für jedes Teilprojekt gesondert Metadaten generiert werden und ein gesonderter Studienreport erstellt wird. Die Teilprojekte schließen einzeln eine Übergabevereinbarung mit *Qualiservice* und können damit auch ergänzende Sicherheitsvorkehrungen für die Archivierung und Nutzungsbedingungen für die sekundäre Verwendung definieren.

► Schließlich können auch Forschungsinstitute ihre Studien bei *Qualiservice* archivieren. Durch das flexible Datenmanagementsystem können Workflows und Nutzungsbedingungen ausgehandelt werden, die dem Bedarf der Institute Rechnung trägt. Das Bestreben von Instituten, ihre Forschung zu archivieren, aber zunächst nur über die Veröffentlichung von Metadaten für die Scientific Community sichtbar zu machen ist über *Qualiservice* möglich. Neben dem Effekt, dass damit die Daten auch institutsintern nachnutzbar bleiben, erlaubt die internationale Veröffentlichung der Metadaten Datensuchenden sich über *Qualiservice* an die Institute zu wenden, um zu prüfen, ob eine Sekundärnutzung in besonders begründeten Fällen für Gastwissenschaftler*innen möglich ist.

5. Das Informationssystem PANGAEA

Dienstleister für die Dateninfrastruktur von *Qualiservice* ist PANGAEA. Als multidisziplinäres Informationssystem werden seit ca. 25 Jahren schwerpunktmäßig geo- und biowissenschaftliche Daten archiviert und publiziert. Zu einem kleinen Teil wurden auch Daten mit sozialwissenschaftlichem Bezug adressiert. Mit *Qualiservice* wird dieser Anteil ausgebaut.

Abbildung 1: Drei-Schichten-Architektur von PANGAEA

Quelle: Eigene Darstellung

Das System entspricht einer Drei-Schichten-Architektur⁹ (Abb. 1). Die relationale Datenbank im Backend ist vollständig normalisiert und erlaubt im Zusammenspiel mit den Middleware-Komponenten beliebige Erweiterungen im Daten- und Metadatenbereich um neue Datendomänen. Die Workflows zur Datenannahme, -archivierung und -publikation können den jeweiligen Anforderungen der Communities angepasst werden. Insbesondere können sensible Daten in flexibler Weise geschützt werden (Gruppen, Einzelnutzer*innen, temporäre Zugänge – nach Freigabe durch Kurator*innen und Autor*innen).

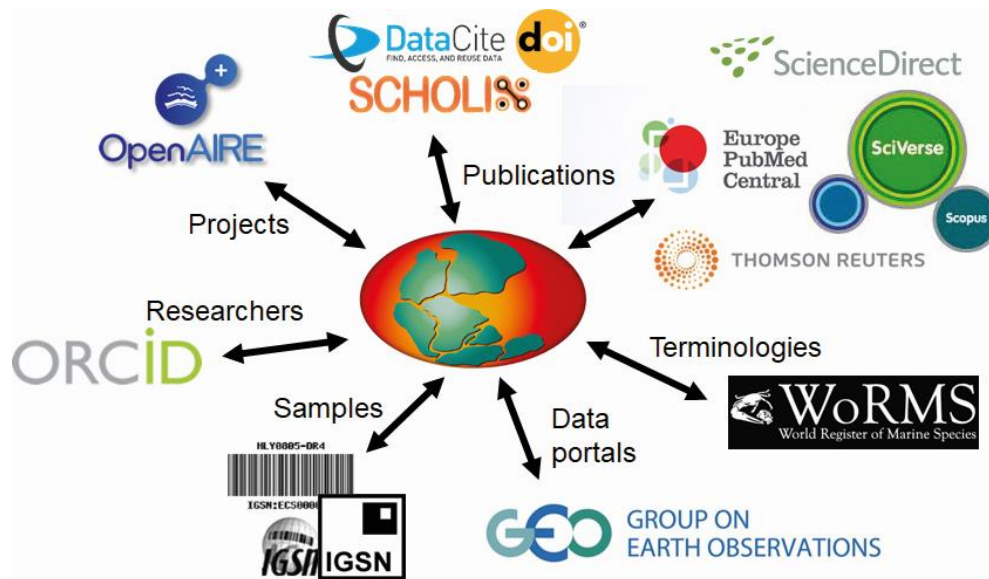
PANGAEA ist Mitglied des “International Council for Science World Data System (ICSU-WDS)” und ist als “World Data Center (WDC)” zertifiziert und akkreditiert. Das Datenzentrum wird entsprechend den FAIR Prinzipien betrieben. Zu archivierende Daten werden durch Domainexpert*innen qualitätsgeprüft sowie strukturell und semantisch über das “Editorial System” harmonisiert. Grundlage sind Inhalts-Standards und relevante Terminologien der jeweiligen Communities (Diepenbroek et al. 2017). Für Datenabgaben wird ein Ticketsystem genutzt,¹⁰ welches einerseits die effiziente Kommunikation zwischen Datenautor*innen und -kurator*innen ermöglicht und gleichzeitig als Dokumentation von Datenabgaben dient. Dabei sind allerdings Originaldaten, die *Qualiservice* zugeordnet sind, online nicht zugreifbar. Für die Nachnutzung von geschützten Daten ist eine Nutzendenregistrierung notwendig. Die Authentifizierung kann über ORCID¹¹ erfolgen.

⁹ <https://de.wikipedia.org/wiki/Schichtenarchitektur> (Zugriff am 10.06.2018).

¹⁰ <https://issues.pangaea.de/> (Zugriff am 10.06.2018) – als Software wird Atlassian Jira genutzt: <https://de.atlassian.com/software/jira> (Zugriff am 10.06.2018).

¹¹ <https://orcid.org/> (Zugriff am 10.06.2018) – ORCID ermöglicht u. a. die eindeutige Zuordnung von Nutzer*innen zu Publikationen von Daten und Literatur.

Abbildung 2: Vereinfachte Darstellung der Interoperabilität von PANGAEA



Quelle: Eigene Darstellung

PANGAEA ermöglicht eine weitgehende Interoperabilität der archivierten Datenbestände (Abb. 2). Hervorzuheben ist das “cross-linking” zwischen Literatur und Daten. Der mit maßgeblicher Beteiligung von PANGAEA aufgebaute Dienst SCHOLIX¹² ermöglicht den Verlagen die Referenzierung von Daten, die mit Zeitschriftenaufsätzen relationiert sind. Das “cross-linking” hat einen signifikanten Effekt auf die Zitationshäufigkeit von Aufsätzen (Piwowar/Day/Fridsma 2007). Zu nennen ist ferner die Verbreitung von Metadaten in zahlreichen z. T. auch Community-spezifischen Datenportalen. Für *Qualiservice* ist die Integration in das GESIS Datenportal und den UK Data Service¹³ geplant.

¹² <http://www.scholix.org/> (Zugriff am 15.05.2018).

¹³ <https://www.ukdataservice.ac.uk/> (Zugriff am 15.05.2018).

Literaturverzeichnis

- Betancort Cabrera, Noemi und Elmar Haake (2014): Das QualiService-Metadatenchema, Version 1.1. QualiService Technical Reports, 2014/01. <http://nbn-resolving.de/urn:nbn:de:gbv:46-00103643-13> (Zugriff am 10.06.2018).
- Diepenbroek, Michael; Uwe Schindler; Robert Huber; Stéphane Pesant; Markus Stocker; Janine Felden; Melanie Buss und Matthias Weinrebe (2017): Terminology supported archiving and publication of environmental science data in PANGAEA. *Journal of Biotechnology* 261, 177–186. <https://doi.org/10.1016/j.jbiotec.2017.07.016> (Zugriff am 15.05.2018).
- Hoyle, Larry und the DDI Qualitative Data Working Group: Louise Corti; Arofan Gregory; Agustina Martinez; Joachim Wackerow; Eirik Alvar; Noemi Betancort Cabrera; Damien Gallagher; Tobias Gebel; Jani Hautamaki; Arja Kuula; Steve McEachern und Cornelia Zuell (2013): A Qualitative Data Model for DDI. DDI Working Paper Series. <https://www.ddialliance.org/sites/default/files/AQualitativeDataModelForDDI.pdf> (Zugriff am 10.06.2018).
- Imeri, Sabine (2017): Open Data? Zum Umgang mit Forschungsdaten in den ethnologischen Fächern. In: Jonas Kratzke und Vincent Heuveline (Hrsg.): *E-Science-Tage 2017: Forschungsdaten managen*. Heidelberg, heiBOOKS. <http://doi.org/10.11588/heibooks.285.377> (Zugriff am 10.06.2018).
- Kretzer, Susanne (2013): Infrastruktur für qualitative Forschungsprimärdaten – Zum Stand des Aufbaus eines Datenmanagementsystems von QualiService. In: Dennis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die qualitative Sozialforschung*. Berlin, SCIVERO, 93–110. <https://www.ratswd.de/publikationen/forschungsinfrastrukturen-qualitative-sozialforschung> (Zugriff am 10.06.2018).
- Medjedović, Irena (2011): Secondary Analysis of Qualitative Interview Data: Objections and Experiences. Results of a German Feasibility Study. *Forum Qualitative Sozialforschung / Forum Qualitative Social Research* 12(3), Art.10. <http://nbn-resolving.de/urn:nbn:de:0114-fqs1103104> (Zugriff am 15.05.2018).
- Medjedović, Irena und Andreas Witzel (unter Mitarbeit von Reiner Mauer und Oliver Watteler) (2010): *Wiederverwendung von qualitativen Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte*. Wiesbaden, VS-Verlag für Sozialwissenschaften.
- Opitz, Diane und Reiner Mauer (2005): Erfahrungen mit der Sekundärnutzung von qualitativem Datenmateriale – Erste Ergebnisse einer schriftlichen Befragung im Rahmen der Machbarkeitsstudie zur Archivierung und Sekundärnutzung qualitativer Interviewdaten. *Forum Qualitative Sozialforschung / Forum: Qualitative Research* 6(1), Art. 43. <http://dx.doi.org/10.17169/fqs-6.1.510> (Zugriff am 15.05.2018).
- Piwowar Heather A; Roger S. Day und Douglas B. Fridsma (2007): Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLOS ONE* 2(3), e308. <https://doi.org/10.1371/journal.pone.0000308> (Zugriff am 15.05.2018).
- von Unger, Hella (2018). *Forschungsethik, digitale Archivierung und biographische Interviews*. In: Helma Lutz; Martina Schiebel und Elisabeth Tuiider (Hrsg.): *Handbuch Biographieforschung*. 2. Auflage, Wiesbaden, VS Verlag für Sozialwissenschaften, 681–693.
- von Unger, Hella (2014). *Forschungsethik in der qualitativen Forschung: Grundsätze, Debatten und offene Fragen*. In: Hella von Unger; Petra Narimani und Rosaline M'Bayo (Hrsg.): *Forschungsethik in der qualitativen Forschung. Reflexivität, Perspektiven, Positionen*. Wiesbaden, VS Verlag für Sozialwissenschaften, 15–39.

Susanne Kretzer ist wissenschaftliche Rätin und Mitarbeiterin in der Methodenabteilung des SOCIUM – Forschungszentrums Ungleichheit und Sozialpolitik an der Universität Bremen. Seit vielen Jahren engagiert sie sich für den Aufbau des Forschungsdatenzentrums „Qualiservice“. Auf der Basis der Vorläufer-Organisation „Archiv für Lebenslaufforschung“ entwickelte sie im Austausch mit der Scientific Community einen Workflow für eine entsprechende Infrastruktur, die die Anforderungen von Datengebenden und Sekundärnutzer/innen berücksichtigt und im Rahmen eines aktuellen DFG-Infrastrukturprojekts implementiert wird.

Dr. Michael Diepenbroek leitet am MARUM das Datenzentrum PANGAEA. Sein Hintergrund liegt im Bereich wissenschaftliches Datenmanagement, Datenpublikation und Informatik. Er ist und war in zahlreichen nationalen und internationalen Projekten beteiligt und war führend beim Ausbau von PANGAEA als multidisziplinäres und zertifiziertes Informationssystem mit Akkreditierungen als Weltdatenzentrum durch ICSU (2001) und der WMO (2011). Von 2007–15 war er Vice-Chair des ICSU World Data System (WDS) Scientific Committee. Seit 2013 koordiniert er das DFG finanzierte Projekt GFBio (21 Partner).

FDZ für qualitative Forschungsdaten in der Arbeits- und Industriesoziologie: das interdisziplinäre Zentrum eLabour

Wolfgang Dunkel und Heidemarie Hanekop

Abstract

In dem Beitrag wird der aktuelle Stand des im Aufbau begriffenen interdisziplinären Kompetenzzentrums für IT-basierte qualitative Forschung in der Arbeitssoziologie (eLabour – s. www.elabour.de) vorgestellt.

Ausgangspunkt sind die Anforderungen an das Forschungsdatenmanagement qualitativer Studien aus der Arbeits- und Industriesoziologie (AIS). Bislang besteht ein großes Defizit darin, solche Datensätze auch über ihre Zeit der Entstehung und Primärauswertung hinaus für aktuelle Forschungsfragen des Faches zu nutzen.

eLabour ist ein Zusammenschluss von Forschungseinrichtungen aus der AIS mit einschlägig erfahrenen IT-Partnern und verfolgt das Ziel, dieses Defizit abzubauen. Dies geschieht zum einen durch den Aufbau einer Infrastruktur, in der Forschungsdaten aufbereitet, archiviert und mit Hilfe intelligenter Such- und Auswahlverfahren sowie IT-basierter Analysemethoden genutzt werden können. Und dies geschieht zum anderen durch die Weiterentwicklung von Forschungsstrategien und Methoden für Sekundäranalysen zum Wandel von Arbeit in zeithistorischer Perspektive, die die Datenaufbereitung und die sekundäranalytische Auswertung qualitativer Studien einschließt. Aus der Verknüpfung von Infrastruktur und Forschungsmethoden resultiert ein Vorschlag für das Forschungsdatenmanagement qualitativer Studien, der weitreichende Möglichkeiten für Sekundäranalysen eröffnet, die über die reine Re-Analyse einer Studie hinausgehen. In diesem Sinne kann er Anregung und vielleicht auch Vorbild für Forschungsfelder auch jenseits der Arbeits- und Industriesoziologie sein. Ziel des (BMBF-geförderten) Verbundes ist die Etablierung einer Institution, die zugleich als Forschungsdatenzentrum für qualitative Daten über Arbeit wie auch als Kompetenzzentrum für Sekundäranalysen in der AIS dient.

1. Das FDZ eLabour

Das BMBF-geförderte Verbundprojekt eLabour (www.elabour.de) hat im Herbst 2015 seine Arbeit aufgenommen und verfolgt das Ziel, den reichen Bestand an qualitativen, empirischen Studien, die in den zurückliegenden Jahrzehnten in der arbeits- und industriesoziologischen Forschung entstanden sind, nachhaltig verfügbar zu machen und mit neuen IT-basierten Methoden zu erschließen. Damit wird die Grundlage für Forschungsfragestellungen geschaffen, die den Wandel von Arbeit in einer längeren historischen Perspektive in den Blick nehmen und mit Sekundäranalysen qualitativer arbeitssoziologischer Studien untersuchen.

Um dieses Ziel zu erreichen, wurde ein interdisziplinärer, mehrere Standorte umfassender Verbund zusammengestellt, der führende Forschungseinrichtungen der Arbeits- und Industriesoziologie (AIS) mit Partnern aus der IT zusammenbringt. Die Initiatoren aus der Arbeitssoziologie sind:

- das Soziologische Forschungsinstitut Göttingen (SOFI),
- das Institut für Sozialwissenschaftliche Forschung - ISF München,
- die Arbeits-, Industrie- und Wirtschaftssoziologie an der Universität Jena,
- die Sozialforschungsstelle an der TU Dortmund.

Da die Institute bereits seit Jahrzehnten in der Forschung tätig sind, kann so eine große Bandbreite arbeits- und industriesoziologischer Forschung aus über 40 Jahren abgedeckt werden. Die Arbeitssoziologie wird ergänzt durch Partner aus der Informationswissenschaft und Informatik, die über große Erfahrung auf dem Feld der Digital Humanities verfügen:

- das Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO),
- die Forschungsabteilung der Staats- und Universitätsbibliothek Göttingen (SUB),
- die Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG),
- das L3S Research Center an der Universität Hannover.

SUB und GWDG bringen ihre Kompetenzen im Bereich Forschungsdateninfrastruktur und Datenmanagement ein, zudem stellen sie die Schnittstelle zu anderen Forschungsinfrastrukturen im Bereich der Digital Humanities wie den Digitalen Forschungsinfrastrukturen für die Geistes- und Kulturwissenschaften (DARIAH-DE) und dem Humanities Data Centre (HDC) her und realisieren mögliche Synergieeffekte aus der Kooperation mit diesen und weiteren Infrastrukturprojekten. Das Forschungszentrum L3S verfügt im Bereich Methodenentwicklung und Suchverfahren über einschlägige Erfahrungen und Kompetenzen, unter anderem durch seine Beteiligung an dem vom SOFI koordinierten Projektverbund „Gute Arbeit“ nach dem Boom. Pilotprojekt zur Längsschnittanalyse arbeitssoziologischer Betriebsfallstudien mit neuen e-Humanities-Werkzeugen (reSozIT)“.

Anknüpfend an diese Erfahrungen hat sich eLabour zum Ziel gesetzt, dauerhaft ein Kompetenzzentrum und Datenarchiv zu etablieren, über das der stetig wachsende Bestand an qualitativen empirischen Studien aus der Arbeits- und Industriesoziologie zugänglich und für qualitative Sekundäranalysen nutzbar gemacht werden kann.

Die Forschungsdaten bieten nicht nur umfangreiches Quellenmaterial, sondern ermöglichen empirisch fundierte, gegenwartsbezogene Analysen der Entwicklung von Arbeit im Kontext des gesellschaftlichen Wandels. Damit können qualitative Einblicke in die Entwicklung von Arbeit, die Arbeitsbedingungen und die Situation der Arbeitenden im Kontext der betrieblichen Organisation verbunden werden mit der Untersuchung von Unternehmen und Branchen im nationalen und

internationalen Zusammenhang, um so Triebfedern für und Wirkung von Veränderungen auf unterschiedlichen Ebenen in ihrem Wechselspiel zu analysieren. Eine solche Art der Nachnutzung geht über die Re-Interpretation einzelner Studien hinaus, indem sie Material aus unterschiedlichen Studien kombiniert und empirische Längsschnittperspektiven auch für die qualitative Forschung und damit eine neue Forschungslinie eröffnet.

In der ersten Förderphase (2015 bis 2018) wurde die eLabour-Forschungsumgebung durch die Verbundpartner entwickelt und getestet. Dementsprechend werden die Datensätze in dieser Phase ausschließlich von den am Projekt beteiligten Forschungsinstituten genutzt. In der zweiten Förderphase nach Abschluss der Entwicklung der Forschungsinfrastruktur (2018 bis 2020) wird eLabour hingegen für einen weiteren Nutzerkreis geöffnet. Darüber hinaus wird in diesem Zeitraum für das Zentrum eine dauerhafte Organisationsstruktur etabliert (Vereinsgründung) und der Regelbetrieb als FDZ für qualitative Forschungsdaten in der Arbeitsforschung aufgenommen. Gleichzeitig werden neue arbeitssoziologische Sekundäranalyseprojekte zum Wandel von Arbeit durchgeführt, um diese Forschungslinie auszubauen und die Methodik qualitativer Längsschnittstudien weiterzuentwickeln.

Das Zentrum eLabour wird deshalb unterschiedliche Funktionen erfüllen:

- Es ist ein Verbund (zukünftig: ein Verein) arbeitssoziologischer Forschungsinstitute, die Datensätze einstellen und auf Services von eLabour zurückgreifen können.
- Es ist ein Forschungsdatenzentrum (FDZ) mit einer für diese Art von Daten eigens entwickelten technischen Infrastruktur, die von GWDG und SUB an der Uni Göttingen nachhaltig betrieben wird.
- Es ist ein Kompetenzzentrum für qualitative Sekundäranalysen in der Arbeitsforschung, in dem diese Forschungslinie und geeignete Methoden für historisierende Längsschnittperspektiven weiterentwickelt und erprobt werden.
- Und es ist ein Netzwerk von NutzerInnen, das sich zusammensetzt aus WissenschaftlerInnen aus der Arbeits- und Industriesoziologie, der Organisationssoziologie, der Zeitgeschichte und aus weiteren Disziplinen, die mit den qualitativen Daten, der Infrastruktur und den Instrumenten, die eLabour bereit stellt, Sekundäranalysen durchführen.

Um diese Funktionen erfüllen zu können, sind umfangreiche Entwicklungsarbeiten zu leisten, da das Forschungsdatenmanagement qualitativer Datensätze mit großen Herausforderungen konfrontiert ist. Auf zwei dieser Herausforderungen werden wir in den folgenden Abschnitten näher eingehen:

Erstens auf die spezifische Datenstruktur der qualitativen Forschung in der AIS, die sich durch sehr heterogene Datentypen, Methoden und Archivierungsbestände auszeichnen.

Zweitens auf den Umgang mit den überwiegend sehr sensiblen personen- und organisationsbezogenen Daten. Hier sind flexible, IT-basierte Lösungen für das kritische Spannungsverhältnis zwischen Datenschutz, Datenqualität und Aufbereitungsaufwand anzustreben.

In zwei weiteren Kapiteln werden dann die Forschungsinfrastruktur von eLabour sowie dessen zukünftige Angebote für verschiedene Gruppen von NutzerInnen vorgestellt.

2. Anforderungen an die Aufbereitung und Qualität archivierter qualitativer Daten in der Arbeitsforschung und die Lösungen von eLabour

Im Rahmen der für die AIS typischen Fallstudienforschung handelt es sich überwiegend um qualitative Daten, die mit unterschiedlichen Methoden erzielt worden sind und deren Bedeutung auf den jeweils spezifischen Fallkontext bezogen ist. Hieraus ergeben sich für das Forschungsdatenmanagement folgende Anforderungen:

2.1. Offenheit der Methoden und Heterogenität der qualitativen Forschungsdaten in der AIS

Eine Herausforderung für die zielgerichtete Datenaufbereitung ergibt sich aus der Methode der Fallanalyse selbst und deren prinzipieller Offenheit in Bezug auf Erhebungsverfahren und das Datenmaterial. Da Fallstudien darauf ausgerichtet sind, die untersuchten Fälle durch die flexible Kombination geeigneter Methoden und genutzter Materialien möglichst genau zu erfassen, sind die Forschungsdaten sehr heterogen, sodass sich für jeden Fall abhängig von der Fragestellung und den Zugangsmöglichkeiten eine unterschiedliche Kombination von Datentypen ergeben kann. Die Heterogenität der Erhebungsmethoden stellt sich bereits innerhalb von Projekten als Herausforderung dar. Diese wird umso größer, wenn man versucht Forschungsdaten projektübergreifend und aus unterschiedlichen Forschungslinien und Zeiten zusammenzuführen und zu kategorisieren.

Die Verfügbarkeit umfangreicher Datensätze in ihrem weitgehend originalen Zustand hat aus heutiger Sicht Vor- und Nachteile. Nachteil ist der zweifellos erhebliche Aufwand für das Sichten, Sortieren, Dokumentieren und Digitalisieren des Materials, das z.T. in den Archiven der Institute mehr oder weniger gut erhalten ist. Von Vorteil ist, dass damit die Logik der Primärprojekte, das damalige Forschungsinteresse und der Entstehungsprozess der Daten gut erkennbar sein können und teilweise umfangreiches Kontextmaterial, z. B. zu Betrieben, Branchen, Ereignissen und zeitgenössischen Debatten (Presse) vorhanden ist. Allerdings variiert die Art und Qualität des Materials im Zeitverlauf hinsichtlich der verwendeten Methoden, methodischen Vorgehensweisen sowie der Art und Form der Ablage: Während in den älteren Studien mit wenigen Ausnahmen Interviews nicht aufgezeichnet, sondern nur (nach Notizen und Erinnerung) protokolliert worden sind, zunächst sogar handschriftlich (z. B. bei der klassischen Studie von Kern und Schumann 1970), stellen seit den 1980er Jahren wörtlich transkribierte Interviews meist den Kern der Empirie dar. Hieran wird zugleich die Weiterentwicklung von Methoden erkennbar. Außerdem variieren Art und Umfang der Interviews über die Zeit und zwischen den Instituten bzw. Forschungslinien.

2.2. Kontextualisierung qualitativer Forschungsdaten zum Wandel von Arbeit

Um Sekundäranalysen durchführen zu können, ist es wichtig, den Kontext des erhobenen empirischen Materials (Interviews, Beobachtungen, Gruppengespräche etc.) zu kennen, um das, was gesagt worden ist und beobachtet wurde, besser in den jeweiligen institutionellen und historischen Kontext einordnen und interpretieren zu können. Hierbei geht es erstens um den studienbezogenen Kontext der Erhebung im Rahmen der Primärstudie selbst sowie zweitens um den zeithistorischen Kontext. Die studienbezogene Kontextualisierung muss im Zuge der Aufbereitung aus dem verfügbaren Material rekonstruiert werden und informiert darüber wie das Sample zustande gekommen ist, wie die Fälle definiert und zusammengesetzt sind, welche Leitfäden zu welchen Interviews gehören usw. Nicht selten muss solches Wissen der PrimärforscherInnen nachträglich ermittelt und dokumentiert werden, z. B. aus dem Material selbst oder durch Interviews mit PrimärforscherInnen. Zukünftige Studien sollten solche Informationen möglichst bei der Erhebung begleitend dokumentieren und explizieren. Trotzdem wird einerseits ein Teil des (Erfahrungs-)Wissens der PrimärforscherInnen verloren gehen, z. B. weil es an die konkrete sinnliche Erfahrung im Feld gebunden ist oder weil sensible personen-

oder unternehmensbezogene Informationen besonderen Schutz erfordern. Daher werden Sekundäranalysen qualitativer Studien in der AIS in aller Regel nicht die gleiche Interpretationstiefe aufweisen können wie eine eigene Primärerhebung; der Zugang zum Kontext bleibt notwendigerweise eingeschränkt. Andererseits bieten studienübergreifende Sekundäranalysen, die Forschungsdaten aus mehreren Studien kombinieren, eine auf der Einzelstudienebene kaum realisierbare Chance breiterer historischer oder thematischer Reichweite. Ihr besonderes Potential erweist sich dort, wo empirische Einzelstudien an ihre Grenzen stoßen.

2.3. Lösungsmöglichkeiten

Um ein derart heterogenes Material in einem digitalen Archiv für die studienübergreifende und vergleichende Sekundäranalyse zu erschließen, ist eine vereinheitlichende, standardisierende Beschreibung bis zu einem gewissen Grad unerlässlich. Gleichzeitig soll die besondere Qualität der qualitativen Daten so weit wie möglich erhalten bleiben und der Kontext der Primärstudie abgebildet werden. Zudem muss der Aufwand für die Aufbereitung der Daten bewältigbar bleiben und ein sinnvolles Verhältnis zum Ertrag gewahrt werden. Das Datenmanagement ist daher eine Gratwanderung zwischen notwendiger Standardisierung von Strukturen, Prozessen und Workflows einerseits, ohne die kein effizientes und professionelles Datenmanagement auskommt, und den Anforderungen an Heterogenität, Kontextualisierung und Qualität der Forschungsdaten andererseits, ohne die eine inhaltlich ertragreiche Re-Interpretation nicht möglich ist.

In eLabour werden folgende Maßnahmen zur Qualitätssicherung qualitativer Forschungsdaten ergriffen:

- Rückgriff auf authentisches Primärmaterial, sorgfältige Aufbereitung,
- Erstellung umfassender, flexibler, teil-automatisiert generierter Metadaten,
- Erschließung und Dokumentation von Kontextinformationen (und -wissen),
- Erschließung von (Zwischen-)Berichten und Primärmaterial aus den Studien,
- Ergänzung durch zusätzliches (zeithistorisches) Kontextmaterial.

3. Rechtliche und ethische Anforderungen an den Datenschutz von qualitativen Daten aus der Arbeitsforschung – der Lösungsansatz von eLabour

Eine zentrale Herausforderung ist der Umgang mit den sensiblen qualitativen Daten arbeitssoziologischer Fallstudien, die oft rechtlich in besonderer Weise schutzbedürftig sind. Dem wird in eLabour durch einen eigens entwickelten Datenschutz- und Anonymisierungsprozess Rechnung getragen. Wann aber ist eine Person nicht mehr identifizierbar, welche Informationen müssen hierzu entfernt werden? Denn neben den juristisch definierten personenbezogenen Daten enthalten qualitative Interviews und Beobachtungen oft auch für die ProbandInnen (und gelegentlich auch für InterviewerInnen) hochsensible Informationen mit zum Teil erheblichen wirtschaftlichen, sozialen und juristischen Schadensrisiken. Je offener die Methode, umso weniger vorhersehbar ist die Wahrscheinlichkeit, dass solche sensiblen Informationen enthalten sind, d. h. es muss im Prinzip jedes Interview manuell geprüft werden. Auch die Verpflichtung zur informierten Einwilligung (siehe Ethik-Kodex von DGS und BDS 2017) und das nicht selten für die Interviewsituation entscheidende Vertrauen zwischen ForscherIn und ProbandIn werfen nicht nur gravierende rechtliche, sondern auch ethische Fragen auf, die zu verbreiteten Vorbehalten von PrimärforscherInnen in Bezug auf die Weitergabe qualitativer Daten für die wissenschaftliche Nachnutzung führen. Personenbezogene und sensible Informationen in qualitativen Interviews sind oft nur mit großem Aufwand und gravierenden Qualitätsverlusten hinreichend (vollständig oder faktisch) anonymisierbar.

Die Anonymisierungsproblematik beschränkt sich in der Arbeits- und Industriesoziologie allerdings nicht nur auf Personen, sondern schließt auch betriebliche und soziale Kontexte mit ein: Interviews in der AIS sind typischerweise in einen organisationalen Kontext eingebunden und finden oft am Arbeitsort statt. In den Interviews mit Beschäftigten eines bestimmten Betriebes geht es auch um deren subjektive Sicht auf betriebliche Probleme, auf KollegInnen und Vorgesetzte, um soziale Beziehungen, Rollen und Konflikte. Solche Informationen sind oft für betriebliche Insider sensibel und de facto nicht anonymisierbar, daher muss durch strikte Zugangsbeschränkungen und Kontrollen verhindert werden, dass Insider Kenntnis von den Daten erlangen.

Damit gewinnen Datenschutz und entsprechende Anonymisierungsverfahren für die Aufbereitung der Daten einen besonderen Stellenwert: Sie sollen Schadensrisiken minimieren und Schäden abwenden sowie die Re-Identifikation der Befragten verhindern. Die sichere Anonymisierung qualitativer Interviews macht daher im Zweifel eine Einzelfallprüfung der Dokumente erforderlich. Darüber hinaus sollten auch soziale und ethische Aspekte explizit einbezogen werden. Zum einen betrifft dies rechtliche und ethische Bedenken von Primärforschenden, die aus einer Vertrauensbeziehung zwischen WissenschaftlerInnen und Befragten resultieren, zweitens forschungspraktische Bedenken von Forschungseinrichtungen und Primärforschenden, die eine mögliche Gefährdung des zukünftigen Feldzugangs fürchten (für die Forschungseinrichtungen eine zentrale Ressource).

Das Datenschutzkonzept kann sich daher nicht darauf beschränken, den Anforderungen des Datenschutzes zu genügen, sondern es geht um eine Gratwanderung zwischen Datenveränderungen zum Schutz der untersuchten Personen und Organisationen und der Nutzbarkeit und Qualität der so veränderten Daten mit dem Ziel wissenschaftlich ertragreicher und valider Sekundärforschung. Und drittens muss der Aufwand für diese Maßnahmen berücksichtigt werden, damit das Gesamtziel nicht an nicht zu bewältigenden Bearbeitungsaufgaben der DatenhalterInnen scheitert.

3.1. Datenschutzkonzept und -prozess von eLabour

Das eLabour-Datenschutzkonzept für qualitative Daten in der AIS umfasst Verfahrensregeln, Praktiken und einen IT-gestützten Workflow für die Risikoklassifikation, Anonymisierung und Pseudonymisierung sowie einen IT-basierten Freigabeprozess (für DatenhalterInnen) und Zugangsmöglichkeiten für WissenschaftlerInnen (NutzerInnen). Um der besonderen Sensibilität der qualitativen Daten gerecht zu werden, sieht eLabour eine im Zweifel bis auf das einzelne Dokument bezogene Risikoanalyse vor, aus der die notwendigen Datenschutzmaßnahmen abgeleitet werden. Anschließend erfolgt ein differenzierter Freigabeprozess, in dem die Zugangsmöglichkeiten für wissenschaftliche Nutzer festgelegt werden. Auch diese können ggf. bis auf die Ebene einzelner Dokumente hinunter differenziert vergeben werden. Das Datenschutzkonzept beinhaltet Regeln, Praktiken und einen IT-gestützten Workflow für die Risikoklassifikation, Anonymisierung bzw. Pseudonymisierung, den Freigabeprozess durch die Datenhalter sowie die Bereitstellung von Zugangsmöglichkeiten für WissenschaftlerInnen:

- In eLabour werden folgende Maßnahmen in Bezug auf Fragen des Datenschutzes getroffen: Risikoanalyse der Originaldaten und der Daten vor der Freigabe für die Nutzung,
- Beteiligung von PrimärforscherInnen und DatenhalterInnen,
- Definition von sechs Risikoklassen, auf deren Grundlage definiert wird, wie Anonymisierung- und Pseudonymisierung durchgeführt werden,
- Freigabemöglichkeiten und -workflow, Nutzungsverträge mit skalierbaren Auflagen (Kontrolle, Zweck).

eLabour möchte in der qualitativen Arbeitssoziologie für flexible Standards, Regeln und Praktiken für die Dokumentation und Archivierung von Forschungsdaten werben und diese verbreiten. Dazu gehören die Bereitstellung von geeigneten IT-gestützten Arbeitsprozessen und nutzerfreundlichen, gut zugänglichen und effizienten Infrastrukturangeboten des Zentrums eLabour. Sie sollen eine nachhaltige Archivierung, Nachnutzung und Interoperabilität aktuell erhobenen Materials vereinfachen und möglichst bereits in der Erhebungsphase angewandt werden.

4. Forschungsinfrastruktur von eLabour

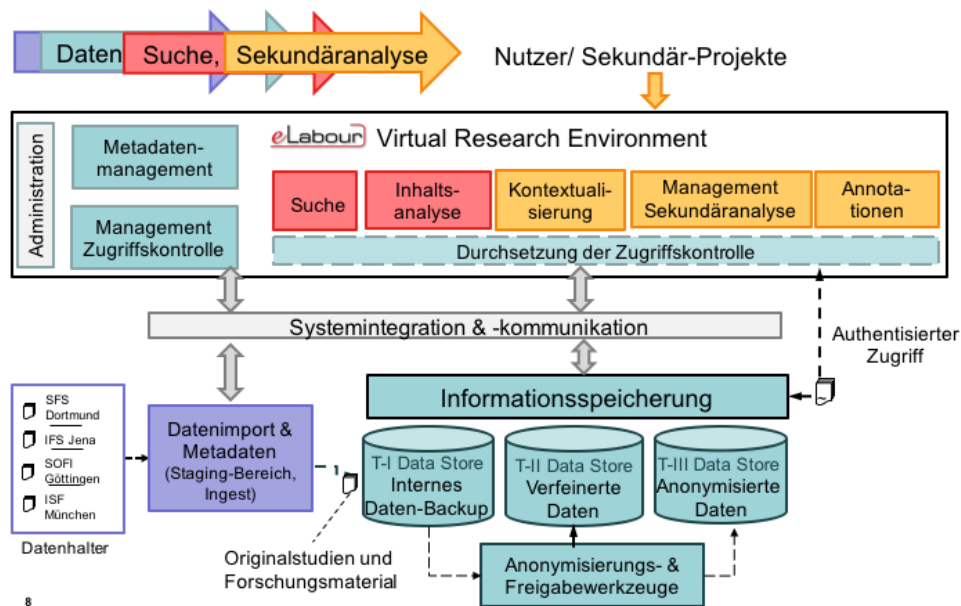
Die in eLabour entwickelte Forschungsinfrastruktur bietet Lösungen für das Forschungsdatenmanagement von qualitativen Studien an. Eine grundlegende Funktion von eLabour ist die Bereitstellung eines sicheren Forschungsdatenarchivs auf dem aktuellen Stand von Technik und Datenschutz. eLabour bietet Leistungen sowohl für datenhaltende Forschungseinrichtungen für die sichere Datenaufbereitung in der eLabour Infrastruktur als auch für die datennutzenden WissenschaftlerInnen. Dementsprechend umfasst die Forschungsdateninfrastruktur drei unterschiedliche Archive mit je speziellen Zugangs- und Rollenmodellen: erstens die nur für die jeweiligen DatenhalterInnen zugänglichen Archive zur Sicherung der Originaldaten; zweitens ein Archiv für die interne Aufbereitung der Daten in eLabour mit striktem Rechtemanagement und Vertraulichkeitsverpflichtung für MitarbeiterInnen; drittens das Forschungsdatenarchiv, das den Zugang für WissenschaftlerInnen mit Nutzerverträgen ermöglicht. Vor der Weitergabe der Daten in das dritte Archiv steht ein umfangreicher Freigabeprozess.

In technischer Hinsicht bietet eLabour folgende Funktionalitäten:

- IT-basierter Prozess für das Einlesen der Originaldaten,
- sichere Langzeitarchivierung der Originalforschungsdaten,
- IT-Werkzeuge für Datenaufbereitung, Dokumentation und Metadatenerzeugung,
- IT-basierter Prozess für Risikoklassifikation, Datenschutzmaßnahmen und Freigabe der Forschungsdaten,
- Zugangskontrolle mit transparentem Rollen- und Rechtemanagement,
- Archiv der für die Nachnutzung freigegebenen Forschungsdaten (FDZ),
- Webbasierte Nutzerplattform für Suche, Zugang, Verarbeitung und den Download von Forschungsdaten,
- Intelligente Such- und Auswertungswerkzeuge sowie Schnittstellen.

4.1. Aufbau der Forschungsinfrastruktur

Abbildung 3: Aufbau der Forschungsinfrastruktur



Quelle: Eigene Darstellung

5. Zukünftiges Angebot von eLabour

Das Kompetenzzentrum eLabour wird 2019 als Forschungsdatenzentrum (FDZ) in den operativen Betrieb gehen und damit WissenschaftlerInnen den Zugang zu einem relevanten Teil von AIS Studien seit den 1970er Jahren öffnen. Zugleich bietet das FDZ Forschungseinrichtungen und WissenschaftlerInnen eine geeignete und sichere Infrastruktur für die Aufbereitung, Archivierung und Nachnutzung qualitativer Daten aus der AIS (und darüber hinaus).

Das weiterreichende Ziel ist, eLabour als Kompetenzzentrum für qualitative Forschungsdaten und Sekundäranalysen über Arbeit zu etablieren. Es wird von einschlägigen Forschungseinrichtungen aus der AIS getragen, die „ihre“ Forschungsdaten hier gemeinsam für die Nachnutzung aufbereiten und zugänglich machen. Gleichzeitig soll der so entstehende gemeinsame Datenpool für die zukünftige Forschung zum Wandel von Arbeit genutzt werden, um für die Institute neue Forschungsmöglichkeiten zu erschließen.

Für Forschungseinrichtungen und WissenschaftlerInnen wird durch die Infrastrukturpartner (insbesondere die GWDG) eine dauerhafte Forschungsdateninfrastruktur mit umfangreichen IT-basierten Werkzeugen zur Unterstützung aller Prozesse von der Dokumentation, Aufbereitung, Kontextualisierung und sicheren Archivierung bis zu Zugangs- und Analysemöglichkeiten bereitgestellt und gemeinsam weiterentwickelt. Diese Leistungen sind eingebettet in ein Kompetenzzentrum, das die Zusammenarbeit, den Austausch und die Beratung von Forschungseinrichtungen und WissenschaftlerInnen fördert und organisiert.

Im Mittelpunkt der Arbeit des Kompetenzzentrums stehen die folgenden Schwerpunkte:

- Anforderungen, Methoden und Werkzeuge des Forschungsdatenmanagements,
- Empfehlungen zur Digitalisierung von älteren Datenbeständen,¹
- Empfehlungen, IT-basierte Prozesse und Beratung bei der digitalen Erschließung von Studien,
- Empfehlungen und IT-basierte Prozesse zur Archivierung und Dokumentation aktueller Studien,
- Beratung im Hinblick auf digitale Verfahren der Anonymisierung und Pseudonymisierung von Forschungsdaten,
- Suche, Zusammenführung und Aufbereitung von Forschungsdaten für qualitative Sekundäranalysen mit variierenden Designmöglichkeiten für Längs- und Querschnittsauswertungen.

Mit dem für 2019 geplanten Regelbetrieb der Forschungsinfrastruktur soll eLabour um weitere Einrichtungen und WissenschaftlerInnen erweitert und für WissenschaftlerInnen zugänglich gemacht werden, die nicht an dem BMBF-geförderten Verbundprojekt teilgenommen haben. Dabei werden über die Arbeits- und Industriosozilogie hinaus WissenschaftlerInnen aus weiteren qualitativ arbeitenden Bereichen der Soziologie, der Zeitgeschichte, der Wirtschaftswissenschaften und anderer Fachdisziplinen angesprochen. Neben der Eröffnung von Nutzungsmöglichkeiten von Forschungsdaten wird ein Angebot für NutzerInnen aufgebaut, das folgende Themen beinhaltet:

- Informationen zur praktischen Nutzung der Forschungsinfrastruktur von eLabour,
- Unterstützung bei Fragen des Datenschutzes,
- Informationen zu den archivierten Datensätzen, die über die Metadaten hinausgehen,
- Beratung zu Fragen der sekundäranalytischen Nachnutzung und zum Einsatz von eHumanities-Werkzeugen.

Damit verfolgt eLabour in den kommenden Jahren eine ambitionierten Zielsetzung: Das Kompetenzzentrum soll als ein fester Bestandteil der Forschungslandschaft und der Forschungsdateninfrastruktur etabliert werden – im Sinne einer Infrastruktur zur Archivierung qualitativer Daten, im Sinne eines Dienstleistungsangebots auf dem Feld des Forschungsdatenmanagements und im Sinne einer Ressource, mit der unterschiedliche Gruppen von NutzerInnen innovative Forschung zu Themen im Bereich von Arbeit betreiben können .

¹ Forschungsdaten aus älteren Studien, die digitalisiert werden müssen, können aufgrund der z. T. doch nur begrenzten OCR-Erkennungsraten in der Regel nicht umfassend mit IT-Verfahren durchsucht und ausgewertet werden. Andererseits sind die Datenschutzanforderungen bei älteren Studien meist geringer, weil das Re-Identifikationsrisiko mit dem Alter der Daten abnimmt.

Literaturverzeichnis

DGS und BDS (2017): Ethik-Kodex der Deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologinnen und Soziologen (BDS). <https://www.soziologie.de/de/die-dgs/ethik/ethik-kodex/> (Zugriff am 01.10.2018)

Kern, Horst und Michael Schumann (1970): Industriearbeit und Arbeiterbewußtsein, Teil I und II. Frankfurt am Main, Europäische Verlagsanstalt.

Dr. Wolfgang Dunkel hat in München Soziologie studiert und war als Wissenschaftler von 1987 bis 1993 an der LMU München am SFB 333 „Entwicklungsperspektiven von Arbeit“ (Projekt „Alltägliche Lebensführung“) sowie von 1994 bis 1996 Projektkoordinator und Methodenberater beim Münchner Forschungsverbund Public Health tätig. Seit 2001 forscht er am ISF München, ist Mitglied des Institutsvorstands und arbeitet auf den Feldern Dienstleistungsforschung, Arbeit und Gesundheit sowie qualitative Methoden. Er war bis 2017 im Vorstand der DGS-Sektion Arbeits- und Industriesoziologie tätig und ist Mitglied des Wissenschaftlichen Beirats der Hans-Böckler-Stiftung. Seit 2015 vertritt er das ISF München im Verbundprojekt eLabour.

Heidemarie Hanekop hat in Göttingen Sozialwissenschaft mit Zusatzfach Informatik studiert. Von 1983 bis 1985 war sie wissenschaftliche Mitarbeiterin in einem medienwissenschaftlichen Projekt an der Universität Göttingen. Seit 1986 ist sie als Wissenschaftlerin am SOFI Göttingen tätig, wo sie bis Mitte 1990 die Projekte bei IT-gestützten Methoden betreute. Seit 1996 bearbeitet sie Projekte im Bereich der Internet- und Innovationsforschung zu Themen wie Online-Plattformen, Mobiles Internet, Wissenschaftskommunikation, Open Access, Open Source und kollaborative Innovationsprozesse. Seit 2015 koordiniert sie das Verbundprojekt eLabour.

Forschungsdateninfrastrukturen für audio-visuelle Daten der Qualitativen Sozialforschung – Bedarf und Anforderungen¹

Hubert Knoblauch und René Wilke

Abstract

Wie sich in der langjährigen Forschungserfahrung bei der Videoanalyse zeigt, mangelt es einem wachsenden Kreis von Forschenden, die Video-Daten in sozialen Feldern erheben, an technisch kompetenten und fachwissenschaftlich adäquaten Möglichkeiten, ihre Daten zu sichern, zu archivieren und gegebenenfalls zur Sekundäranalyse zur Verfügung zu stellen. In diesem Beitrag soll ein an der TU Berlin angesiedeltes DFG-Projekt vorgestellt, umrissen und zur Diskussion gestellt werden, das eine solche auf videographische Daten zugeschnittene Archivierung explorieren, die Entwicklung der Methoden und Techniken im Bereich *qualitative Sozialforschung und Videoanalyse* fördern sowie zur Nachhaltigkeit der in diesem Bereich erhobenen und analysierten Video-Daten beitragen soll.

¹ Dieser Beitrag basiert maßgeblich auf einem Vortrag anlässlich des Workshops des Rates für Sozial- und Wirtschaftsdaten (RatSWD) mit dem Titel „Archivierung und Zugang zu qualitativen Daten“, der vom 27. bis 28. April 2018 an der Universität Bremen stattfand.

1. Einleitung

Der Titel dieses Beitrags ist angelehnt an den Namen eines bei der Deutschen Forschungsgemeinschaft (DFG) beantragten Forschungsprojekts, das wir zusammen mit Bernt Schnettler und seinem Lehrstuhl für Kultur- und Religionssoziologie der Universität Bayreuth (UBT) sowie der Universitätsbibliothek (UB) und dem IT-Service-Center (tubIT) der Technischen Universität Berlin geplant haben: „Entwicklung einer Forschungsdateninfrastruktur für audio-visuelle Daten der Qualitativen Sozialforschung“, kurz: aviDa. Dieses Projekt ist, nach einer langen Vorbereitung und Umarbeitungsphase, kürzlich bewilligt worden, sodass wir es an dieser Stelle, gleichsam *en passant*, vorstellen können. In der Hauptsache möchten wir diesen kurzen Beitrag allerdings dafür nutzen, auf grundsätzliche Herausforderungen und Probleme hinzuweisen, die sich rund um die Archivierung und ggf. Sekundäranalyse von Videodaten ergeben. Dazu möchten wir zunächst klären, welche audio-visuellen Daten der Qualitativen Sozialforschung in Forschungsdateninfrastrukturen, wie z. B. in aviDa, überhaupt konkret archiviert werden sollen, sodass wir (2) mit einer kurzen Skizze der *Besonderheiten* videographischer Daten fortfahren werden. Anschließend werden wir (3) auf die Frage der *Bedarfe* eingehen, ob also und warum wir eine Archivierung benötigen. Weil wir diese Frage ja offenkundig bejahen, werden wir (4) einige Hinweise darauf geben, welchen *Anforderungen* eine solche Archivierung unserer Auffassung nach zu genügen hat. Dabei werden wir Formulierungen und Thesen subsumieren und weiterentwickeln, die in den letzten Jahren entwickelt und einschlägig veröffentlicht wurden (z. B. Knoblauch und Solga 2011; Knoblauch 2013). Hierbei werden wir abschließend auch einige der Herausforderungen benennen, die unser Projekt „aviDa“ konkret fokussiert.

2. Videographische Daten

Gerade angesichts der explosiven Ausbreitung von Videodaten, insbesondere seit der Gründung des damals höchst innovativen Portals YouTube im Jahr 2005, sollte man daran erinnern, dass Video und Film davor in einem weitaus beschränkteren Gebrauch waren. Das gilt sowohl für den Alltag als auch für die Wissenschaft. In der Tat handelt es sich bei dem, was später in Geistes- und Sozialwissenschaften als Videoanalyse bezeichnet wird, um einen Ansatz, der sich erst seit den späten 1970er Jahren langsam zu entwickeln begann (vgl. Wilke 2018: 486–489). Auch wenn es zuvor schon sehr berühmte Analysen gab, die auf audio-visuellen Aufnahmen gründeten, etwa die frühen Beispiele des ethnographischen Films, von Robert J. Flaherty bis Timothy Asch (vgl. Marks 1995; vgl. auch Reichert 2007), die Filmaufzeichnungen des Entwicklungspsychologen Arnold Gesell in den 1930er Jahren oder auch die „First Five Minutes“ von Robert E. Pittenger, Charles F. Hockett und John J. Danehy (1960), kommt es erst mit der Entwicklung leistungsfähiger Batterien und alltagsfreundlicher technischer Designs zu dem, was wir heute als Videoanalyse bzw. „Videographie“ (Tuma/Schnettler/Knoblauch 2013) bezeichnen.

Bei der Videographie handelt sich um die Analyse von Videoaufzeichnungen im Rahmen dessen, was in Anlehnung an die Konversationsanalyse *natürliche Situation* genannt wird. Um das Missverständnis des Naturalismus oder Positivismus zu vermeiden, sollte man genauer sagen: Es handelt sich um audio-visuelle Protokollierungen von sozialem Handeln in Situationen, die nicht, wie das wissenschaftliche Experiment, das Interview oder der Fragebogen, von den untersuchenden Wissenschaftler/-innen selbst vorentworfen oder hergestellt werden, sondern den Produktionsbedingungen der untersuchten Handelnden folgen. Damit diese Bedingung erfüllt wird, verfolgen videographisch arbeitende Wissenschaftler/-innen ähnliche Strategien wie in der teilnehmenden Beobachtung bzw. in der soziologischen Ethnographie, nur dass zur Protokollierung

und Analyse (zuweilen besonderer Foki in den Feldern) Videogeräte eingesetzt werden (vgl. Knoblauch 2001).

Dieser ethnographische Zugang hat verschiedene methodologische Gründe, die auch mit der Frage der visuellen Analyse zu tun haben. Vor allem aber unterscheidet er die Videographie ebenso deutlich von anderen audio-visuellen Methoden (z. B. Medienproduktanalyse) wie er videographische Daten als Datensorte ganz entschieden von den zahlreichen anderen audio-visuellen Datensorten (z. B. Online-Tutorials) abzugrenzen erlaubt. Videographische Daten zeichnen sich dadurch aus, dass die Forschenden die Kontexte und Bedingungen der Datenerhebung aufgrund ihrer ethnographischen Arbeit sehr genau angeben können. Diese Angaben sind für die spätere Archivierung von Interesse, weil sie eine besondere Qualität und Dichte der Metadaten erlauben.²

Entsprechende videographische Erhebungen finden in einer großen Reihe unterschiedlicher Disziplinen statt, die von der Anthropologie, (Europäischen) Ethnologie und Erziehungswissenschaft über die Informatik und Kulturwissenschaften bis zur Politikwissenschaft, Soziologie und den Science and Technology Studies (STS) reichen. „Videographie“ wird häufig von eigens installierten Videolaboren betreut, wie sie sich mittlerweile an zahlreichen internationalen (z. B. Christian Heath, King's College, London; Lorenza Mondada, Universität Basel) und deutschen Universitäten finden (Bernt Schnettler, Bayreuth; Christian Meyer, Konstanz; Jürgen Raab, Landau). Die Videographie ist inzwischen an vielen Standorten und in sehr unterschiedlichen Disziplinen Teil der Ausbildung in den Qualitativen Methoden. Sie hat hier mittlerweile Lehrbuchstatus erreicht (z. B. Tuma/Schnettler/Knoblauch 2013; Heath/Hindmarsh/Luff 2010; Dinkelacker und Herrle 2009; Knoblauch und Schnettler 2007), belegt eigene Sessions in Veranstaltungen nationaler und internationaler Organisationen (z. B. International Sociological Association (ISA), European Sociological Association (ESA), Deutsche Gesellschaft für Soziologie (DGS)) und mobilisiert gerade hierzulande eine sehr aktive Forschungscommunity mit regelmäßigen Nachwuchstreffen.

3. Bedarfe

Schon die wachsende Zahl der videographischen Arbeiten wirft, bedenkt man, dass wir schon seit über 30 Jahren Videodaten erheben, die Frage auf, wie mit den Daten umgegangen werden soll. Auch wenn es bereits disziplinäre oder datensortenspezifische Ansätze und Best Practices-Beispiele für Konzepte der Nachnutzung qualitativer Daten gibt, verdeutlicht in Bezug auf audio-visuelle Forschungsdaten etwa die von Kai Wörner an der Fakultät für Geisteswissenschaften der Universität Hamburg durchgeführte Professorenbefragung zum Thema Forschungsdatenmanagement, dass sich „in vielen

² Aus diesem Grund sollen mit aviDa nur solche Datensätze adressiert werden, (a) die von Forschenden selbst im Feld methodisch erhoben wurden. Das schließt alle lediglich von den Massenmedien aufgezeichnete Materialien aus, da diese hochgradig unübersichtlich, in ihrer Qualität schwer bestimmbar und zumeist nur unzulänglich mit Metadaten versehen sind. Berücksichtigt werden sollen zudem nur solche Daten, (b) für deren Erhebung Forschende selbst Feldkontakte hergestellt und ausführliche Feldinformationen (ethnographische Daten) erhoben haben. Dies schließt die systematische Erhebung von Metadaten mit ein, die nicht nur den Kontext der Daten, sondern auch analytische Merkmale (Sampling, Kodierung etc.; ausführlich in Tuma/Schnettler/Knoblauch 2013) und die Einwilligungen bzw. Nutzungsrechte umfassen. Drittens (c) sollen ausschließlich Daten aufgenommen werden, die im Rahmen von wissenschaftlichen Untersuchungen erhoben wurden und daher konkrete wissenschaftliche Fragestellungen verfolgen sowie ausgewiesene Analysemethoden verwenden.

[...] Fächern Sammlungen von Videomaterial [finden], für die es bisher keine fortgeschrittene Repositoriumslösung gibt“ (Wörner 2015: 3).³

Deutschlandweit gibt es mittlerweile zwar mehrere Datenarchive für qualitative Forschungsdaten, wie etwa das Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO) in Bielefeld (Gebel und Liebig 2013), das Datenzentrum für qualitative Daten in den Sozialwissenschaften (QualiService) in Bremen (Kretzer 2013), das Archiv „Deutsches Gedächtnis“ in Hagen (Leh 2013) und eLabour – Interdisziplinäres Zentrum für IT-basierte qualitative arbeitssoziologische Forschung (Dunkel und Hanekop 2018). Allerdings zeigt sich, dass sich die Archivierung qualitativer Forschungsdaten, die sich national wie international schon weit aus dem explorativen Stadium herausentwickelt hat und rasch fortschreitet (Corti 2013; Krügel und Ferrez 2013), weitgehend auf sprachlich-schriftliche Datensorten wie Interviews und Transkriptionen oder Audioaufzeichnungen gesprochener Sprache beschränkt. Erste darüberhinausgehende Ansätze werden derzeit am Forschungsdatenzentrum (FDZ) Bildung am Deutschen Institut für Internationale Pädagogische Forschung (DIPF) in Frankfurt am Main (Bambey und Rittberger 2013) und am „Archiv für Gesprochenes Deutsch“ (FDZ AGD) des Instituts für Deutsche Sprache (IDS) in Mannheim beispielgebend entwickelt und praktiziert. Beide FDZ sind vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditiert. Die dortigen Ansätze sind bisher jedoch sehr stark disziplinär (etwa ausschließlich auf erziehungswissenschaftliche Daten, wie auch im videobasierten Fallarchiv Hilde der Universität Hildesheim) oder sprachspezifisch (weitgehend gesprochenes Deutsch, wie im IDS) beschränkt. Grundsätzlich bleibt die Frage der Archivierung audio-visueller Forschungsdaten der Qualitativen Sozialforschung bislang weitgehend ausgespart, sodass davon auszugehen ist, dass „große Datenmengen für eine Nachnutzung verloren gehen“ (Smioski 2013, Absatz 34). Enden diese Daten in vergessenen Schubladen irgendwelcher Universitätsschränke? Oder werden sie zu einer Art Daten- und Speichermedienmuseum, das die Wände von Arbeitszimmern ziert? Wie also, stellt sich daher die Frage, sollten videographische Daten zukünftig archiviert werden – und vorrangig: Warum überhaupt sollte man sie archivieren? Wir sehen dafür wenigstens die folgenden drei wichtigen Gründe.

3.1. Nachhaltigkeit

Die Empfehlungen von DFG, BMBF und Wissenschaftsrat zur nachhaltigen Sicherung und Nachnutzung von Forschungsdaten wurden vom RatSWD aufgenommen und die Forderung nach einer Archivierung von Daten der Qualitativen Sozialforschung ausdrücklich befürwortet. Die Empfehlungen wurden im Rahmen mehrerer sozialwissenschaftlicher Tagungen diskutiert (z. B. Huschka et al. 2013) und es wird allgemein eine umsichtige Einführung von Dateninfrastrukturen für qualitative Daten gefordert, die sowohl die Rechte der Forschenden als auch die der Erforschten berücksichtigt (Schaar 2016). Auch wenn die Archivierung nicht zwingend gemacht werden darf, da die Weitergabe der Daten u. a. auch die Frage des geistigen Eigentums berührt, so ist es dennoch notwendig, für die qualitative Forschung eine Infrastruktur zu schaffen, um Forschungsdaten sichern, nachvollziehbar und nachnutzbar machen zu können.

³ Unsere eigenen Recherchen in den einschlägigen internationalen Online-Repositorien-Verzeichnissen OpenDOAR (Directory of Open Access Repositories), ROAR (Registry of Open Access Repositories) sowie re3data (Registry of Research Data Repositories) ergaben, dass keiner der insgesamt 77 für die Begriffskombination(en) aus ‚Sozialwissenschaften‘ („social science“ und/oder „humanities“) und ‚audio-visuelle Daten‘ („audio-visual data“ und/oder „video“) automatisch generierten Einträge auf ein sozialwissenschaftliches Repositorium für audio-visuelle Forschungsdaten verweist.

3.2. Forschungsdatenmanagement und Güte

Die Archivierung von Forschungsdaten ist kein isolierter Akt, der erst posthoc ausgeführt wird, um die Daten nach getaner Arbeit für die Ewigkeit bereitzustellen. Vielmehr stellt das Management, gerade vor dem Hintergrund methodologischer Beweis- und Belegpflichten,⁴ der Forschungsdaten immer schon auch einen ersten Archivierungsschritt dar. Im Rahmen eines Repositoriums für videographische Daten, wie die geplante Forschungsdateninfrastruktur aviDa, sollte daher eine an diese vorgängigen Arbeitsprozesse von Forscher/-innen anschlussfähige, digitale und webbasierte Repräsentationsform audio-visueller Forschungsdaten sowie ein elaboriertes Metadaten-Konzept⁵ (vgl. nestor-Arbeitsgruppe Media 2016: 74 f.) entwickelt werden, das alle relevanten kontextuellen sowie methodologisch-analytischen Metadaten eines Datensatzes umfasst und abbildet. Insbesondere müssen in diesem Zuge für jeden Datensatz sämtliche Informationen erfasst werden, die für die „intersubjektive Nachvollziehbarkeit“ (Behnke/Baur/Behnke 2006: 444) des Erhebungsprozesses notwendig sind. So kann u. a. auch die Beurteilung von Forschungsdaten und -ergebnissen entlang der etablierten Gütekriterien Qualitativer Forschung (vgl. Steinke 2010) ermöglicht und die Methodenreflexion durch Primär- und Sekundäranalyt/-innen videographischer Daten gefördert werden.

3.3. Sekundäranalyse

Wie bei den natürlichsprachlichen Audiodaten, die etwa im IDS gesammelt werden, um die Formen und Veränderungen der deutschen Sprache zu dokumentieren, können auch die videographischen Daten für Sekundärnutzungen zur Verfügung gestellt werden. Diese Sekundärdatenverwendung ist im Feld der Qualitativen Sozialforschung zwar noch umstritten, doch möchten wir auf einige positive Beispiele aus der Forschung am Fachgebiet für Allgemeine Soziologie der TU Berlin verweisen. Hier wurde bereits im Jahr 2003 das erste deutsche Labor für qualitative sozialwissenschaftliche audio-visuelle Daten eröffnet. Seither ist es entscheidend an der Methodenentwicklung in der Qualitativen Sozialforschung beteiligt und darf als eines der internationalen Zentren für Videographie gelten (z. B. Knoblauch/Tuma/Schnettler 2015; Knoblauch 2012; Knoblauch 2011; Knoblauch und Tuma 2011). Das Videolabor ist mit einem festen Modul ein wichtiger Teil der soziologischen Methodenausbildung an der TU Berlin. Neben der Betreuung zahlreicher Qualifikationsarbeiten bildet es das methodische Zentrum für eine Reihe drittmittelfinanzierter Forschungsprojekte von DFG, ESF und VolkswagenStiftung („Publikumsemotionen in Sport und Religion“, DFG seit 2016; „Bildkommunikation in der Wissenschaft am Fallbeispiel der Computational Neuroscience“, DFG 2013–2017; „Tod und tote Körper“, VolkswagenStiftung 2008–2015; „Emotionalisierung der Religion“, FU Exzellenzcluster „Languages of Emotion“, 2011–2014; „PowerPoint. Die kommunikative Konstruktion der Wissensgesellschaft“, DFG 2010–2012; EUROQUAL. „Qualitative Research in the Social Sciences“, ESF 2006–2010; „Performanz visuell unterstützter mündlicher Präsentationen“, DFG 2004–2007). Aus diesen Forschungskontexten verfügt das Labor über eine

⁴ Die Daten belegen, dass die Arbeit durchgeführt wurde; sie stellen die empirische Basis der Forschung dar. Die Bedeutung der Belegpflicht wird noch dadurch verstärkt, dass Videoanalysen ja keineswegs nur Fallanalysen einzelner Auszüge darstellen; im Falle der ‚analytischen Induktion‘ von Katz (2001) und unseres an der Grounded Theory orientierten ethnographischen Samplings (Tuma/Schnettler/Knoblauch 2013) besteht ein Korpus keineswegs nur aus den wenigen Fällen, die in den zumeist nur papiernen Veröffentlichungen in sehr gekürzter Form zugänglich sind. Der Zugang zu einem breiten Korpus erst kann die „quick and dirty“-Analysen von systematischen Feldanalysen unterscheiden. Dieser Korpus kann aber nur in Form eines elaborierten Forschungsdatenrepositoriums zufriedenstellend zugänglich gemacht werden.

⁵ Die Erstellung von Metadaten kann unseres Erachtens selbst zur wissenschaftlichen Reflexion der qualitativen Forschung und damit zur Fortentwicklung ihrer reflexiven Methodologie (Knoblauch 2000) beitragen, also der Klärung der Frage, wie die Forschung praktisch betrieben wird.

große Anzahl audio-visueller Daten, die bislang grob erfasst und rudimentär sortiert wurden. Dank dieser Datenbestände und der intensiven Nutzung bildet das Videolabor einen idealen Ort für Sekundäranalysen im Bereich videographischer Daten. Derzeit werden drei umfangreiche Korpora des Labors in Projekten nachgenutzt bzw. für die Nachnutzung vorbereitet. Hierzu zählen audio-visuelle Daten zu religiösen Großveranstaltungen,⁶ zu Kontrollräumen⁷ und zu PowerPoint-Präsentationen.⁸ Es muss an dieser Stelle allerdings nochmals wiederholt werden, dass diese Daten nur nutzbar sind, weil sie von Wissenschaftler/-innen methodisch kontrolliert erhoben wurden und vor allem, weil sie mit den für die neue Untersuchung relevanten Metadaten versehen sind. Darauf werden wir auch unten (siehe 4.) nochmals zurückkommen.

4. Anforderungen

Vor dem Hintergrund der in 3. genannten Bedarfe fand bei der DFG bereits im Jahr 2010 eine Veranstaltung über Datenarchivierung statt, in der auch Archive qualitativer Methoden diskutiert wurden.⁹ Die Debatte aufgreifend, formulierten Hubert Knoblauch und Heike Solga (2011), die dem Anliegen sehr positiv gegenüberstand, schon 2011 ein erstes Papier. In der Folgezeit wurden Anforderungen an ein Forschungsdatenrepositorium für qualitative Daten formuliert, die an dieser Stelle z.T. in nur leicht veränderter und aktualisierter Form subsummiert werden sollen (vgl. Knoblauch 2013: 29ff). Unser Hauptaugenmerk liegt hierbei allerdings auf der Berücksichtigung der Besonderheiten Qualitativer Sozialforschung im Bereich videographischer Ansätze.

Daher gilt für die Einführung qualitativer Forschungsdatenrepositorien im Allgemeinen und audio-visueller Repositorien im Besonderen einerseits zwar sicherlich, dass von den bestehenden quantitativen Forschungsinfrastrukturen gelernt werden kann (vgl. Solga 2013: 19–24). Doch sollte bei der Entwicklung und Einführung von Archiven andererseits, auf die (aus der oben angedeuteten Besonderheit der videographischen Methoden abgeleiteten) besonderen Erfordernisse qualitativer Forschung mit audio-visuellen Daten geachtet werden. Im Folgenden möchten wir daher auf einige dieser Besonderheiten hinweisen und die Anforderungen andeuten, die sich daraus für die Entwicklung von Forschungsdateninfrastrukturen für audio-visuelle Daten der Qualitativen Sozialforschung ergeben. Gleichzeitig wollen wir dabei erläutern, welche Maßnahmen in aviDa ergriffen werden sollen, um den beschriebenen Herausforderungen und Problemen zu begegnen.

Die Einführung von Infrastrukturen für audio-visuelle Daten kann keineswegs dem einheitlichen Modell quantitativer Infrastruktureinrichtungen folgen. Während diese sich an einer einheitlichen Datensorte (numerischer Form) orientieren und prinzipiell mathematisch-statistische Verfahren verwenden, zeichnen sich qualitative Daten sozusagen naturgemäß durch ihre deutlich größere Vielfalt aus. Standardisierung, als ein weitgehend durchgängiges Merkmal des Umgangs mit quantitativen Daten, ist nur begrenzt dienlich, wenn es um qualitative Forschung geht. Da sich die qualitativen Daten durch eine höhere Vielfalt hinsichtlich ihrer Medialität, Materialität und Zeichenform, ihrer kontextuellen Spezifität und ihres höheren Interpretationsbedarf deutlich von quantitativen (bzw.

⁶ Nachnutzung im Rahmen des Forschungsprojekts „Publikumsemotionen in Sport und Religion“ im Rahmen des DFG-Sonderforschungsbereichs 1171 „Affective Societies“.

⁷ Nachnutzung im Rahmen des Forschungsprojekts „Zentren der Koordination“ im Rahmen des DFG-Sonderforschungsbereichs 1265 „Re-Figuration von Räumen“.

⁸ Geplante Nachnutzung in einem vorgesehenen Forschungsprojekt zur Wissenschaftstheorie im Rahmen des DFG-Sonderforschungsbereichs 1171 „Affective Societies“.

⁹ Hubert Knoblauch war zum damaligen Zeitpunkt Kollegiat für qualitative empirische Sozialforschung und wurde später zum ständigen Gast im RatSWD ernannt. Nach zwei Jahren in dieser Rolle initiierte er eine dauerhafte Beteiligung durch eine Vertreterin bzw. einen Vertreter Qualitativer Methoden im RatSWD. Diese Position nimmt derzeit Jörg Strübing ein.

„standardisierten“) Daten unterscheiden. Deswegen sollte zwischen den besonderen Merkmalen von qualitativen Daten, also den jeweiligen Datensorten, und dem allgemeinen Begriff digitalisierter Daten unterschieden werden.

Weil und sofern die qualitative Sozialforschung nicht-standardisierte Daten erhebt, besteht eine besondere Herausforderung in der Entwicklung standardisierter Metadaten, die es erlauben, Daten gezielt wieder zu finden, adäquat zu kontextualisieren und sekundär zu nutzen. In diesem Sinne sollen auch die einzelnen Elemente von aviDa im Rahmen eng verzahnter methodischer wie technischer Arbeitspakete sowie im Sinne eines nutzerzentrierten Entwicklungsansatzes (vgl. nestor-Arbeitsgruppe Media 2016: 70; Rat für Informationsinfrastrukturen 2017: 28) konzipiert, erarbeitet und schrittweise in das bestehende Forschungsdatenrepositorium der TU Berlin (DepositOnce) implementiert werden. Dies soll in enger Abstimmung zwischen dem Videolabor des Fachgebiets Allgemeine Soziologie (TUB) und dem Videoanalyse-Labor am Lehrstuhl für Kultur- und Religionssoziologie der Universität Bayreuth (UBT) als zentrale Testnutzer einerseits sowie den technischen Entwicklern der Universitätsbibliothek und des IT-Service-Centers der TU Berlin andererseits erfolgen.

Eine bereits in einigen Feldern praktizierte Lösung zum Umgang mit der Vielfalt qualitativer Daten kann darin bestehen, Forschungsinfrastrukturen für qualitative Methoden an der Art der zu archivierenden Datensorte festzumachen. Diese Vorgehensweise liegt bislang auch daher nahe, da sich sowohl die wissenschaftlichen Anforderungen an unterschiedliche Datensorten wie auch die technischen Anforderungen (z. B. Softwareprogramme) stark nach Datensorten unterscheiden. Videographische Daten unterscheiden sich nicht nur in ihrer Produktion, sondern auch in ihrer Medialität und Modalität von Interviews, Fragebögen oder auch massenmedialen Daten oder Internetdaten wie z. B. nutzergenerierte Inhalte. Diese Unterschiede der Datensorten müssen sowohl rechtlich wie auch ethisch beachtet werden: Die Forderung nach „informed consent“ lässt sich nicht auf diskursive Mediendaten oder öffentliche audio-visuelle Daten übertragen, bei denen wissenschaftlicher Forschung mindestens dieselben Rechte eingeräumt werden wie etwa den Massenmedien.

Videographische Daten schließen ethnographische Daten ein. Eine befriedigende Archivlösung für diese Daten muss daher dazu in der Lage sein, die digitale Verknüpfung verschiedener zu archivierender Datensorten, die im videographischen Forschungsprozess erhoben wurden, zu ermöglichen. Daher wurde die zugrundeliegende Softwarelösung (DSpace) für aviDa so gewählt, dass sie grundsätzlich mit verschiedensten Arten von Datentypen, Dateiformaten, Metadatenschemata und Klassifikationen umgehen kann. Außerdem wird das Standard-Metadatenschema Dublin Core im Rahmen der technischen Entwicklung um disziplinspezifische Felder ergänzt und erweitert werden.

Viele Daten der qualitativen Sozialforschung sind in außerordentlichem Maße sensibel. Sie zeichnen sich dadurch aus, dass (entweder) die Forschenden sehr entscheidend an ihrer Herstellung beteiligt sind (etwa Feldprotokolle) und (oder) dass sie sehr stark kontextuell verankert (z. B. Teilnehmende Beobachtung) und entsprechend schwer anonymisierbar sind (z. B. audio-visuelle Daten). Vor dem Hintergrund der Sensibilität dieser Daten ist es sehr verständlich, dass sich qualitative Forscher/-innen schon im Vorfeld der hier besprochenen Initiative gegen eine mögliche Verpflichtung zur Bereitstellung ihrer Daten ausgesprochen haben. Während die Bereitstellung der Daten von der Forschungsförderung ermöglicht und von den Forschungsinfrastruktureinrichtungen (hinsichtlich etwa technischer Qualitätsanforderungen) geprüft werden sollte, muss die Entscheidung über die Bereitstellung der Daten bei den Forschenden liegen, die für die Datenerhebung verantwortlich zeichnen (Archivierungsangebot und -möglichkeit, nicht Pflicht!). Zusätzlich sollte zwischen

verschiedenen Arten der Archivierung unterschieden werden: (a) Daten, die keinerlei weitere Verwendung erlauben; (b) Daten, deren Metadaten recherchierbar sind, die aber (derzeit) nicht weiter verwendet werden können; (c) Daten, die nur eingeschränkt verwendet werden dürfen; (d) Daten, die in der wissenschaftlichen Community frei verwendet werden dürfen, und schließlich (e) Daten, die ohne Einschränkungen öffentlich zugänglich sind.

Natürlich stellen auch datenschutzrechtliche Vorgaben große Anforderungen an die Entwicklung qualitativer und insbesondere audio-visueller Forschungsdatenrepositorien. Neben den Rechten der Forschenden an ihrer eigenen Forschung gilt es selbstverständlich die Rechte der Beforschten zu wahren. Eine zentrale Aufgabe der Entwicklung entsprechender Archive wie aviDa besteht daher darin, ein elaboriertes Rechte- und Zugriffsmanagement zu entwickeln, das die Datenschutzrechte der videographierten Akteure nicht verletzt und zugleich den Voraussetzungen der qualitativen Videoanalyse genügt. Hierfür kommen Embargofunktionen und graduelle Zugänge ebenso in Frage wie die Einbeziehung verschiedener technischer Methoden der Anonymisierung der (audio-visuellen) Daten, die im Rahmen kleiner Nutzungsstudien auch an die Bedürfnisse der Forschungscommunity angepasst werden müssen.

Nicht nur die Datensorten, auch die Methoden der qualitativen Sozialforschung zeichnen sich durch eine enorme Vielfalt aus, die sich stark in Netzwerken organisiert und von einer großen Innovationsdynamik geprägt ist. Um die Schaffung teurer Datenfriedhöfe zu vermeiden, deren Daten von den Forschenden nicht genutzt werden, ist der durchgängige, forschungspraktische und methodenpluralistische Einbezug der qualitativ Forschenden in die Entwicklung der Infrastruktur sicherzustellen. Nur so können die nachhaltige Akzeptanz und die Dissemination in der Forschungscommunity erreicht werden. Im Rahmen von aviDa ist daher eine weite Beteiligung der Forschungscommunity im Rahmen gemeinsamer Datensitzungen mit Testnutzer/-innen sowie von Workshops vorgesehen. Trotzdem beschränkt sich das Projekt in seiner ersten Phase bewusst auf Qualitative Sozialforschung, um der bereits in diesem Bereich ausgeprägten Methodenvielfalt Rechnung tragen zu können. Eine Öffnung für Daten aus weiterer qualitativer Forschung mit audio-visuellen Daten ist erst in einer späteren Förderphase des Projekts vorgesehen.

Ebenso vielfältig wie die Datensorten und Forschungsmethoden im Bereich der qualitativen Forschung, insbesondere auch innerhalb der Videographie und anderer visueller Verfahren, sind auch die technischen Formate, in denen audio-visuelle Daten erhoben, abgespeichert und aufbereitet werden. In der videographischen Praxis gibt es keine technischen Standards, die die Kompatibilität der Formate und Codecs der Forschungsdaten a priori sicherstellen. Daher müssen bei der Entwicklung eines Repositoriums für audio-visuelle Forschungsdaten der Qualitativen Sozialforschung auch Tools entwickelt werden, die eine automatische Transkodierung im Rahmen des Uploads auf die entsprechende Plattform ermöglichen. Auf diese Weise wird nicht allein der Herausforderung durch die Vielfalt an Dateiformaten etc., sondern auch der kurzen Halbwertszeit von Formaten und Codecs begegnet. Schon heute ist es so, dass ein nicht kontinuierlich betreuter Datenkorpus, bereits nach wenigen Jahren, häufig nur unter Schwierigkeiten erneut ausgelesen werden kann. Ein Repositorium, das, wie für aviDa vorgesehen, ein Transkodierungstool anbietet, kann so zur Vereinheitlichung und kontinuierlichen Aktualisierung der technischen Formate beitragen.

In diesem Kontext muss – last but not least – auch das Erfordernis der Nachhaltigkeit von Forschungsdateninfrastrukturen betrachtet werden. Die Entwicklung entsprechender Lösungen, wie sie hier besprochen werden, müssen nachhaltig konzipiert werden, sodass ihre technische Betreuung und Finanzierung auf Dauer gewährleistet sind. Hierzu bieten sich vor allem Kooperation mit öffentlichen Expert/-inneneinrichtungen aus den Bereichen Archivierung und Informatik an, wie etwa, in unserem

Fall, mit der Universitätsbibliothek (UB) sowie dem IT-Service-Center (tubIT) der TU Berlin. Auf Grundlage dieser institutionellen Einbettung soll aus sozialwissenschaftlicher Perspektive die dauerhafte Server-Pflege und weitere technische Betreuung des geplanten Repositoriums sichergestellt werden. Mit den Videolaboren der Allgemeinen Soziologie der TU Berlin sowie am Lehrstuhl für Kultur- und Religionssoziologie der Universität Bayreuth wird gleichzeitig eine nachhaltige sozialwissenschaftlich-methodologische Betreuung von aviDa garantiert. Auf diese Weise soll gesichert werden, dass die aufzubauende Lösung zum Zweck der Archivierung und Nachnutzbarkeit von audio-visuellen Daten über die übliche Laufzeit von Drittmittelprojekten hinaus Bestand hat und so kontinuierlich weiterentwickelt werden kann, um der methodologischen Dynamik der Forschungscommunity dauerhaft gerecht werden zu können.

Literaturverzeichnis

- Bambey, Doris und Marc Rittberger (2013): Das Forschungsdatenzentrum (FDZ) Bildung des DIPF: Qualitative Daten der empirischen Bildungsforschung im Kontext. In: Denis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die Qualitative Forschung*. Berlin, SCIVERO Verlag, 63–72.
- Behnke, Joachim; Nina Baur und Nathalie Behnke (2006): *Empirische Methoden der Politikwissenschaft*. Paderborn u. a., Ferdinand Schöningh.
- Corti, Louise (2013): Infrastructures for qualitative data archiving. In: Denis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die Qualitative Forschung*. Berlin, SCIVERO Verlag, 35–62.
- Dinkelacker, Jörg und Matthias Herrle (2009): *Erziehungswissenschaftliche Videographie. Eine Einführung*. Wiesbaden, VS Verlag für Sozialwissenschaften.
- Dunkel, Wolfgang und Heidemarie Hanekop (2018): FDZ für qualitative Forschungsdaten in der Arbeits- und Industriesoziologie: das interdisziplinäre Zentrum eLabour. RatSWD Working Paper 267/2018. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD), 36–46. <https://doi.org/10.17620/02671.35>.
- Gebel, Tobias und Stefan Liebig (2013): Die Dokumentation qualitativer Interviews im Rahmen von Organisationsstudien. In: Denis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die Qualitative Forschung*. Berlin, SCIVERO Verlag, 73–92.
- Heath, Christian; John Hindmarsh und Paul Luff (2010): *Video in Qualitative Research*. Los Angeles u. a., Sage Publications.
- Huschka, Denis; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die Qualitative Forschung*. Berlin, SCIVERO Verlag.
- Katz, John (2001): Analytic Induction. In: Neil J. Smelser und Paul B. Baltes (eds.): *International Encyclopedia of the Social and Behavioral Sciences*. Oxford, Elsevier Science, 1–480.
- Knoblauch, Hubert (2001): Fokussierte Ethnographie. *sozialersinn* 1/2001, 123–141.
- Knoblauch, Hubert (2011): Videography. In: Michael Stausberg und Steven Engler (eds.): *The Routledge Handbook of Research Methods in the Study of Religion*. London, Routledge, 433–444.
- Knoblauch, Hubert (2012): Video analysis and videography. *Qualitative Research* 12(3), 251–254.
- Knoblauch, Hubert (2013): Einige Anforderungen an Forschungsinfrastrukturen aus der Sicht der qualitativen Forschung. In: Denis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die Qualitative Forschung*. Berlin, SCIVERO Verlag, 27–34.
- Knoblauch, Hubert und Bernt Schnettler (2007): Videographie. Erhebung und Analyse qualitativer Videodaten. In: Renate Buber und Hartmut H. Holzmüller (Hrsg.): *Qualitative Marktforschung. Konzepte, Methoden, Analysen*. Wiesbaden, Gabler Verlag, 583–599.
- Knoblauch, Hubert und Heike Solga (2011): Thesen zur Handhabung quantitativer und qualitativer in Forschungsinfrastrukturen der Sozialwissenschaften: Ist eine Integration möglich? RatSWD Working Paper 190/2011. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). https://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_190.pdf (Zugriff am 18.06.2018).
- Knoblauch, Hubert und René Tuma (2011): Videography. An Interpretive Approach to Video-Recorded Micro-Social Interaction. In: Eric Margolis und Luc Pauwels (eds.): *The Sage Handbook of Visual methods*. Los Angeles, Sage, 414–430.
- Knoblauch, Hubert; René Tuma und Bernt Schnettler (2015): *Videography: Introduction to Interpretive Video-Analysis of Social Situations*. Frankfurt am Main, Peter Lang.
- Kretzer, Susanne (2013): Vom Nutzen des Datasharing für die Lehre in der qualitativen Sozialforschung. In: Denis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die Qualitative Forschung*. Berlin, SCIVERO Verlag, 153–166.
- Krügel, Sybil und Eliane Ferrez (2013): Sozialwissenschaftliche Infrastrukturen für die qualitative Forschung – Stand der Integration von qualitativen Daten bei DARIS (FORS). In: Denis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): *Forschungsinfrastrukturen für die Qualitative Forschung*. Berlin, SCIVERO Verlag, 113–126.

- Leh, Almut (2013): Das Archiv „Deutsches Gedächtnis“ und seine Bestände: Herkunft – Erschließung – Nutzung. In: Denis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): Forschungsinfrastrukturen für die Qualitative Forschung. Berlin, SCIVERO Verlag, 127–136.
- Marks, Dan (1995): Ethnography and Ethnographic Film. From Flaherty to Asch and after. *American Anthropologist*, New Series 97(2), 339–347.
- nestor-Arbeitsgruppe Media (2016): Leitfaden für die digitale Langzeitarchivierung audiovisueller Medien. Nestor-materialien 19. <http://nbn-resolving.de/urn:nbn:de:0008-2016102107> (Zugriff am 18.06.2018).
- Pittenger, Robert E.; Charles F. Hockett und John J. Danehy (1960): The first five minutes: A sample of microscopic interview analysis. Oxford, Paul Martineau.
- Rat für Informationsinfrastrukturen (2017): Entwicklung von Forschungsdateninfrastrukturen im internationalen Vergleich. Bericht und Anregungen. Göttingen. <http://www.rfii.de/download/rfii-fachbericht-laenderanalysen-2017/> (Zugriff am 12.10.2018).
- Reichert, Ramón (2007): Im Kino der Humanwissenschaften. Studien zur Medialisierung wissenschaftlichen Wissens. Berlin, Transcript.
- Schaar, Katrin (2016): Was hat die Wissenschaft beim Datenschutz künftig zu beachten? Allgemeine und spezifische Änderungen beim Datenschutz im Wissenschaftsbereich durch die neue Europäische Datenschutzgrundverordnung. RatSWD Working Paper 257/2016. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). http://www.ratswd.de/dl/RatSWD_WP_257.pdf (Zugriff am 18.06.2018).
- Smioski, Andrea (2013): Archivierungsstrategien für qualitative Daten [54 Absätze]. *Forum Qualitative Sozialforschung* / *Forum: Qualitative Social Research* 14(3), Art. 5.
- Solga, Heike (2013): Was lässt sich von den Forschungsinfrastrukturen in der empirischen Sozial- und Wirtschaftswissenschaft für qualitative Forschungsinfrastrukturen lernen? In: Denis Huschka; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.): Forschungsinfrastrukturen für die Qualitative Forschung. SCIVERO Verlag, Berlin, 19–26.
- Steinke, Ines (2010): Gütekriterien qualitativer Forschung. In: Uwe Flick; Ernst von Kardoff und Ines Steinke (Hrsg.): *Qualitative Forschung. Ein Handbuch*. 8. Auflage, Reinbek, Rowohlt, 319–331.
- Tuma, René; Bernt Schnettler und Hubert Knoblauch (2013): Videographie. Einführung in die interpretative Videoanalyse sozialer Situationen. Wiesbaden: Springer VS.
- Wilke, René (2018): Das Interpretations-Bild. In: Christine Moritz und Michael Corsten (Hrsg.): *Handbuch Qualitative Videoanalyse*. Wiesbaden, VS Verlag für Sozialwissenschaften, 485–500.
- Wörner, Kai. (2015): Auswertung der Professorenfrage zum Konzept eHumanities 2020+. Universität Hamburg. https://www.dokserv.gwiss.uni-hamburg.de/servlets/MCRFileNodeServlet/mir_derivate_00000027/umfrage-dok.pdf (Zugriff am 12.10.2018).

Prof. Dr. Hubert Knoblauch, is Professor of General Sociology at the Technical University of Berlin. 1997 Christa-Hoffmann-Riem-Award for Qualitative Research; 2003–2005 Chair ESA RN Qualitative Methods; Board Member of FQS, Qualitative Sociology, Qualitative Review etc. Research topics range from social theory to the sociology of religion and knowledge and, of course, qualitative methods, particularly ethnography and videography. Publications in Qualitative Methods include: H. Knoblauch, B. Schnettler, R. Tuma: Videography. New York: Lang 2014. H. Knoblauch, A. Baer, E. Laurier, S. Petschke & B. Schnettler, eds, Visual Analysis. New Developments in the Interpretative Analysis of Video and Photography. Forum: Qualitative Social Research, 9(3) (2008).

René Wilke ist wissenschaftlicher Mitarbeiter in dem DFG-Projekt „aviDa – Entwicklung einer Forschungsdateninfrastruktur für audio-visuelle Daten der Qualitativen Sozialforschung“. Dort ist er insbesondere mit der Projektkoordination betraut. Zuvor war er wissenschaftlicher Mitarbeiter in dem DFG-geförderten Forschungsprojekt „Bildkommunikation in der Wissenschaft am Fallbeispiel der Computational Neuroscience“, wo er insbesondere mit der Videographie betraut war. In seinem Promotionsprojekt beschäftigt er sich empirisch mit Fragen der wissenssoziologischen Kommunikationsforschung.

Arbeitsschwerpunkte: Videographie, wissenssoziologische Kommunikationsforschung, visuelle Soziologie, fokussierte Ethnographie

Bereitstellung und Nachnutzung qualitativer Daten in der Bildungsforschung. Das Forschungsdatenzentrum (FDZ) Bildung am DIPF

Doris Bambey, Alexia Meyermann, Maike Porzelt und Marc Rittberger

Abstract

Die pluridisziplinäre empirische Bildungsforschung stellt den Bezugsrahmen für die Services des FDZ Bildung und anderer Datenzentren mit Relevanz für das Forschungsgebiet dar und wird vor diesem Hintergrund in wesentlichen Aspekten skizziert. Betrachtet wird die dynamisierte Forschungsentwicklung v. a. der letzten Dekade und in diesem Kontext die Bedeutung der qualitativen Bildungsforschung, aktuelle förderpolitische Regelungen der Datennachnutzung sowie die Diskurse und Stellungnahmen der Fachcommunity zum nachhaltigen Umgang mit qualitativen Forschungsdaten.

Die wesentliche Maßnahme zur Entwicklung einer übergreifenden und koordinierten Forschungsdateninfrastruktur im Forschungsgebiet stellt der Verbund Forschungsdaten Bildung (VerbundFDB) dar. Er wird von den drei Kernpartnern DIPF, GESIS und IQB und weiteren bildungsrelevanten Datenzentren getragen. Mit Blick auf seine Beiträge zur Professionalisierung des Datenmanagements in den Forschungsprojekten wird der implementierte Verbund-Workflow zur förderierten Archivierung und Bereitstellung von qualitätsgesicherten Forschungsdaten der qualitativen und quantitativen Forschung dargestellt.

Das FDZ Bildung ist Partner des VerbundFDB und hat in diesem Kontext die Verfügbarmachung von qualitativen Forschungsdaten der Bildungsforschung resp. der videobasierten Schul- und Unterrichtsforschung im Fokus. Dargelegt werden das am FDZ Bildung praktizierte Vorgehen der Aufbereitung und Dokumentation qualitativer Forschungsdaten sowie die rechtssichere Organisation der Zugangswege zu den Daten und ihrer Nachnutzung.

Abschließend werden relevante Handlungsfelder des FDZ Bildung bzw. infrastrukturelle und nutzungsbezogene Anforderungen an die Datenzentren, die qualitative Forschungsdaten bereitstellen, benannt.

1. Forschungsdaten der Bildungsforschung

Die Bildungsforschung zeichnet sich durch eine ausgeprägte interdisziplinäre Ausrichtung aus, indem sich verschiedene Fachrichtungen, wie die Erziehungswissenschaft, Psychologie, Soziologie, Ökonomie, Philosophie, Rechtswissenschaft, Linguistik etc., mit jeweils spezifischen Fragestellungen in das Forschungsgebiet einbringen. Eine entsprechend multiperspektivische Sicht entspricht dem Erkenntnisinteresse, „Voraussetzungen, Prozesse und Ergebnisse von Bildung über die Lebensspanne, und zwar innerhalb und außerhalb von (Bildungs-)Institutionen und im gesellschaftlichen Kontext“ (vgl. Prenzel 2006) zu verstehen und zu optimieren.

Ansätze der qualitativen empirischen Bildungsforschung sind im Forschungsfeld seit langem fest verankert und können gerade in den letzten Jahrzehnten auf einen sichtbaren Bedeutungszuwachs in allen Segmenten des Erziehungs- und Bildungswesens verweisen (vgl. u. a. die Beiträge in Friebertshäuser/Langer/Prenzel 2013; Schäffer und Dörner 2012; Schweppe 2003). Auch in dem Ende 2007 gestarteten ersten BMBF-Rahmenprogramm zur Förderung der empirischen Bildungsforschung ist der hohe Stellenwert qualitativer Forschungsansätze unübersehbar. In diesem Sinne hat das Förderprogramm mit seinem Schwerpunkt in der Schul- und Unterrichtsforschung v. a. zu einer deutlichen Zunahme qualitativer, videogestützter Verfahren zur Untersuchung unterrichtlicher Lehr-Lernprozesse beigetragen.¹

Bei den erhobenen Daten der qualitativen Bildungsforschung handelt es sich im Wesentlichen um mediengestützte Dokumentationen aus der pädagogischen Praxis auf Basis video- bzw. bildbasierter oder auditiver Verfahren und deren Verschriftlichung in Form von Transkripten sowie Beobachtungsprotokolle (bspw. ethnografische Beobachtungsprotokolle), Interviews, Gruppendiskussionen oder auch Daten und Dokumente, deren Kombination aus verschiedenen Herkunftsquellen eine neue Qualität mit Blick auf die sekundäranalytische Nachnutzung darstellen.²

Die Bildungsforschung produziert somit einen vielfältigen Datenbestand, dessen Analysepotential in vielen Fällen im Rahmen des ursprünglichen Forschungsprojektes, in dem die Daten erhoben werden, kaum ausgeschöpft werden kann.³ Daneben kommt der Nachnutzung von Daten der Bildungsforschung auch aus forschungspraktischer Sicht eine hohe Bedeutung zu. So sind einzelne Bereiche des Forschungsfeldes – wie etwa die Schul- und Unterrichtsforschung – mit dem Problem eines per se schwierigen Feldzuganges konfrontiert, der sich nicht nur zeit- und kostenintensiv gestalten kann, sondern aufgrund der politisch-administrativen Umfeldbedingungen nicht selten auch gänzlich zu scheitern droht. Entsprechende Hürden entstehen insbesondere dadurch, dass aufgrund der verstärkten Forschungsaktivitäten in diesem Feld der Zugang zu den Schulen von Seiten der zuständigen Genehmigungsbehörden zusehends stärker reglementiert wird (vgl. Stanat 2014: 1).⁴ Auch ist dem Tatbestand Rechnung zu tragen, dass die (videogestützte) Unterrichtsbeobachtung einen per se

¹ Das 2017 aufgelegte zweite Rahmenprogramm geht explizit über das thematische Spektrum der Schulbildung hinaus und adressiert verstärkt non-formale und informelle Bildungs- und Lernprozesse.

² Ein prototypisches Beispiel für einen komplexen Mixed-Methods-Ansatz mit qualitativem und quantitativem Datenoutput stellt die Studie „DESI: Deutsch-Englisch-Schülerleistung International“ dar. Online unter: www.forschungsdaten-bildung.de/studiendetails.php?id=3 (Zugriff am 07.06.2017).

³ Siehe hierzu die entsprechenden Stellungnahmen aus den Fachgesellschaften der Bildungsforschung zur Archivierung und Nachnutzung quantitativer und qualitativer Daten, insb. Stanat (2014) und DGfE (2017).

⁴ Erhebungen an Schulen sind in Deutschland in der Regel genehmigungspflichtig und auf Landesebene bei den zuständigen Schulaufsichtsbehörden zu beantragen. Einen Überblick zu den Regelungen in den Bundesländern gibt der Verbund Forschungsdaten Bildung (VerbundFDB): <https://www.forschungsdaten-bildung.de/genehmigungen> (Zugriff am 16.03.2018). Exemplarisch sei die Praxis des Landesverwaltungsamtes des Bundeslandes Sachsen-Anhalt angeführt, das ein streng regulatives Verfahren als notwendig erachtet, „um durch ein Übermaß von Untersuchungen die Erfüllung des Erziehungs- und Bildungsauftrages der Schule nicht zu behindern“ (<https://bildung.sachsen-anhalt.de/schulen/schulinhalte-und-qualitaet/sonstige-inhalte/empirische-untersuchungen/>, Zugriff am 22.5.2018).

eher invasiven Ansatz darstellt und Forschungsaktivitäten im Schulkontext mit der grundsätzlichen Problematik einhergehen, dass es sich bei Schüler/innen um in der Regel minderjährige Untersuchungspersonen handelt. Gerade für Nachwuchswissenschaftler/innen oder kleinere Projekte, die über keine umfangreichen finanziellen und personellen Ressourcen verfügen, um einen aufwändigen Feldzugang zu bewältigen, aber auch generell aus Gründen der Wissenschaftseffizienz und nicht zuletzt aus ethischen Gesichtspunkten heraus kann der Rückgriff auf bereits mit hohem organisatorisch-technologischen Aufwand erhobene Daten somit einen großen Vorteil darstellen.

2. Der Verbund Forschungsdaten Bildung und das FDZ Bildung des DIPF

Die Forschungsdateninfrastruktur in der Bildungsforschung zeichnet sich durch eine hohe Diversifizierung und Spezialisierung aus. Dieses an sich hohe Maß an fachlicher Einschlägigkeit der einzelnen Datenzentren ist jedoch, aufgrund der Heterogenität und Fragmentiertheit der Angebotsstruktur, mit einer problematischen Unübersichtlichkeit verbunden. Dies führt dazu, dass sich der Nutzen und Mehrwert der vorhandenen Ressourcen für die Forschung nicht in ausreichendem Maße entfalten kann. Vor diesem Hintergrund hat der Verbund Forschungsdaten Bildung (VerbundFDB) seit Ende 2013 begonnen, auf der Basis eines Netzwerkes mit den Partnern Deutsches Institut für Internationale Pädagogische Forschung (DIPF; agiert als Koordinationsstelle), GESIS – Leibniz-Institut für Sozialwissenschaften und Institut zur Qualitätsentwicklung im Bildungswesen (IQB) einen Workflow zur förderierten Archivierung und Bereitstellung von Forschungsdaten der Bildungsforschung aufzubauen. Zu den Leistungen des VerbundFDB gehören eine FDZ-übergreifende Suche nach Datenbeständen, ein FDZ-übergreifendes Archivierungsangebot sowie speziell für Bildungsforschende entwickelte Informations- und Schulungsangebote zum Forschungsdatenmanagement.⁵ Das Modell der „verteilten Archivierung“, das der Verbund realisiert, zielt darauf, dass Forschende Daten an den Verbund übermitteln. Dieser leitet die Daten an das jeweils zuständige FDZ weiter, welches dann die fachspezifische Archivierung und Verfügbarmachung des betreffenden Bestandes übernimmt. Zuständig für qualitative Daten im VerbundFDB ist das FDZ Bildung am DIPF.⁶

Das FDZ Bildung wurde nach einer Projekteinführungszeit 2012 am DIPF in Frankfurt am Main eingerichtet, um den Zugang zu qualitativen Daten der Bildungsforschung unabhängig vom institutionellen Kontext ihrer Entstehung nachhaltig zu sichern und damit dazu beizutragen, die bis dato vorhandene Verfügbarkeitslücke in diesem Bereich zu schließen.

Das inhaltliche Profil des Angebotes beim FDZ Bildung zeichnet sich durch einen besonderen Schwerpunkt im Bereich der Schul- und Unterrichtsforschung aus und fokussiert hierbei v. a. auf Verfahren der videobasierten Unterrichtsforschung. Darüber hinaus besteht jedoch ein grundsätzliches Ziel der Angebotsentwicklung darin, über den spezifischen Schwerpunkt Schule hinaus, die gesamte Bandbreite der empirischen Bildungsforschung im Kontext des lebenslangen Lernens zu adressieren und neben formalen Lernarrangements auch non-formale und informelle Lernszenarien zu berücksichtigen.⁷ Das zeitliche Spektrum der am FDZ Bildung verfügbaren Datenbestände umfasst im Wesentlichen aktuellere Untersuchungen insbesondere der letzten zwei Dekaden, wie etwa die großen

⁵ Eine ausführliche Beschreibung des Projektes und seines Arbeitsprogrammes findet sich unter: <https://www.forschungsdaten-bildung.de/arbeitsprogramm> (Zugriff am 16.03.2018).

⁶ Siehe hierzu die Ergebnisse der ersten Projektphase in Meyermann et al (2017)

⁷ Siehe hierzu die Leitlinien des FDZ Bildung: <https://www.fdz-bildung.de/leitlinien-fdz> sowie die Beschreibung seines inhaltlichen Profils: <https://www.fdz-bildung.de/collection-policy-fdz> (Zugriff am 16.03.2018). Neben Daten der qualitativen Bildungsforschung stellt das FDZ Bildung auch Erhebungsinstrumente (Fragebogen, Items, Skalen, Testinstrumente) der quantitativen Schulqualitätsforschung bereit, auf die im vorliegenden Zusammenhang jedoch nicht näher eingegangen wird

nationalen Studien „Pythagoras: Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“ oder „DESI – Deutsch-Englisch-Schülerleistungen-International“, aber auch eine in historischer Sicht relevante Sammlung von AV-Datenbeständen aus Unterrichtsaufzeichnungen, die in den 1970er bis 1990er Jahren in der DDR und der Bundesrepublik Deutschland durchgeführt wurden und aus disparat verteilten Quellen stammen.

Im Einzelnen umfasst das Angebot des FDZ Bildung folgende Daten- und Materialarten:

- Audiovisuelle Forschungsdaten (z. B. Videoaufzeichnungen von beobachteten Unterrichtssituationen sowie Audioaufnahmen von Interviews mit Lehrkräften);
- textuelle Forschungsdaten (v. a. Transkripte der beobachteten Unterrichtsstunden und geführten Interviews);
- Kontextmaterialien zu den beobachteten Settings (z. B. Memos, Sitzpläne, Tafelbilder);
- Analysematerial (Kodierungen für videografierte Situationen aus niedrig inferenten Bewertungen von Unterrichtssequenzen sowie Ratings, Ratingskalen aus hoch inferenten Bewertungsansätzen von Unterrichtsqualität);
- Dokumentationsmaterial (Interviewleitfäden, Manuale der Videografie, Beschreibungen des Vorgehens bei der Transkription etc.).

3. Datendokumentation und Kontextualisierung als Voraussetzung für die Datennutzung

Das FDZ Bildung sieht in einer umfassenden Kontextualisierung eine Grundvoraussetzung, um Forschungsdaten nachvollziehbar und somit nachnutzbar zu machen.⁸ Diese Dokumentation des Kontexts ist erforderlich, um eine Brücke zwischen Primärforschern/-forscherinnen und Sekundärnutzern/-nutzerinnen zu schlagen und zu ermöglichen, „den institutionellen, theoretischen und methodischen Zugang der Studie, den Prozess der Datenerhebung, Entscheidungen im Verlauf des Forschungsprozesses oder die Reflexionen der PrimärforscherInnen nachzuvollziehen“ (Smioski 2013: Absatz 15). Vor diesem Hintergrund ist das FDZ Bildung bestrebt, neben den eigentlichen Roh- bzw. Primärdaten auch ein möglichst breites Spektrum an zusätzlichen Auswertungs- und Dokumentationsmaterialien zur Verfügung zu stellen.⁹

Die Umsetzung des Anspruches einer systematischen Kontextualisierung von Forschungsdaten gestaltet sich deshalb mitunter schwierig, weil Forschende bis dato eher in Ausnahmefällen eine umfangreichere Dokumentation zugehöriger Materialien zusammen mit den Daten übermitteln. So kommt es vor, dass dem Datenzentrum zwar die Transkripte geführter Interviews vorliegen, aber keinerlei weitere Informationen z. B. zur konkreten Beobachtungs- oder Befragungssituation (wie Beschreibung des Settings, Interviewleitfaden etc.) oder darüber, was im Anschluss an die Erhebung mit den Daten geschehen ist, etwa Hinweise dahingehend, wie diese transkribiert und anonymisiert wurden. Solche Informationen werden zum Teil in den Projekten gar nicht verschriftlicht und

⁸ Die Kontextabhängigkeit qualitativer Daten und deren Auswirkung auf Sekundäranalysen wird an verschiedenen Stellen diskutiert, siehe beispielsweise Medjedović 2014.

Auch die erziehungswissenschaftlichen Fachgesellschaften sehen in einer umfassenden Kontextualisierung der qualitativen Daten deshalb eine der Grundvoraussetzungen, um diese archivieren und nachnutzen zu können (vgl. Stanat 2014; DGfE 2017).

⁹ Eine intensive Diskussion über die Notwendigkeit, die interpersonale Nachvollziehbarkeit insbesondere von qualitativen Forschungsdaten durch Publikationen, Kontextmaterialien und reichhaltige Metadaten zu steigern, fand im Juli 2017 im Rahmen eines Workshops zur Machbarkeit qualitativer Sekundäranalysen statt. Der Workshop wurde gemeinsam von der DGfE und dem VerbundFDB ausgerichtet (vgl. Bambey/Meyermann/Porzelt 2017).

übermittelt, da sie im Forschungsprozess selbst jederzeit informell abrufbar bzw. den Akteuren des primären Forschungsprojektes auch ohne schriftliche Fixierung geläufig sind.

Dabei kann sich die Datendokumentation über den späteren Verwendungszweck für eine Sekundärnutzung hinaus auch für den Forschungsprozess generell sehr positiv auswirken, da auch innerhalb des ursprünglichen Forschungsvorhabens der Kontext der Datenerhebung und die Relationen zwischen den Daten interpersonell nachvollziehbar sein sollten (vgl. Smioski 2013: Absatz 23). Unmittelbar evident ist der beträchtliche Nutzen einer umfassenden Datendokumentation nicht zuletzt mit Blick auf Forschungsszenarien, die die Arbeit mit erhobenen Daten über verschiedene Standorte hinweg vorsehen oder im Falle von Mitarbeiterwechseln während der Projektlaufzeit.¹⁰

Wichtige Meta- und Kontextinformationen, die Nutzer/innen für die Nachvollziehbarkeit bzw. Zweitnutzung vorliegender Daten benötigen, entstehen auf unterschiedlichen Ebenen des Forschungsprozesses (vgl. Bishop 2006), wie Abbildung 1 illustriert. Relevante Informationen bieten sowohl der übergeordnete Studienkontext (Makro-Ebene oder extra-situationaler Kontext) und der methodische Rahmen der Datenerhebung als auch der konkrete Kontext, in dem die Daten entstanden sind (Mikro-Ebene oder situationaler Kontext). Darüber hinaus werden technisch-formale Informationen zu den entstandenen Datenobjekten und den an diesen durchgeführten Aufbereitungsschritten festgehalten.

Abbildung 4: Beispiel für Kontextinformationen der unterschiedlichen Ebenen

Ebene Studie (Makro)	Ebene Setting (Mikro)	Ebene Objekt
• Hintergrund der Studie (Projektleiter/in, Mitarbeiter/innen, Funding etc.) • Beschreibung der Methodik und des Prozesses der Datenerhebung (Erhebungsinstrumente, Stichprobenauswahl/Feldzugang, Orte und Zeitpunkte der Erhebung etc.) • Rechtliche Aspekte (Vertraulichkeit, Zugangsregelungen und Nutzungsbedingungen, Einverständniserklärungen)	• Metainformationen über das Gespräch als soziale Situation oder das Setting (z. B. Interview-situation) • Informationen zu den untersuchten Einheiten / Probanden (Alter, Geschlecht, Bildung etc.)	• Datenaufbereitung (z. B. Vorgehen beim Transkribieren und Anonymisieren) • Modifikationen an den Daten, Versionierungen • Anzahl und Beziehungen der Dateien (z. B. Aufzeichnungen) • Informationen zur Datei selbst (Dateiname, Format, Größe etc.) • Hinweise auf sensible Inhalte in den Daten und wie mit diesen Daten umzugehen ist

Quelle: in Anlehnung an Bishop 2006

3.1. Auffindbarkeit der Daten

Die im FDZ Bildung angewandten und speziell an den Bedarfen der Bildungsforschung (mit Fokus auf die Schul- und Unterrichtsforschung) ausgerichteten formalen und inhaltlichen Beschreibungselemente sind die Grundlage für die Auffindbarkeit der Daten. Übersetzt in ein definiertes Set an Metadaten können Interessierte mittels verschiedener Retrieval- und

¹⁰ Informationsmaterialien, Webinare und Schulungsangebote, die dabei unterstützen, die Datendokumentation bereits frühzeitig im Forschungsprozess selbst zu verankern, finden sich auf dem Portal des Verbundes Forschungsdaten Bildung: <https://www.forschungsdaten-bildung.de/> (Zugriff am 16.03.2018).

Browsingzugänge gezielt nach den in der Datenbank enthaltenen Studien, Erhebungen und Datenobjekten recherchieren.¹¹ Die audiovisuellen Forschungsdaten des FDZ Bildung können mit dem verwendeten Metadatenset sehr feingranular recherchiert werden; Unterrichtsaufzeichnungen beispielsweise auf der Ebene einer durchgeführten Unterrichtsstunde. D. h. dass Nutzer/innen die Daten auf einer sehr spezifischen Ebene und über verschiedene Datenkollektionen hinweg nicht nur über eine Freitextsuche, sondern auch anhand spezieller Facetten wie z. B. Sozialform, Klassenstufe, Schulform, Unterrichtsthema und Unterrichtsfach etc. durchsuchen können. Eine weitere wichtige Funktion der Metainformationen besteht darin, den Nutzern auch bei der Orientierung, ob und welche Daten für die eigene Fragestellung relevant sein könnten, Unterstützung zu leisten.

Um die Sichtbarkeit und Auffindbarkeit der Daten des FDZ Bildung zu erweitern, werden diese z. T. parallel in unterschiedliche Suchportale eingebettet. Während die differenzierte, spezifische und tiefe Suche exklusiv über das Portal des FDZ Bildung¹² möglich ist, werden die zugehörigen Projekt- und Studienangaben auch über das Portal des VerbundFDB nachgewiesen und sind somit auch im weiteren Rahmen der Bildungsforschung insgesamt sichtbar (www.forschungsdaten-bildung.de/studienliste.php). Das FDZ Bildung optimiert die Auffindbarkeit seiner Daten zudem, indem es für gesamte Datenkollektionen oder einzelne Unterrichtsaufzeichnungen sogenannte persistente Identifikatoren (Digital Object Identifier – DOI) vergibt, die eine dauerhafte Auffindbarkeit- und Zitierbarkeit der Daten ermöglichen.¹³

4. Datenweitergabe: Umgang mit persönlichen und sensiblen Daten

Orientiert an den Prinzipien von Open Science bzw. Open Data strebt auch das FDZ Bildung einen grundsätzlich freien Zugang zu seinen Inhalten an. Hiervon ausgenommen sind jedoch solche Inhalte, die einem besonderen datenschutzrechtlichen und/oder ethischen Schutzgebot unterliegen.

Die qualitativen Daten, die dem FDZ Bildung zur Sicherung und Bereitstellung übergeben werden, können in ihrem Personenbezug und damit in ihrer datenschutzrechtlichen Sensibilität erheblich voneinander abweichen. So existieren im Bestand des FDZ neben anonymisierten Transkripten von Unterrichtsstunden und Interviews auch Audio- und vor allem Videomaterialien mit einem hohen Personenbezug.

Um alle Materialien trotz des unterschiedlich ausgeprägten Personenbezugs für eine datenschutzrechtlich konforme Sekundärnutzung bereitstellen zu können, hat das FDZ Bildung in Zusammenarbeit mit einer Rechtsanwaltskanzlei ein differenziertes Zugriffskonzept erarbeitet und über ein Rechtsgutachten fundiert (Scheller 2017).¹⁴

Auf die Forschungsdaten des FDZ Bildung kann derzeit über drei Differenzierungsstufen zugegriffen werden:

1. Materialien, die überhaupt **keinen Personenbezug** enthalten (z. B. Kontextmaterialien wie Interviewleitfäden, Manuale für die Transkription oder Videografie), sind für jede/n Interessierte/n frei zugänglich.

¹¹ Die einzelnen Datensatznachweise im strukturierten Überblick: www.fdz-bildung.de/studienliste.php (Zugriff am: 16.03.2018).

¹² Suche im Datenbestand des FDZ Bildung: <https://www.fdz-bildung.de/datenarchiv> (Zugriff am 16.03.2018).

¹³ Die Registrierung erfolgt über da|ra: <http://www.da-ra.de> (Zugriff am 16.03.2018).

¹⁴ Unveröffentlichte Rechtsexpertise zum DIPF-Projekt „Videomaterialien“ der Rechtsanwälte Goebel & Scheller, Dezember 2009, Bad Homburg vor der Höhe. Diese Expertise beruht auf dem Rechtsstand, der bis zum 25. Mai 2018 galt. Es ist vorgesehen, die Expertise auf die neue, ab dem 25. Mai 2018 geltende Rechtslage umzustellen.

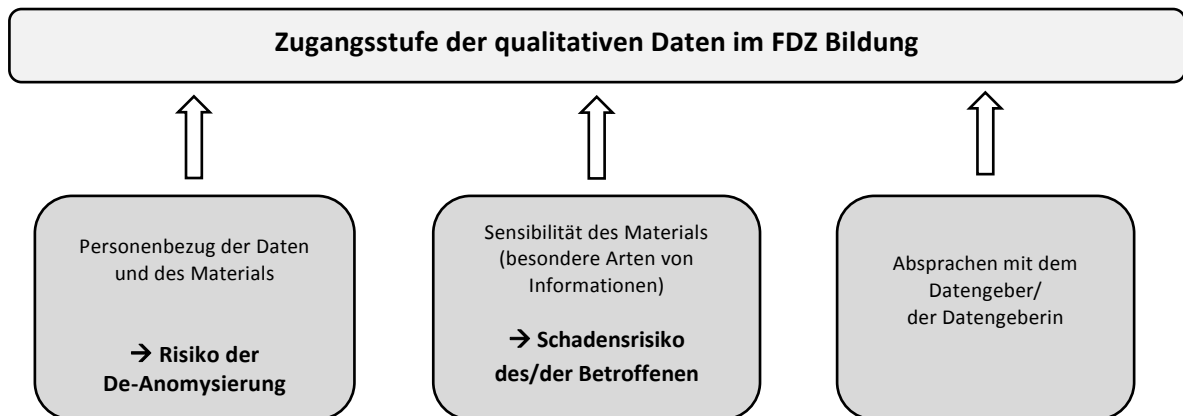
2. **Anonymisierte Daten** (z. B. Transkripte aufgezeichneter Unterrichtsstunden und/oder von Interviews mit Lehrkräften und Schülern), deren Personenbezug durch entsprechende Anonymisierungsmaßnahmen entfernt wurde, möglichst unter Erhalt des Sinngehaltes und ohne Zerstörung des Analysepotentials, sind für jede/n Interessierte/n nach Registrierung zugänglich. Bei der Registrierung müssen die Nutzer/innen allgemeinen Nutzungsbedingungen zustimmen, die zu einem datenschutzrechtlich und ethisch sorgsamem Umgang mit den Daten verpflichten und beispielsweise Versuche zur (Re)-Identifikation untersagen.¹⁵
3. **Daten mit Personenbezug** (vor allem Audio- und Videomaterial) sind nur nach Registrierung und zusätzlicher Antragstellung zugänglich. Das Antragsverfahren umfasst hierbei die Beschreibung des Forschungsvorhabens, für das die Daten zwingend notwendig sind, die Identitätsfeststellung der Antragsteller/innen mittels PostIdent-Verfahren sowie den Abschluss eines schriftlichen Datennutzungsvertrags.

Insbesondere für die nicht anonymisierten Videomaterialien der dritten Stufe, die derzeit über das FDZ zur Verfügung gestellt werden, liegen häufig keine expliziten Einverständniserklärungen der Teilnehmer/innen vor. Dies ist gerade bei den historischen Datenkollektionen des FDZ Bildung, wie den Unterrichtsaufzeichnungen aus der DDR, dem seinerzeitigen Entstehungskontext geschuldet. Aber auch bei Videoaufzeichnungen aus neueren Studien kann eine dezidierte Einverständniserklärung fehlen. Um dieses wertvolle Material dennoch zur Verfügung stellen zu können, kann der Zugang nur in einer strikt reglementierten Form (Stufe 3) und auf Grundlage eines gesetzlichen Erlaubnistatbestandes erfolgen. Das sogenannte Forschungsprivileg besagt, dass Daten auch ohne Einwilligung für Forschungszwecke verarbeitet werden dürfen, „wenn es zur Durchführung wissenschaftlicher Forschung erforderlich ist, das wissenschaftliche Interesse an der Durchführung des Forschungsvorhabens das Interesse des Betroffenen an dem Ausschluss der Zweckänderung erheblich überwiegt und der Zweck der Forschung auf andere Weise nicht oder nur mit unverhältnismäßigem Aufwand erreicht werden kann“ (§ 14 BDSG a. F.). Der neue § 27 Abs. 1 BDSG2017 fordert, dass die Verarbeitung für wissenschaftliche Forschungszwecke erforderlich ist und die Interessen der Verantwortlichen an der Verarbeitung die Interessen der betroffenen Person an einem Ausschluss der Verarbeitung erheblich überwiegen. Das Forschungsprivileg ist somit an enge Voraussetzungen geknüpft, deren Vorliegen im Einzelfall geprüft werden muss.

Neben dem Personenbezug der Daten sind zwei weitere Faktoren für die Wahl der Zugangsstufe im FDZ Bildung maßgeblich: die Sensibilität des Inhaltes (und damit die Einschätzung eines möglichen Schadensrisikos für die Betroffenen bei Re-Identifizierung) sowie individuelle Absprachen mit den Datengebern. Erst die Betrachtung aller drei Faktoren eines spezifischen Datenbestandes definiert die Rahmenbedingungen bei der Entscheidung für eine Zugangsstufe (vgl. Abb. 2).

¹⁵ Die Nutzungsbedingungen des FDZ Bildung sind über die Website einsehbar:
<https://www.fdz-bildung.de/nutzungsbedingungen-fdz-bildung> (Zugriff am 16.03.2018).

Abbildung 5: Faktoren zur Bestimmung der Zugangsstufen am FDZ Bildung.



Quelle: Eigene Darstellung

5. Schluss

Die Frage, wer das Angebot des FDZ Bildung zu welchen Zwecken nutzt, ist von hohem Interesse sowohl für die bedarfsorientierte Weiterentwicklung des Dienstes als auch im Hinblick auf einen hinlänglich fundierten und differenzierten Nachweis des konkreten Mehrwertes, den das Datenangebot für die Fachcommunity darstellt. Die systematische Erfassung von Nutzungsszenarien wird im FDZ Bildung in den kommenden Jahren eine wesentliche Rolle spielen. Diesbezüglich bieten insbesondere die Nutzungsverträge, die die Forschenden mit dem FDZ Bildung zur Datennutzung abschließen, wertvolle Informationen für entsprechende Analysen, etwa über den Status der Forschenden und die mit der Datennutzung verfolgten wissenschaftlichen Ziele. Erste Analysen von ca. 120 Nutzungsverträgen zeigen, dass das Datenangebot des FDZ Bildung in hohem Maße durch den wissenschaftlichen Nachwuchs genutzt wird. Dies zeigt, dass die Verfügbarkeit von Primärdaten in einer Domäne mit z. T. schwierigem Feldzugang (wie eingangs beschrieben), gerade auch für solche Forschende von vitalem Interesse ist, die wissenschaftlich noch nicht etabliert sind und weniger über das hierzu erforderliche Netzwerk, die Ressourcen und die Reputation verfügen, um selbst Primärerhebungen durchzuführen. Neben diesen Aspekten kann das feststellbare Interesse des wissenschaftlichen Nachwuchses an der Daten(nach)nutzung durchaus auch als ein Anhaltspunkt für einen sich abzeichnenden Paradigmenwechsel dahingehend gewertet werden, dass Sekundärforschung als eine probate Option des Forschens zunehmend Akzeptanz findet.

Literaturverzeichnis

- Bambey, Doris; Alexia Meyermann und Maike Porzelt (2017): Potentiale der Sekundärforschung von qualitativen Daten – ein Workshopbericht. *forschungsdaten bildung informiert* 7/2017. https://www.forschungsdaten-bildung.de/files/fdb-informiert_nr-7.pdf (Zugriff am 16.03.2018).
- Bishop, Libby (2006): A Proposal for Archiving Context for Secondary Analysis. In: *Methodological Innovations Online* 1(2), 10–20. DOI: 10.4256/mio.2006.0008 (Zugriff am 16.03.2018).
- DGfE (2017): Stellungnahme der DGfE zur Archivierung, Bereitstellung und Nachnutzung qualitativer Forschungsdaten in der Erziehungswissenschaft. http://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2017.09_Archivierung_qual._Daten.pdf (Zugriff am 16.03.2018).
- Friebertshäuser, Barbara; Antje Langer und Annedore Prengel (Hrsg.) (2013): *Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft*. Weinheim/Basel, Beltz.
- Medjedović, Irena (2014): *Qualitative Sekundäranalyse. Zum Potential einer neuen Forschungsstrategie in der empirischen Sozialforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Meyermann, Alexia; Doris Bambey; Thomas Ebel; Marcus Eisentraut; Malte Jansen; Poldi Kuhl; Reiner Mauer; Claudia Neuendorf; Lisa Pegelow; Maike Porzelt; Marc Rittberger; Thomas Schwager; Petra Stanat und Jessica Trixa (2017): Schlussbericht zum Verbundprojekt „Sicherung und Nachnutzung der Forschungsdaten des Rahmenprogramms zur Förderung der empirischen Bildungsforschung“. <https://doi.org/10.2314/GBV:897124898> (Zugriff am 10.10.2018).
- Prenzel, Manfred (2006): *Bildungsforschung zwischen pädagogischer Psychologie und Erziehungswissenschaft*. In: Hans Merckens (Hrsg.): *Erziehungswissenschaft und Bildungsforschung*. Wiesbaden, VS Verlag für Sozialwissenschaften.
- Schäffer, Burkhard und Olaf Dörner (Hrsg.) (2012): *Handbuch qualitative Erwachsenen- und Weiterbildungsforschung*. Opladen/Berlin/Toronto, Barbara Budrich.
- Scheller, Jürgen (2017): Rechtliche Rahmenbedingungen der Verwendung von Videos in der Schul- und Unterrichtsforschung. Diskrepanzen zwischen Datenschutzrecht, Förder- und Genehmigungsaufgaben. *forschungsdaten bildung informiert* 5/2017. <https://www.forschungsdaten-bildung.de/files/fdb-informiert-nr-5.pdf> (Zugriff am 16.03.2018).
- Schweppe, Cornelia (Hrsg.) (2003): *Qualitative Forschung in der Sozialpädagogik*. Opladen, Leske und Budrich.
- Smioski, Andrea (2013): Archivierungsstrategien für qualitative Daten. In: *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research* 14(3), Art. 5. <http://nbn-resolving.de/urn:nbn:de:0114-fqs130350> (Zugriff am 16.03.2018).
- Stanat, Petra (2014): *Bereitstellung und Nutzung quantitativer Forschungsdaten in der Bildungsforschung: Memorandum des Fachkollegiums „Erziehungswissenschaft“ der DFG*. http://dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/forschungsdaten_memorandum_fk_109.pdf (Zugriff am 16.03.2018).

Dr. Doris Bambey hat an der Universität Marburg Soziologie, Politikwissenschaft, Psychologie und Literaturwissenschaft studiert und ist wissenschaftliche Dokumentarin. Promoviert wurde sie an der TU Darmstadt. Sie leitet den Arbeitsbereich Forschungsdaten Bildung am DIPF, bei dem auch das Modellvorhaben Verbund Forschungsdaten Bildung angesiedelt ist. Zuvor war sie verantwortlich für den Aufbau und die Betreuung des Fachportals Pädagogik sowie des fachlichen Open-Access-Publikationsservers pedocs und weitere Projekte zur Entwicklung von fachlichen Forschungs- und Informationsinfrastrukturen, z. B. für das Projekt Langzeitarchivierung erziehungswissenschaftlicher Ressourcen (LZA Pädagogik) im Rahmen des Verbundprojektes Digital Preservation for Libraries (DP4Lib).

Alexia Meyermann ist wissenschaftliche Mitarbeiterin im Informationszentrum Bildung im Arbeitsbereich Forschungsdaten Bildung am DIPF. Zuvor war sie wissenschaftliche Mitarbeiterin am Datenservicezentrum Betriebs- und Organisationsdaten (DSZ-BO) der Universität Bielefeld.

Maike Porzelt hat an der Humboldt-Universität zu Berlin Bibliotheks- und Informationswissenschaft studiert und ist seit 2014 wissenschaftliche Mitarbeiterin am Deutschen Institut für Internationale Pädagogische Forschung (DIPF) im Projekt Verbund Forschungsdaten Bildung sowie am Forschungsdatenzentrum (FDZ) Bildung. Ihre Arbeitsschwerpunkte liegen im Bereich der Dokumentation, Datenaufbereitung und der Bereitstellung qualitativer Daten der Bildungsforschung sowie der Beratung zum Forschungsdatenmanagement.

Prof. Dr. Marc Rittberger ist Direktor des Informationszentrum Bildung (IZB) des DIPF, Stellvertretender Geschäftsführender Direktor des DIPF und hat eine Professur für Informationsmanagement an der Hochschule Darmstadt inne. Nach seinem Studium der Physik und Informationswissenschaft in Karlsruhe und Konstanz wurde er in Konstanz in Informationswissenschaft promoviert. Am DIPF leitet er verschiedene Projekt zur Entwicklung von Forschungs- und Informationsinfrastrukturen (im Kontext von Fachportal Pädagogik, Forschungsdaten Bildung, Deutscher Bildungsserver sowie Technology-Based Assessment). Er forscht zu Information Retrieval und Informationsqualität.

Archivierung und Verantwortung. Zum Stand der Debatte über den Umgang mit Forschungsdaten in den ethnologischen Fächern

Sabine Imeri

Abstract

In den ethnologischen Fächern stehen dem Interesse, Forschungsdaten langfristig zu sichern, mit Blick auf Nachnutzungsszenarien bzw. -routinen eine Reihe Unklarheiten und Vorbehalte gegenüber: Kontrovers diskutiert werden etwa Formen notwendiger Kontextualisierung sowie – angesichts großer Heterogenität der Datenformen – der Auswahl von Daten zur Archivierung. Besondere Bedeutung kommt Fragen zu Datenschutz und Persönlichkeitsrechten zu, während im Schnittfeld zu forschungsethischen Debatten auch zu erwartende Auswirkungen auf die Feldforschung selbst thematisiert werden. Will man eine breite Akzeptanz von Datenarchivierung und Nachnutzung erreichen, sind die Anforderungen an Professionalität und Vertrauenswürdigkeit künftiger Datenrepositorien entsprechend hoch: Erforderlich sind Lösungen mit ausgeprägter Sensibilität für die Eigenheiten der Wissensproduktion vor allem in der ethnografischen Forschung, wie etwa die geringe Standardisierung, die Individualität der Zugänge und die enge Bindung der forschenden Person an ihr Datenmaterial, aber auch die spezifische Verknüpfung von methodischen und forschungsethischen Fragen. Besondere Bedeutung wird hier Lösungen für den Umgang mit sensiblen Daten, wie sie in sehr vielen ethnografischen Forschungen entstehen, zukommen. Benötigt werden deshalb in erster Linie fachlich spezialisierte Repositorien, die möglichst flexible Prozesse der Archivierung und kontrollierte Prozesse der Nachnutzung genauso anbieten können wie fachspezifische Beratung.

1. Einleitung

In den ethnologischen/kulturanthropologischen Fächern im deutschsprachigen Raum sind Fragen des Datenmanagements bisher wenig diskutiert worden, erst nach und nach ist eine systematischere Auseinandersetzung mit Chancen und Problemen der langfristigen Datenarchivierung und Möglichkeiten der Nachnutzung von Daten erkennbar. Zumindest in der Breite der Fächer wird vor allem mit Blick auf die Bereitstellung von Daten zur Nachnutzung bisher wenig intrinsischer Bedarf artikuliert, nicht zuletzt, weil der eigenen Feldforschung vor allem in Qualifikationsarbeiten ein hoher Stellenwert zugesprochen bzw. die Erhebung eigenen empirischen Materials als wichtiges Qualitätsmerkmal wissenschaftlicher Arbeit angesehen wird. Insofern wird das Thema – vor allem wenn es um formalisierte und standardisierte Prozesse von Datenmanagement geht – eher von außen an die Fächer herangetragen, in Gestalt von Anforderungen, die Forschungsförderinstitutionen und wissenschaftspolitische Gremien formulieren.¹

Die Wissensstände unter den Forschenden sowie die aktuellen Praxisformen und Erfahrungen mit der Datenarchivierung sind entsprechend äußerst heterogen.² Die Fachgesellschaften – in erster Linie die Deutsche Gesellschaft für Sozial- und Kulturanthropologie (DGSKA) und die Deutsche Gesellschaft für Volkskunde (dgv), auf europäischer Ebene neuerdings auch die European Association of Social Anthropologists (EASA) – haben aber inzwischen Diskussions- und Abstimmungsprozesse angestoßen, die zu Positionspapieren oder fachspezifischen Empfehlungen und damit zu mehr Handlungs- und Sprechfähigkeit in der Debatte führen sollen. Trotz vieler Unklarheiten und offener Fragen ist aber die grundsätzliche Bereitschaft erkennbar, Daten zu erhalten und künftig in Repositorien zugänglich zu machen – sofern bestimmte Bedingungen erfüllt sind (Imeri und Danciu 2017: 24f).

2. Notizen zur ethnografischen Forschungspraxis

Aus der Perspektive von Datenmanagementprozessen sind Daten aus ethnografischer Forschung schwierige Daten. Das hängt wesentlich mit der Art und Weise ihrer Entstehung, mit fachspezifischen Zugängen und Forschungsstilen sowie sich daraus ergebenden forschungsethischen Problemen und Anforderungen zusammen, die hier schlaglichtartig skizziert werden sollen.

EthnologInnen betreiben nicht nur, aber in vielen Forschungsszenarien teilnehmende Beobachtung, eine Methode der Datengewinnung, die auf „Begegnung, Interaktion und der sozialen Teilnahme am Alltagsleben unterschiedlicher Menschen“ beruht (Knecht 2013: 83). Umfassender bezeichnet Feldforschung den gesamten Forschungsprozess, in dem teilnehmende Beobachtung mit Interviewmethoden, Kartierungen und anderen Formen der Dokumentation, aber auch Archivstudien kombiniert wird. Entsprechend wird in den Ethnologien ein dezidiert Methodenpluralismus zur Geltung gebracht. Weil sie in *real-world settings* stattfinden, sind Feldforschungen wesentlich situations- wie beobachterabhängig und gekennzeichnet durch wenig Standardisierung, flexible und vor allem feldspezifische Forschungsstrategien. Ein solches Vorgehen ist notwendig kontingent und erfordert in aller Regel Umwege, Verzweigungen und Anpassungen von Fragestellung und

¹ Dieser Beitrag stützt sich auf eine online-Umfrage, Interviews und viele informelle Gespräche, die im Rahmen des Fachinformationsdienstes (FID) „Sozial- und Kulturanthropologie“ mit Forschenden unterschiedlicher Statusgruppen durchgeführt worden sind (vgl. Imeri und Danciu 2017). Der FID ist an der Universitätsbibliothek der Humboldt-Universität zu Berlin angesiedelt und wird von der DFG im Programm „Fachinformationsdienste für die Wissenschaft“ gefördert. Er ist „zuständig“ für die Fächergruppen Ethnologie und Europäische Ethnologie, deren Perspektiven und Zugänge, Forschungsstile und Forschungspraktiken große Schnittmengen aufweisen, aber nicht in eins fallen.

² In einzelnen Segmenten kann das anders sein, wie das Archiv DOBES (Documentation of Endangered Languages) zeigt, zu dem auch EthnologInnen beitragen: <http://dobes.mpi.nl/> (vgl. Widlok 2013).

Forschungsdesign. Ethnografische Feldforschung wird deshalb häufig als „messy process“ beschrieben. (z. B. Bernard 2002: 66). Bezeichnet ist damit jedoch eine „methodische Programmatik ethnografischer Offenheit und Relationalität“ (Knecht 2013: 86), die sich im Verlauf der Forschung immer wieder neu an zuvor nicht vorhersehbare Entwicklungen des Gegenstandes anzupassen vermag – ein notwendiges Verfahren, um Komplexität und Dynamik sozialer Situationen und Prozesse angemessen erfassen und beschreiben zu können. Zwar ermöglichen diese Offenheit und Flexibilität besonders intensive Einsichten, mit Blick auf fachübergreifende Regulierungen zum Datenmanagement entstehen aber gerade daraus immer dann Probleme, wenn bereits vor einer Forschung Festlegungen zu den erwarteten Daten und deren Verbleib nach der Forschung getroffen werden sollen, etwa wenn ein Datenmanagementplan regelmäßig bereits als Teil von Forschungsanträgen eingereicht werden muss.

Forschungen in den Ethnologien sind oft ausgesprochen ressourcenintensiv. Unabhängig davon, ob die Forschung einen langen stationären Feldaufenthalt einschließt oder zeitlich diskontinuierlich, temporalisiert abläuft (Welz 2013), muss in vielen Fällen ein erheblicher Aufwand betrieben werden, um Forschungsfelder überhaupt etablieren zu können: Es müssen Fremdsprachen und/oder Dialekte erlernt werden, Sondierungsaufenthalte vor der eigentlichen Forschung sind notwendig, insgesamt verbunden mit hohen Investitionen in den Aufbau sozialer Netzwerke vor Ort, häufig auch mit ausgeprägter emotionaler Involviertheit (vgl. Stodulka 2014).

Nicht in allen Fällen lässt sich deshalb von einer „Studie“ mit Anfang und Ende sprechen: Weil sich solche intensiven Feldbeziehungen nicht beliebig oft im ForscherInnenleben neu etablieren lassen, weil emotionale Bindungen an Menschen vor Ort dazu auffordern oder verpflichten können, wiederholt Zeit zu investieren oder weil die Zwänge akademischer Arbeitswelten keine andere Arbeitsform zulassen, kehren Forschende oft an den Ort früherer stationärer Forschung zurück oder es werden *permanent fieldsites* unterhalten, die zu kürzeren Feldaufenthalten mehrfach auch über Jahre aufgesucht werden (vgl. Welz 2013). Damit verbunden ist die Frage, wann eine Forschung als abgeschlossen betrachtet werden kann, und wann Daten an ein Repositorium gegeben werden können oder wollen. Zwar wird, wie in anderen Wissenschaften, auch in den Ethnologien überwiegend projektförmig gearbeitet – im Hinblick auf Förderzeiten, zeitliche Rahmen für Qualifikationsarbeiten oder im Hinblick auf ein Ergebnis. Aber auch dort erarbeitetes Material wird unter Umständen über lange Zeiträume temporalisierter Forschung ergänzt, erweitert, wiederholt be- und umgearbeitet. Die Heterogenität der Förderstrukturen und deren Konsequenzen für Bedingungen und Möglichkeiten der Datenarchivierung sollten also im Blick behalten werden.

Ganz gleich aber, welche konkrete Form eine Feldforschung annehmen kann: In jedem Fall sind Forschende auf die Kooperation und Zustimmung der Akteure im Feld angewiesen, denn sie suchen in der Regel die Lebensräume der ProtagonistInnen ihrer Forschung persönlich auf. Ein großer Teil ethnografischer Daten entsteht entsprechend in einem Prozess sukzessiven Lernens, in dem Forschende und Akteure im Forschungsfeld mehr oder weniger langfristige Beziehungen aufbauen. In diesem Prozess werden die Gegenüber der Forschenden nicht als „TeilnehmerInnen“ angesehen, die als „Erhebungseinheit“ konzipiert und als „Stichprobe“ ausgewählt werden, wie es in empirischen Studien benachbarter Sozial- und Verhaltenswissenschaften oft der Fall ist (vgl. z. B. Akremi 2014). Es ist die forschende Person, der die Teilnahme am Alltagsgeschehen ermöglicht und Einblicke in Lebenszusammenhänge anderer Menschen gewährt werden. Weil dies in wechselseitiger Einflussnahme geschieht, tragen die ProtagonistInnen genuin zur Datengewinnung bei. Insofern gelten Daten aus ethnografischer Forschung weithin als von Forschenden und Beforschten ko-produziert (Fabian [1983] 2014).

Solche Forschungsbeziehungen sind „fragile Gebilde“ (Breidenstein et al. 2015: 62), in denen Zurückhaltung oder auch Misstrauen überwunden, wechselseitig Vertrauen entwickelt und kontinuierlich erhalten werden muss (vgl. Weißköppel 2006). In diesen Forschungsbeziehungen verbürgen sich Forschende persönlich für einen verantwortungsvollen Umgang mit Daten, nicht nur mit Blick auf Auswertung und Publikation, sondern auch auf deren Verbleib und Verwendung nach dem Abschluss einer Forschung. Sie tun dies auf der Grundlage forschungsethischer Standards und im Kontext kontinuierlicher Fachdebatten (vgl. von Unger/Dilger/Schönhuth 2016; Barker and Plemmons 2016).

3. Ethnografische Daten

3.1. Annäherungen an den Datenbegriff

Für viele ethnografisch Forschende spielt der Begriff „Daten“ bisher eine eher untergeordnete Rolle; häufig wird er vermieden zugunsten von Begriffen wie „Material“, „Dokumente“ oder „Unterlagen“ (vgl. Lehmann/Stodulka/Huber 2018; Markham 2013). Insofern „Daten“ als verfügbare Ware aufgefasst werden, wird der Begriff auch dezidiert abgelehnt (vgl. Pels et al. 2018). Es ist vorläufig offen, wie sich die verschiedenen Begriffe zueinander verhalten und welcher Datenbegriff zur Anwendung kommt, wenn im Forschungsprozess aus „Material“ „Daten“ werden. Will man „Daten“ aus ethnologischer Forschung langfristig archivieren, dann wird man nicht umhin können, den Datenbegriff stärker zu reflektieren, Unterschiede zu anderen Wissenschaften, aber auch zur allgemeineren Begriffsverwendung zu markieren. Denn diese Unklarheiten verweisen auch darauf, dass die in der Debatte um Forschungsdaten regelmäßig anzutreffenden Unterscheidungen von (objektiven) Rohdaten oder Primärdaten, prozessierten Daten und Interpretationen, die am Ende der Datenanalyse stehen, nicht ohne Weiteres auf die ethnografische Forschung übertragen werden können, deren Daten „nicht außerhalb einer interpretierten sozialen Realität“ existieren (Breidenstein et al. 2015: 114). Wie jede wissenschaftliche Erkenntnis sind auch Forschungsdaten in kommunikativen Prozessen sozial hergestellt, kontextgebunden, theoriebeladen und Produkte medialer Dispositive schon in der Erzeugung (Knorr Cetina 1988). Gleichwohl gelten „Daten“ häufig als Fakten, eine machtvolle Vorstellung von *hard numbers*, beobachterunabhängig, mit direktem Bezug zur Realität, gereinigt von Störung und Rauschen (vgl. Reichert 2018: 18ff). Ähnlich werden auch im informatorischen Datenbegriff Bedingungen und Rahmungen der Datenproduktion reduziert oder eliminiert. Es sind aber gerade die Dimensionen Kontextualität und Referentialität, die im ethnografischen Forschungsprozess zentral sind (Koch 2018). Die Kontextualisierung von Daten ist also unverzichtbar für die Interpretation: „Datasets don’t speak for themselves“ (Lederman 2016: 261). Es geht dabei nicht nur um die Dokumentation allgemeiner Projekthintergründe oder auch forschungsstrategischer Entscheidungen, obwohl schon solches projektspezifische Wissen gerade in Einzelforschungen oder kleinen Forschungsteams häufig implizit bleibt (Smioski 2013: Abs. 14). Vielmehr wirft die enge Bindung der Datenerhebung, der Bearbeitung und Auswahl für die Analyse an Fragestellungen, an theoretische wie methodische Perspektiven sowie feldspezifische Bedingungen und Möglichkeiten Fragen nach unterschiedlichen Ebenen von „Kontext“ auf. „Kontext“ wäre auch das in die Forschungsbeziehung eingebettete relationale, implizite, auch intuitive Wissen sowie das für die ethnografische Wissensproduktion wichtige körperlich-sinnliche Erleben der Forschenden, das sich nur begrenzt datenförmig abbildet (vgl. Okely 2007). Eine unmittelbare, umstandslose Verwendung ethnografischer Daten für eine Sekundärforschung wird deshalb nicht ohne weiteres möglich sein. Um die Rückbindung komplexer Forschungs- und Erhebungskontexte an die Daten zu ermöglichen, werden also – auch mit Blick auf zeitliche und finanzielle Ressourcen – angemessene Verfahren und Formate für die Dokumentation dieser Kontexte entwickelt werden müssen. Metadaten werden hier kaum genügen, auch weil sich zwischen dem methodologischen Anspruch, spezifische

Kontexte von Forschungsfeldern detailliert zu beschreiben, datenschutzrechtlichen Anforderungen an Anonymisierungsstrategien und notwendigerweise offenen Metadaten ein Spannungsfeld aufbaut. Gleichzeitig müssen Status und Wert solcher Daten für eine Sekundärforschung reflektiert werden, wenn sie eben nicht wie selbst erhobenes Material verwendet werden können.

3.2. Datenformen, Heterogenität und Komplexität

Insgesamt entsteht in ethnografischen Forschungen eine große Vielfalt an Datenformen und auch technischen Formaten: Beobachtungsprotokolle, Interviewtranskripte, Feldnotizen, Fotografien, audiovisuelles Material, aber auch Karten, Dokumente, Facebook- und Twittereinträge u. a. (vgl. Imeri und Danciu 2017: 10) als Ergebnis unterschiedlicher Praktiken des Beobachtens und Dokumentierens sowie der Dokumentensammlung. Diese Heterogenität kennzeichnet nahezu jede Ethnografie, weil daraus die Möglichkeit erwächst, die Vielstimmigkeit eines Feldes sichtbar und die Komplexität sozialer, sozio-technischer und sozio-natürlicher Ordnungen angemessen beschreibbar zu machen. Dazu trägt wesentlich bei, dass unterschiedliches, oft auch mehrsprachiges Material aufeinander Bezug nimmt, sich wechselseitig kommentiert, erweitert und ergänzt (Amann und Hirschauer 1997: 16).

Künftig wird etwa vermehrt damit zu rechnen sein, dass Kommunikationsdaten aus sozialen Netzwerken, Foren, Blogs und Webseiten Forschungsdaten werden, nicht nur im Kontext dezidierter Internetethnografie bzw. Digitaler Ethnografie (z. B. Hine 2015), sondern weil ihre Nutzung Teil alltäglicher Praxis vieler Forschungsfelder ist. In dem Maße, in dem sich digitale Technik und Praxisformen in Forschungsfeldern manifestieren, können auch neue Datenarten entstehen (z. B. Slama 2016), wenn etwa im Sinne von Post-Fieldwork Fieldwork (Cohen 1992) die Forscherin dann weniger Briefe und Anrufe von ForschungspartnerInnen erreichen, als vielmehr WhatsApp-Nachrichten und Handyfilme. Für die Datenarchivierung bedeutet das, dass auf die Eigenheiten der verschiedenen Datentypen zugeschnittene Strategien und Verfahren kontinuierlich entwickelt sowie konkrete Bezüge – und damit die Komplexität von Datenkorpora – möglichst nachvollziehbar gehalten werden müssen.

Gleichwohl wird nach wie vor nicht in jeder Forschung jedes Material digital erzeugt, vielmehr durchdringen sich digitale und nicht-digitale Praktiken der Beobachtung und Dokumentation. Die Gründe dafür können feldspezifisch sein, etwa wenn Technikeinsatz in einer Feldsituation nicht oder nur teilweise möglich ist oder vermieden werden soll (vgl. z. B. Röttger-Rössler 2004: 289f). Es sind aber – gerade mit Blick auf handschriftliche Feldnotizen – auch etablierte und bewährte Arbeitsroutinen, die im Laufe eines ForscherInnenlebens nicht oder nur geringfügig verändert werden. Auch weil Handschriftliches oft ebenso wenig oder nicht vollständig digitalisiert wird wie gedruckte Dokumente, ist die Forschungspraxis überdies von einer Reihe von Übergängen zwischen digitalem Material und solchem, das nicht digital erzeugt worden ist, gekennzeichnet. Diese Übergänge können – wenn zum Beispiel bei der Transkription von Notizen einzelne Passagen übersetzt werden – mit ersten Bearbeitungsschritten verbunden sein.

Kontrovers diskutiert wird überdies auch die „Archivierungsfähigkeit“ einzelner Materialarten, vor allem von Feldnotizen und Feldtagebüchern, die in der ethnografischen Forschung zwar weit verbreitet sind, in der Art und Weise, wie Forschende dieses spezifische Dokumentationsformat einsetzen, aber enorm variieren. Als nicht nur individuelles, sondern oft sehr persönliches Material, in dem sich auch emotionale Zustände der Forschenden abbilden können, können sie für die Erkenntnisproduktion hohen Wert haben, gleichzeitig lassen sie die Verwobenheit der forschenden Person mit ihrem Material besonders deutlich hervortreten. Entsprechend ist *data sharing* hier häufig mit besonderen Vorbehalten und Ängsten verbunden (vgl. Jackson 1990).

Insgesamt ist hier die Frage berührt, welches Material bei welchem Bearbeitungsstand archiviert werden soll. Vollständigkeit wird aus Kapazitätsgründen und wegen mangelnder Ressourcen nicht das Ziel sein können, zumal die für viele Disziplinen wichtige Frage der Replikation in den Ethnologien derzeit eine untergeordnete Rolle spielt. Es werden also Strategien und Kriterien der Auswahl von für die Langzeitarchivierung geeigneten Daten entwickelt werden müssen.

Darauf, wie eng Prozesse der Archivierung qualitativer Daten mit Forschungsprozessen verknüpft sind, ist schon an anderer Stelle hingewiesen worden (z. B. Smioski 2013). Insgesamt aber wirft die Auseinandersetzung mit Möglichkeiten und Anforderungen an Datenarchivierung und Nachnutzung Fragen danach auf, wie „das Digitale“ Teil von Forschungspraxis wird bzw. geworden ist und zu weiteren Veränderungen führen wird. Insofern ist hier auch eine Reflexion der methodischen, methodologischen und auch epistemologischen Konsequenzen notwendig, die die Digitalisierung ethnografischer Forschung haben wird.

4. Rechtliche Fragen und Forschungsethik

In ethnografischen Forschungen entstehen regelmäßig Daten, die im Rahmen der Datenschutzgesetze den „besonderen Kategorien personenbezogener Daten“ zuzuordnen sind: Daten, „aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen [...] Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung“ (Art. 9 DSGVO, Abs. 1). Entsprechend sind Datenschutz und der Schutz von Persönlichkeitsrechten unmittelbar auch mit Fragen des Datenmanagements verbunden. Wie die in der Forschungsbeziehung zugesicherte Vertraulichkeit – insbesondere bei der Weitergabe von Daten – gewahrt bleiben kann, ist eine Frage, die aber auch forschungsethisch beantwortet werden muss. Nicht zuletzt, weil Daten (hoch-)sensibel sein können ohne zwingend Personenbezüge aufzuweisen, etwa mit Blick auf vulnerable Gruppen oder Milieus am Rand der Legalität, im Kontext von Migration, politischem Aktivismus o. Ä. In beiden Fällen können Erkenntnisse und Informationen, die Forschenden zuteilwerden, unter Umständen zu Risiken werden – für die beforschten ProtagonistInnen und für Forschende selbst – und das, z. B. in eskalierenden Konflikten, auch noch lange nach dem Ende einer Forschung.

Mit Blick auf offene Feldsituationen und der spezifischen Form ko-produzierter Daten stehen EthnologInnen dem Instrument der informierten Einwilligung im Sinne einer formalen, wenig flexiblen, vorab und schriftlich dokumentierten Erklärung – für die Forschung selbst und für die Datenarchivierung – häufig skeptisch gegenüber. Einverständnis und Freiwilligkeit der Beteiligung sind in jeder ethnografischen Forschung auch forschungsethisch geboten. Einverständnis wird jedoch als permanente Aufgabe und dynamischer, reflexiver Prozess der Aushandlung verstanden, in feldspezifischer Form und in aller Regel ohne standardisierte Vereinbarung (vgl. Albrow und Plemmons 2016). Schriftliche Einwilligungserklärungen können in verschiedenen Milieus und Gesellschaften nicht angemessen, nicht praktikabel und im Grunde wertlos sein, etwa weil rechtliche Konzepte der westlichen Welt nicht vollständig übersetzbar sind (vgl. Fluehr-Lobban 1994). Beschrieben werden aber auch Forschungskontexte, in denen schon die schriftliche Fixierung eines Forschungszwecks nicht möglich ist, wenn dort z. B. Krankheitsbilder erwähnt werden, die mit Stigmata verbunden sind, wie etwa eine HIV/AIDS-Infektion (Dilger 2017). Ähnliches kann gelten, wenn illegale Aktivitäten im Forschungsfeld eine Rolle spielen, wenn beforschte Gruppen Repressionen ausgesetzt sind oder auch, wenn Personen in machtvollen Positionen sich zu Hintergrundgesprächen bereiterklären. Während hier – verbunden mit der Befürchtung, dass Feldzugänge erschwert oder verunmöglicht werden – auch grundlegende Widersprüche zu den geforderten Eindeutigkeiten von Erklärungen und Begutachtungen auftreten können, liegt der Umgang mit ethischen Dilemmata, die dabei entstehen können, letztlich in

der Verantwortung der forschenden Person: „It is the quality of the consent, not its format, which is relevant“ (Albro und Plemmons 2016: 120).

Entsprechend sind mit Blick auch auf die Datenarchivierung ausreichend flexible Regelungen und Formen notwendig, die einerseits Rechtssicherheit bieten und andererseits Forschungsprozesse möglichst wenig beeinträchtigen. Antworten auf die Frage, wie sich komplexes ethnografisches Material so anonymisieren lässt, dass gleichzeitig die Interpretierbarkeit erhalten bleibt, wird hier eine zentrale Rolle zukommen. Zu bedenken ist dabei, dass sich Konzepte von „Privatheit“ oder „Sensibilität“ dynamisch entwickeln und Schaden durch zugängliche Daten auch in der Zukunft entstehen könnte – auch wenn künftige Entwicklungen nicht vollständig antizipiert werden können (vgl. Cliggett 2016: 245). Sensibel können Daten aber auch mit Blick auf die Forschenden selbst sein. Denn das intensive „im Feld-Sein“, die Teilnahme der Forschenden am Geschehen berührt deren ganze Person, die „keine unabhängige Größe darstellt, sondern bei der Abrechnung der Forschungsleistungen mit all ihren menschlichen Stärken und Schwächen voll zu Buche schlägt“ (Simon 2015: 93). Und das bedeutet eben auch, dass Forschende selbst in Forschungsdaten als Personen erkennbar werden.

Allgemein ist abzusehen, dass die datenschutzgerechte Aufbereitung für Langzeitarchivierung und Nachnutzung vor allem für Material mit hohem Schutzbedarf sehr komplex sein kann und dass man es hier mit einem Balanceakt zwischen Bürokratie und Machbarkeit zu tun hat, der weiter reflektiert werden muss. Hier sind bisher kaum orientierende *best practices* vorhanden. Erste Berichte aus internationalen Kooperationsprojekten verweisen überdies darauf, dass zusätzlich Schwierigkeiten auftreten können, wenn Daten grenzüberschreitend genutzt werden sollen, auch hier besteht weiterer Klärungsbedarf. Unklar ist zudem, wie der zusätzliche Dokumentations- und Begründungsaufwand mit Blick auf finanzielle wie zeitliche Ressourcen bewältigt werden und wie ein Kostenrahmen für Datenarchivierung eigentlich aussehen kann.

Bestenfalls andiskutiert sind außerdem Fragen des Urheberrechts und der Nutzungsrechte. Auch hier müssen Konsequenzen aus der spezifischen Konzeptionierung ko-produzierter Daten bedacht werden. Darüber hinaus wird eine Auseinandersetzung mit *cultural property rights* notwendig, die etwa im Diskurs um Bestimmung, Erhalt und Vermarktung von Kulturerbe zentral sind, und im Zusammenhang mit ethnografischen Forschungsdaten und ihrer Verfügbarmachung eine Rolle spielen können (vgl. z. B. Widlok 2013). Solche Daten können nicht nur in Auseinandersetzungen um die kulturelle, sondern z. B. auch um die territoriale Integrität beforschter Gruppen bedeutsam sein. Der Wert und die Relevanz von Daten, die im weitesten Sinne kulturelles Erbe betreffen, können sich verändern oder überhaupt erst in der Langzeitperspektive sichtbar werden. Deshalb können generelle Schutzfristen und Restriktionen des Zugangs problematisch werden und müssen besonders sorgfältig geprüft werden (Leopold 2008). Mindestens Zugriffsrechte sollten in solchen Fällen jedenfalls nicht nur WissenschaftlerInnen, sondern – das ist forschungsethisch ohnehin oft geboten – auch Mitgliedern der sog. *source communities* eingeräumt werden.

5. Fazit

Regelrechte Publikationen von Daten aus ethnografischen Forschungen wird es in vielen Fällen nicht geben können. Ob und welche Daten dauerhaft archiviert und zur Nachnutzung bereitgestellt werden können, muss in Abhängigkeit vom jeweiligen Forschungsfeld und dem konkreten Material sorgfältig abgewogen werden. Die letztliche Entscheidung darüber sollte in der Verantwortung der Forschenden, ggf. im Austausch mit beforschten Personen oder Gruppen, liegen und nicht bei fachlichen oder überfachlichen Kommissionen. Die Entscheidung darüber sollte nicht förderrelevant sein.

Weil die Forschung heterogen ist, müssen Datenrepositorien insgesamt sehr flexibel und mit Blick auf die Ressourcenausstattung auch in der Lage sein, gemeinsam mit Forschenden Festlegungen zu den einzelnen Forschungen zu treffen. Datenrepositorien sollten z. B. ein gestaffeltes Rechtemanagement implementieren, sog. Safe Rooms einrichten, Zugangskontrolle gewährleisten, Datennutzungsverträge abschließen. Viele Forschende wollen z. B. informiert werden, wenn ihre Daten zu Nachnutzung freigegeben werden. Auch aus der Perspektive der Nachnutzung kann die Kontaktaufnahme zu den Primärforschenden wichtig sein, etwa um zusätzliche Informationen über das Material einzuholen. Es sollte Exit-Strategien geben, also die Möglichkeit, Daten zurückzuziehen, eingeräumt werden. Darüber hinaus ist fachspezifische Beratung, etwa zur Vorbereitung der Daten, zu Anonymisierungskonzepten, zur Metadatenerstellung u. Ä. unabdingbar, auch wenn sich Wissensstände auf dem Weg von Weiterbildungen oder dem entsprechenden Ausbau von Curricula möglicherweise angleichen werden.

Es kommt insgesamt darauf an, fachspezifische Zugänge, Forschungsstrategien und forschungsethische Grundsätze zu akzeptieren, und unter diesen Maßgaben geeignete Verfahren zu etablieren und technische Lösungen zu entwickeln, die die Archivierung multimodaler Daten und deren Nachnutzung erlauben. Gebraucht werden deshalb professionalisierte, im Umgang mit solchen Daten versierte und ausreichend sensibilisierte (Fach-)Repositorien oder Datenzentren, deren Existenz und Arbeitsfähigkeit überdies langfristig abgesichert sein muss.

Literaturverzeichnis

- Akreml, Leila (2014): Stichprobenziehung in der qualitativen Sozialforschung. In: Nina Baur und Jörg Blasius (Hrsg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden, VS Verlag für Sozialwissenschaften, 265–282.
- Amann, Klaus und Stefan Hirschauer (1997): Die Befremdung der eigenen Kultur. Ein Programm. In: Hirschauer Stefan und Klaus Amann (Hrsg.): Die Befremdung der eigenen Kultur. Zur ethnographischen Herausforderung soziologischer Empirie. Frankfurt am Main, Suhrkamp, 7–52.
- Albro, Robert und Dena Plemmons (2016): Obtain Informed Consent and Necessary Permission. In: Alex W. Barker und Dena Plemmons (eds.): Anthropological ethics in context. An ongoing dialogue. Walnut Creek (California), Left Coast Press Inc., 119–144.
- Barker, Alex W. und Dena Plemmons (eds.) (2016): Anthropological ethics in context. An ongoing dialogue. Walnut Creek (California), Left Coast Press Inc.
- Bernard, H. Russell (2002): Research Methods in Anthropology. Qualitative and Quantitative approaches. 3rd Edition. Walnut Creek (California), AltaMira Press.
- Breidenstein, Georg; Stefan Hirschauer; Herbert Kalthoff und Boris Nieswand (2015): Ethnografie: Die Praxis der Feldforschung. Konstanz, UVK Verlagsgesellschaft mbH, 2. Auflage.
- Cliggett Lisa (2016): Preservation, Sharing and Technological Challenges of Longitudinal Research in the Digital Age. In: Roger Sanjek und Susan W. Tratner (eds.): eFieldnotes: The Makings of Anthropology in the Digital World. Philadelphia, University of Pennsylvania Press, 231–250.
- Cohen, Anthony (1992): Post-Fieldwork Fieldwork. Journal of Anthropological Research 48, 339–354.
- Dilger, Hansjörg (2017): Ethics, Epistemology and Ethnography: The Need for an Anthropological Debate on Ethical Review Processes in Germany. Sociologus 67 (2), 191–208.
<https://doi.org/10.3790/soc.67.2.191> (Zugriff am 19.08.2018).
- Fabian, Johannes ([1983] 2014): Time and the Other. How Anthropology Makes Its Object. New York, Columbia University Press.
- Fluehr-Lobban, Carolyn (1994): Informed Consent in Anthropological Research: We Are Not Exempt. Human Organization 53(1), 1–10.
- Hine, Christine (2015): Ethnography for the Internet. Embedded, Embodied and Everyday. London u. a., Bloomsbury.
- Imeri, Sabine und Ida Danciu [Mitarb.] (2017): Open Data. Forschungsdatenmanagement in den ethnologischen Fächern. Auswertung einer Umfrage des Fachinformationsdienstes Sozial- und Kulturanthropologie an der Universitätsbibliothek der Humboldt-Universität zu Berlin 2016. Teil I: Statistiken.
<http://www.evifa.de/cms/ueber-evifa/forschungsdatenmanagement/> (Zugriff am 17.05.2018).
- Jackson, Jean (1990): I am a Fieldnote. Fieldnotes as a Symbol of Professional Identity. In: Roger Sanjek (ed.): Fieldnotes. The Makings of Anthropology. Ithaca (New York), Cornell University Press, 3–33.
- Knecht, Michi (2013): Nach Writing Culture, mit Actor-Network: Ethnographie/Praxeographie im Feld der Wissenschafts-, Medizin- und Technikanthropologie. In: Sabine Hess; Johannes Moser und Maria Schwertl (Hrsg.): Europäisch-ethnologisches Forschen. Neue Methoden und Konzepte. Berlin, Reimer Verlag, 79–106.
- Knorr Cetina, Karin (1988): Das naturwissenschaftliche Labor als Ort der „Verdichtung“ von Gesellschaft. Zeitschrift für Soziologie 17 (2), 85–101.
- Koch, Gertraud (2018): Zur meta-kulturellen Produktion der Digitalisierung von Kulturerbe in Gedächtnisinstitutionen. Alltag - Kultur - Wissenschaft. Beiträge zur Europäischen Ethnologie (5).
- Lederman, Rena (2016): Archiving Fieldnotes? Placing ‘Anthropological Records’ Among Plural Digital Worlds. In: Roger Sanjek und Susan W. Tratner (eds.): eFieldnotes. The Makings of Anthropology in the Digital World. Philadelphia, University of Pennsylvania Press, 251–271.
- Lehmann, Jörg; Thomas Stodulka und Elisabeth Huber (2018): H2020 Project K-PLEX: WP4 Report on Data, Knowledge Organisation and Epistemics [Research Report]. Freie Universität Berlin.
<https://hal.archives-ouvertes.fr/hal-01761214> (Zugriff am 27.07.2018).

- Leopold, Robert (2008): The Second Life of Ethnographic Fieldnotes. *Ateliers d'anthropologie – Revue éditée par le Laboratoire d'ethnologie et de sociologie comparative* 32: L'ethnologue aux prises avec les archives. <https://doi.org/10.4000/ateliers.3132> (Zugriff am 24.07.2018).
- Markham, Annette (2013): Undermining 'data': A critical examination of a core term in scientific inquiry. *First Monday* 18(10). <https://doi.org/10.5210/fm.v18i10.4868> (Zugriff am 17.07.2018).
- Okely, Judith (2007): Fieldwork Embodied. *The Sociological Review* 55(1_suppl), 65–79. <https://doi.org/10.1111/j.1467-954X.2007.00693.x> (Zugriff am 20.08.2018).
- Pels, Peter; Igor Boog; J. Henrike Florusbusch; Zane Kripe; Tessa Minter; Metje Postma; Margaret Sleeboom-Faulkner; Bob Simpson; Hansjörg Dilger; Michael Schönhuth; Anita von Poser; Rosa Cordillera A. Castillo; Rena Lederman and Heather Richards-Rissetto (2018): Data management in anthropology: the next phase in ethics governance? *Social Anthropology* 26(3), 391–413. <https://doi.org/10.1111/1469-8676.12526> (Zugriff am 20.08.2018).
- Reichert, Ramón (2018): Big Data als Boundary Objects. Zur medialen Epistemologie von Daten. In: Thorben Mämecke; Jan-Hendrik Passoth und Josef Wehner (Hrsg.): *Bedeutende Daten. Medien – Kultur – Kommunikation*. Wiesbaden, Springer VS, 17–33. https://doi.org/10.1007/978-3-658-11781-8_2 (Zugriff am 17.08.2018).
- Röttger-Rössler, Birgitt (2004): Die kulturelle Modellierung des Gefühls. Ein Beitrag zur Theorie und Methodik ethnologischer Emotionsforschung anhand indonesischer Fallstudien. Münster, LIT Verlag.
- Simon, Michael (2015): Ethnologische Anmerkungen zu Bernd Riekens ‚Gesprächen mit Einheimischen‘ in Galtür. In: Bernd Rieken (Hrsg.): *Wie bewältigt man das Unfassbare? Interdisziplinäre Zugänge am Beispiel der Lawinenkatastrophe von Galtür*. Münster u. a., Waxmann, 93–105.
- Slama, Martin (2016): File Sharing and (Im)Mortality: From genealogical Records to Facebook. In: Roger Sanjek und Susan W. Tratner (eds.): *eFieldnotes. The Makings of Anthropology in the Digital World*. Philadelphia, University of Pennsylvania Press, 94–109.
- Smioski, Andrea (2013): Archivierungsstrategien für qualitative Daten [54 Absätze]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 14(3), Art. 5. <http://nbn-resolving.de/urn:nbn:de:0114-fqs130350> (Zugriff am 23.08.2018).
- Stodulka, Thomas (2014): „Feldforschung als Begegnung – Zur pragmatischen Dimension ethnographischer Daten. *Sociologus* 64(2), 179–206.
- Weißköppel, Cordula (2006): Gemischte Gefühle. Prekäre Dynamiken in der Forschung mit politischen Flüchtlingen. *EthnoScripts: Interaktion im Feld* 8(2), 124–144.
- von Unger, Hella; Hansjörg Dilger und Michael Schönhuth (2016): Ethikbegutachtung in der sozial- und kulturwissenschaftlichen Forschung? Ein Debattenbeitrag aus soziologischer und ethnologischer Sicht. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 17(3), Art. 20. <http://dx.doi.org/10.17169/fqs-17.3.2719> (Zugriff am 17.08.2018).
- Welz, Gisela (2013): Die Pragmatik ethnografischer Temporalisierung. Neue Formen der Zeitorganisation in der Feldforschung. In: Sabine Hess; Johannes Moser und Maria Schwertl (Hrsg.): *Europäisch-ethnologisches Forschen. Neue Methoden und Konzepte*. Berlin: Reimer Verlag, 39–54.
- Widlok, Thomas (2013): The Archive Strikes Back: Effects of Online Digital Language Archiving on Research Relations and Property Rights. In: Mark Turin; Claire Wheeler und Eleanor Wilkinson (eds.): *Oral Literature in the Digital Age Archiving Orality and Connecting with Communities*, Cambridge, Open Book Publishers, 3–19. <https://www.openbookpublishers.com/product/186/> (Zugriff am 23.08.2018).

Dr. des. Sabine Imeri war von 2006 bis 2014 als wissenschaftliche Mitarbeiterin und in verschiedenen DFG-geförderten Forschungsprojekten zur Wissensgeschichte der Volkskunde am Institut für Europäische Ethnologie in Berlin tätig. Promotion 2015. Seit 2016 ist sie als wissenschaftliche Mitarbeiterin beim Fachinformationsdienst Sozial- und Kulturanthropologie an der Universitätsbibliothek der Humboldt-Universität zu Berlin für den Bereich Forschungsdatenmanagement zuständig.

Probleme der Archivierung und sekundären Nutzung ethnografischer Daten

Christian Meier zu Verl und Christian Meyer

Abstract

Ethnografische Forschung basiert auf der Umsetzung eines „Methodenpluralismus“ (Sombart), der eine Vielzahl unterschiedlicher und äußerst heterogener Formen objektivierte Wissens als „Artefakte“, „Konstrukte“, „Spuren“, „Momentaufnahmen“ und „Repräsentationen“ erzeugt, die im Verlauf der Forschung bisweilen in „Daten“ transformiert werden, bisweilen aber auch implizit bleiben und als „Orientierungswissen“ (Elwert 2003) zum Verstehen dessen verwendet werden, was andererseits als „Daten“ isoliert wird. Die produzierten Wissensobjekte umfassen u. a. schriftliche Notizen, Protokolle und Tagebücher, visuelle Skizzen, Fotografien und Dokumente oder Gegenstände des Feldes, technische Aufzeichnungen von Gesprächen, sozialen Ereignissen, Performanzen oder Interviews nebst deren multimodalen Transkriptionen sowie semi-theoretische und semi-methodologische Vermerke.

Nur zum Teil werden die ethnografischen Erfahrungen jedoch in dieser Weise objektiviert; ein Gutteil schlägt sich allein auf der Ebene präreflexiver Vertrautheit im Gedächtnis und Körper des Ethnografen in Form von episodischem oder prozeduralem Wissen nieder, das z. T. nur situativ abrufbar ist. Das ethnografische Wissen entsteht letztlich durch die Verknüpfung, Vernetzung und Kombination der unterschiedlichen Wissensformen zu einem umfassenden und kohärenten Bild oder Narrativ. Unter den qualitativen Methoden der Soziologie ist die Ethnografie diejenige Methode, die am konsequentesten nicht-reduktiv operiert und den aspektuellen und dimensional Reichtum des erforschten Gegenstands bewahrt.

Diese Verfasstheit der ethnografischen Forschung stellt das Anliegen der Archivierung und Dokumentation vor große Herausforderungen, die mit diesem Artikel angesprochen werden.

1. Einleitung

Die gegenwärtige Diskussion über Vorteile und Nachteile einer Archivierung und sekundären Nutzung wissenschaftlicher Daten erfasst zunehmend auch Bereiche der Soziologie und insbesondere der empirischen Sozialforschung, die mit qualitativen Methoden soziale Phänomene erforscht (vgl. u. a. Huschka et al. 2013). Mit diesem Artikel stellen wir die Ethnografie und deren Wissensobjekt – ethnografische Daten – in den erkenntnistheoretischen Mittelpunkt und gehen möglichen Problemen der Archivierung und sekundären Nutzung ethnografischer Daten nach. Wir möchten v. a. auf einige bisher kaum betrachtete epistemologische, methodologische und forschungspraktische Probleme hinweisen, die einem allzu einfachen Zugriff auf ethnografische Daten zum Zwecke einer Archivierung und sekundären Nutzung entgegenstehen.

Ethnografische Forschung basiert auf der Umsetzung eines Pluralismus qualitativer Methoden (teilnehmende Beobachtung, Interview, technische Aufzeichnungen, Sammlung von Dokumenten und Artefakten), der eine Vielzahl unterschiedlicher und äußerst heterogener Formen objektivierten Wissens als „Artefakte“, „Konstrukte“, „Spuren“, „Momentaufnahmen“ und „Repräsentationen“ erzeugt, die im Verlauf der Forschung bisweilen in „Daten“ transformiert werden, oft aber auch implizit bleiben und als „Orientierungswissen“ (Elwert 2003) zum Verstehen dessen verwendet werden, was als „Daten“ isoliert wurde. Die produzierten Wissensobjekte umfassen u. a. schriftliche Notizen, Protokolle und Tagebücher, visuelle Skizzen, Fotografien und Dokumente oder Gegenstände des Feldes, technische Aufzeichnungen von Gesprächen, sozialen Ereignissen, Performanzen oder Interviews nebst deren multimodalen Transkriptionen sowie semi-theoretische und semi-methodologische Vermerke. Nur zum Teil werden die ethnografischen Erfahrungen jedoch in dieser Weise objektiviert; ein Gutteil schlägt sich allein auf der Ebene präreflexiver Vertrautheit im Gedächtnis und Körper des Ethnografen in Form von episodischem oder prozeduralem Wissen nieder, das z.T. nur situativ abrufbar ist. Das ethnografische Wissen entsteht letztlich durch die Verknüpfung, Vernetzung und Kombination der unterschiedlichen Wissensformen zu einem umfassenden und kohärenten Bild oder Narrativ. Unter den qualitativen Methoden der Soziologie ist die Ethnografie diejenige Methode, die am konsequentesten nicht-reduktiv operiert und den aspektuellen und dimensionalsten Reichtum des erforschten Gegenstands bewahrt. Diese Verfasstheit der ethnografischen Forschung stellt das Anliegen der Archivierung, aber auch der sekundären Nutzung vor große Herausforderungen.

In diesem Beitrag vertreten wir die These, dass eine Archivierung, Dokumentation und sekundäre Nutzung ethnografischer Daten in diesem Sinne nicht gelingen kann und sogar scheitern muss, da das präreflexive ethnografische Wissen der Ethnografin sich einer technischen Archivierung versperrt. Letztlich müsste man – um ethnografische Daten verständlich zu archivieren – auch den Körper der Ethnografin mit zu den Akten legen. Aber auch in diesem Fall kann man das präreflexive Wissen der Ethnografin kaum konstant – d. h. unveränderbar und dadurch referenzierbar – halten, da sich die Ethnografin und ihr Wissen in einem fortwährenden zeitlichen Wandel befinden.

Um diese These zu begründen, werden wir zunächst auf den „Methodenpluralismus“ der Ethnografie eingehen (2.). Wir werden zeigen, dass die Ethnografie selbst keine Methode darstellt, sondern vielmehr als eine grundlegende Strategie des Forschens angesehen werden muss, mit der sich die Sozialwissenschaftlerin (unbekannten) sozialen Phänomenen nähert. Aus diesem „Methodenpluralismus“ lässt sich – wenn wir uns auf die Ebene epistemischer Wissensobjekte bewegen – auch ein „Datenpluralismus“ der Ethnografie ableiten (3.). Dies bedeutet auch, dass im Rahmen von Ethnografien Daten mit unterschiedlichen epistemologischen und methodologischen „Reichweiten“ produziert werden, die selbst wiederum zu epistemisch unterschiedlich konfigurierten Gegenständen führen. Diese Pluralismen lenken unseren Blick auf den Körper der Ethnografin, die mit ihrem

reflexiven, aber auch präreflexiven Orientierungswissen über das Forschungsfeld jene Heterogenität der Repräsentationen sozialer Wirklichkeit praktisch „zusammendenken“ kann (4.). Der Körper der Ethnografin stellt daher in den meisten methodologischen Varianten der Ethnografie das primäre „Forschungsinstrument“ dar, das immer nur durch andere extra-korporale Instrumente ergänzt, aber nie ersetzt werden kann. Einige dieser Varianten der Ethnografie werden wir in diesem Artikel kurz skizzieren (5.), bevor wir uns den Problemen einer Archivierung und sekundären Nutzung ethnografischer Daten im Detail zuwenden (6.). In unserem Fazit heben wir abschließend fünf entscheidende Probleme hervor und empfehlen mit Blick auf diese Probleme einen methodologisch reflektierten und freiwilligen Umgang mit Formen der Archivierung und sekundären Nutzung ethnografischer Daten (7.).

2. Der ethnografische „Methodenpluralismus“

Ethnografische Forschung ist eine „methodenplurale“ Strategie. Sie kann daher methodisch nicht auf „teilnehmende Beobachtung“ und die im Zuge einer Beobachtung entstandenen Beobachtungsprotokolle verengt werden. Neben dem Protokollieren von Beobachtungen fertigen Ethnografinnen – je nach Forschungsinteresse – Interviews an, sammeln Dokumente und andere Artefakte des Feldes, führen (ethnografische) Gespräche und produzieren Audio- und Videoaufzeichnungen von sozialen Ereignissen, die während ihrer Feldaufenthalte stattfinden. Wie dieser „Methodenpluralismus“ durch die Ethnografinnen selbst gewichtet und welche Rolle einer einzelnen Methode im Verlauf der Forschung beigemessen wird, entscheidet nicht nur das zugrunde gelegte sozialtheoretische Verständnis und dessen methodische Konsequenz, sondern auch die forschungspraktischen Bedingungen des Feldes. Ist es der Ethnografin z. B. möglich, in einer Organisation als Praktikantin eigentätig zu werden, oder darf sie nur als „stummer Schatten“ eine Sachbearbeiterin beobachtend verfolgen? Kann sie Interviews oder Gespräche mit unterschiedlichen (Schlüssel-)Informantinnen über sie interessierende Phänomene führen? Hat sie Zugang zu Dokumenten und Artefakten, die die Mitglieder der Organisation produzieren? Darf sie technisch-registrierende Aufzeichnungen (Audio- oder Videoaufzeichnungen) machen?

Eine Besonderheit der Ethnografie, die nicht unmittelbar aus diesem „Methodenpluralismus“ hervorgeht, ist der (mitlaufende) Aufbau von Deutungskompetenzen durch eine langfristige Anwesenheit im Feld und die damit verbundenen umfassenden Erfahrungen sowie das zunehmende Vertrautwerden mit den verschiedensten Feldsituationen. Dabei entstehen graduelle Unterschiede von Deutungskompetenzen durch die Dauer (Monate oder Jahre) und Intensität (eher aktiv teilnehmend oder eher passiv beobachtend) der Anwesenheit der Ethnografin im Feld.

3. Datenpluralismus der Ethnografie

Der skizzierte „Methodenpluralismus“ der Ethnografie hat direkte Konsequenzen für unsere Diskussion von Problemen der Archivierungen und sekundären Nutzung ethnografischer Daten. Erstens generieren Ethnografinnen im Verlauf ihrer Forschung eine Vielzahl von empirischen Materialien bzw. Wissensobjekten. Diese Objekte werden mit Hilfe unterschiedlicher „Zugriffsmethoden“ generiert und nur mit Hilfe weiterer Methoden analysiert, interpretiert und dabei letztlich auch in ethnografische Daten transformiert. Zweitens sind ethnografische Daten, wie sozialwissenschaftliche Daten im Allgemeinen, relationale Wissensobjekte, die ihre wissenschaftliche Relevanz überhaupt nur in Bezug auf wissenschaftliche Problemstellungen, Fragestellungen oder andere Wissensobjekte ausbilden können. Empirische Wissensobjekte sind daher keineswegs mit empirischen Daten gleichzusetzen, vielmehr müssen die Objekte erst durch die praktische Arbeit der Ethnografin aufwendig in Daten transformiert werden (vgl. Meier zu Verl 2018). Drittens handelt es sich bei den

Objekten um Externalisierungen des Wissens der Ethnografin, die ihr dabei helfen, sich nicht nur an die miterlebten ethnografischen Situationen, in denen sie entstanden sind, zu erinnern, sondern auch, um sich selbst *ex post factum* nach mit ihnen in Zusammenhang stehenden, relevant erscheinenden weiteren Ereignissen befragen zu können. Dieses neue Wissen wird meist in neuen Beschreibungen weitergehend elaboriert. Ethnografische Wissensobjekte werden daher durch das reflexive und präreflexive Wissen der Ethnografin in Relation zueinander sowie in Relation zum wissenschaftlichen Diskurs gebracht. Sie werden als wissenschaftlich unterdeterminierte Proto-Daten schrittweise durch die praktische Arbeit der Ethnografin in Daten-Kandidaten und schließlich in wissenschaftlich relevante Daten überführt.

Das sozialwissenschaftliche Interesse der Ethnografin besteht nun darin, ihre empirischen Materialien – schriftliche Notizen, Beobachtungsprotokolle, Feldtagebücher, visuelle Skizzen, Fotografien, technische Aufzeichnungen von Gesprächen, sozialen Ereignissen, Performanzen, Interviews, Dokumente und Artefakte des Feldes sowie semi-theoretische und semi-methodologische Vermerke – in ihrem sinnlichen und sinnhaften Reichtum zu bewahren. Die unterschiedlichen Sinndimensionen der Wissensobjekte dürfen nicht einem „Datenreduktionismus“ in Gestalt der Numerisierung und Metrisierung zum Opfer fallen, sondern müssen als lebensweltliche Komplexität der Objekte erhalten und bewahrt werden. Die Ethnografie erkennt damit die „Vor-Interpretiertheit der Welt“ durch die zu erforschenden Akteure an und versucht ihre Konstrukte zweiten Grades adäquat in Hinblick auf die ihnen zugrunde liegenden Konstrukte ersten Grades zu gestalten (vgl. Schütz 1953).

Dies bedeutet, dass wir – wenn wir von ethnografischen Wissensobjekten sprechen – zunächst von im Verlauf der Forschung produzierten wissenschaftlich unterdeterminierten Proto-Daten, Konstrukten, Spuren oder auch Momentaufnahmen und Repräsentationen ausgehen. Was zum wissenschaftlich relevanten Datum wird, kann nicht vorab wissenschaftstheoretisch festgelegt werden, sondern ist kontingent und ergibt sich erst aus der situiert verkörperten Forschung und deren (impliziten) Praktiken der Ethnografin. Dabei werden nicht alle Proto-Daten und Daten-Kandidaten in ratifizierte Daten transformiert, sondern ein großer Teil dieser Wissensobjekte dient als Kontext- und Orientierungswissen, mit dem wiederum andere und weitere Erkenntnisse über das Forschungsfeld generiert und finale Daten interpretiert werden. Was im Verlauf einer ethnografischen Forschung Datum und was Kontextwissen ist, bleibt kontingent, verschiebt sich fortlaufend und wird (allenfalls) in Veröffentlichungen als Momentaufnahme des Forschungsprozesses (schriftlich) vorübergehend festgelegt. Datum und Kontextwissen sind in der Ethnografie daher nur analytisch, nicht aber ontologisch zu unterscheiden.

4. Der Körper der Ethnografin

Das primäre Forschungsinstrument der Ethnografie ist der Körper der Ethnografin, in dem sich jenes Orientierungswissen anreichert, das es ihr ermöglicht, empirisches Material als Datum *von etwas* zu verstehen. Dabei entsteht ethnografisches Wissen als implizites, aber prinzipiell explikationsfähiges, episodisches und prozedurales Erfahrungswissen durch die aktive Teilnahme an, aber auch durch schlichte Kopräsenz in Forschungssituationen. V. a. aufgrund seines impliziten, präreflexiven und präpropositionalen Charakters ist Orientierungswissen schwer zu dokumentieren und wird dementsprechend selten durch Ethnografinnen erschöpfend objektiviert und versprachlicht. Im Forschungsprozess lassen sich aber immer wieder Situationen lokalisieren, in denen Ethnografinnen auf ihr implizites Orientierungswissen zurückgreifen, um ihr empirisches Material, das ja lebensweltlich gesättigt und daher sozialwissenschaftlich unterdeterminiert ist, in wissenschaftlich relevante Daten über die soziale Wirklichkeit zu transformieren. Dies findet häufig in der Datenanalyse am Schreibtisch oder in Datensitzungen statt, in denen sich Ethnografinnen ihren

Wissensobjekten zuwenden und ihr Wissen über und Verständnis von diesen Objekten reflektieren und so dokumentierbar machen. Dabei „befragt“ sich die Ethnografin häufig selbst als „Informantin“ und macht so ihr implizites Wissen über die untersuchten Phänomene in Auseinandersetzung mit empirischem Material für sich und andere Kolleginnen explizit. Z. B. führen Ethnografinnen Videoanalysen durch, in denen sie ihr implizites Wissen über die (technisch-registrierte) flüchtige soziale Wirklichkeit visuell sichtbar und damit explizit machen können. Dokumentiert werden können dann sowohl die Prozesse des Hervorbringens von Wissen (z. B. durch eine Aufnahme der Sitzung) als auch einzelne konkrete Wissens Elemente und -objekte (z. B. neue Erkenntnisse über die soziale Wirklichkeit in Form von Begriffen, Kategorien, Thesen etc.).

Die Bedeutung des Körpers der Ethnografin als „Forschungsinstrument“ (Amann und Hirschauer 1997) und dessen Teilnahme und Kopräsenz in Forschungssituationen wird in den einzelnen, hier diskutierten Varianten der Ethnografie methodologisch unterschiedlich bewertet, so dass wir diese Varianten kurz erörtern, bevor wir uns den Problemen der Archivierung, Anonymisierung und sekundären Nutzung ethnografischer Daten zuwenden.

5. Varianten der Ethnografie

Im Folgenden werden wir uns verschiedenen methodologischen Varianten der Ethnografie zuwenden. Dabei verwenden wir eine analytische Einteilung von methodologischen Varianten der Ethnografie, die auf den hier verfolgten Fokus „Archivierung und sekundäre Nutzung ethnografischer Daten“ zugeschnitten ist. Freilich können wir die einzelnen Varianten der Ethnografie an dieser Stelle nicht umfassend beleuchten, sondern nur einzelne Aspekte – insbesondere die Rolle des Körpers als „Datenarchiv“ – hervorheben (vgl. Meier zu Verl und Meyer 2017, Meier zu Verl 2018). Die vier methodologischen Varianten der Ethnografie, die wir behandeln, sind: (1) die naturalistische, (2) die textuell-semiotische, (3) die korporal-koprärente und (4) die ethnomethodologische Ethnografie.

Zu (1): Malinowski (1922) begründete nicht nur die „teilnehmende Beobachtung“ als eine Forschungsmethode der Ethnografie, sondern prägte auch das frühe Verständnis von Ethnografie als eine naturalistische Unternehmung. Die Vorstellung, dass ein jeder Ethnograf auch teilnehmender Beobachter in der zu erforschenden fremden Kultur sein muss, um deren kulturelles Leben mit dem eigenen Geist zu bezeugen und am eigenen Körper nachzuvollziehen, veränderte die ethnografische Forschung Anfang des 20. Jahrhunderts. Der durch eine Teilnahme im Sinne eines Dabeiseins ablaufende Prozess der Enkulturation ist methodologisch möglich, weil sowohl der Ethnograf als auch die zu erforschenden „Fremden“ über biologisch gleiche Anlagen, aber kulturell differenzierte Eigenschaften verfügen, die im Erwachsenenalter wieder (mehr oder weniger) umgeformt werden können. Dies geschieht, indem von außen „Energien“ in die kulturell bedingte Sinneswahrnehmung oder das abstrakte Denken „fließen“ (vgl. Rivers 1901). Mögliche ethnografische Erfahrungen werden daher durch den Grad der Teilnahme an der zu erforschenden Kultur geprägt und schreiben sich als erlerntes Wissen und erworbene Praktiken in den Körper und Geist des Ethnografen ein. Daneben werden die Erforschten durch den naturalistischen Ethnografen beobachtet und interviewt, um zu umfassenden Verhaltens- und Situationsbeschreibungen über die erforschte (fremde) Kultur zu gelangen. Für Malinowski ist daher nicht nur die (distanzierte) Teilnahme im Sinne eines Dabeiseins am kulturellen Leben, das Beobachten und Befragen der zu Erforschenden wichtig, sondern auch deren Dokumentation im Medium der Schrift. Mit dieser Form der Beobachtung ergibt sich für ihn die Möglichkeit der Reflexion der eigenen (sich teilweise vollziehenden) Enkulturation und Wirkung auf den Untersuchungsgegenstand. Ziel der naturalistischen Forschung ist es, eine holistische und funktionale Beschreibung der zu erforschenden Kultur zu erarbeiten, die aus den vorherigen Verhaltens- und Situationsbeschreibungen herausgearbeitet werden muss.

Zu (2): Die textuell-semiotische Ethnografie positionierte Geertz (1973) als ein Gegenprogramm zu positivistischen und auch behavioristischen Verständnissen in der Anthropologie und Ethnologie, die sich – wie Malinowski – darauf beschränken, beobachtbares Verhalten in Medien der Schrift, der Fotografie etc. zu dokumentieren. Geertz geht es dem entgegen um ein hermeneutisches Verstehen fremder kultureller Symbolsysteme. Die Sozialwelt – deren kleinste Einheit für ihn die symbolische Handlung ist – entfaltet sich für den Ethnografen als ein aus sinnhaftem Handeln bestehender, zu interpretierender Text, den er „lesen“ muss, um sich seinen Sinn zu erschließen. Der textuell-semiotische Ethnograf ist in diesem Sinn in einem „hermeneutischen Zirkel“ zwischen dem eigenen kulturellen Vorverständnis der zu erforschenden Kultur und der fremden zu erforschenden Kultur eingespannt. Damit geht es der textuell-semiotischen Ethnografie nicht um körperliche Erfahrungen in einem Enkulturationsprozess oder um die „Nachahmung“ beobachteter Tätigkeiten, sondern um ein ethnografisches Überführen eines fremden Diskurses in den eigenen. Um eine fremde Kultur lesen zu können, führt der textuell-semiotische Ethnograf Interviews, betreibt historische Forschung und protokolliert aktuelle Beobachtungen. Damit können die Netze des gemeinten Sinns der zu erforschenden Akteure hin zu einer dichten Beschreibung kondensiert werden. Geertz fordert daher von der Ethnografie ein Verstehen anderer Kulturen, ohne dass sich Ethnografen an diesen Kulturen aktiv beteiligen, um implizit-verkörperte Erfahrungen zu erwerben. Die ethnografischen Beschreibungen gehen damit bereits in ihrer Entstehung über ein noch uninterpretiertes Datum – das eine „dünne“ Beschreibung wäre – hinaus. In dichten Beschreibungen sind Daten und deren Interpretationen untrennbar miteinander „verschmolzen“, da sie die Motivstrukturen eines Phänomens explizieren.

Zu (3): Für Goffman (1989), aber auch Hirschauer (2001), Kalthoff (2003) und Kollegen (Breidenstein et al. 2013), ist es der korporal-kopräsenten Ethnografie nicht möglich, in den zu erforschenden Kulturen *a priori* wie in einem fremden Manuskript zu lesen. Das durch die Ethnografin erfahrene Soziale muss in dieser Variante überhaupt erst einmal zur Sprache gebracht werden. Diesem Verständnis nach dient die Ethnografie dazu, die „Schweigsamkeit des Sozialen“ (Hirschauer 2001) durch körperliche Kopräsenz der Ethnografin zu durchbrechen. Die korporal-kopräsente Ethnografie betont – im Gegensatz zur textuell-semiotischen Ethnografie – die epistemische Bedeutung des Körpers der Ethnografin, der als ein teilnehmender Beobachter sozialer Situationen durch kein „extrakorporale[s]“ Forschungsinstrument“ (Amann und Hirschauer 1997: 25) ersetzt werden kann. In diesem Sinne muss die Ethnografin in sozialen Situationen die zu erforschende Kultur selbst erleben, um sie dann „abseits“ von ihr im Medium der Schrift sukzessive zur und in eine Sprache zu bringen. Das situativ erworbene ethnografische Orientierungswissen der korporal-kopräsenten Ethnografie entsteht v. a. durch körperliche Kopräsenz – im Sinne eines Dabeiseins – in sozialen Situationen, die im Schreibprozess reflexiv dokumentiert werden. Die Ethnografin beobachtet Situationen und nimmt an diesen teil, indem sie anwesend ist, aber weniger dadurch, dass sie selbst zu einem gleichberechtigten Akteur im situativen Geschehen wird. Daneben führt die korporal-kopräsente Ethnografin auch Interviews mit den an der Forschung teilnehmenden Akteuren. Ihre „Deutungskompetenz“ erwirbt sie sich jedoch durch ihre körperlich-kopräsente Anwesenheit in sozialen Situationen und nicht etwa durch das Lesen empirischer Materialien.

Zu (4): Die ethnomethodologische Ethnografie spricht sich stärker als alle bisher diskutierten Varianten nicht nur für eine Teilnahme aus, sondern dafür, dass die Ethnografin zu einem gleichberechtigten Akteur – einem „member“ – der zu erforschenden Praxis wird. Teilnahme heißt in der ethnomethodologischen Ethnografie, dass der Ethnograf zu einem Mitglied an der das soziale Geschehen tragenden Kollektivität wird und sich so sein ethnografisches Orientierungswissen in Form von Akteurskompetenzen aneignet. Daneben fertigen ethnomethodologische Ethnografen v. a. Video- und Audioaufzeichnungen sozialer Geschehen an, um retrospektiv anhand ihres Orientierungswissens und den technischen Aufzeichnungen die konstitutiven Praktiken sozialer Phänomene zu

rekonstruieren. Mit ihrem Postulat einer „unique adequacy requirement of methods“ (Garfinkel und Wieder 1992) geht die ethnomethodologische Ethnografie über eine naturalistische Vorstellung von Teilnahme und Erkenntnis hinaus. Dieses Postulat schlägt eine reflexive Anwendung von zu erforschenden Praktiken vor, um diese Praktiken wissenschaftlich und lebensweltlich adäquat zu beschreiben. Der Ethnograf ist also nicht mehr nur korporal-kopräsenter „Beisitzer“, sondern selbst „kompetenter Teilnehmer“ jener Praxis, die es selbstreflexiv zu erforschen gilt. Der Ethnograf muss sich daher in Datensitzungen an sein verkörpertes implizites Praxiswissen erinnern, es befragen, es abermals durchführen und anschließend anhand der vorhandenen technisch-registrierten Aufzeichnungen für sich und andere explizieren. Damit „zerfällt“ auch der Datenbegriff der ethnomethodologischen Ethnografie in zwei nur analytisch voneinander zu trennenden Teile: a) das Praxis- oder Orientierungswissen und b) die (adäquaten/praktisch selbst-reflexiven) Beschreibungen von zu erforschenden Praktiken. Ethnomethodologische Daten sind daher allenfalls „aids to a sluggish imagination“ (Garfinkel 1967: 38) oder auch „a source of insight into what we ‚already know‘“ (Lynch 2002: 535) und bilden erst in der Zusammenführung von verkörpertem Praxiswissen und Repräsentationen jener Praxis eine Einheit.

Wie wir sehen konnten, haben die hier diskutierten methodologischen Varianten der Ethnografie unterschiedliche Grundlagen. Diese unterschiedlichen Grundlagen sind bereits auf der Ebene der Erkenntnistheorie und Sozialtheorie angelegt und treten nicht wie oftmals unterstellt wird erst auf der Ebene der angewandten Methoden zu Tage. In diesem methodologischen Sinne pflegt die Ethnografie nicht nur einen „Methodenpluralismus“, sondern auch einen erkenntnis- und sozialtheoretischen Pluralismus, der sich direkt auf die Konstitution ethnografischer Daten auswirkt.

6. Probleme der Archivierung, Anonymisierung und sekundären Nutzung

Aus den hier dargelegten methodologischen Aspekten der Ethnografie (der „Methoden- und Datenpluralismus“, die Besonderheiten des Körpers der Ethnografin und die unterschiedlichen methodologischen Varianten der Ethnografie) ergibt sich für die Fragestellung, wie ethnografische Daten technisch archiviert, anonymisiert und für die sekundäre Nutzung aufbereitet werden können, ein überaus komplexes und differenziertes Bild. In den nachfolgenden Abschnitten werden wir die aufgrund der sozialtheoretischen und methodischen Besonderheiten der Ethnografie entstehenden Probleme diskutieren.

Zunächst ist festzuhalten, dass Ethnografinnen ihr implizites und explizites ethnografisches Wissen in unterschiedlichen Medien „archivieren“. Mit Ausnahme der textuell-semiotischen Ethnografie ist in den Ethnografien der Körper der Ethnografin das primäre „Forschungsinstrument“. Aus einer archivarischen Perspektive stellt der Körper der Ethnografin daher das verkörperte „Zentralarchiv“ ethnografischer Forschung dar. Alle anderen extra-korporalen Materialien werden durch den Körper der Ethnografin bzw. ihr verkörpertes Orientierungswissen referenz- und bedeutungsfähig gemacht und gehalten (oder verworfen). Die extra-korporalen, technisch erzeugten Materialien sind von „außen“ gesehen nur unzusammenhängend wirkende Elemente, die erst durch das Wissen der Ethnografin zu einem Gesamtbild „zusammengefügt“ werden können. Ethnografin und (technisches) Archiv sind daher nicht voneinander zu trennen. Der Körper der Ethnografin ist konstitutiver Teil des Archivs, das analytisch gesehen aus einem körperlichen und technischen Teil besteht. Ethnografisches Wissen entsteht also erst durch die Kombination, Verknüpfung, Referenzierung und Verdichtung jener Materialien, die nur mit Hilfe des ethnografisch erworbenen Orientierungswissens möglich sind.

In diesem Sinne erscheint eine „technische Lösung“ der Archivierung ethnografischer Daten z. B. eine Digitalisierung aller extra-korporalen Materialien nicht sinnvoll, da sie das notwendige implizite

Orientierungswissen nicht mitführen kann. Ohne dieses Wissen fehlt aber die Möglichkeit, die einzelnen Materialien zu verstehen, in Beziehung zueinander und in Relation zu einem wissenschaftlichen Problem zu setzen. Für Dritte, die über das ursprüngliche ethnografische Orientierungswissen der Ethnografin nicht verfügen, sind diese technisch gespeicherten und archivierten Materialien weitgehend unverständlich. Selbst nach einer erschöpfenden Anreicherung der empirischen Materialien durch „Meta-Daten“ bleibt ein Verstehen weiterhin unmöglich, da diese Anreicherungen aus „Selbstbefragungen“ – z. B. in Datensitzungen – immer situiert und spezifisch auf ein konkretes sozialwissenschaftliches Problem und nicht allgemein auf wissenschaftliche Fragen bezogen sind. Dies gilt v. a. für verkörpertes, präreflexives oder vertrautes, als bekannt vorausgesetztes Wissen, das nicht vollumfänglich in technischen Medien archiviert werden kann. Dieses Wissen muss aufwendig „zur Sprache gebracht“ werden bzw. durch die Konfrontation mit technisch-registrierten Daten (z. B. Videos) „erinnert“ und reflexiv gemacht werden. Ein solches Reflexivmachen ist jedoch zeitintensiv und kontextspezifisch und eignet sich daher kaum für standardisierte Verfahren der Archivierung. Dieses Problem gilt v. a. für die Anwendung der korporal-kopräsenten und ethnomethodologischen Ethnografie und ihrer sozialtheoretischen Überzeugungen. In der textuell-semiotischen Ethnografie wiederum sind Beobachtungen und Interpretationen nicht voneinander zu trennen und können daher auch nicht getrennt voneinander archiviert werden. Die Möglichkeiten einer sekundären Nutzung derartiger Materialien wurde bislang noch nicht erprobt.

Die empirischen Materialien enthalten außerdem oft orts- und personenbezogene Elemente, die für ein näheres Verständnis und auch für das Aufeinanderbeziehen unterschiedlicher Wissensobjekte notwendig sind. Auch wenn diese Elemente in den Publikationen punktuell z. B. durch Pseudonyme ersetzt werden können, würde eine „flächendeckende“ Pseudonymisierung a) zu einer Bedeutungsverschiebung führen, b) die Aussagekraft der Materialien verringern und c) zu einem Arbeitsaufwand führen, den die Ethnografin, die eigentlich immer nur punktuell pseudonymisiert, kaum neben ihrer Forschung bewältigen könnte.

Ein letztes Problem sehen wir mit Blick auf die intime soziale Beziehung zwischen Ethnografinnen und Teilnehmerinnen (Beforschten), deren Vertrauensbeziehung geschützt werden muss. Mit einer Verpflichtung zur Archivierung und Bereitstellung von „primären Forschungsdaten“ wäre diese Beziehung gefährdet. Will man die Aussagekraft möglichst weitreichend erhalten, sollte keine Anonymisierung, durchgeführt werden. Der Schutz der ethnografischen Beziehung zwischen Ethnografin und Teilnehmerin ist dann gefährdet bzw. die ethnografische Forschung scheitert bereits beim Versuch der Etablierung eines Feldzugangs. Will man alternativ eine Pseudonymisierung vollumfänglich und systematisch durchführen, stellt sich die Frage, ob dieser Aufwand der Aufbereitung nicht den zu erwartenden Nutzen übersteigt. Ethnografie ist immer ein kollaboratives Unternehmen, dessen „Schlüsselinformantinnen“ geschützt werden müssen, und dies unabhängig von zu erwartendem wissenschaftlichen Erkenntniswert.

7. Fazit

Im Fazit sollen fünf bereits angesprochene Probleme noch einmal besonders hervorgehoben werden, die die Archivierbarkeit und sekundäre Nutzbarkeit ethnografischer Daten massiv einschränken.

Erstens sind ethnografische Daten kontextspezifische bzw. kontextsensible Dokumente sozialer Wirklichkeit, die auf ein spezifisches sozialwissenschaftliches Problem hin ausgerichtet sind. Daten sind also keine allgemeinen oder objektiven Repräsentationen sozialer Wirklichkeit, sondern Artefakte, Konstrukte und Spuren, die erst in einem transformativen Prozess verobjektiviert werden.

Dementsprechend sind ethnografische Daten nur eingeschränkt zur Bearbeitung anderer wissenschaftlicher Probleme nutzbar. Sekundäre Nutzungen können diese Kontextspezifizität kaum umgehen.

Zweitens müsste ein „Sekundär-Ethnograf“ den Grad an praktischer Kompetenz mitbringen, den auch der „Primär-Ethnograf“ im Verlauf seiner Teilnahme an den beobachteten Forschungssituationen erworben hat. Diese praktische und verkörperte Kompetenz ist notwendig, um die (archivierten) empirischen Materialien adäquat zu verstehen. Dazu muss der „Sekundär-Ethnograf“ entweder eine eigene ethnografische Feldforschung durchführen oder bereits vor der sekundären Nutzung ethnografischer Daten diese Kompetenzen besitzen. Andernfalls würde ihm das notwendige Orientierungswissen fehlen.

Drittens gilt, dass für einzelne ethnografische Daten, die nur ein allgemeines Orientierungswissen voraussetzen, eine sekundäre Nutzung denkbar wäre. Oder mit Schütz (1932) gesprochen: Je anonymer (im phänomenologischen Sinne) die in den Daten dokumentierten Handlungen bereits sind, desto einfacher sind diese Daten für sekundäre Nutzungen und damit für andere Problem- und Fragestellungen nutzbar. Dies trifft z. B. für technisch-registrierte Audioaufzeichnungen zu, die in der Konversationsanalyse bereits seit den 1960er Jahren als Daten auch sekundär genutzt werden.

Viertens sind Daten, für deren Verständnis spezifisches und nur bedingt dokumentierbares Orientierungswissen notwendig ist, nur eingeschränkt nutzbar. Dazu trägt auch die Tatsache bei, dass ethnografische Daten häufig nur in stark vorinterpretierter Form und damit auf ein wissenschaftliches Problem spezifizierter Form vorliegen.

Fünftens ergeben sich arbeitsökonomische Herausforderungen, wenn Ethnologinnen ihre empirischen Materialien und ihre Analysen mit diesen Materialien noch stärker als bisher dokumentierten müssten. Hier stehen Arbeitslast bzw. Finanzierungsbedarf und zu erwartender Erkenntnisgewinn durch sekundäre Nutzungen in keiner ausgewogenen Relation. Daneben ergeben sich große ethische Herausforderungen, sollten Ethnologinnen ihre gesamten Materialien, in denen unvermeidbar personenbezogene Elemente enthalten sind und für deren Verständnis auch enthalten sein müssen, archivieren und teilen. Eine Anonymisierung und Pseudonymisierung ist für ethnografische Daten immer nur in Teilen, aber nie für die Gesamtheit des ethnografischen Datenkorpus möglich, ohne dass die Aussagekraft massiv reduziert würde.

Es ist also keinesfalls zu empfehlen, dass eine Datenarchivierung von den Organisationen der Forschungsförderung standardmäßig verlangt wird oder ein Rechtfertigungszwang einseitig für eine Nicht-Archivierung erzeugt wird. Vielmehr müssen beide Varianten, Archivierung und Nicht-Archivierung, überdacht und begründet werden. Für den Fall, dass eine Archivierung sinnvoll und für Sekundäranalysen nützlich erscheint, muss der dazu notwendige hohe Aufwand vollumfänglich finanziert werden. Wie gezeigt wurde, ist dieser bei ethnografischen Forschungen besonders hoch und darf in seinem finanziellen, personalen und infrastrukturellen Ausstattungsbedarf nicht unterschätzt werden.

Literaturverzeichnis

- Amann, Klaus und Stefan Hirschauer (1997): Die Befremdung der eigenen Kultur. Ein Programm. In: Stefan Hirschauer und Klaus Amann (Hrsg.): Die Befremdung der eigenen Kultur. Zur ethnographischen Herausforderung soziologischer Empirie. Frankfurt am Main, Suhrkamp, 7–52.
- Breidenstein, Georg; Stefan Hirschauer; Herbert Kalthoff und Boris Nieswand (2013): Ethnografie. Die Praxis der Feldforschung. Konstanz, UVK-Verlag.
- Elwert, Georg (2003): Feldforschung. Orientierungswissen und kreuzperspektivische Analyse. Berlin, Schiler.
- Garfinkel, Harold (1967): Studies in Ethnomethodology. Englewood Cliffs (NJ), Prentice Hall.
- Garfinkel, Harold und D. Lawrence Wieder (1992): Two Incommensurable, Asymmetrically Alternate Technologies of Social Analysis. In: Graham Watson und Robert M. Seiler (eds.): Text in Context. Contributions to Ethnomethodology. New York, Sage, 175–206.
- Geertz, Clifford (1973): The Interpretation of Cultures. New York, Basic Books.
- Goffman, Erving (1989): On Fieldwork. *Journal of Contemporary Ethnography* 18(2), 123–132.
- Hirschauer, Stefan (2001): Ethnographisches Schreiben und die Schweigsamkeit des Sozialen. Zu einer Methodologie der Beschreibung. *Zeitschrift für Soziologie* 30(6), 429–451.
- Huschka, Denis; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.) (2013): Forschungsinfrastrukturen für die qualitative Sozialforschung. Berlin, SCIVERO.
- Kalthoff, Herbert (2003): Beobachtende Differenz. Instrumente der ethnografisch-soziologischen Forschung. *Zeitschrift für Soziologie* 32(1), 70–90.
- Lynch, Michael (2002): From naturally occurring data to naturally organized ordinary activities. Comment on Speer. *Discourse Studies* 4(4), 531–537.
- Malinowski, Bronislaw (1922): Argonauts of the Western Pacific. An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea. London, Routledge & Kegan Paul Ltd.
- Meier zu Verl, Christian (2018): Daten-Karrieren und epistemische Materialität. Eine wissenschaftssoziologische Studie zur methodologischen Praxis der Ethnografie. Stuttgart, J.B. Metzler.
- Meier zu Verl, Christian und Christian Meyer (2017): Die zwei Körper des Ethnografen. Konstanz. Unveröffentlichtes Manuskript.
- Rivers, William H. R. (1901): Vision. In: Alfred C. Haddon (ed.): Reports of the Cambridge Anthropological Expedition to Torres Straits (Vol. II). Cambridge, Cambridge University Press, 8–132.
- Schütz, Alfred (1932): Der sinnhafte Aufbau der sozialen Welt. Eine Einleitung in die verstehende Soziologie. Wien, Springer.
- Schütz, Alfred (1953): Common-Sense and Scientific Interpretation of Human Action. *Philosophy and Phenomenological Research* 14(1), 1–38.

Dr. Christian Meier zu Verl ist wissenschaftlicher Mitarbeiter an der Professur für Allgemeine Soziologie und Kulturosoziologie sowie am Exzellenzcluster „Kulturelle Grundlagen von Integration“ der Universität Konstanz. Seine Forschungsschwerpunkte sind Kulturosoziologie, Interaktionssoziologie, Wissenschafts- und Techniksoziologie und Qualitative Methoden. Buchpublikation: Daten-Karrieren und epistemische Materialität (Stuttgart 2018)

Prof. Dr. Christian Meyer ist Professor für Allgemeine Soziologie und Kulturosoziologie an der Universität Konstanz. Zuvor hatte er Professuren an den Universitäten Würzburg (Spezielle Soziologie und Qualitative Methoden der Sozialforschung) und Duisburg-Essen (Kommunikationswissenschaft) inne. Von 2012 bis 2014 war er Mitglied im Vorstand der DGS-Sektion „Qualitative Methoden“. Seine Forschungsschwerpunkte sind Kulturosoziologie, Interaktionssoziologie, Sozialtheorie und Qualitative Methoden. Neuere Buchpublikationen: Culture, Practice, and the Body (Stuttgart 2018), Intercorporeality (Hg., New York 2017) und Moving Bodies in Interaction-Interacting Bodies in Motion (Hg., Amsterdam 2017).

Archivierung und Nachnutzung qualitativer Daten aus forschungsethischer Perspektive

Hella von Unger

Abstract

Dieser Beitrag diskutiert die Potentiale und Grenzen der digitalen Archivierung und Sekundärnutzung von qualitativen Daten aus forschungsethischer Perspektive. Ein Schwerpunkt liegt auf offenen Fragen der Schadenvermeidung – sowohl für die Studienteilnehmenden als auch für die Forschenden. Weil die Möglichkeiten der Anonymisierung qualitativer Daten stark eingeschränkt sind, sich mögliche Risiken nur begrenzt antizipieren lassen und der potentielle Nutzen in einem fraglichen Verhältnis zu dem Aufwand steht, erscheint die digitale Archivierung qualitativer Daten zu diesem Zeitpunkt nicht generell, sondern nur für bestimmte Datensorten, in begründeten Fällen und bei Sicherstellung einer angemessenen Vorbereitung und Begleitung forschungsethisches vertretbar. Welche Datentypen dies sind, muss näher bestimmt werden. Um die Debatte und die Dateninfrastruktur weiter zu entwickeln, ist es zudem notwendig, die offenen Fragen, z. B. nach dem Erkenntnisgewinn durch Nachnutzung sowie nach den Risikoeinschätzungen und Schadenpotential für Studienteilnehmende und Forschende, empirisch zu untersuchen.

1. Einleitung

Die digitale Archivierung und Sekundärnutzung qualitativer Daten sind sowohl mit vielversprechenden Möglichkeiten als auch mit erheblichen Problemen verbunden und werden in der Fachdebatte seit geraumer Zeit kontrovers diskutiert (Corti/Witzel/Bishop 2005; Hirschauer 2014; Huschka et al. 2013; Jagodzinski/Schuman/Witzel 2005; Kluge und Opitz 1999; Mruck et al. 2000; von Unger 2015). Dabei stehen oft forschungspraktische und datenschutzrechtliche, teilweise auch methodologische Fragen im Mittelpunkt. Dieser Beitrag nimmt eine *forschungsethische* Perspektive auf das Thema ein. Es wird zunächst geklärt, wodurch sich diese Perspektive auszeichnet. Anschließend werden Fragen der Bereitstellung und Nachnutzung qualitativer Forschungsdaten im Hinblick auf zentrale ethische Grundsätze diskutiert. Abschließend nehme ich eine zusammenfassende Abwägung vor.

2. Forschungsethische Reflexivität

Eine forschungsethische Perspektive begreift Forschung als sozialen Prozess und fragt nach den Werten und Prinzipien, die das Forschungshandeln leiten. Ethische Fragen werden in der qualitativen Forschung lebendig diskutiert (Canella und Lincoln 2011; Iphofen und Tolich 2018; Hammersley und Traianou 2012; Hopf 2010 [2000]; Roth und von Unger 2018; Siri 2013; von Unger/Narimani/M'Bayo 2014). Die forschungsethische Perspektive, die hier vertreten wird, zeichnet zudem aus, dass sie nicht nur den wissenschaftlichen Erkenntnisgewinn als höchsten oder gar einzigen Wert der Forschung proklamiert, sondern weitere Werte und Prinzipien berücksichtigt. Vor dem Hintergrund der historischen Entwicklung der Debatte (von Unger und Simon 2016) geht es darum, Forschung als sozialen Prozess zu adressieren, der Konsequenzen für das Leben der beteiligten Personen und der Gesellschaft haben kann – und dies bei der Gestaltung der Forschungsbeziehungen zu berücksichtigen. Ein zentrales forschungsethisches Anliegen ist es beispielsweise, Schaden zu vermeiden und die Rechte und Anliegen der Teilnehmer/innen zu achten. Das heißt, nicht jedes Vorgehen, das interessante wissenschaftliche Ergebnisse verspricht, ist angemessen – es muss auch im Hinblick auf die mit ihm einhergehenden Risiken und antizipierbaren Folgen vertretbar sein. Allerdings stellt auch die Gewährleistung der wissenschaftlichen Qualität ein zentrales ethisches Anliegen dar, dem ein fundamentaler Stellenwert zukommt. Eine forschungsethische Einschätzung beinhaltet also immer Abwägungen und Gewichtungen von Prinzipien – bezogen auf einen konkreten Fall oder eine spezifische Forschungssituation. Da sich ethische Fragen oft erst im Verlauf des Forschungsprozesses ergeben und nur eingeschränkt vorab antizipieren lassen, sprechen wir von einer forschungsethischen Reflexivität, die im Prozess sensibel auch auf unvorhergesehene Wendungen reagiert und das Forschungshandeln fortwährend einer kritischen Prüfung unterzieht (von Unger/Narimani/M'Bayo 2014; von Unger 2018). Auch vor diesem Hintergrund ist eine einmalige Vorabprüfung ethischer Fragen (z. B. im Rahmen einer Begutachtung durch eine Ethikkommission) nicht ausreichend.¹

3. Kodifizierte Prinzipien und Grundsätze

Ethische Prinzipien und Grundsätze liegen auch in kodifizierter Form vor, z. B. in Form von Ethik-Kodizes der Fachgesellschaften. In Deutschland verabschiedeten beispielsweise der Berufsverband Deutscher Soziologinnen und Soziologen (BDS) und die Deutsche Gesellschaft für Soziologie (DGS)

¹ Die Begutachtung ethischer Aspekte durch Kommissionen, wie dies beispielsweise in angloamerikanischen Ländern auch für die sozialwissenschaftliche Forschung üblich ist, kann hier nicht weiter vertieft werden. Insbesondere qualitativ Forschende üben an diesen Verfahren scharfe Kritik (vgl. Canella und Lincoln 2011; Hammersley und Traianou 2012; van den Hoonaard 2011; von Unger/Dilger/Schönhuth 2016).

Anfang der 1990er Jahre einen gemeinsamen Ethik-Kodex, der sich an dem Kodex der American Sociological Association (ASA) orientierte und seitdem nur minimal überarbeitet wurde (DGS und BDS 2017). Zu den zentralen Grundsätzen dieses Kodex gehören: 1) das Streben nach Objektivität und Integrität der Forschenden und 2) der Schutz und die Rechte der Teilnehmenden, wie z. B. Risikoabwägung und Schadensvermeidung, die Freiwilligkeit der Teilnahme, die informierte Einwilligung und die Vertraulichkeit und Anonymisierung der Daten. Diese Grundsätze liefern eine hilfreiche Orientierung für das Forschungshandeln, werfen jedoch auch Fragen und Probleme auf – nicht nur aber auch in der qualitativen Forschungspraxis (von Unger 2014). Sie müssen grundsätzlich projektbezogen interpretiert, auf ihre Relevanz für die jeweilige Forschungssituation geprüft und mit-/bzw. gegeneinander abgewogen werden. Zudem sollten sie, ähnlich wie andere Dokumente dieser Art, als „living document“² behandelt und unter Beteiligung einer breiteren Fachgemeinschaft weiter entwickelt werden. So findet sich beispielsweise in dem soziologischen Kodex bislang kein Hinweis auf Fragen der Archivierung und Sekundärnutzung.

4. Archivierung und Sekundärnutzung – was spricht dafür, was dagegen?

Sozialwissenschaftliche Forschungsdaten, die von den Studienteilnehmenden für wissenschaftliche Zwecke zur Verfügung gestellt und in der Regel mit hohem Aufwand erhoben werden, sind ein wertvolles Gut, das verantwortlich bearbeitet, angemessen aufbewahrt und möglichst umfassend genutzt werden sollte. *Data sharing* innerhalb der Fachcommunity, also das Bereitstellen von empirischem Material zur Nutzung durch andere Wissenschaftler/innen, eröffnet in dieser Hinsicht die Möglichkeit, bereits erhobenes Material umfassender zu nutzen. Eine Struktur der digitalen Archivierung, die dies auch längerfristig ermöglicht, verspricht eine größere Nachhaltigkeit in der Nutzung und zusätzliche Erkenntnisgewinne – zumindest auf den ersten Blick. Die Frage, ob sich qualitative Daten zur Sekundärnutzung überhaupt eignen, ist jedoch umstritten – insbesondere für solche Varianten der qualitativen Forschung, wie z. B. die Ethnografie oder Grounded Theory-Studien, die einen starken, methodologisch begründeten Fokus auf die Erhebung und Analyse von *Primärdaten* legen (vgl. Hirschauer 2014). An dieser Stelle wird deutlich, dass es notwendig ist, in dem vielfältigen Spektrum der qualitativen Forschung zu differenzieren und zwischen verschiedenen Ansätzen und Datenformen zu unterscheiden.

Mit der Archivierung von sozialwissenschaftlichen Daten wird zudem häufig die Hoffnung verbunden, die Forschung könne dadurch transparenter und glaubwürdiger werden. Doch während dies für quantitative Forschungsdaten zutreffend sein mag, wo sich Auswertungen rechnerisch prüfen lassen, die Daten entsprechend beschaffen sind und Sekundäranalysen traditionell einen hohen Stellenwert haben, stellt sich dies bei qualitativen Daten und Analysen anders dar. In der qualitativen Forschung lassen sich Daten und Ergebnisse nicht in der gleichen Weise replizieren (Strübing 2018) und das methodische Vorgehen fußt auf anderen erkenntnistheoretischen Grundlagen. Bei allen Differenzen zwischen qualitativen Ansätzen: die Ergebnisse sind Interpretationen, die qualitativen Gütekriterien entsprechend im empirischen Material verankert und intersubjektiv nachvollziehbar sein müssen, beispielsweise durch angemessene Zitation von Auszügen aus Rohdaten und Beschreibungen des methodischen Vorgehens in den Veröffentlichungen. Da sich Interpretationen jedoch nicht ‚nachrechnen‘ lassen und die Bereitstellung des Rohmaterials in vielen Fällen aus methodologischen Gründen als wenig sinnvoll, aus forschungspraktischen Gründen als wenig praktikabel und aus

² Vgl. bspw. die überarbeitete kanadische Richtlinie (TCPS 2)
<http://www.pre.ethics.gc.ca/eng/archives/draft-preliminaire/Default/> (Zugriff am 10.9.2018).

forschungsethischen Gründen als hoch problematisch erachtet wird, muss die Glaubwürdigkeit der Forschung auf anderem Weg hergestellt werden.³

Wenn die nachhaltige Nutzung, ein zusätzlicher Erkenntnisgewinn und eine bessere Nachvollziehbarkeit für die Archivierung und Sekundärnutzung qualitativer Daten angeführt werden, welche forschungsethischen Gründe sprechen nun genau dagegen? Neben den bereits erwähnten methodologischen Zweifeln an der Durchführbarkeit und Ergiebigkeit von Sekundäranalysen, sind weitere forschungsethische Grundsätze zu berücksichtigen, die verstärkt die Rechte und das Wohlergehen der beteiligten Personen in den Blick nehmen.

5. Risikoeinschätzung und Schadenvermeidung

Ein zentrales ethisches Anliegen ist es, Forschung so zu gestalten, dass den Teilnehmenden möglichst kein Schaden aus der Teilnahme entsteht, ihre Selbstbestimmung gefördert wird und die Risiken, die mit der Forschung einhergehen, gering gehalten werden und vertretbar sind. Die Forschenden sind aufgefordert, potentiellen Schaden zu antizipieren und zu reduzieren (z. B. DGS und BDS 2017, §2). Mögliche Gefahren können nicht nur körperliche und psychische Wohlbefinden der Teilnehmer/innen betreffen (wie z. B. die Möglichkeit der Re-Traumatisierung durch Erzählungen). Weit zentraler für die meisten sozialwissenschaftlichen Daten, insbesondere für solche mit Personenbezug, sind negative soziale Folgen, die mit dem Bekanntwerden der Daten verbunden sind; diese können beispielsweise Stigmatisierungs- und Exklusionsprozesse umfassen, die auch ökonomische (z. B. Verlust der Beschäftigung oder Einnahmequelle) oder rechtliche (z. B. straf-, zivil-, versicherungs- und aufenthaltsrechtliche) Folgen für die Beteiligten nach sich ziehen können. Allerdings besteht generell das Problem, dass Risiken nur begrenzt antizipierbar sind. Zudem stellt sich die Frage, wer Risiken und insbesondere die Risiken für andere eigentlich realistisch und angemessen einschätzen kann? Im Hinblick auf das vorliegende Thema sind die möglichen Risiken, die mit einer digitalen Archivierung und Nachnutzung von qualitativen Daten einhergehen, schwer einzuschätzen. Grundsätzlich lassen sich Risiken nur begrenzt antizipieren und die Einschätzung, welche Risiken „im Alltag üblich“ (DGS und BDS 2017, §2(5)) sind, hängt stark vom Standpunkt der Betrachterin ab. Die Perspektiven der Teilnehmenden selbst sollten bei der Einschätzung möglicher Risiken mit einbezogen werden. Eine besondere Schwierigkeit besteht zudem in der Beschaffenheit qualitativer Daten: Diese können i. d. R. nur mit sehr hohem Aufwand, unter großen Verlusten der Aussagekraft und nie vollständig anonymisiert werden, wodurch die Möglichkeiten der Schadensminimierung durch Anonymisierung erheblich eingeschränkt sind (siehe unten). Hinzu kommt die zukünftige Entwicklung und Vernetzung digitaler Technologien, die beispielsweise durch Gesichtserkennungen oder geographische Lokalisierungen die Anonymisierung von visuellen Daten ad absurdum führen kann.

In der jüngeren, internationalen Diskussion werden zunehmend auch solche Risiken adressiert, die nicht nur für die Teilnehmenden (also die Proband/innen, Interviewpartner/innen, Interlokutor/innen etc.) sondern auch für die Forschenden mit einer Studie verbunden sein können. Dies trifft zum Beispiel auf Forschung in Krisengebieten zu und auf weitere Forschungskontexte, die Risiken für das Wohlergehen der Forschenden beinhalten können (z. B. emotional beanspruchende Forschung (Bahn und Weatherill 2012)). Das heißt, auch für die in forschender Funktion Beteiligten sind mögliche

³ Selbstverständlich wird auch in der qualitativen Forschung und Methodenausbildung überprüft, ob Daten tatsächlich erhoben, angemessen aufbereitet, ausgewertet und zitiert werden. Dies geschieht jedoch bereits im Erhebungs- und Forschungsprozess, z. B. durch Besprechungen des Materials in Interpretationsgruppen und Forschungswerkstätten. In der Methodenausbildung erfordert das *data sharing*, das beispielsweise in Lehrforschungsprojekten stattfindet, nicht nur eine intensive forschungsethische Reflexivität, sondern auch ein sorgfältiges Datenmanagement (https://www.qualitative-sozialforschung.soziologie.uni-muenchen.de/ressourcen/hinweise_qualitativ1/index.html, Zugriff am 10.09.2018.).

Risiken und negative Auswirkungen mit zu berücksichtigen und geeignete Gegenmaßnahmen zu ergreifen. Welche Risiken gehen nun mit der Archivierung und Sekundäranalyse qualitativer Daten *für die Forschenden* einher? Auch hier hängt die Beantwortung der Frage stark davon ab, welche Daten wie erhoben wurden. Qualitative Daten, die mithilfe von „Nah-Methoden“ (Breuer/Mruck/Roth 2002: 3) generiert werden, also Interviews, Feldnotizen, Beobachtungsprotokolle, Postskripte, u. Ä. geben oft sehr viel über die Forschenden preis. Stellt reflektierte Subjektivität ein Gütekriterium der Forschung dar, so ist es erforderlich, diese Interaktivität und Subjektivität im Forschungsprozess zu dokumentieren, um sie einer kritischen Analyse unterziehen zu können. Wird auf Basis konstruktivistischer Annahmen erkenntnistheoretisch von dem Einschluss der Beobachterin in den Erkenntnisprozess ausgegangen, muss dem im Erhebungsprozess Rechnung getragen werden. Diese Daten geben infolge dessen viel über die Forschenden preis und sind im Grunde nicht anonymisierbar. Die damit verbundenen Risiken sind auch im Hinblick auf Forschungsteams und insbesondere für die Beteiligung von abhängig beschäftigten Mitarbeiter/innen (oder sogar Studierenden) nicht zu unterschätzen.

6. Informierte Einwilligung

Der Grundsatz der informierten Einwilligung ist datenschutzrechtlich festgeschrieben und erscheint vor dem Hintergrund seiner geschichtlichen Entwicklung auf den ersten Blick wenig kontrovers (von Unger und Simon 2016). Fragen ergeben sich jedoch spätestens bei der praktischen Umsetzung: Wie viel Information ist ausreichend – und in welcher Form soll das Einverständnis eingeholt werden? Eine digitale Archivierung und Sekundärnutzung von Daten erscheint nur dann vertretbar, wenn sie auf der expliziten und informierten Einwilligung der Teilnehmenden beruht. Mit dieser Voraussetzung potenzieren sich jedoch grundlegende Probleme des *informed consent*. Aufgrund der größeren Flexibilität und der eingeschränkten Planbarkeit explorativer Prozesse ist es in der qualitativen Forschung oft nicht möglich, zu Beginn genau zu bestimmen, wie der Forschungsprozess verlaufen und zu welchen Resultaten er führen wird. Durch das Prinzip der Offenheit können die Inhalte, die im Rahmen der informierten Einwilligung vermittelt werden sollen, nicht in der gleichen Form wie bei quantitativen Studien vorab festgelegt und kommuniziert werden. Zudem hat die schriftliche Einverständniserklärung ihre Tücken und so wird das informierte Einverständnis in vielen qualitativen Studien angemessener als iterativer und dialogischer Prozess verstanden (Narimani 2014, von Unger 2018; Warin 2011). Zudem weisen kritische Stimmen darauf hin, dass das Konzept des informierten Einverständnisses einen Mittelschichts-Bias hat (M'Bayo 2014) und für manche Varianten der qualitativen Forschung, wie beispielsweise die ethnografische Feldforschung, nur wenig passend und angemessen ist (Bell 2014). Die digitale Archivierung und Sekundärnutzung verschärft nun diese Probleme: Forschende sind angehalten, formal und schriftlich die informierte Einwilligung der Teilnehmenden einzuholen, wobei sie nur sehr wenig darüber sagen können, in welchen Kontexten die Daten zukünftig Verwendung finden werden. Das heißt, sie benötigen einen noch größeren Vertrauensvorschuss als bei regulärer qualitativer Forschung ohne Nachnutzung und gleichzeitig sind dafür schlechtere Voraussetzungen gegeben, denn der Prozess ist noch formalisierter, da die Einwilligung schriftlich einzuholen ist. Damit ergeben sich nicht zu unterschätzende Risiken für die Qualität und Durchführbarkeit der Forschung, insbesondere für Studien mit marginalisierten Personengruppen, bei denen solch gut gemeinte Prozesse auf Unverständnis und Skepsis stoßen können (von Unger 2018). Allerdings steht auch im Hinblick auf die Auswirkungen einer schriftlichen Einwilligung, die eine digitale Archivierung und Nachnutzung der Daten einschließt, eine empirische Untersuchung noch aus, die genauer nach Studienkontexten, beteiligten Personen und Datensorten unterscheidet.

7. Vertraulichkeit und Anonymisierung

Auch bei den Grundsätzen Vertraulichkeit und Anonymisierung – zwei der wesentlichen Schutzmaßnahmen, die in der Sozialforschung ergriffen werden, um die Beteiligten, ihre Daten und ihre Rechte zu schützen – ergeben sich nicht zu unterschätzende Herausforderungen. Inwiefern ist der Grundsatz der Vertraulichkeit noch gegeben, wenn wir die Daten weitergeben? Diese Frage ist eng mit Fragen der Anonymisierung verbunden. In der qualitativen Forschung haben Indexikalität, Kontingenz und Kontextualität einen hohen methodologischen Stellenwert – d. h. qualitativ Forschende arbeiten immer mit Aussagen, Deutungen und Beobachtungen *in einem spezifischen Kontext*. Die Daten machen nur Sinn und beziehen ihrer Aussagekraft aus diesem Kontext – dieser kann also nicht ohne Weiteres unkenntlich gemacht werden. Praktiken der Anonymisierung bewegen sich daher in dem Spannungsfeld zwischen forschungsethischen und datenschutzrechtlichen Vorgaben einerseits und den jeweils gültigen methodologischen Erfordernissen der qualitativen Forschungsrichtung und sowie ggf. den Wünschen und Anliegen der beteiligten Personen andererseits.

Forschungspraktisch gehen Strategien der Anonymisierung über die Löschung von Personennamen und das Verwenden von Pseudonymen hinaus. Saunders und Kolleg/innen (2015) beschreiben ausführlich, wie sie Anonymisierungen (von Personennamen, Ortsnamen, religiöser und kultureller Zugehörigkeit, Berufstätigkeit, familiären Beziehungen und anderen Informationen, die die Identität der Person preisgeben könnten) in einer qualitativen Interviewstudie vorgenommen haben – immer in dem Bemühen, eine kontext-sensible Balance zwischen dem Schutz der Interview-Partner/innen (in diesem Fall: Angehörige von Menschen im Wachkoma) und dem Bewahren der Reichhaltigkeit und der Aussagekraft der Daten zu finden. Sie weisen darauf hin, dass es von Vorteil ist, die Form und das Ausmaß der Anonymisierung auch mit den Interview-Partner/innen abzusprechen, denn auch gut gemeinte Schutzstrategien können von den Betroffenen als unethisch empfunden werden. Aber auch Forschende, die dies tun, sind nicht vor Überraschungen gefeit – einer ihrer Forschungsteilnehmer hat beispielsweise bei einem sorgfältig anonymisierten Open-Access-Artikel sein eigenes Pseudonym ausgetüftelt und den Online-Artikel anschließend mit seiner Facebook-Seite verlinkt (Saunders/Kitzinger/Kitzinger 2015, S. 629–30).

Qualitative Daten weisen durch ihre besondere Beschaffenheit und Kontextdichte die Eigenart auf, dass sie, auch wenn sie sorgfältig anonymisiert sind, allein dadurch, *wie* eine Interaktion oder ein Ort beschrieben wird oder wie eine Lebensgeschichte erzählt wird, Rückschlüsse auf die Person möglich machen können – insbesondere für Personen, die diese Orte und Personen gut kennen. Es bedarf also maßgeschneiderter Lösungen und des Eingeständnisses, dass in Zeiten des Internets, wo sich die Grenzen zwischen Privatheit und Öffentlichkeit verschieben, Strategien der Anonymisierung in vielfältiger Hinsicht unterlaufen werden können. Mit den verfügbaren Informationen und Technologien ist es ein Leichtes, über die Kombination bestimmter Schlagwörter in einer Suchmaschine Rückschlüsse auf beteiligte Personen und Institutionen zu ziehen. Daher wird auch gefragt, ob der Grundsatz der Anonymisierung überhaupt noch praktikabel und in jedem Fall angebracht und zielführend ist (Tilley und Woodthorpe 2011).

8. Fazit: Digitale Archivierung auf dem Prüfstand

Zusammenfassend kann festgehalten werden, dass einerseits viel für die Möglichkeiten der digitalen Aufbewahrung und Nachnutzung von qualitativen Daten spricht: Diese sind ein kostbares Gut und können möglicherweise auch Jahre später noch gewinnbringend für Sekundäranalysen genutzt werden. Allerdings steht der Beleg für diese Hoffnung noch aus und es ist wahrscheinlich, dass sich manche Datensorten dazu besser eignen (z. B. Sammlungen von frei verfügbaren Dokumenten, die im Rahmen einer Diskursanalyse ausgewertet wurden) als andere (z. B. ethnografisches Material). Auch das

Anliegen, die empirische Arbeit von qualitativ Forschenden zu überprüfen, ist grundsätzlich im Hinblick auf die Rechenschaftspflicht von Forschenden, deren Arbeit ja größtenteils durch öffentliche Gelder und Infrastrukturen ermöglicht wird, durchaus legitim. Könnte man die Rohdaten einsehen, so hoffen manche, wären die Vorgehensweisen und Ergebnisse (noch) besser nachvollziehbar. Auf der anderen Seite bestehen grundlegende Probleme. Ganz zentral stellen sich Fragen der Anonymisierung, denn die meisten qualitativen Daten sind nur mit sehr hohem Aufwand und großen Verlusten an Datenqualität und heuristischem Wert anonymisierbar. Eine Möglichkeit, mit diesem Problem umzugehen, könnte lauten, nicht-anonymisierte Daten zur Verfügung zu stellen. Dies wirft jedoch weitere forschungsethische und datenschutzrechtliche Fragen auf: Ist das gesetzeskonform? Wer darf diese Daten einsehen und nutzen? Wie kann Schaden und Missbrauch vermieden werden? Ein weiteres Problem besteht im Hinblick auf die informierte Einwilligung der Teilnehmer/innen und den Auswirkungen, die es hätte, wenn die Teilnehmenden nicht nur für den einen spezifischen Studienkontext, sondern für weitere, in der Zukunft liegende Untersuchungen durch Wissenschaftler/innen, die sie nicht kennen, ihr generelles Einverständnis geben müssten. Manch ein/e Teilnehmer/in würde dies abschrecken. Hinzu kommen methodische Bedenken, inwiefern unter diesen Umständen – wenn unklar ist, wer die Daten für welche Zwecke nutzt – anders erzählt (und gefragt) wird? All diese Fragen erfordern empirische Untersuchungen.

Trotz der vielfältigen Probleme und offenen Fragen könnte die digitale Archivierung und Sekundärnutzung qualitativer Daten forschungsethisch dennoch vertretbar sein, nämlich wenn die Risiken für die Beteiligten relativ gering und der Nutzen – auch im Hinblick auf den Erkenntnisgewinn – besonders hoch wären. Das Streben nach wissenschaftlich hochwertiger Forschung stellt ein zentrales forschungsethisches Anliegen dar. In diesem Zusammenhang stehen die Aussicht auf Erkenntnisgewinne und methodisch-methodologische Bedenken, dass die prospektive Archivierung und Sekundäranalyse der Daten den Erhebungsprozess und damit die Qualität der Forschung negativ beeinflusst, relativ unversöhnlich gegenüber. Hier gilt es allerdings die Vielfalt der Forschungssituationen zu berücksichtigen und zu fragen, für welche Situationen und Daten diese Einschränkungen nicht gegeben sind.

Vor dem Hintergrund der hier dargestellten Erwägungen erscheint es verfehlt, oder zumindest verfrüht, eine digitale Archivierung und Bereitstellung von qualitativen Daten grundsätzlich zu empfehlen. Auch die Sektionen Biographieforschung und Qualitative Methoden der Deutschen Gesellschaft für Soziologie haben 2014 eine gemeinsame Resolution verabschiedet, die besagt, dass es sich bei der digitalen Archivierung um eine freiwillige Option handeln muss, die nur mit der informierten Einwilligung der Teilnehmer/innen, unter sachkundiger Anleitung (z. B. zu den Möglichkeiten der Re-Kontextualisierung des Materials) und bei ausreichenden Ressourcen empfehlenswert ist. Bestehende Archive (z. B. Qualiservice in Bremen) archivieren bislang vor allem leitfadengestützte Interviewdaten. Möglicherweise erfordern visuelle Daten, die beispielsweise mittels fokussierter Videographien erhoben wurden (vgl. Knoblauch und Wilke in diesem Band: 47–58) ganz andere Verfahren und Strukturen. Soziologische Studien, die die Möglichkeiten und Grenzen der digitalen Archivierung und Sekundärnutzung von weiteren Datensorten empirisch untersuchen, wären ein großer Gewinn, um manche der zur Zeit noch offenen Fragen datengestützt zu beantworten. Denn eine generelle Absage ist ebenso verfehlt wie eine unkritische Übernahme der technologischen Möglichkeiten, die für quantitative Forschung entwickelt wurden. Wir leben und forschen im digitalen Zeitalter und es ist durchaus denkbar, dass es für einige Studienkontexte machbar, sinnvoll und nützlich sein kann, die erhobenen Daten (oder Teile des Datenkorpus) so aufzubereiten, dass sie – mit Einwilligung der Teilnehmer/innen und Beschränkungen des Zugangs und der Nutzung – für andere einsehbar und für Sekundärnutzungen verfügbar sind.

Literaturverzeichnis

- Bahn, Susanne und Pamela Weatherill (2012): Qualitative social research: a risky business when it comes to collecting 'sensitive' data. *Qualitative Research* 13(1), 19–35.
- Bell, Kirsten (2014): Resisting commensurability: Against informed consent as an anthropological virtue. *American Anthropologist* 116(3), 511–522.
- Breuer, Franz; Katja Mruck und Wolff-Michael Roth (2002): Subjektivität und Reflexivität: Eine Einleitung. *Forum Qualitative Sozialforschung* 3(3), <http://nbn-resolving.de/urn:nbn:de:0114-fqs020393> (Zugriff am 10.9.2018).
- Cannella, Gaile S. und Yvonna S. Lincoln (2011): Ethics, research regulations and critical social science. In: Denzin, Norman K. Denzin und Yvonne S. Lincoln (eds.): *The Sage Handbook of Qualitative Research*. Los Angeles, Sage, 81–89.
- Corti, Louise; Andreas Witzel und Libby Bishop (2005): Sekundäranalyse qualitativer Daten. *Forum Qualitative Sozialforschung* 6(1) <http://nbn-resolving.de/urn:nbn:de:0114-fqs0501495> (Zugriff am 10.09.2018).
- DGS und BDS (2017): Ethik Kodex der Deutschen Gesellschaft für Soziologie (DGS) und des Bundesverbandes deutscher Soziologinnen und Soziologen (BDS). <https://www.soziologie.de/de/die-dgs/ethik/ethik-kodex.html> (Zugriff am 10.09.2018).
- Hammersley, Martyn und Anna Traianou (2012): *Ethics in qualitative research. Controversies and contexts*. London, Sage.
- Hirschauer, Stefan (2014): Sinn im Archiv? Zum Verhältnis von Nutzen, Kosten und Risiken der Datenarchivierung. *Soziologie* 43(3), 300–312.
- Hopf, Christel (2010 [2000]). *Forschungsethik und qualitative Forschung*. In Uwe Flick; Ernst von Kardorff und Ines Steinke (Hrsg.): *Qualitative Forschung. Ein Handbuch*. 8. Auflage, Reinbek, Rowohlt, 589–600.
- Huschka, Denis; Hubert Knoblauch; Claudia Oellers und Heike Solga (Hrsg.) (2013): *Forschungsinfrastrukturen für die qualitative Sozialforschung*. Berlin: SCIVERO.
- Iphofen, Ron und Martin Tolich (eds.) (2018): *The Sage Handbook of qualitative research ethics*. London, Sage.
- Jagodzisinski, Wolfgang; Karl F. Schuman und Andreas Witzel (2005): Archivierung und Sekundärnutzung qualitativer Interviewdaten – eine Machbarkeitsstudie. Abschlussbericht zum DFG-Projekt. Bremen, Qualiservice. www.qualiservice.org/fileadmin/templates/qualiservice/DFG_Abschlussbericht.pdf (Zugriff am 10.09.2018).
- Knoblauch, Hubert und René Wilke (2018): Forschungsdateninfrastrukturen für audio-visuelle Daten der Qualitativen Sozialforschung – Bedarf und Anforderungen. RatSWD Working Paper 267/2018. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD), 47–58. <https://doi.org/10.17620/02671.35>.
- Kluge, Susann und Diane Opitz (1999): Die Archivierung qualitativer Interviewdaten. *Forschungsethik und Datenschutz als Barrieren für Sekundäranalysen?* *Soziologie* 4, 48–63.
- M'Bayo, Rosaline (2014): Keine Wahl haben und doch eine treffen: Ethische Herausforderungen in der HIV/Aids-Forschung mit afrikanischen Frauen. In: Hella von Unger; Petra Narimani und Rosaline M'Bayo (Hrsg.): *Forschungsethik in der qualitativen Forschung: Reflexivität, Perspektiven, Positionen*. Wiesbaden, VS Verlag für Sozialwissenschaften, 115–132.
- Mruck, Katja; Louise Corti; Susanne Kluge und Diane Opitz (2000): Text. Archiv. Re-Analyse. *Forum Qualitative Sozialforschung* 1(3), <http://nbn-resolving.de/urn:nbn:de:0114-fqs000311> (Zugriff am 10.09.2018).
- Narimani, Petra (2014): Zustimmung als Prozess: Informiertes Einverständnis in der Praxisforschung mit von Ausweisung bedrohten Drogenabhängigen. In: Hella von Unger; Petra Narimani und Rosaline M'Bayo (Hrsg.): *Forschungsethik in der qualitativen Forschung: Reflexivität, Perspektiven, Positionen*. Wiesbaden, VS Verlag für Sozialwissenschaften, 41–58.
- Roth, Michael und Hella von Unger (eds.) (2018): *Research ethics in qualitative research*. *Forum Qualitative Social Science Research*. Im Erscheinen.
- Saunders, Benjamin; Jenny Kitzinger und Celina Kitzinger (2015): Anonymising interview data: Challenges and compromise in practice. *Qualitative Research* 15(5), 616–632.

- Siri, Jasmin (2013): Sense and Sensibility #1: Der Grundsatz der Anonymisierung unter digitalen Bedingungen. Sozblog (Blog der Deutschen Gesellschaft für Soziologie). <http://soziologie.de/blog/2013/06/sense-and-sensibility-1-der-grundsatz-der-anonymitat-unter-digitalen-bedingungen/> (Zugriff am 10.09.2018).
- Strübing, Jörg (2018): Problem, Lösung oder Symptom? Zur Forderung nach Replizierbarkeit von Forschungsergebnissen. *Forschung & Lehre* 2/18. <https://www.forschung-und-lehre.de/zur-forderung-nach-replizierbarkeit-in-der-forschung-328/> (Zugriff am 10.09.2018).
- Tilley, Liz und Kate Woodthorpe (2011): Is it the end for anonymity as we know it? A critical examination of the ethical principle of anonymity in the context of 21st century demands on the qualitative researcher. *Qualitative Research* 11(2), 197–212.
- van den Hoonaard, Will C. (2011): *The seduction of ethics. Transforming the social sciences*. Toronto etc., University of Toronto Press.
- von Unger, Hella (2014): Forschungsethik in der qualitativen Forschung: Grundsätze, Debatten und offene Fragen. In: Hella von Unger; Petra Narimani und Rosaline M'Bayo (Hrsg.): *Forschungsethik in der qualitativen Forschung: Reflexivität, Perspektiven, Positionen*. Wiesbaden, VS Verlag für Sozialwissenschaften, 15–39.
- von Unger, Hella (2015): Qualitative Forschung / Forschungsethik / Streitpunkt: Digitale Archivierung. Sozblog (Blog der Deutschen Gesellschaft für Soziologie) 16.8.2015. <http://soziologie.de/blog/2015/08/qualitative-forschung-forschungsethik-streitpunkt-digitale-archivierung/> (Zugriff am 10.09.2018).
- von Unger, Hella (2018): Ethische Reflexivität in der Fluchtforschung. Erfahrungen aus einem soziologischen Lehrforschungsprojekt. *Forum Qualitative Sozialforschung* 19(3), Art. 6. <http://dx.doi.org/10.17169/fqs-19.3.3151> (Zugriff am 11.9.2018).
- von Unger, Hella; Hansjörg Dilger und Michael Schönhuth (2016): Ethikbegutachtung in der sozial- und kulturwissenschaftlichen Forschung? Ein Debattenbeitrag aus soziologischer und ethnologischer Sicht. *Forum Qualitative Sozialforschung* 17(3), <http://nbn-resolving.de/urn:nbn:de:0114-fqs1603203> (Zugriff am 10.09.2018).
- von Unger, Hella und Dagmar Simon (2016): Ethikkommissionen in den Sozialwissenschaften. Historische Entwicklungen und internationale Kontroversen. RatSWD Working Paper 253/2016. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://www.ratswd.de/publikation/working-paper/1847> (Zugriff am 10.09.2018).
- Warin, Jo (2011): Ethical Mindfulness and Reflexivity: Managing a Research relationship With Children and Young People in a 14-Year Qualitative Longitudinal Research (QLR) Study. *Qualitative Inquiry* 17(9), 805–814.

Prof. Dr. Hella von Unger ist Professorin für Soziologie mit dem Schwerpunkt qualitative Methoden der empirischen Sozialforschung an der LMU München. Sie forscht zu qualitativen Methodologien, partizipativer Forschung, Soziologie von Gesundheit und Krankheit, Ethnizität und Migration. Sie ist außerdem Vorsitzende der Ethikkommission der sozialwissenschaftlichen Fakultät der LMU.

Zentrale Ergebnisse und Empfehlungen aus dem Workshop für Archivierung und Sekundäranalyse qualitativer Forschungsdaten

Betina Hollstein und Jörg Strübing¹

Die nachfolgenden Empfehlungen und Forderungen sind Ergebnisse der im Plenum und in verschiedenen Arbeitsgruppen des interdisziplinären Workshops zu „Archivierung und Zugang zu qualitativen Daten“ geführten Diskussionen, an denen sowohl qualitativ-interpretativ Forschende unterschiedlicher methodischer Ausrichtung als auch Expertinnen und Experten von Archiven für qualitative Forschungsdaten teilgenommen haben.

Folgende Kontextüberlegungen bilden die Basis der Diskussionen auf dem Workshop:

- Öffentlich finanzierte Forschung sollte nachhaltig sein, die Sekundärnutzung der entstandenen Daten zusätzlich zur originären Forschung verfolgt dieses Ziel.
- Zur Sicherung guter wissenschaftlicher Praxis und der Nachvollziehbarkeit von Forschungsleistungen fordern wissenschaftliche Zeitschriften bei der Publikation von Forschungsarbeiten zunehmend auch die Zugänglichmachung des Datenmaterials bzw. entsprechender Kodierungen.
- Der RatSWD sieht seine Zuständigkeit auch in der Archivierung und dem Zugang zu qualitativen Daten und dem damit anzuschließenden kulturellen Wandel, was ebenfalls durch den Wissenschaftsrat und die RatSWD-Aktivitäten bestätigt wird.
- Die Diskussion in den Sozialwissenschaften hat sich bislang besonders auf qualitative Interviews bezogen, sollte jedoch zukünftig stärker auch andere qualitativ-empirische Materialien (ethnographische Daten, Bild- und Videomaterial) einbeziehen und die datenschutzrechtlichen und forschungsethischen Probleme sowie mögliche Nachnutzungspotentiale adressieren.

Die nachstehenden Empfehlungen und Forderungen geben, ohne Vollständigkeit zu beanspruchen, den Tenor der Diskussion wieder.

Zentrale Anforderungen sind hervorgehoben.

¹ Die vorliegende schriftliche Zusammenfassung wurde von Betina Hollstein und Jörg Strübing erstellt. Die Empfehlungen und Forderungen aus dem Workshop spiegeln nicht unbedingt die Meinung des RatSWD wider.

1. Voraussetzungen und Einschränkungen

- **Die Beurteilung des Archivierungsbedarfs und der Möglichkeiten für wissenschaftlich-methodisch anerkannte sekundäranalytische Zugriffe muss von den qualitativ Forschenden und den entsprechenden Fachdisziplinen ausgehen.**
- Bei Zielkonflikten gilt ein Primat der Primärforschung. Wo Archivierungsabsichten und potentielle Sekundäranalysen den Erfolg der Primärforschung (z. B. durch Einschränkungen bei Feldzugang und Vertrauensbildung) gefährden, hat die Primärforschung Vorrang.
- Qualitative Sozialforschung operiert mit einem anderen Datenbegriff als andere Felder empirischer Forschung. Das Material liegt in vielfältiger Ausprägung vor und ist – im Sinne von Daten – unterschiedlich verfügbar. Dies muss bei Konzepten für die Archivierung und Verfügbarmachung berücksichtigt werden.
- Die Archivierung von qualitativen Daten ist insofern mit Qualitätsverlusten verbunden, als Kontextinformationen verloren gehen bzw. nur eingeschränkt für die Nachnutzung zur Verfügung stehen. Dies betrifft aber je nach Materialsorte auch wichtige Informationen aus den Daten selbst, die je nach (Sekundär-)Fragestellung Einfluss auf die Deutung des verbliebenen Datenmaterials haben.
- Ursächlich hierfür sind insbesondere die aufgrund von datenschutzrechtlichen Bestimmungen und forschungsethischen Erwägungen notwendigen Maßnahmen zur Anonymisierung bzw. Pseudonymisierung.
- Da nicht alle zukünftigen Forschungsfragen, die in Sekundäranalysen bearbeitet werden könnten, antizipierbar sind, lassen sie sich zum Zeitpunkt der Archivierung nur bedingt in die Archivierungsstrategien (Aufbereitung, Metadaten, Breite und Tiefe der Materialakquise) einbeziehen.
- Mit der Archivierung des Datenmaterials werden zugleich die Vorgehens- und Verhaltensweisen von Forschenden archiviert. Dies kann als Element der Qualitätssicherung gesehen werden, zugleich aber werden Primärforschende bei Nachnutzungen auch besonders exponiert. Darüber hinaus stellen sich bei allen Verfahren, bei denen Erhebung und Auswertung kaum voneinander zu trennen sind (insbesondere bei ethnographischen Studien), nicht nur grundsätzliche Fragen der Nutzung fremden geistigen Eigentums, sondern auch zu den Grenzen der Explizierbarkeit von kontextuellem Erfahrungswissen.
- Ansätze, die bei quantitativen Daten erfolgreich angewandt werden, können nicht ohne weiteres auf qualitative Daten übertragen werden; vielmehr müssen Archivierungslösungen und Sekundärnutzungsstrategien der Spezifik qualitativ-interpretativer Forschung Rechnung tragen.

2. Bedarfe und Anonymisierung

- Pauschale, standardisierte Archivierungskonzepte sind nicht zielführend und werden den speziellen Charakteristika qualitativer Daten nicht gerecht, vielmehr sind sowohl medien- als auch inhaltspezifische Lösungen erforderlich.
- Die konkrete Ausgestaltung der Anonymisierungs- bzw. Pseudonymisierungsverfahren ist im Einzelfall zu prüfen. Flexible Ansätze zur Anonymisierung sowie transparente Zugangsbeschränkungen und sorgfältig formulierte Datennutzungsverträge können hier Ansätze sein. Auch die Verschlüsselung von Rohdaten für einen definierten Zeitraum (sog. Embargoperiode) ist denkbar – immer vorausgesetzt, dass es möglich ist, im Primärforschungsprozess einen so

weitreichenden informed consent zu erzielen, dass Archivierung ermöglicht wird, ohne Feldzugänge und zukünftige Forschungsmöglichkeiten zu gefährden.

- Zu beachten sei auch, dass nicht alle erhobenen qualitativen Daten notwendigerweise archivierungswürdig sind. Geeignete Kriterien in Abhängigkeit von Datentypus und Aufbereitungslevel sind noch zu entwickeln. Grundsätzlich besteht jedoch Interesse sowohl am Ausgangsmaterial als auch an dessen analytischen Bearbeitungsstufen.
- Die Konsequenzen einer möglichen De-Anonymisierung von personenbezogenen Informationen (z. B. zur interviewten Person und zu von ihr genannten Personen) sind nicht immer klar abschätzbar (z. B. im Hinblick auf Forschung mit Minoritäten), müssen aber sehr ernst genommen werden. Hier gilt ein Primat des Schutzes der Teilnehmenden.
- Forschungsethische Einschätzungen zur Unbedenklichkeit bzw. getroffene Risikoabwägungen können sich im Zeitablauf ändern (z. B. bei sich verändernden politischen Kontexten). Gleichzeitig kann eine De-Anonymisierung direkten Nutzen bringen, z. B. wenn dokumentierte langjährige Eigentums- oder Nutzungsrechte etwa an Ressourcen (wie Land, Zugang zu Frischwasser, Weideflächen, Fischbestände etc.) im Konfliktfall mit Hilfe der durch Forschende erhobenen qualitativen Daten nachgewiesen und ggf. juristisch durchgesetzt werden könnten. Über die Freigabe archivierter Daten für Nachnutzungen muss vor diesem Hintergrund immer wieder neu befunden werden.

3. Lösungen

- **Für Forschungsdatenzentren (FDZ) ist die Zusammenarbeit mit einem kompetenten, erfahrenen und serviceorientierten IT-Partner sowie der interdisziplinäre Austausch (z. B. mit der Medizin) und der internationale Austausch (z. B. mit dem UK Data Archive) wichtig, um von bereits bestehenden Erkenntnissen im Forschungsdatenmanagement zu profitieren (z. B. im Hinblick auf Metadatenschemata und technische Standards).**
- **Ein Netz aus disziplinär aufgestellten, datentypspezifischen und themenorientierten Forschungsdatenzentren entwickelt über die gut durchdachte (z. B. metadatenorientierte) Verknüpfung und ein einheitliches Zugangsportal einen Mehrwert.**
- **Zur Archivierung und Nachnutzung von Daten müssen entsprechend formulierte informierte Einwilligungserklärungen vorliegen, bei denen die Anforderungen der EU-Datenschutz-Grundverordnung berücksichtigt werden müssen. Dabei ist darauf zu achten, dass der Rapport mit dem Feld nicht gefährdet, die Möglichkeit zur Primärforschung also nicht beeinträchtigt wird.**
- **Zur Erhaltung des Nachnutzungspotentials sind die Kontexte der Studien und der erhobenen Daten ausführlich zu dokumentieren. Forschungsdatenzentren sollten diesbezügliche Beratungsleistungen anbieten.**
- **Entscheidungen über Art und Intensität der (in der Regel sehr zeitaufwändigen und personalintensiven) Anonymisierung/Pseudonymisierung des Datenmaterials müssen sich strikt am Ziel des Schutzes der Befragten/Forschenden sowie der mittelbar im Datenmaterial aufgenommenen Personen orientieren. Um die Nachnutzbarkeit der Daten zu erhöhen, sollten auch aufwändigere flexible Ansätze zur Pseudonymisierung in Erwägung gezogen werden.**

- Eine Entscheidung für die Datenarchivierung zu Beginn eines Forschungsprojektes kann einerseits die Qualität der archivierten Daten erhöhen, da dann schon bei der Datengewinnung ein Bewusstsein für eine Archivierung und mögliche Nachnutzung besteht, andererseits sind nichtintendierte Faktoren (wie z. B. Selbstzensur von Forschenden) nicht auszuschließen, um die eigene Angreifbarkeit z. B. in bestimmten Punkten des Forschungsprozesses zu vermeiden.
- In vielen Fällen sind Entscheidungen über die Archivierungsmöglichkeiten und Sekundärnutzungsperspektiven in qualitativ interpretativen Projekten erst im Verlauf oder gar ex post zu treffen, weil die Dynamik der Feldbeziehungen und der Entwicklung der Fragestellungen und Datengewinnungsformen nicht zuverlässig antizipierbar ist.
- Es ist von Fall zu Fall konkret zu entscheiden, ob und ggf. welche archivierten Daten welchen Personenkreisen in welcher Form zur Verfügung gestellt werden können. Archivierung kann also nicht automatisch auch Zugänglichkeit für Sekundärnutzungen bedeuten.
- In den Datenzentren könnte im Einvernehmen mit den Primärforschenden die Überprüfung der Datenqualität und des Archivierungsbedarfs verortet sein. Dies setzt einschlägige fachliche und methodische Expertise voraus.
- Um den Schutz der sensiblen personenbezogenen Daten und den Vertrauensschutz gegenüber den an der Primärstudie Teilnehmenden zu gewährleisten, sind strenge Prüfungen der Nutzenden unabdingbar.
- Es ist über verschiedene Formen von Autorenschaften (Stichwort: Einbezug der Interviewten) nachzudenken, die dann ggf. nach Analyseebene gestaffelt sind (weitreichend aufbereitete Daten dürften eher nur noch die Autorenschaft von Forschenden haben).

4. Kultur und Community

- **Die Anforderungen an eine Forschungsinfrastruktur müssen von den Bedarfen der Fach-Communities (kanalisiert durch die Fachgesellschaften) geleitet sein und auch von diesen entwickelt werden. Politik oder forschungsfördernde Institutionen sollten Standards nur im Einklang mit den Bedarfen der Fach-Communities setzen, um letztlich auch die tatsächliche Nachnutzung zu fördern.**
- **Ob und inwiefern die Sekundärnutzung qualitativer Forschungsdaten einen Mehrwert für die Sozialforschung hat, entscheidet sich in der Praxis. Eine Kultur der Sekundärnutzung muss aus der Entwicklung, Erprobung und Etablierung entsprechender Praktiken heraus entstehen. Hilfreich sind dabei Leuchtturmprojekte, insbesondere exemplarische Sekundäranalysen, die Nutzungsszenarien in der Breite verdeutlichen.**
- **Eine Intensivierung der Vernetzung von Personen und Institutionen im Themengebiet sowie z. B. auch ein Datenportal zur Verbesserung der Auffindbarkeit von qualitativen Daten ist wünschenswert. Hier bietet sich eine enge Kooperation mit dem RatSWD an.**
- **Die einzelnen Communities sollten sich auf grobe Verschlagwortungsregeln einigen und dann in einen interdisziplinären Austausch gehen, um interdisziplinär verständliche Metadaten generieren zu können.**
- **Auch in der Lehre muss das Thema der Archivierung stärker berücksichtigt werden, was den Einbezug in die regulären Curricula auf allen Qualifikationsebenen des Wissenschaftsbetriebs erfordert, wie auch die Finanzierung entsprechender Professuren im Bereich der qualitativen Methoden notwendig macht.**

- **Die verstärkte Forschung zu Archivierung und Sekundärnutzung sollte auch Handreichungen und best practice-Beschreibungen produzieren.**
- Allerdings ist auch ein kultureller Wandel innerhalb der Forschungsgemeinde notwendig, sodass die im Forschungsdatenmanagement erbrachten Leistungen gewürdigt werden (etwa durch entsprechende Zitierung der bereitgestellten Forschungsdaten und Studienreports in darauf basierenden Publikationen).
- Gleichzeitig eröffnet dieser Aspekt ein Spannungsverhältnis in der Zuordnung der Autorenschaft zwischen Forschenden und Beforschten. So dürfte die alleinige Autorenschaft ggf. erst für Datenmaterial gelten, das sehr weitgehend aufbereitet und analysiert ist.
- Der RatSWD sollte sich intensiv mit dem Themenkomplex der Archivierung von und des Zugangs zu qualitativen Daten befassen und die Interessen der qualitativen Sozialforschung vertreten. Es wird auch kritisch die Frage nach einer angemessenen Repräsentation der qualitativen Forschung im RatSWD und zwar auch auf Seiten der sog. Datenproduzierenden aufgeworfen.

5. Politik und Förderung

- **Eine zentrale Forderung der Teilnehmenden betrifft den erhöhten Ressourcenbedarf an Geld, Personal und Zeit in Forschungsprojekten, der mit einer hochwertigen Archivierung qualitativer Daten verbunden ist. Dieser muss von den Forschungsfördernden eingeplant werden – und zwar zusätzlich, also nicht zu Lasten der Fördersummen für die eigentliche Forschung.**
- **Infrastrukturen wie Forschungsdatenzentren (FDZ) müssen langfristig finanziert sein, um eine professionelle Archivierung und den Datenservice dauerhaft sicherzustellen.**
- **Die Vorteile der Sekundärnutzung von Forschungsdaten lassen sich über Leuchtturmprojekte gut kommunizieren. Hier sind Drittmittelgebende notwendig, die gezielt Sekundäranalysen fördern.**
- **Die Finanzierung von Professuren im Bereich der qualitativen Methoden sichert den hochwertigen Einbau in die akademische Lehre und damit den Kulturwandel ab. Darin ist eine Daueraufgabe der Universitäten/Fakultäten zu sehen.**

Impressum

Herausgeberin:

Prof. Regina T. Riphahn Ph.D.

Kontakt:

Rat für Sozial- und Wirtschaftsdaten (RatSWD)

Chausseestraße 111

10115 Berlin

office@ratswd.de

www.ratswd.de

Redaktion:

Mathias Bug, Dr. Tim Deeken, Lea Salathé

Berlin, Oktober 2018

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01UW1402 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt, sofern nicht anders ausgewiesen, beim RatSWD.

doi: 10.17620/02671.35

Zitationsvorschlag für Einzelbeiträge:

<Nachname, Vorname> (2018): <Titel des jwlg. Beitrags>. RatSWD Working Paper 267/2018. Berlin, Rat für Sozial und Wirtschaftsdaten (RatSWD), <Seitenzahl>. <https://doi.org/10.17620/02671.35>

Zitationsvorschlag für den Gesamtband:

RatSWD [Rat für Sozial- und Wirtschaftsdaten] (Hrsg.) (2018): Archivierung und Zugang zu qualitativen Daten. RatSWD Working Paper 267/2018. Berlin, Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://doi.org/10.17620/02671.35>

■ **Der Rat für Sozial- und Wirtschaftsdaten (RatSWD)** berät seit 2004 die Bundesregierung und die Regierungen der Länder in Fragen der Forschungsdateninfrastruktur für die empirischen Sozial-, Verhaltens- und Wirtschaftswissenschaften. Im RatSWD arbeiten acht durch Wahl legitimierte Vertreterinnen und Vertreter der sozial-, verhaltens- und wirtschaftswissenschaftlichen Fachdisziplinen mit acht Vertreterinnen und Vertretern der wichtigsten Datenproduzenten zusammen.

Er versteht sich als institutionalisiertes Forum des Dialoges zwischen Wissenschaft und Datenproduzenten und erarbeitet Empfehlungen und Stellungnahmen. Der RatSWD engagiert sich für eine Infrastruktur, die der Wissenschaft einen breiten, flexiblen und sicheren Datenzugang ermöglicht. Solche Daten werden von staatlichen, wissenschaftsgetragenen und privatwirtschaftlichen Akteuren bereitgestellt. Der RatSWD hat 31 Forschungsdatenzentren akkreditiert, deren Kooperationen er fördert.

RatSWD Working Paper Series

www.ratswd.de