

Sinn im Archiv?

Zum Verhältnis von Nutzen, Kosten und Risiken der Datenarchivierung

Stefan Hirschauer

Die qualitative Sozialforschung wird seit ein paar Jahren mit einem Modell der Archivierung und Sekundärnutzung von Daten konfrontiert, das sich in der quantitativen Sozialforschung bewährt hat (Heaton 2004; Mauthner, Parry 2013).¹ Jetzt soll das Modell irgendwie auf qualitative Daten übertragen werden. Der Wissenschaftsrat (2012) hat dazu einige Empfehlungen abgegeben. Sie beinhalten 1. die Erhebung und Publikation von Forschungsdaten als eigenständige Forschungsleistung anzuerkennen und andere Publikationen mit den Daten zu verlinken; 2. die erforderlichen personellen und finanziellen Kapazitäten bereits bei der Konzeption von Projekten und der Beantragung von Fördermitteln einzuplanen; 3. eine Empfehlung an die Deutsche Forschungsgemeinschaft, ein Programm auszuschreiben, das Modellprojekte für Forschungen fördert, die quantitative und qualitative Daten kombinieren. Denn Gewinne der Archivierung verspricht sich der Wissenschaftsrat vor allem in diesem Bereich. Auch Hubert Knoblauch und Heike Solga (2011) haben im Rahmen des Rates für sozialwissenschaftliche Daten vertreten, dass ein Bedarf an qualitativen Datenban-

¹ Dieser Artikel beruht auf einem Vortrag auf der Tagung der Sektion Biographieforschung »Was geschieht mit unseren Daten? Datenarchivierung und Forschungsethik in (inter)nationaler Perspektive« der am 10. und 11. April 2014 an der Goethe-Universität Frankfurt am Main stattfand. (Siehe dazu den Bericht auf Seite 331 in diesem Heft, *Anm. der Red.*) Kommentierungen erhielt der Beitrag durch Ruth Ayaß, Christian Meyer, Jörg Strübing, Herbert Kalthoff, Georg Breidenstein und Boris Nieswand.

ken wesentlich aus einer Integration beider Datensorten folge, die zu neuen Erkenntnissen führen könne. Weitere Gründe, die für die Archivierung sozialwissenschaftlicher Daten angeführt werden, sind die Sicherung der Überprüfbarkeit und Transparenz von Forschungsergebnissen, die Nutzbarmachung von Datensätzen für Lehrzwecke und die Eröffnung der Chance, Sekundäranalysen in der Primärforschung offen gebliebener Themen durchführen zu können.

Die den Empfehlungen des Wissenschaftsrates folgende Diskussion hat sich recht schnell auf organisatorisch-praktische Fragen sowie auf Fragen der Wahrung von forschungsethischen Grundsätzen gerichtet. *Vor* diesen beiden Fragen liegt aber eine Erörterung der Prämissen des Wissenschaftsrates. Welchen wissenschaftlichen Sinn hat die Archivierung von Daten überhaupt und welchen wissenschaftlichen Unsinn könnte sie fördern?

Sinn und Unsinn der Datenarchivierung

Unbedingt sinnvoll erscheint eine massenhafte Archivierung wissenschaftlicher Daten, wenn man sie als wertvolle Kulturgüter betrachtet. Dies scheint in einigen Fächern tatsächlich der Fall: Die gesamte Ethnologie ist nicht nur in ihren Sammlungen, sondern auch in ihren Ethnografien von einem empirischen Rettungsmotiv durchdrungen. So wie der Klimawandel Inseln überspült, so verschluckt die Globalisierung das kulturhistorische Erbe der Menschheit, heißt es. Museale Sammlungen von Artefakten und Videoaufzeichnungen von Kulturtechniken und Sprachen sollen das Schlimmste verhüten. Ein ähnliches Rettungsmotiv kann die zeithistorische Forschung beanspruchen: Sind Zeitzeugen erst einmal verstorben, stehen sie für *Oral History* nicht mehr zur Verfügung. Mit ihrem Körper entschwindet auch ein Teil des kulturellen Gedächtnisses. Das gibt der Bewahrung ihrer Erinnerungsspuren in Datensätzen einen unmittelbaren Sinn. Denselben archivarischen Sinn kann auch ein Teil der Biografieforschung beanspruchen (Rosenthal 2013). Sie hat hier eine Schnittmenge mit der Geschichtswissenschaft.

Diesem Sinn steht allerdings das Risiko eines archivarischen Unsinnns gegenüber. Die bloße Anhäufung von Daten ist ebenso sinnlos wie eine bloße Ansammlung ethnologischer Artefakte in staubigen Regalen und dunklen Speicherkammern. Auf die Ausstellung kommt es an, also auf die

hochselektive, öffentliche Verfügbarmachung zu bestimmten Gelegenheiten. Dieser Ausstellung entsprechen in der Wissenschaftskommunikation die Aufsätze und Bücher, in denen wir unseren Daten eine analytisch geordnete und kommunikativ verstehbare Form geben. Erst in dieser Form werden tatsächliche kulturelle Erinnerungsleistungen hergestellt – auch solche über untergegangene Inseln und Epochen. Will man über diese Publikationen hinaus Erinnerungen ermöglichen, so ist die entscheidende Frage, wie hoch man das Potenzial von Sammlungen für eventuelle spätere Verfügbarmachungen einschätzt.

Ist es zum Beispiel sinnvoll, wie an der Universität Mainz kürzlich angestrebt, die Fernsehprogramme der wichtigsten zwölf Sender rund um die Uhr aufzuzeichnen, damit der Nachwelt der Zugriff auf diesen Teil unserer kulturellen Produktivität nicht verloren geht? Natürlich sind Fragestellungen denkbar, für die so eine Zugriffsmöglichkeit in 20 oder 50 Jahren hoch interessant wären. Aber welchen Speicheraufwand wollen wir in dieser Zeit dafür betreiben? 12 Sender, 24 Stunden, 365 Tage, 20 Jahre ... Wer meint, dass die Erinnerungsfunktion unserer Publikationen, also unserer, mittels Analyse und kommunikativer Darstellung aufbereiteten Daten, nicht ausreichend ist, der muss sich die Frage stellen, wie sich Aufwand und Potenzial der Speicherung von unanalysierten Rohdaten zueinander verhalten.

Der Fall des Fernseharchivs verweist auf einen zweiten archivarischen Unsinn. Soziologinnen sind ja keine Ethnologen und Historiker. Die Biografieforschung ist hier mit der zeithistorischen Bedeutung mancher narrativer Interviews eher in einer Sonderrolle. Der strategische Vorteil der meisten Sozialwissenschaften gegenüber der (klassischen) Ethnologie und Geschichtswissenschaft liegt darin, dass sie inmitten ihrer zeitgenössischen, sprudelnden Datenquellen sitzen. Unser Aufwand, *frische* Daten für innovative Fragestellungen zu generieren – ob durch Interviews, teilnehmende Beobachtung, Audio- und Videotakes oder das Anlegen von Diskurskorpora – ist vergleichsweise gering. Das macht die Frage umso dringlicher, wie sich der Aufwand der Archivierung eigentlich zu ihrem potenziellen Nutzen verhält.

Es gibt aber neben dem Rettungsmotiv noch eine weitere Sinnquelle der Datenarchivierung, die aus der sozialen Organisation der Forschung stammt. Sinnvoll ist die Datenarchivierung in dem Maße, wie die Forschung insgesamt *arbeitsteilig* organisiert ist. Prototyp in den Sozialwissenschaften ist die international vergleichende Umfrageforschung. Wer repräsentative Aussagen über die Verhältnisse in zehn europäischen Ländern

machen will, braucht entweder einen personenstarken Forschungsverbund oder Zugriffsmöglichkeiten auf für bestimmte Fragen vergleichbare Datensätze aus den Nachbarländern. Die Vergleichsmöglichkeit mit ähnlichen Studien ist in der standardisierten Forschung aber noch aus einem anderen Grund ein hochrangiges Gut: nicht erst um die Reichweite von Datensätzen zu erhöhen, sondern schon um deren grundlegende Sinnschwäche zu beheben, braucht es den Vergleich. Isolierte numerische Daten über eine Population – sagen wir eine hochsignifikante Verteilung von 62 zu 38% der Befragten – besagen nichts, wenn man sie nicht in Relation zu anderen Populationen oder zu früheren Zuständen als »hoch« oder »niedrig« qualifizieren kann. Eines der Hauptmotive der Datenarchivierung entstammt dieser Sinnschwäche quantitativer Daten.

Die Ausgangslage qualitativer Daten ist hier genau entgegengesetzt. Anders als das Kreuz auf dem Fragebogen, das sich in der numerischen Aggregation zu einem bloßen Zählwert verflüchtigt, sind Anekdoten im Interview, Streitgespräche auf einer Tonaufzeichnung oder ein protokolliertes Dominanzverhalten von vornherein semantisch dicht, sie leiden nicht unter Sinnschwäche, sondern unter Sinnfülle. Sie haben viel zu viele Bedeutungen, als dass sie für die Forschung unkontextualisiert und unanalysiert – als Rohdatum – handhabbar wären. Während die Bedeutung einer Zahl als »viel« oder »wenig« nur durch einen Vergleich als ein »mehr« oder »weniger« hergestellt werden kann, kann die Bedeutung einer sprachlichen Äußerung nur durch ihre Kontextierung hergestellt werden, also durch eine deutende Spezifikation der lokalen Bedingungen, unter denen sie getan wurde. Die wesentliche Leistung qualitativer Forschung liegt nicht in der Kumulation von Daten, sondern in der Reduktion der Datenkomplexität. Sie folgt hier einer anderen Logik als die standardisierte Forschung. Sie schließt nicht einer »Datengewinnung« deren Analyse an, sondern stellt in einer theorieorientierten Analyse den Wert bestimmter Informantenäußerungen *als Datum* erst her.

Was ist ein »Datum«?

Aber welchen Begriff des Datums setzen die Bemühungen zur Datenarchivierung eigentlich voraus? Die unkritische Übernahme der Datenarchivierung für die qualitative Forschung importiert auch einen Datenbegriff, des-

sen Passung noch gar nicht diskutiert worden ist. Es fehlt an einer differenzierten Diskussion der Eigenschaften, die unterschiedliche Datentypen haben. Das, was wir »Daten« nennen, sind Zeichen von sehr unterschiedlicher Qualität: Es sind Transkriptionen mündlicher Äußerungen, die bestimmte Informanten in hochspezifischen sozialen Situationen auf bestimmte Fragereize hin machen. Es sind hochselektive, minutenlange Videomitschnitte von Verhaltensweisen, die Personen in bestimmten Situationen in Anwesenheit einer Kamera an den Tag legen. Es sind schriftliche Äußerungen in einer Diskursöffentlichkeit, die für einen bestimmten Zeitraum, bestimmte Medien und eine spezifische Frage zu einem Korpus kompiliert wurden, usw.

Der unhandlichste Fall für den Datenbegriff der Datenarchivierung ist wohl die Ethnografie (Breidenstein et al. 2013). Es ist schwer zu sagen, was dort überhaupt als »Datum« gelten soll. Die Vorstellung eines kleinen, verlässlichen Sinnquantums, das sich aus seinem Kontext herauslösen, archivieren und unbeschädigt in den wissenschaftlichen Diskurs übertragen ließe, ist hier nicht recht anwendbar. Eine gute, zitierfähige Beschreibung, wie sie etwa Clifford Geertz (1973) vom balinesischen Hahnenkampf angefertigt hat, ist viel *mehr* als eine Sammlung von Daten, nämlich eine hochgradig interpretierte Darstellung eines Ausschnittes kulturellen Lebens. Die Beobachtungsprotokolle, die dieser Beschreibung zugrunde liegen, also eine Sammlung von einzelnen Szenen, sind ebenfalls *mehr*, als der Datenbegriff der standardisierten oder auch anderer qualitativer Verfahren zuließe: Sie sind zwar weniger idealtypisch als die Beschreibung, aber immer noch mehrfach editierte, interpretationsgebundene Darstellungen. Den Protokollen wiederum liegen *Fieldnotes* zugrunde. Diese muss man nun freilich in Verdacht haben, viel *weniger* als ein Datum zu sein. Die Ideosynkrasien ihrer situativen Produktion und ihre handschriftliche Form stehen ihrer produktiven Nutzung in der eigenen Forschung überhaupt nicht im Wege, aber sie machen sie für eine dekontextualisierte Nutzung durch andere unbrauchbar. Die Feldnotizen bekommen ihre Bedeutung nämlich nur in Interaktion mit dem körperlichen Gedächtnisspeicher – oder soll man sagen dem physiologischen »Datenarchiv« –, das ein Ethnograf in eine beobachtete Situation mitgebracht hat. Es sind die langfristig gemachten, kumulierten Erfahrungen in einem Feld, die den zusätzlich erzeugten Schriftspuren überhaupt erst ihren Sinn geben. Diese Erfahrungen sind gar keine Daten, aber sie sind wegen ihrer langfristigen kontinuierlichen Entstehungsgeschichte meist viel wichtiger als die hochselektiven technischen Aufzeich-

nungen einzelner Situationen, da sie Orientierungswissen und Deutungskompetenz in Bezug auf die technischen Aufzeichnungen verleihen. Was soll man also archivieren von den Forschungen einer Ethnografin? Ihre 20 technischen Mitschnitte, ihre 30 Interviewtranskripte, die Notizen der 50 informellen Gespräche, die sie führte, die 120 Notizzettel, auf denen ihre Handschrift soziale Praktiken festhielt, oder nicht besser doch – geeignete Scantechniken vorausgesetzt – ihr Gehirn?

Der Datenbegriff im Wort Datenarchivierung setzt einen Forschungsprozess voraus, den es in den meisten Fällen qualitativer Sozialforschung so gar nicht gibt. Dieser Prozess trennt eine bloße Erhebung und Aufbereitung von Daten von deren Analyse und Interpretation. Solche sinnarmen Daten kann man tatsächlich leicht archivieren, das heißt, dekontextualisieren. Das Gros der qualitativen Sozialforschung stellt dagegen ihre Daten erst *durch* ihre Analysen her.

Soviel zur Frage des archivarischen Sinns und Unsinn. Der wissenschaftliche Wert von Rohdaten für spätere Studien oder gar andere Disziplinen, wie sie sich der Wissenschaftsrat vorstellt, ist eine ungeprüfte Prämisse. Sie wäre erst einmal für einzelne Fächer, für spezifische Verfahren und sogar für einzelne Studien differenziert zu erörtern. Zur Zeit wird dieser wissenschaftliche Wert von den Fürsprechern der Datenarchivierung ohne jede Diskussion vorausgesetzt. Sie extrapolieren damit einfach nur einen bestimmten Fall empirischer Sozialforschung.

Nutzen gering, Schaden beträchtlich. Eine Güterabwägung

Es ist schwierig, zu den Empfehlungen des Wissenschaftsrates Stellung zu nehmen, ohne dass diese Diskussion über den Sinn und Unsinn der Datenarchivierung verfahrensspezifisch geführt worden ist, das heißt eine Diskussion über den potenziellen Nutzen, die potenziellen Schäden und den zeitlichen und finanziellen Aufwand, den eine Datenarchivierung für einen spezifischen Datentyp mit sich bringen würde. Vor dem Hintergrund der Diversität von Datentypen, die auf den Tagungen der Sektion Methoden der qualitativen Sozialforschung in den letzten zehn Jahren präsentiert wurden, lässt sich vorläufig – und noch ohne jede Berücksichtigung der Empfindlichkeit einzelner Forschungsfelder – folgende grobe Abschätzung

der potenziellen Kosten, Nutzen und Schäden durch Archivierung qualitativer Daten machen:

Datentypen	Potenzieller Nutzen	Potenzieller Schaden	Kosten/Aufwand
Biografische Narrationen	moderat	sehr groß	sehr groß
Leitfadeninterviews	themenabhängig	gering	moderat
Ethnografische Protokolle	sehr gering	sehr groß	sehr groß
Konversationsdaten	groß	groß	moderat
Videodaten	gering	sehr groß	sehr groß
Amtliche Dokumente	gering	sehr groß	sehr groß
Öffentliche Diskurse	eher gering	gering	gering

Grundsätzlich lautet die Einschätzung: Der Nutzen einer Archivierung qualitativer Daten ist weit geringer als der quantitativer Daten. Dafür ist der Schaden, den eine obligatorische Archivierung anrichten würde, enorm. Der Nutzen ist aus vier Gründen geringer:

1. Die Datenarchivierung dient in der standardisierten Forschung auch der Replizierbarkeit von Studien. Diese Anforderung folgt aus dem Gütekriterium der Reliabilität, also der personenunabhängigen Verlässlichkeit eines Datums, was eine grundsätzliche Vergleichbarkeit von Forschungssituationen impliziert. Eben diese ist in qualitativen Studien von vornherein ausgeschlossen und wird auch gar nicht angestrebt. Selbst die Wiederholung eines Interviews mit einem Informanten durch dieselbe Interviewerin im selben Wohnzimmer wäre ein anderes Interview. Selbst die Beobachtung derselben Schulklasse im selben Fach durch denselben Ethnografen wäre eine Beobachtung anderer Szenen. Es gibt einen unvermeidbaren Unikat-Charakter qualitativer Studien. Ihre Vertrauenswürdigkeit beruht auf ganz anderen Aspekten der Forschung als auf der Austauschbarkeit von Personen.
2. Überprüfbarkeit und Transparenz ließen sich durch die Datenarchivierung ebenfalls kaum steigern. Die einfache Zugänglichkeit von Datensätzen ist nur dann mit Transparenzgewinnen verbunden, wenn sich die kollegiale Kontrolle auf Einsicht in Indikatorenbildung und Operationalisierungen und auf ein »Nachrechnen« beschränkt. Die Qualitätskontrolle in der qualitativen Sozialforschung ist dagegen eine ausgedehnte

Prozesskontrolle: Sie beginnt mit der Deutungskontrolle durch die Informanten (am ausgeprägtesten in der Ethnografie), setzt sich fort mit der Konkurrenz von Lesarten in Datenanalyseteams, mit dem Zeigen von Daten in Publikationen, deren Deutungen Leser überzeugen oder nicht überzeugen, und endet natürlich (wie in der standardisierten Forschung) bei der Selektion durch Fachzeitschriften, die drucken oder ablehnen, und durch Leser, die zitieren oder vernachlässigen. Ein zusätzliches Kontrollangebot durch Einsicht in die Datenmassen, die einer Veröffentlichung zugrunde liegen, würde fast immer an knapper Aufmerksamkeit scheitern. Denn:

3. Studien mit geringer Fallzahl erzeugen eine um Vieles *größere* Datenmenge. Textdaten in der Biografieforschung füllen schnell Tausende von Seiten, eine exemplarische Videografie kann schnell 50 Terabites umfassen. Eine Orientierung in einem solchen Korpus kann von Dritten kaum geleistet werden, ohne dass man sie zu Kooperationspartnern des Projektes macht. Denn die primäre Leistung qualitativer Forschung liegt, wie gesagt, nicht in der Kumulation von Fällen, sondern in der Reduktion ihrer Komplexität.
4. Schließlich ist die Wiederverwertbarkeit qualitativer Daten auch dadurch stark eingeschränkt, dass sie durch ihre Dekontextualisierung meist ihren Sinn verlieren. Wenn man nicht weiß, *wer* es ist, der diese archivierte Äußerung *wem* gegenüber in *welcher* Situation im Kontext *welcher* Beziehungsgeschichte und institutionellen Einbettung *wann* getan hat, weiß man schlicht nicht, was diese Äußerung bedeutet. Wenn man aber umgekehrt jedes einzelne Datum für ein Archiv kontextualisiert, damit es für Dritte verstehbar wird, müsste die Forschung in manchen Fällen personell gedoppelt werden. Der Grund liegt, wie gesagt, im Forschungsdesign, das Datengewinnung und -analyse nicht separiert, sondern alternieren lässt.

Diese Gründe machen eine Archivierung qualitativer Daten nur für ein kleines Segment der Forschung überhaupt *potenziell* sinnvoll (etwa bei der Konversationsanalyse oder in Studien mit Leitfadeninterviews), für das Gros hätte sie eher musealen Wert. Die Wiederverwendbarkeitswahrscheinlichkeit ist insgesamt sehr gering. Trotzdem ist sie größer als Null. Ich sehe tatsächlich drei – bescheidene – Fälle:

1. Ein Datasharing für Lehrzwecke (Kretzer 2013) kann im Einzelfall durchaus sinnvoll sein. Wenn es sich um forschendes Lernen im Rah-

- men grundständiger Studiengänge handeln soll, wäre es allerdings absurd, nicht auch die Datengewinnung zu lehren und zu lernen, inklusive ihrer präzisen Spezifikation für die jeweils verfolgte Fragestellung.
2. In einigen Fällen kann auch eine Sekundärnutzung mancher Daten durchaus sinnvoll sein. Die bestehende soziale Praxis ist hier ein gelegentlicher Datenaustausch zwischen Projekten, in dem auf eine persönliche Anfrage Datensätze freundlich überlassen werden. Denkbar ist, diese – seltene – dezentrale Praxis in Eigenregie und auf der Basis von Vertrauens- und Kooperationsbeziehungen zu einer freiwilligen, zentralen Archivierung bestimmter Daten auszudehnen. Wer dies tut, nimmt allerdings einen erheblichen Aufwand der Bearbeitung seiner Daten in Kauf, denn man kann über einem konstant gehaltenen, qualitativen Datenkorpus, der für eine bestimmte Frage angelegt wurde, nicht einfach die Fragestellung »austauschen«. Es braucht dafür nicht nur eine umfassende Anonymisierung von Daten, sondern vor allem eine weit umfänglichere *Kontextierung* für die KollegInnen: Einleitungen, Verlinkungen, Erläuterungen, Kommentare, Aus- und Umarbeitungen von Daten, um diese überhaupt konservierbar, das heißt, unabhängig vom Gehirn des Produzenten zu machen. Der komplementäre Aufwand entsteht auch auf Seiten der Sekundärnutzer. Daher ist im Einzelfall zu fragen: Lohnen die Mühen der »Reanimation«, wenn die Produktion *neuer* Daten so einfach ist?
 3. Auch mögliche Gewinne der Datenarchivierung für die *Mixed Method* Forschung kann es durchaus geben. Aber die besondere Förderempfehlung des Wissenschaftsrats an die DFG hätte sich der Tatsache zu stellen, dass *Mixed Methods* zum einen nur einen kleinen Teil der empirischen Sozialforschung ausmachen; zum anderen, dass diese vorwiegend anwendungsbezogene Forschung sicher nicht zu den soziologischen Forschungsfronten gehört. Die avancierte quantitative Forschung braucht keine qualitative Ergänzung, sie sucht eher Befreiung davon, noch empirische Forschung sein zu müssen. Ihr Anliegen heißt Modellbildung und Simulation. Und die avancierte qualitative Forschung – etwa in der Narrations- oder Konversationsanalyse – braucht keine quantitative Ergänzung. Ihr Anliegen ist die theoriehaltige, multimodale Sinnrekonstruktion.

Soviel zum bescheidenen Nutzen. Der *Schaden* einer verallgemeinerten Datenarchivierung lässt sich dagegen klar benennen. Er liegt im Risiko einer Zerrüttung des Rapports, also des Vertrauensverhältnisses unserer Infor-

mantInnen zu uns. Dieses hängt zunächst wie in aller Forschung an einer förmlichen Zusicherung von *Anonymität*. Eine Anonymisierung umfangreicher und hoch kontextualisierter Daten ist aber entweder nur mit großem Aufwand und unter Verfälschung von Details (etwa in der Biografieforschung) möglich, oder es ist gar nicht möglich (wie bei Videodaten und wie bei allen Daten gegenüber Angehörigen), ohne dass die Daten ihre Aussagekraft und damit Analysefähigkeit verlieren.

Zur formalisierten Zusicherung kommt in qualitativer Forschung aber auch noch die Gestaltung eines Arbeitsbündnisses, das heißt, von Gesprächsbeziehungen. Das Problem, das hier auftaucht, ist, dass eine pauschalisierte Datenarchivierung schon die *Glaubhaftigkeit* unserer Anonymitätszusicherungen untergräbt. Viele InformantInnen lassen eine Aufzeichnung ihrer Äußerungen oder Verhaltensweisen nur dann zu, wenn man ihnen die Vorläufigkeit ihrer Speicherung zu unmittelbaren Analysezwecken zusichert: Nach der Auswertung werden Ton- und Bildspur von ihnen und zu ihnen gelöscht, so dass die Forschung sozial folgenlos bleibt – ein Desiderat, das besonders in empfindlichen Feldern (etwa im Privatleben, bei Angehörigen von Minderheiten, etwa Migranten, oder bei illegalen Praktiken) virulent ist. In diesem Sinne bemängelt der Ethikkodex der DGS zu Recht das Fehlen eines Zeugnisverweigerungsrechts für Sozialwissenschaftler.

Eine pauschalisierte Datenarchivierung würde diese Bedingung des Rappports grundsätzlich ändern. Aus forschungsethischen Gründen müssten wir die explizite Zustimmung unserer Informanten zur Archivierung ihrer Daten einholen. Wenn wir das aber tun, implantieren wir in den Aufbau einer Vertrauensbeziehung eine Publikationsdrohung. Ich würde zum Beispiel sinngemäß sagen:

»Guten Tag, ich komme von der Universität Mainz und forsche über Paarbeziehungen. Ich würde gern ein Interview zu Ihrer Ehe mit Ihnen führen. Die Daten würden selbstverständlich anonymisiert und im Prinzip auch vertraulich behandelt. Allerdings würden sie für die soziologische Öffentlichkeit archiviert, damit – Sie verstehen? – auch die Kolleginnen in Deutschland und im deutschsprachigen Ausland freien Zugriff auf sie haben. Wären Sie damit einverstanden?«

Ein Patient, dem man Risiken eines chirurgischen Eingriffs mitteilt, um sein informiertes Einverständnis zu erhalten, verzichtet fast nie auf eine Operation; ein Informant, den man auf Veröffentlichungsrisiken für sein Leben hinweist, hat nur einen Grund mehr, an einer Studie nicht teilzunehmen. Er darf die archivarische Verewigung und unkontrollierbare Öffnung

seiner Äußerungen für unbekannte Dritte als eine Art putativer ›Auslieferung‹ auffassen. Dieses Problem einer Schädigung des Arbeitsbündnisses mit den InformantInnen existiert in standardisierten Erhebungen so nicht, denn Standardisierung bedeutet auch, dass die Forschung alles tut, um sich gegen jene dichten Sozialbeziehungen abzupuffern, auf denen die qualitative Forschung gerade beruht. Es geht hier nicht um gewisse Einbußen in Rücklaufquoten, sondern um die Pflege der sozialen Basis der Forschung. Schon die öffentliche Diskussion um Archivierungsforderungen kann hier Schäden erzeugen, weil die Datenarchivierung in ihrem Grundgedanken einer ›Vorratsdatenspeicherung‹ entspricht, mit der sich Missbrauchssorgen verbinden. Dass die Sozialwissenschaften keine Geheimdienste sind, bedeutet nicht, dass sie von Informanten für so ›arglos‹ gehalten werden, wie sie sich selbst erscheinen, sondern, dass sie nicht so fest im Sattel sitzen wie die NSA. Diese Wissenschaften beruhen auf kündbaren Sozialbeziehungen. Ihr wertvollstes Gut sind nicht die einmal erlangten Daten, sondern die höchst labile Bereitschaft von Personen, auch zukünftig an sozialwissenschaftlichen Erhebungen teilzunehmen.

Das Hauptproblem einer flächendeckenden Datenarchivierung ist also, dass die Archivierung der Daten ihre *Produktion* untergraben kann. Das zentrale Problem ist nicht, dass die Archivierung unseren *Informanten* schaden kann – und dies ist bereits ein beträchtliches Problem –, sondern dass sie deren Vertrauen so untergräbt, dass dies unserer *Forschung* schaden kann.

Fazit

Die Sekundärnutzung von Daten der qualitativen Sozialforschung ist ein Anliegen, das nur in Teilen dieser Forschung überhaupt begrenzt sinnvoll ist.² Um es umzusetzen, wird man einen Aufwand treiben müssen, der diesen Sinn stark relativiert. Wenn die Datenarchivierung aber zur Regelanforderung an qualitative Sozialforschung wird, dann setzt sie die primäre Erzeugung solcher Daten aufs Spiel.

2 Auszuloten ist hier die Entwicklung von Archivdaten angemessenen Forschungsfragen. Welche qualitativen Längsschnittdesigns sind denkbar? Könnte man einige Privatleute Egodokumente fürs Archiv erzeugen lassen, so wie sie ihre Körper postmortal für Forschungszwecke freigeben?

Wie steht es vor diesem Hintergrund mit den zwei eingangs genannten Gründen des Wissenschaftsrates für die Datenarchivierung? Dessen Empfehlung, die Erhebung und Publikation von Daten als eigenständige Forschungsleistung anzuerkennen, ist aus Sicht der qualitativen Sozialforschung absurd. Das Aufstellen von Kameras, die Durchführung, Aufzeichnung und Transkription von Gesprächen mit zigtausenden von Seiten Datenmaterial ist wie die Totalerhebung von Fernsehproduktionen eine wissenschaftlich sinnlose Hilfstätigkeit. Sie mit Geldmitteln zu fördern, überschätzt die knappste Ressource in der Wissenschaftskommunikation: die Aufmerksamkeit von Lesern.

Die Empfehlung, bei der Konzeption, Beantragung und Bewilligung von Projekten personelle Kapazitäten für die Datenarchivierung einzuplanen, kann man entweder als eine Aufforderung an die DFG sehen, qualitativer Forschung pauschal einen Datenarchivierungs-Overhead einzuräumen, der die aufwändige Anonymisierung und Kontextierung unserer Daten finanziert. Oder man betrachtet die flächendeckende Datenarchivierung als eine zu vermeidende Mittelverschwendung. Richard Münch (2007) hat darauf aufmerksam gemacht, dass es bereits einen Zielkonflikt zwischen der Publikation und der *Beantragung* von Projekten gibt. Die zur Bedienung von Produktivitätsindikatoren »herausgehauenen« Anträge gehen auf Kosten dessen, was man noch aus der Forschung publizieren kann. Eine Archivierungspflicht würde diese Zeitbudgetproblematik verschärfen. Wenn nach den Wissenschaftsevaluierern nun auch die Bibliothekare ihre Standards durchsetzen, dann wird die als Ressourcenschonung gedachte Datenarchivierung in eine große Zeit- und Geldverschwendung münden. Wer hat daran ein Interesse?

Es gibt keinen Grund für eine Fetischisierung von Daten. Weder besteht der Sinn von Sozialforschung in der Analyse von Daten – er besteht in der Analyse sozialer Situationen, Prozesse und Strukturen *mithilfe* von Daten. Noch besteht der Sinn von Forschung in der Produktion von Daten, er besteht in der Produktion von Publikationen, in denen Begriffe gebildet und Zusammenhänge *anhand* von Daten verstanden und erklärt werden. Dass dabei (wie bei anderen Konsumprozessen) auch Unmengen an »Abfall« entstehen, ist kein Schaden. Im Gegenteil: Es ist ein Segen, dass die meisten Daten nach ihrer Gewinnung und analytischen Verarbeitung aus unserem Gedächtnis und unseren Dateien verschwinden. Das macht den Kopf frei für die Erfindung neuer und besserer Forschungsfragen.

Literatur

- Breidenstein, G., Hirschauer, S., Kalthoff, H., Nieswand, B., 2013: Ethnografie. Die Praxis der Feldforschung. Konstanz: UVK, UTB.
- Geertz, C., 1973: Deep Play – Bemerkungen zum balinesischen Hahnenkampf. In C. Geertz, Dichte Beschreibung. Beiträge zum Verstehen kultureller Systeme. Frankfurt am Main: Suhrkamp, 202–260.
- Heaton, J., 2004: Reworking Qualitative Data. London: Sage.
- Huschka, D., Knoblauch, H., Oellers, C., Solga, H. (Hg.) 2013: Forschungsinfrastrukturen für die qualitative Sozialforschung. Wiesbaden: Springer VS.
- Knoblauch, H., Solga, H., 2011: Thesen zur Handhabung quantitativer und qualitativer Daten in Forschungsinfrastrukturen der Sozialwissenschaften. RatSWD Working Paper Nr. 190. www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_190.pdf (letzter Aufruf 20.3.2014).
- Kretzer, S., 2013: Vom Nutzen des Datasharing für die Lehre in der qualitativen Sozialforschung. In D. Huschka, H. Knoblauch, C. Oellers, H. Solga (Hg.), Forschungsinfrastrukturen für die qualitative Sozialforschung. Wiesbaden: Springer VS, 153–164.
- Mauthner, N., Parry, O., 2013: Open access digital data sharing: policies, principles and practices. *Social Epistemology*, 27. Jg., Heft 1, 47–67.
- Münch, R., 2007: Die akademische Elite. Zur sozialen Konstruktion wissenschaftlicher Exzellenz. Frankfurt am Main: Suhrkamp.
- Rosenthal, G., 2013: Was geschieht mit unseren Daten? Ein Plädoyer für eine Diskussion über die Möglichkeiten der Sekundärnutzung von autobiographischen Materialien. Newsletter / Rundbrief Nr. 64 der Sektion Biographieforschung der DGS, 44–45.
- Wissenschaftsrat 2012: Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. www.wissenschaftsrat.de/download/archiv/2359-12.pdf (letzter Aufruf 20. März 2014).