Datenschutz im Forschungsdatenmanagement

Watteler, Oliver; Ebel, Thomas

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
Verlag Barbara Budrich

Empfohlene Zitierung / Suggested Citation:
Watteler, O., & Ebel, T. (2019). Datenschutz im Forschungsdatenmanagement. In U. Jensen, S. Netscher, & K. Weller (Hrsg.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten: Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten* (S. 57-80). Opladen: Verlag Barbara Budrich. https://doi.org/10.3224/84742233.05

Auszug aus dem Buch:

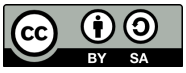
Uwe Jensen
Sebastian Netscher
Katrin Weller (Hrsg.)

Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten

Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten

Verlag Barbara Budrich
Opladen • Berlin • Toronto 2019

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
http://dnb.d-nb.de abrufbar.



Dieses Buch steht im Open-Access-Bereich der Verlagsseite zum kostenlosen Download bereit (https://doi.org/10.3224/84742233).
Eine kostenpflichtige Druckversion (Print on Demand) kann über den Verlag bezogen werden. Die Seitenzahlen in der Druck- und Onlineversion sind identisch.

ISBN 978-3-8474-2233-4 (Paperback)
eISBN 978-3-8474-1260-1 (eBook)
DOI 10.3224/84742233

Umschlaggestaltung: Bettina Lehfeldt, Kleinmachnow – www.lehfeldtgraphic.de
Lektorat: Nadine Jenke, Potsdam
Satz: Anja Borkam, Jena – kontakt@lektorat-borkam.de
Titelbildnachweis: Foto: Florian Losch
Druck: paper & tinta, Warschau
Printed in Europe

4. Datenschutz im Forschungsdatenmanagement[1]

Oliver Watteler und Thomas Ebel

In den Sozialwissenschaften wird oft mit personenbezogenen und eventuell sensiblen Daten von Studienteilnehmer/innen geforscht. Erhebung, Verarbeitung und Nutzung von personenbezogenen Daten unterliegen dann datenschutzrechtlichen und forschungsethischen Bestimmungen.

Ziel dieses Kapitels ist es, sozialwissenschaftlichen Forscher/innen die grundlegenden Regelungen im Bereich des Datenschutzes aufzuzeigen und Anleitungen aus der Praxis zu bieten, wie diese Regelungen konkret umgesetzt werden können.

Zu diesen Zwecken werden im ersten Abschnitt (4.1) dieses Kapitels forschungsethische und datenschutzrechtliche Aspekte in der sozialwissenschaftlichen Forschung erörtert. Daran anschließend diskutiert Abschnitt 4.2 den Datenschutz in den unterschiedlichen Phasen eines Forschungsprojekts. Im dritten Abschnitt (4.3) werden Anonymisierungsmöglichkeiten für quantitative Daten vorgestellt und anhand von Fallbeispielen aus der Praxis verdeutlicht. Abschließend gehen wir auf häufig wiederkehrende Fehler von Forscher/innen in der Umsetzung datenschutzrechtlicher Bestimmungen ein (Abschnitt 4.4).

4.1 Der Datenschutz in der sozialwissenschaftlichen Forschung

Fragen des Datenschutzes betreffen Daten, die von Individuen oder über sie erhoben werden (*human subject research*). Solche Daten fallen u.a. in Befragungen an. Für unsere Zwecke unterscheiden wir drei große Bereiche von Individualdaten: Erstens den Bereich der amtlichen Statistik, zweitens den Bereich prozess-produzierter Daten, zu denen man auch internetbasierte Daten (z.B. Social-Media-Daten) zählen kann, und drittens Daten aus eigenen Befragungen (Watteler 2017: 127-136).

Die Daten werden u.a. durch Messungen, also die Zuordnung von Zahlen zu Objekten oder Ereignissen, oder über qualitativ orientierte Verfahren, wie offene Interviews oder Beobachtung, gewonnen (ebd.). Unabhängig davon, welche Methoden verwendet werden, beziehen sich die Untersuchungen, mit denen wir uns an dieser Stelle befassen, auf Personen. Beim Umgang mit diesen Personen, der Beobachtung ihres Verhalten, der Erfragung ihrer Ansichten und Einstellungen oder der Darstellung ihrer sozialen und wirtschaftlichen Lebensbedingungen sind besondere forschungsethische Rahmenbedingungen zu beachten. Die Personen dürfen nicht geschädigt werden und es sind rechtliche Regelungen und freiwillige Verpflichtungen zu ihrem Schutz zu beachten (Graumann 2006: 253-256). Dazu zählen die Persönlichkeitsrechte, wie das Recht auf informationelle Selbstbestimmung, das Recht am eigenen Bild oder das Recht auf Privat- und Intimsphäre (Palandt 2008: 1216-1226). Ein besonderes Schutzrecht ist der Datenschutz (RatSWD 2016; Höhne 2010; Watteler 2010; Bethlehem 2009; Wirth 2003; Metschke/Wellbrock 2002; Wirth 1992), wonach „die Grundrechte und Grundfreiheiten natürlicher Personen und insbesondere deren Recht auf Schutz

1 Wir danken dem externen Datenschutzbeauftragten von GESIS, Harald Eul, und Wolfgang Jagodzinski für ihre Anmerkungen und Änderungsvorschläge.

https://doi.org/10.3224/84742233.05

personenbezogener Daten“ zu wahren sind (Datenschutz-Grundverordnung (DSGVO) Art. 1 Abs. 2). Neu ist infolge der DSGVO, dass der Grundrechtsschutz und nicht mehr der Schutz der Daten im Vordergrund steht.[2] Mit natürlichen Personen sind im Rahmen des seit 25. Mai 2018 geltenden Rechts lebende Menschen gemeint. Die *Handreichung Datenschutz* des Rates für Sozial- und Wirtschaftsdaten (RatSWD 2017) bietet einen sehr guten Überblick über das Thema. Wir konzentrieren uns daher im Folgenden auf praktische Hinweise zum Umgang mit Daten, die in sozialwissenschaftlichen Projekten bei Personen erhoben worden. Wir beginnen mit einem Überblick über die aktuelle rechtliche Situation.[3]

Grundsätzlich kann man drei Ebenen des Datenschutzrechts unterscheiden:

1. die Ebene der Grundrechte, die in der deutschen Verfassung sowie in der Charta der Grundrechte der Europäischen Union verankert sind,
2. die Ebene der Einzelgesetze (national wie international) und
3. die Ebene der Regulierungen etwa über wissenschaftliche Fachgremien.

Schaukasten 4.1: Definition personenbezogene/-beziehbare Daten

Personenbezogene Daten

Personenbezogene Daten wurden im Bundesdatenschutzgesetz bisher definiert als „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener)“ (§ 3 Abs. 1 BDSG). Personenbezogene Daten sind in Zukunft „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person […] beziehen“ (Art. 4 Abs. 1 DSGVO). Identifizierbar ist eine natürliche Person, wenn sie direkt oder indirekt unter Zuhilfenahme anderer Daten und Merkmale identifiziert werden kann (ebd.).

Besondere Arten personenbezogener Daten

Das BDSG, die EU-Datenschutzrichtlinie 95/46/EG und die DSGVO heben einige Merkmale natürlicher Personen als besonders schutzwürdig hervor. Dies waren im bisherigen BDSG „Angaben über die rassische und ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftszugehörigkeit, Gesundheit oder Sexualleben“. Die DSGVO übernimmt in Artikel 9 (Verarbeitung besonderer Kategorien personenbezogener Daten) diese Aufzählung und erweitert sie explizit um genetische und biometrische Daten (Art. 9 Abs. 1 DSGVO).

Personenbeziehbare Daten

Wenn eine Identifizierung nur indirekt, beispielsweise durch die Kombination mehrerer Merkmalswerte in den Daten, durch Zusatzwissen oder die Kombination mit externen Datenquellen möglich erscheint, wurde bisher der Begriff der Personenbeziehbarkeit verwendet (Metschke/Wellbrock 2002: 15). Den Unterschied zwischen personenbezogenen und personenbeziehbaren Daten macht die DSGVO nicht mehr.

Quelle: Eigene Darstellung

Auf der Ebene der Grundrechte steht dem Recht auf Freiheit der Forschung das Recht des Einzelnen auf informationelle Selbstbestimmung gegenüber. Beide sind im Grundgesetz festgeschrieben oder wurden durch Auslegung des Bundesverfassungsgerichts (BVerfG) genauer bestimmt. Diese beiden Grundrechte wurden auch in die Charta der Grundrechte der Europäischen Union übernommen. So bezieht sich Artikel 8 ausdrücklich auf den *Schutz personenbezogener Daten* und umreißt die Rahmenbedingungen für den Datenschutz, etwa die Verarbeitung von Daten für festgelegte Zwecke und die Einwilligung zur Verarbeitung durch betroffene Personen. Artikel 13 der Charta bestimmt, dass die Forschung frei ist.

2 Bis zum Inkrafttreten der DSGVO war unter Datenschutz der Schutz einer natürlichen Person davor zu verstehen, „durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt“ zu werden (Bundesdatenschutzgesetz (BDSG-alt) § 1 Abs. 1).

3 Man muss berücksichtigen, dass sich zum Zeitpunkt der Erstellung dieses Textes (Sommer 2018) das Datenschutzrecht vor allem durch das Inkrafttreten der DSGVO in Bewegung befindet. Daher können an dieser Stelle keine abschließenden Aussagen zum Datenschutzrecht getroffen werden.

Für den Fall, dass solche Grundrechte in Konflikt geraten, hat der Gesetzgeber Regelungen zu schaffen, die dem Grundsatz der *praktischen Konkordanz* entsprechen. Beispiele für Regelungen sind die inhaltlichen und formalen Anforderungen an rechtswirksame Einwilligungen in die Verarbeitung personenbezogener Daten. In diversen Gesetzen (u.a. DSGVO, Bundes- und Landesdatenschutzgesetze) wurden daher Voraussetzungen festgelegt, unter denen personenbezogene Daten für Forschungszwecke verarbeitet werden dürfen (zur Definition personenbezogener bzw. -beziehbarer Daten s. Schaukasten 4.1). Ferner ist in der Forschung seit Längerem das Konzept des *informed consent* bekannt, welches die Entscheidung über eine freiwillige Teilnahme an Forschungsprojekten von ausreichender Information über das Vorhaben abhängig macht.

Auf der Ebene der Einzelgesetze gibt es seit dem Inkrafttreten der DSGVO einen EU-weit einheitlichen Rahmen für das Datenschutzrecht. Die DSGVO löste im April 2016 die EU-Datenschutzrichtlinie 1995/46/EG ab und ist unmittelbar in allen EU-Mitgliedsstaaten anwendbar. Die Vorschriften der DSGVO dürfen vom nationalen Gesetzgeber nur abgeändert oder konkretisiert werden, wenn und soweit der europäische Gesetzgeber sogenannte *Öffnungsklauseln* vorgesehen hat.

Einer der Anwendungsbereiche, die Änderungen sowohl auf europäischer als auch auf nationaler Ebene erfuhren, ist die Forschung. Auf der Basis der in der DSGVO genannten Öffnungsklauseln wurde das Bundesdatenschutzgesetz (BDSG-alt) im Juli 2017 durch das *Gesetz zur Anpassung des Datenschutzrechts*, kurz DSAnpUG-EU, reformiert (im Folgenden BDSG-neu). Das Gesamtbild der Änderungen in diesem Fall kann zum jetzigen Zeitpunkt noch nicht abschließend dargestellt werden. Insgesamt ist jedoch davon auszugehen, dass das bisher in der Forschung übliche Vorgehen (informierte Einwilligung, Zweckbindung, Trennung von direkten Identifizierungsmerkmalen von den Befragungsdaten, Risikoabschätzung u.a.) auch weiterhin Gültigkeit haben wird (vgl. Buchner 2016).

Führt man also z.B. eine allgemeine Bevölkerungsumfrage im gesamten Gebiet der Bundesrepublik Deutschland durch, so gelten die DSGVO und das Bundes- bzw. das jeweilige Landesdatenschutzgesetz.[4] Da Daten von Personen erhoben werden, gilt das Datenschutzrecht und das Prinzip des *Verbots mit Erlaubnisvorbehalt*. D.h., dass es keine anlasslose Datensammlung von personenbezogenen Daten geben darf.[5] Dieses Verbot gilt nicht, wenn Personen z.B. ihre informierte Einwilligung erteilen (sogenannter Erlaubnistatbestand). Der Einwilligung kommt somit wie oben bereits erwähnt eine zentrale Bedeutung für die wissenschaftliche Forschung zu (zu den Details s. Schaukasten 4.2).[6]

Schaukasten 4.2: Definition informierte Einwilligung („informed consent“)

Die Einwilligung zur Teilnahme an einer Untersuchung muss freiwillig erfolgen, die Person muss einwilligen können bzw. dürfen, die Informationen müssen verständlich sein und, falls die Einwilligung in Schriftform erfolgt, muss nachgewiesen werden, dass diese Art der Zustimmung vorliegt (Gola 2017 bietet im Kommentar zu Artikel 7 der DSGVO einen Überblick über die rechtliche Situation; Schaar 2017 geht auf die Neuerungen bei der Einwilligung ein). Die Einwilligung gilt als „Ausdruck des aus dem allgemeinen Persönlichkeitsrecht abgeleiteten Rechts auf informationelle Selbstbestimmung“ (Rogosch 2013: 17).

Quelle: Eigene Darstellung

4 Im Einzelfall können bei der Erhebung und Verarbeitung von personenbezogenen oder personenbeziehbaren Daten weitere Gesetze Berücksichtigung finden. Zu diesen Gesetzen zählen etwa das Sozialgesetzbuch (SGB) oder auf Länderebene die Schulgesetze.

5 Das Prinzip wird auch in der aktualisierten Rechtssituation beibehalten (Buchner 2016: 156).

6 Anders liegt die Situation im Falle von Daten der amtlichen Statistik, bei denen die Teilnahme in aller Regel per Gesetz verpflichtend ist (Wirth 2016). Zu prozessproduzierten Daten, wie denen der Deutschen Rentenversicherung, vgl. z.B. Himmelreicher et al. (2017) und Hansen et al. (2012).

Die DSGVO regelt, welche Inhalte eine Einwilligungserklärung enthalten muss. Maßgeblich sind hier die Artikel 6 (Rechtmäßigkeit der Verarbeitung), 7 (Bedingungen für die Einwilligung), 8 (Bedingungen für die Einwilligung eines Kindes in Bezug auf Dienste der Informationsgesellschaft) und 9 (Verarbeitung besonderer Kategorien personenbezogener Daten). Die Betroffenen müssen grundsätzlich freiwillig und informiert einwilligen. Voraussetzung dazu ist, dass sie umfassend über alle wesentlichen Aspekte der Rechtsgrundlagen und Zwecke der Verarbeitung ihrer Daten informiert werden (Schaukasten 4.3 bietet Verweise auf Vorlagen für Einwilligungserklärungen).

Schaukasten 4.3: Beispiele für die Einwilligungserklärung

ADM (Arbeitskreis Deutscher Markt und Sozialinstitute e.V.): Mustererklärung für mündliche oder schriftliche Interviews. https://www.adm-ev.de/datenschutz/ [Zugriff: 09.02.2017].

Forschungsdaten-Bildung: Checkliste für die Erstellung eigener Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen. http://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=FDB_Einwilligung_Checkliste.pdf [Zugriff: 28.07.2016].

RatSWD: Mustererklärungen für die Erhebung personenbezogener qualitativer Daten sowie deren Archivierung. http://www.ratswd.de/dl/RatSWD_WP_238.pdf [Zugriff: 28.07.2016].

QualiService: MUSTER – Einwilligungserklärung zur Erhebung und Verarbeitung personenbezogener Interviewdaten. http://www.qualiservice.org/fileadmin/text/Einverstaendnis2013_08.pdf [Zugriff: 07.06.2018].

Quelle: Eigene Darstellung

Die Einwilligungserklärung sollte mindestens folgende Informationen enthalten (Schaar 2017; Verbund Forschungsdaten Bildung/Rechtsanwälte Goebel & Scheller 2015; Metschke/Wellbrock 2002: 25ff.):

- Identität der verantwortlichen und verarbeitenden Stelle(n), sprich in unserem Fall Leiter/in und Träger/in des Forschungsvorhabens sowie Name und Kontakt der Stellen, an denen Daten verarbeitet werden (z.B. Einrichtungen der Markt- und Meinungsforschung);
- Zweckbestimmung der Verarbeitung der Daten (der Begriff Verarbeitung schließt sämtliche Phasen einer Datennutzung ein, wie etwa das Erheben, Erfassen, die Organisation, das Ordnen, die Speicherung, die Veränderung, das Auslesen, das Abfragen, die Verwendung, die Offenlegung, Verbreitung, die Verknüpfung oder das Löschen (vgl. Art. 4 Abs. 2 DSGVO), sprich in unserem Fall regelmäßig die Forschung als Zweck);
- Kategorien der Empfänger der Daten;
- Art der verarbeiteten Daten und Hinweis auf vertraulichen Umgang mit personenbezogenen Daten;
- Hinweis auf die Rechte der Befragten (Auskunft, Berichtigung, Widerruf und Löschung, die aber eingeschränkt werden können, wenn z.B. durch die Ausübung dieser Rechte die Verwirklichung der Forschungszwecke ernsthaft beeinträchtigt würde) und insbesondere auf die Freiwilligkeit der Angaben und die Widerruflichkeit der Einwilligung mit Wirkung für die Zukunft;
- Bei Erhebung von besonderen Arten personenbezogener Daten sind diese gesondert und ausdrücklich in der Einwilligungserklärung aufzuführen (Art. 9 Abs. 2a DSGVO; BDSG-neu § 27 Abs. 1 (Ausnahme); RatSWD 2017: 21ff.).

Ergänzt werden muss die Einwilligungserklärung um Informationen über die Verarbeitung der Daten. Darunter fallen die Rechtsgrundlagen und Zwecke der Verarbeitung (soweit diese über die Einwilligung hinausgehen), eine eventuelle Datenübermittlung in Länder außerhalb der EU, die Speicher- bzw. Löschfristen der personenbezogenen Daten und das Beschwerderecht bei einer Datenschutzaufsichtsbehörde.

Vor dem Einholen von Einwilligungen ist zudem noch zu klären, ob die Betroffenen einwilligen können (Mindestalter und Einsichtsfähigkeit), inwieweit es weitere Vorschriften gibt, die einzuhalten sind (insbesondere Landesgesetze und Schulgesetze bei Befragungen von Schülern/innen) und wer in den Adressatenkreis der Einwilligenden fällt (beispielsweise

bei Minderjährigen die Eltern).[7] Idealerweise wird zusätzlich darauf hingewiesen, dass die Daten anderen Forscher/innen nur in einer Form zur Nachnutzung zur Verfügung gestellt werden, die keine Rückschlüsse auf die Person der Befragten zulassen.

Ein weiterer Rechtsbereich, der besonders bei der Verarbeitung von internetbasierten Daten von Bedeutung ist, betrifft Lizenzrechte und das fehlende Einverständnis für die Nutzung von Daten für andere Zwecke. Unternehmen, die z.B. sogenannte Social-Media-Dienste wie Chats, Messaging, Info Boards oder den Tausch digitaler Bilder anbieten, behalten sich Rechte an den von ihnen verarbeiteten Inhalten vor. Diese Rechte werden in Allgemeinen Geschäftsbedingungen und in aller Regel in Datenschutzbestimmungen geregelt. Internetbasierte Daten können als prozessproduzierte Daten gesehen werden. Wie im Falle anderer prozessproduzierter Daten, wie denen von Verwaltungen, wurden sie ursprünglich nicht zu Zwecken der Forschung erzeugt (zu Social-Media-Daten s. auch Kapitel 11; Bender et al. 2017; Lane et al. 2014).

Auf der Ebene der Regulierungen haben sich viele Fachverbände freiwillige Verpflichtungen zur Forschungsethik gegeben, die im Wesentlichen ähnliche Ideen aufgreifen. So stellt der Ethikcode der Deutschen Gesellschaft für Soziologie z.B. dar, dass bei der Datenerhebung das Prinzip der informierten Einwilligung gilt und Studienteilnehmer/innen nicht geschädigt werden dürfen. Der Code erwähnt auch ausdrücklich den Schutz vor Re-Identifizierung. Schaar (2017) stellt die Regelungen verschiedener Fachverbände gegenüber. Darüber hinaus richten viele Hochschulen derzeit allgemeine Ethikkommissionen ein, die Forschungsvorhaben im Vorfeld begutachten und Maßnahmen zum Schutz der Untersuchungspersonen vorschlagen (zu Ethikkommissionen in den Sozialwissenschaften vgl. von Unger/ Simon 2016).

Diese knappe Darstellung des Datenschutzrechts verdeutlicht, dass es zwar eine Vielfalt rechtlicher und freiwilliger Regelungen auf verschiedenen administrativen Ebenen gibt, ein möglicher gemeinsamer Kern für viele sozialwissenschaftliche Projekte jedoch die informierte Einwilligung der Untersuchungsperson auf der Basis ausreichender Angaben zum Forschungsvorhaben ist. Wie in der Praxis mit den Daten und dem Datenschutz umgegangen werden kann, um den Schutz der Grundrechte der Untersuchungspersonen bei der Verarbeitung ihrer Daten zu gewährleisten, wird im Folgenden ausführlich dargelegt.

4.2 Datenschutz in typischen Phasen der Forschung

Die folgenden Ausführungen orientieren sich am sogenannten *Lebenszyklus* von Forschungsdaten, wie in Kapitel 2.2 dieses Buches vorgestellt. Im hiesigen Kontext spielen die Phase der Archivierung und Nachnutzung eine zentrale Rolle. Fehler, die eventuell in einer vorherigen Forschungsphase gemacht wurden, lassen sich im Nachhinein zumeist nur noch schwer oder gar nicht korrigieren.

Im Zusammenhang mit datenschutzrelevanten Aspekten unterscheiden wir im Folgenden generell drei zentrale Phasen im Lebenszyklus:

1. Design- und Erhebungsphase,
2. Aufbereitungs- und Analysephase,
3. Veröffentlichungs- oder Archivierungsphase.

7 Eine Checkliste findet sich bei Verbund Forschungsdaten Bildung/Rechtsanwälte Goebel & Scheller (2015).

4.2.1 Design- und Erhebungsphase

In der Designphase werden die grundsätzlichen Entscheidungen über Art und Umfang der Forschungsdaten getroffen, die im Projekt zur Beantwortung einer oder mehrerer Forschungsfragen verwendet werden sollen. Die anschließende Erhebungsphase bestimmt die Güte dieser Daten (vgl. u.a. Diekmann 2007: 186-229). Zentral für den Umgang mit Forschungsdaten prinzipiell und mit Datenschutz im Besonderen ist unseres Erachtens ein gut geplantes Forschungsdatenmanagement. Dies entspricht zum einen der guten wissenschaftlichen Praxis und zum anderen dem Erforderlichkeitsgrundsatz des Datenschutzes (vgl. RatSWD 2017). Die Auseinandersetzung mit diesen Aspekten schon zu Beginn der Forschung erleichtert die spätere Veröffentlichung und Nachnutzung der Daten, wie in Kapitel 3.2.2 ausführlicher dargestellt. Wir beziehen uns hier daher auf eine weitere Publikation des Rates für Sozial- und Wirtschaftsdaten zu diesem Thema (RatSWD 2016). In dieser werden u.a. folgende Leitfragen zum Umgang mit personenbezogenen Daten gestellt, die wir hier übernehmen (s. auch Schaukasten 4.4).

> *„Welche Daten erheben oder verwenden Sie?" (RatSWD 2016: 10)*

Prinzipiell steht es Forscher/innen frei, auf legalem Weg alle für ein Forschungsvorhaben notwendigen Daten zu erheben (s.o.). Je nach Design wird man im Rahmen einer sogenannten Sekundäranalyse auf bereits bestehende Daten zurückgreifen. Erhebungsprogramme wie das Sozio-oekonomische Panel (SOEP), das deutsche Nationale Bildungspanel (NEPS) oder die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) sind Beispiele für breit genutzte Datenbestände. In diesen Programmen und in Datenangeboten, die etwa von den im Rat für Sozial- und Wirtschaftsdaten (RatSWD) zusammengefassten Forschungsdatenzentren oder vom Datenarchiv von GESIS bereitgestellt werden, sind bereits organisatorische Vorkehrungen für eine datenschutzkonforme Nachnutzung von Forschungsdaten getroffen. Eine Nachnutzung von Daten erfüllt auch das datenschutzrechtliche Gebot der Datenminimierung: „Personenbezogene Daten müssen […] c) dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein" (Art. 5 Abs. 1c DSGVO).[8] D.h., Forscher/innen erheben keine eigenen Daten und damit werden zu beforschende Personen weniger oft auf eine Teilnahme angesprochen. Da für die genannten Daten bereits Regelungen getroffen wurden, konzentrieren wir uns im Folgenden auf die Erhebung eigener Daten.

> *„Welche Nachnutzungsmöglichkeit der Daten planen Sie für die Zeit nach dem Projektende? Wie werden die Forschungsdaten innerhalb und nach Ablauf des Projekts genutzt? Werden Daten an mehreren Projektstandorten genutzt?" (RatSWD 2016: 10)*

Bereits zu Beginn eines Projektes sollte man sich Gedanken darüber machen, ob und, wenn ja, wie die Daten nach Abschluss des Projektes weiterverwendet werden sollen. Vom gewählten Szenario hängt das Einverständnis der Befragten ab.

Das Vorgehen beim Umgang mit Daten im Forschungsvorhaben und im Anschluss ist Teil der informierten Einwilligung. Die datenschutzrechtlichen Vorgaben zu Einwilligungen umfassen, wie im ersten Abschnitt dieses Kapitels erwähnt, die Informiertheit, die Zweckbindung und die Freiwilligkeit. Befragte müssen vollständig über die Verwendung ihrer Daten aufgeklärt werden, damit sie wirksam einwilligen können.

Zwischen dem Forschungsinteresse der Wissenschaftler und Wissenschaftlerinnen und datenschutzrechtlichen Vorgaben kann es in der Praxis zu einem Spannungsverhältnis kommen. Zum einen kann zum Zeitpunkt der Erhebung nicht abgeschätzt werden, ob die

8 Im BDSG-alt wurde von Datensparsamkeit gesprochen.

personenbezogenen Daten länger benötigt werden. Zum anderen wären bei sehr strikter Abgrenzung des Zwecks spätere, auch vermeintlich geringfügige Änderungen des Forschungsvorhabens oder Folgestudien ausgeschlossen. Ein wesentlicher Bestandteil der informierten Einwilligung ist daher der Nutzungszweck. Um Schwierigkeiten zu vermeiden, sind für die wissenschaftliche Forschung – im Gegensatz zu anderen Bereichen – weit gefasste Formulierungen der Zweckbindung akzeptabel. Dabei muss darauf geachtet werden, den Zweck hinreichend bestimmt zu umgrenzen und gleichzeitig die Erklärung so zu verfassen, dass eine inhaltliche Ausweitung der ursprünglichen Forschungsfrage möglich ist (Metschke/Wellbrock 2002: 22f.).

Die Verarbeitung der Daten zu Forschungszwecken und die Zusicherung der Verarbeitung ohne Rückschlüsse auf die einzelne Person sind gängige Formulierungen in der Praxis. Im Sinne der Nachnutzung sollte besonders darauf geachtet werden, wie lange welche Daten vorgehalten werden sollen. Das Forschungsvorhaben (Projekt oder Programm) als Nutzungszweck ist von der wissenschaftlichen Forschung per se als Zweck ohne zeitliche Befristung zu unterscheiden (vgl. zur prinzipiellen Offenheit der Forschung aus rechtlicher Sicht Starck 2010; aus wissenschaftstheoretischer Sicht vgl. Chalmers 2007).[9] Ist eine längerfristige Datenerhebung, wie etwa ein Panel geplant, ist das Einbeziehen von Datentreuhändern möglich, der als einziger in der Lage ist, einen Personenbezug herzustellen. Dies kann als Instrument genutzt werden, um den Datenschutz einzuhalten und möglichen Sorgen der Betroffenen zu begegnen.

Soll eine Studie mit minderjährigen Personen durchgeführt werden, ist zu beachten, dass ggf. die Eltern zusätzlich in die Erhebung einwilligen müssen. Ob dies zutrifft, hängt zum einen vom Alter der Befragten bzw. ihrer Einsichtsfähigkeit und zum anderen von den möglichen Schäden, die ihnen durch die Erhebung (und eine mögliche Re-Identifizierung) entstehen könnten. Üblicherweise werden als Mindestalter für das Vorliegen von Einsichtsfähigkeit 14 oder 16 Jahre angenommen (Rogosch 2013: 49). Die DSGVO sieht als Altersgrenze für eine rechtmäßige Verarbeitung von Daten ohne Einwilligung der Eltern das 16. Lebensjahr vor (Art. 8 Abs. 1 DSGVO). Sollen Schüler/innen befragt werden, müssen darüber hinaus spezielle Landesschulgesetze eingehalten werden.

Idealerweise lassen Forscher/innen Befragte zusätzlich in die Archivierung ihrer Daten einwilligen. Auf diese Weise sind einerseits die Befragten umfänglich darüber informiert, was langfristig mit ihren Angaben geschieht, andererseits ist die Archivierung rechtlich abgesichert.

Internetbasierte Forschung stellt Sozialwissenschaftler/innen vor neue Herausforderungen, da es schwierig, wenn nicht sogar unmöglich ist, informierte Einwilligungen zur Teilnahme an einem Forschungsvorhaben einzuholen. Beispielsweise ermöglicht es der Mikroblogging-Dienst Twitter Dritten, mehrere tausend Nachrichten, sogenannte Tweets, pro Tag zu durchsuchen und auszuwerten. Dies wird zwar in den Lizenzbedingungen des Unternehmens dargelegt, allerdings ist fraglich, ob diese Nutzung betroffenen Twitter-Nutzern bewusst ist, wie in Kapitel 11 dargelegt.

Überlegungen, warum personenbezogene Daten erhoben werden müssen, zu welchem Zeitpunkt Pseudonymisierungs- und/oder Anonymisierungsmaßnahmen unternommen werden sollen, sowie das Löschdatum für die personenbezogenen Daten müssen in dem eingangs erwähnten Forschungsdatenmanagementplan festgehalten werden. Falls es schriftliche Einwilligungserklärungen gibt, müssen auch diese vorgehalten werden.

9 Aus der täglichen Praxis kennen die Autoren Beispiele für Einverständniserklärungen, in denen eine Datennutzung ausschließlich im Rahmen des Forschungsprojektes vorgesehen wird. Damit ist unseres Erachtens eine Nachnutzung oder anonymisierte Archivierung ausgeschlossen, da die Befragten ihre Teilnahme auf der Basis der Erklärung machen.

Für den Fall, dass Forscher/innen zur Archivierung ihrer Daten verpflichtet sind oder eine Archivierung freiwillig anstreben, sollten sie sich frühzeitig mit dem Repositorium oder Datenarchiv ihrer Wahl in Verbindung setzen. Diese Einrichtungen helfen bei Fragen rund um datenschutzrechtliche Voraussetzungen oder Schutzmaßnahmen der Datenarchivierung (vgl. dazu ausführlich Kapitel 7.4). Die in Schaukasten 4.4 zusammengefassten Leitfragen des RatSWD (2016: 16) bieten eine beispielhafte Checkliste zur Datensicherheit im Forschungsprojekt an.

Schaukasten 4.4: Checkliste zum weiteren Vorgehen bei der Datenspeicherung und -sicherung im Projekt

„ D.1 Wie werden die Daten während des Forschungsprozesses gespeichert und gesichert?
Leitfragen:
- Werden Versionierungen der Dateien vorgenommen und wie erfolgt dies?
- Wie werden die Daten gesichert, d.h. welche Art von Sicherung wird in welchen Intervallen durchgeführt?
- Ist sichergestellt, dass ausreichende Kapazitäten für Speicherung und Sicherung der Daten zur Verfügung stehen?

D.2 Wie wird der Zugriff auf die Daten verwaltet und werden die Daten vor Zugriffen Unbefugter geschützt?
[…]
Leitfragen:

- Wie wird verhindert, dass Unbefugte auf die Daten zugreifen können?
- Wie werden Daten bei der Übermittlung (z.B. zwischen den im Feld eingesetzten Systemen und den Systemen am Arbeitsplatz oder zwischen Projektmitarbeitenden unterschiedlicher Einrichtungen) geschützt?“

Quelle: Auszugsweise Darstellung nach RatSWD (2016: 16)

4.2.2 Aufbereitungs- und Analysephase

Daten, die direkte Identifikationsmerkmale wie Namen der Befragten, Anschriften oder E-Mail-Adressen enthalten, besitzen einen Personenbezug. Aber auch andere Merkmale, z.B. genaue Angaben zum Beruf oder kleinräumige Angaben zur Wohnregion, können eventuell indirekt mit den Befragten in Verbindung gebracht werden (siehe Abschnitt 4.3). Wie oben erläutert, können in Forschungsvorhaben alle Daten verarbeitet werden, in deren Verarbeitung die Befragten informiert eingewilligt haben. Allerdings sollten während der Verarbeitung bereits Schutzmaßnahmen getroffen werden. Dazu zählt etwa die getrennte Speicherung direkter Identifizierungsmerkmale. Das Datenschutzrecht sieht weitere Maßnahmen wie die Pseudonymisierung und die Anonymisierung vor, um Befragte auch über das Ende des Forschungsvorhabens hinaus zu schützen. Im Folgenden gehen wir zunächst auf die Pseudonymisierung und schlaglichtartig auf den sicheren Umgang mit Daten ein. Es folgt eine längere Darlegung des Konzeptes der Anonymisierung, das für die meisten Formen der Nachnutzung außerhalb des Forschungsvorhabens, etwa durch Dritte, wichtig ist.

Das Konzept der Pseudonymisierung

Die Pseudonymisierung erfährt in der DSGVO eine deutliche Aufwertung, indem sie in verschiedenen Artikeln als zentrales Instrument des Datenschutzes genannt wird, u.a. in Art. 25 Abs. 1, Art. 32 Abs. 1 und Art. 89 Abs. 1 (vgl. Marnau 2016: 430ff.; Schaar 2016: 7ff.). Die DSGVO definiert Pseudonymisierung in Art. 4 Abs. 5 als eine „Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden

können“. Diese wird durch Datentrennung erreicht. Dazu werden identifizierende von sonstigen Angaben getrennt und in einem zweiten Datensatz untergebracht. Beide Datensätze sind durch eindeutige Zuordnungsschlüssel (Pseudonyme) wieder miteinander verknüpfbar. Die Möglichkeit der Feststellung der wahren Identität bleibt erhalten (Gola 2017 Art. 4: 173).

Neben der Pseudonymisierung sollten personelle und technische Maßnahmen genutzt werden, die eine unkontrollierte Verbreitung von Dateien mit sensiblen Inhalten verhindern. Dazu zählt, die Verarbeitung der Daten auf bestimmte Personen zu beschränken. Nur Bearbeiter/innen, die mit den Bestimmungen des Datenschutzes vertraut ist, sollten im Projekt über Zugriffsrechte auf die (in der Regel mindestens pseudonymisierten) Daten verfügen (Häder 2009: 13). Die Daten sollten zudem nicht ungeschützt auf Netzlaufwerken abgelegt werden. In aller Regel gelangen Daten auf Netzlaufwerken nämlich (automatisiert) in Backups des Instituts oder eines Rechenzentrums. Zur Sicherung der personenbezogenen Merkmale bieten sich hier verschiedene Formen der technischen Verschlüsselung an (s. zur Datensicherheit Kapitel 5.3.3; Hammer/Knopp 2015: 507).

Das Konzept der Anonymisierung

Die DSGVO definiert Anonymisierung in Erwägungsgrund 26 derart, dass die Grundsätze des Datenschutzes nicht mehr für „Informationen, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder personenbezogene Daten [gelten], die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.“ Diese Definition wird in der Praxis *absolute Anonymität* genannt (s. Schaukasten 4.5). Die vielleicht unscheinbare Trennung zwischen *nicht* und *nicht mehr* weist jedoch darauf hin, dass Anonymisierung durch die DSGVO weiterhin als aktive Handlung bestimmt wird (so auch Wójtowicz/Cebulla 2017: 187). Das BDSG-neu ergänzt ferner, dass besondere Kategorien personenbezogener Daten zu anonymisieren sind, sobald dies nach Forschungs- oder Statistikzweck möglich ist.

Schaukasten 4.5: Anonymisierungsniveaus

Unabhängig von der Rechtslage unterscheidet man in der einschlägigen Literatur drei Niveaus der Anonymisierung: die formale, die faktische und die absolute (Höhne 2010: 10f.; Metschke/Wellbrock 2002: Kap. 3.3).

Formale Anonymisierung

Die formale Anonymisierung umfasst das Entfernen direkter Identifizierungsmerkmale. Es gilt zu beachten, dass die Daten weiterhin datenschutzrechtlichen Bestimmungen unterliegen.

Faktische Anonymisierung

Faktische Anonymisierung entspricht der Teildefinition des bisherigen alten BDSG (§ 3 Abs. 6) zu Anonymisierung, nach der Einzelangaben „nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können“.

Absolute Anonymisierung

Von absoluter Anonymisierung spricht man, wenn es in jeder Hinsicht ausgeschlossen ist, dass Daten auf eine natürliche Person bezogen werden können.

Quelle: Eigene Darstellung in Anlehnung an Höhne (2010) und Metschke/Wellbrock (2002)

Zwar kennt das neue Datenschutzrecht auf europäischer und nationaler Ebene keine explizite Unterscheidung mehr zwischen *absoluter* und sogenannter *faktischer* Anonymität, allerdings wird im Erwägungsgrund 26 der DSGVO davon gesprochen, dass als *objektive Faktoren*, die „nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden [können], […] Faktoren, wie die Kosten der Identifizierung und der dafür

erforderliche Zeitaufwand, herangezogen werden“ sollten. Diese Bestimmung entspricht nun wiederum dem Ansatz der *faktischen* Anonymität. Da die Unterscheidung zwischen den beiden Ansätzen den Bereich der Archivierung und Nachnutzung von Daten auf Individualebene zu Forschungszwecken betrifft, erscheint uns eine Erläuterung hier wichtig.[10]

Es ist umstritten, ob Verantwortliche, also in unserem Fall die Forscher/innen, jedes mögliche Risiko auch bei der Verarbeitung nicht mehr personenbezogener Daten im Griff haben müssen. Offen ist etwa die Frage, ob jedes Zusatzwissen über die Befragten den Verantwortlichen zuzurechnen ist. Die Mehrheit der Autoren und Autorinnen folgt hier der sogenannten relativen Theorie. Diese geht davon aus, dass man lediglich das Wissen und die Fähigkeiten in Betracht ziehen muss, welche der verarbeitenden Stelle (also etwa dem Forschungsvorhaben) momentan zur Verfügung stehen (Boehme-Neßler 2016: 420; Marnau 2016: 429f.; Karg 2015: 525). Auch die DSGVO folgt dieser relativen Position. Prinzipiell ist das „Bestehen eines theoretischen Restrisikos der möglichen erneuten Individualisierung nach erfolgter Anonymisierung […] von Gesetzes wegen hinzunehmen.“ (Gola 2017 Art. 6: 197)

Bei einer Wiederherstellung des Personenbezugs, also einer möglichen Re-Identifizierung oder De-Anonymisierung, kann unseres Erachtens also nicht von einer *absoluten* Anonymität der hier betrachteten Forschungsdaten für die Nachnutzung ausgegangen werden. Diese Position wird auch von der Artikel-29-Datenschutzgruppe (2014) und dem Europäischen Gerichtshof geteilt (EUGH 2016, Urteil zum Personenbezug dynamischer IP-Adressen; Wójtowicz/Cebulla 2017: 189f.).

Bei der Anonymisierung sind Abwägungen zu treffen. Einerseits sollen Forscher/innen Daten vorhalten, um Replikationen und Sekundäranalysen zu ermöglichen, anderseits soll der Persönlichkeitsschutz der Befragten dadurch gestärkt werden, dass Daten nicht länger als benötigt gespeichert werden. Nicht mehr benötigte personenbezogene Merkmale, wie z.B. Adressangaben oder Telefonnummern, müssen gelöscht werden, sobald der Forschungszweck erreicht ist.

Als problematisch ist prinzipiell jedes Zusatzwissen über Untersuchungspersonen anzusehen, über welches Dritte verfügen könnten, die auf die Daten zugreifen. Nach der oben genannten *relativen Theorie* und den Ausführungen in der DSGVO müssen Forscher/innen dieses Zusatzwissen *nach allgemeinem Ermessen* (Erwägungsgrund 26 DSGVO) momentan überschauen. Zum Zusatzwissen zählen die bei Dritten eventuell vorhandene Teilnahmekenntnis und jede Möglichkeit, über eindeutige Angaben im Datensatz weitere Informationen den Daten hinzuzufügen. Bei dieser möglichen Hinzufügung ist u.a. der *Stand der Technik* zu beachten, der Dritten zur Verfügung steht, um ggf. unrechtmäßig eine De-Anonymisierung durchzuführen. Allerdings ist dieses Konzept dynamisch, sodass für die praktische Arbeit nur von anerkannten Regeln der Technik ausgegangen werden kann (Grundlach/Weidenhammer 2018).

Eine Anonymisierung wird u.a. durch Angaben aus Antworten zu offen gestellten Fragen, durch die Datenerhebung in speziellen Populationen (z.B. sogenannte Eliten) oder die Kombinierbarkeit mit anderen Datenquellen erschwert:

a) Offene Angaben in quantitativen Datensätzen und in Transkripten qualitativer Interviews sind immer mit dem Risiko verbunden, dass der Befragte sich zu seinen persönlichen oder sachlichen Verhältnissen äußert und so zu seiner möglichen Re-Identifikation beitragen kann. Daher müssen diese sehr intensiv geprüft werden und führen zu einem beträchtlichen Arbeitsaufwand bei den die Anonymisierung durchführenden Stellen.

10 Seit Ende Juli 2018 liegen alle überarbeiteten Landesdatenschutzgesetze vor. Acht von 16 Bundesländern haben ihre Gesetze um Definitionen von „Anonymität“ im Sinne „faktischer Anonymität“ ergänzt. Es sind dies Brandenburg, Hessen, Niedersachsen, Nordrhein-Westfalen, Sachsen, Sachsen-Anhalt, Schleswig-Holstein und Thüringen (Landesdatenschutzgesetze 2018).

b) Bei Datenerhebungen in speziellen Populationen sind bereits durch die Begrenzungen, die sich aus den Definitionen der Populationen ergeben, Anonymisierungsmaßnahmen notwendig. Ein Fall spezieller Populationen stellen Professoren/innen dar. In der Regel stehen deren Lebensläufe online zur Verfügung, inklusive Angaben z.B. zur Promotion oder zu persönlichen Lebensverhältnissen (verheiratet, Anzahl Kinder, Wohnort etc.). Enthält eine Studie etwa unter deutschen Archäologieprofessoren/innen Angaben zum Jahr, in denen diese promoviert wurden, kann eine einfache Suchmaschinen-Anfrage bereits zur Identifikation der einzelnen Personen führen.
c) Neben der Kombinierbarkeit von Merkmalen innerhalb eines Datensatzes kann die Verknüpfung von Daten einer Studie mit anderen Datenquellen zu einer Verschärfung der Problemlage führen. Dies kann selbst dann der Fall sein, wenn die eigentlichen Daten an sich ohne jegliches identifizierendes Material zu sein scheinen. Beispielsweise konnten Narayanan und Shmatikov (2008) anonymisierte Personen aus einem frei verfügbaren Datensatz des Online-Streaming-Dienstes Netflix mit hoher Wahrscheinlichkeit Nutzerprofilen des Filmbewertungsportals Internet Movie Database (IMDb) zuordnen. Hierdurch erhöhte sich das Risiko einer möglichen Re-Identifizierung. Das Beispiel zeigt, dass die Einhaltung des Datenschutzes nicht mehr von der erhebenden Stelle kontrolliert werden kann, wenn ein Weg gefunden wird, die publizierten Daten mit anderen, externen Daten zu verbinden. Auf konkrete Maßnahmen der Anonymisierung gehen wir in Abschnitt 4.3 ein.

Die Herausforderungen beim Umgang mit von Personen erhobenen Daten in der Forschung werden durch eine ganzheitliche Betrachtung der Situation gemildert. Ein wichtiger Teil dieser ganzheitlichen Betrachtung ist die Bewertung des möglichen Risikos für die Untersuchungspersonen. Hierbei werden die für eine Veröffentlichung der Daten notwendigen Maßnahmen bestimmt und von der *verarbeitenden Stelle* wird das Risiko einer möglichen Re-Identifizierung bewertet. Geht man von einer *faktischen* Anonymität und einem hinnehmbaren Restrisiko aus, sollte die Wahrscheinlichkeit des Eintretens einer Re-Identifizierung oder einer De-Anonymisierung vernachlässigbar sein. Für eine vollständige Abschätzung eignet sich z.B. der von Bieker et al. (2016) auf der Basis des Standard-Datenschutzmodells (vgl. SDM 2016) sowie der DSGVO erstellte Ablauf für eine Datenschutz-Folgenabschätzung (ausführlich auch Forum Privatheit 2016).

Beispiele für einen sicheren Umgang mit Forschungsdaten, der nicht allein auf die Anonymisierung abzielt, sind Datenerhebungsprogramme wie etwa das SOEP oder das Projekt CILS4EU (s. Schaukasten 4.6). Im Fall des Sozio-oekonomischen Panels haben weder das Deutsche Institut für Wirtschaftsforschung (DIW) noch das Team des SOEP direkten Zugriff auf die Kontaktdaten der Befragten. Diese liegen nur dem Erhebungsinstitut vor. Ferner gelten für den Zugriff auf die Daten besondere vertragliche Regelungen (Frick et al. 2010).

Schaukasten 4.6: Beispiel für Nutzungsbestimmungen ist die Studie Children of Immi-grants Longitudinal Survey in Four European Countries (CILS4EU)

- Der Zugang zur Vollversion (http://dx.doi.org/10.4232/cils4eu.5353.3.3.0) ist nur nach expliziter Zustimmung des Datengebers in die Zwecke der Sekundärnutzung und nach Vertragsabschluss an einem technisch besonders umfangreich gesicherten Gastarbeitsrechner (Secure Data Center) möglich.
- Der Zugang zu einer reduzierten Version (http://dx.doi.org/10.4232/cils4eu.5656.3.3.0) außerhalb von GESIS ist nur nach expliziter Zustimmung des Datengebers in die Zwecke der Sekundärnutzung und nach Vertragsabschluss möglich. Die Daten werden verschlüsselt zum Download freigeschaltet.
- Der CILS4EU-Vertrag stellt einige besondere Anforderungen an die Nachnutzenden:
  - Bezüglich der Anonymität beispielsweise: Alle Versuche einer Re-Identifizierung von Studienteilnehmer/innen sind zu unterlassen, keine Verknüpfung mit anderen Daten auf Individualebene, keine Darstellung von Einzelfällen (einzelne Studienteilnehmer/innen).
  - Bezüglich sonstiger Bedingungen beispielsweise: Bei allen Publikationen auf Grundlage von CILS4EU muss das Projekt referenziert werden, keine Weitergabe an Dritte, keine kommerzielle Nutzung.

Quelle: Eigene Darstellung

Zusammenfassend lässt sich sagen, dass in der Aufbereitungs- und Analysephase die geplanten und notwendigen Maßnahmen zum Datenschutz auch sachgerecht umgesetzt werden sollten. Personenbezogene Merkmale in den Daten sollten getrennt von den Analysemerkmalen gespeichert werden. Je nach Grad des Risikos für den Grundrechtsschutz der Untersuchungspersonen treten informationsverändernde Maßnahmen wie die Pseudonymisierung und Anonymisierung hinzu. Der Grundrechtsschutz kann zudem durch Mittel wie den Zugriffsschutz verstärkt werden. Alle genannten Maßnahmen dienen dem Grundrechtsschutz der Untersuchungspersonen bei der Veröffentlichung und Archivierung. Sie sollten vor der Datenerhebung geplant und ihnen sollte durch die Betroffenen über die informierte Einwilligung zugestimmt werden.

4.2.3 Veröffentlichungs- oder Archivierungsphase

> *„Welche Daten müssen gelöscht werden?“ (RatSWD 2016: 17)*

An dieser Stelle müssen wir zunächst eine Einschränkung machen. Wenn die Einverständniserklärung eine Nutzung nach Abschluss des Forschungsvorhabens ausschließt, dürfen Forschungsdaten nicht archiviert werden. Das gilt auch, wenn diese Daten nach rechtlichen Vorgaben pseudonymisiert oder anonymisiert wurden. Selbst anonyme Daten dürfen nicht verwendet werden, wenn die Einwilligung explizit eine Nutzung nach Projektende ausschließt. Über die Einverständniserklärung wird erstens der *Forschung* als Zweck der Datennutzung zugestimmt und zweitens Untersuchungspersonen eine Wahrung ihrer Grundrechte durch Schutz ihrer Daten (z.B. durch Nutzung ohne Personenbezug) zugesichert (siehe oben).

Archive für Forschungsdaten verfügen – neben ihrer Expertise bezüglich Anonymisierungsmaßnahmen – über eine Reihe weiterer Möglichkeiten, die zur Einhaltung der datenschutzrechtlichen Vorgaben beitragen. Dazu zählen u.a. technische Zugangsbeschränkungen, Verpflichtungen über Nutzungsverträge/-bedingungen und kontrollierte Gastarbeitsrechner. Forscher/innen sollten sich frühzeitig mit dem Archiv ihrer Wahl in Verbindung setzen und mit den Ansprechpartner/innen dort das notwendige Schutzniveau für die einzureichenden Daten diskutieren.

> *„Wie werden die Daten für die Nachnutzung zur Verfügung gestellt?“ (Ebd.)*

Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) empfiehlt, Daten selbst dann in Repositorien zu sichern, wenn keine Verfügbarmachung für Sekundärstudien vorgesehen ist, beispielsweise weil keine Dokumentationsmaterialien erstellt wurden. Durch diese Sicherung wird zumindest gewährleistet, dass die Daten datenschutzkonform sowie sicher und langfristig lesbar vorgehalten werden (RatSWD 2016: 6). Die Archivierung in einem Repositorium oder Archiv hat weitere Vorteile, wenn die eingereichten Daten mit anderen Forscher/innen geteilt werden.

> *„Welchen Beschränkungen unterliegt die Nachnutzung bzw. wer kann die Daten unter welchen Bedingungen nachnutzen?“ (Ebd.)*

Wenn Daten in einem Repositorium bzw. Archiv eingereicht werden sollen, ist mit den dort zuständigen Ansprechpartner/innen zu klären, unter welchen Nachnutzungsbedingungen und Zugangsbeschränkungen die Daten verfügbar sein sollen. Es muss etwa die Frage geklärt werden, wer über welchen Zugangsweg und unter welchen Einschränkungen die Daten nutzen dürfen soll. Einen Einblick in mögliche Bedingungen für die Nachnutzung von Daten aus einem Archiv bietet die Benutzungsordnung des GESIS Datenarchivs (2018).

4.3 Anonymisierungs- und weitere Schutzverfahren

Wie dargestellt, sollte bereits im Forschungsvorhaben mit Maßnahmen wie Datentrennung, Pseudonymisierung und Verschlüsselung gearbeitet werden. Ab wann von anonymisierten Daten gesprochen werden kann, ist im Einzelfall zu bestimmen (s. Abschnitt 4.2.2).

Im Folgenden werden nur einige der gängigsten Anonymisierungsmaßnahmen in den Sozialwissenschaften dargestellt. Einzig auf die Eigenschaft k-anonymity soll abschließend noch einmal gesondert eingegangen werden.

4.3.1 Verfahren der Pseudonymisierung und Anonymisierung

Bei der Pseudonymisierung oder Anonymisierung werden alle Merkmale verändert, entfernt oder gelöscht, die Daten auf eine natürliche Person beziehbar machen. Hierbei ist zwar weiterhin ein (reduzierter) Informationsgehalt für Analysen gegeben, Einzelperson können jedoch nicht mehr identifiziert werden. Eine Auflistung möglicher direkter und indirekter Identifikationsmerkmale findet sich in Schaukasten 4.7. Neben dem Löschen der Daten bzw. ganzer Variablen, das mehr oder weniger selbsterklärend ist, gibt es die Möglichkeit, Variablenwerte zu klassifizieren oder einzelne Werte zu recodieren.

Schaukasten 4.7: Direkte und indirekte Identifizierungsmerkmale

Direkte Identifizierungsmerkmale

a) Namen
b) Anschrift
c) Telefonnummer
d) Kfz-Kennzeichen *(mögliche Identifizierung Fahrzeughalter)*
e) Personalausweisnummer
f) Sozialversicherungsnummer
g) E-Mail-Adresse *(z.B. berufliche mit Vor- und/oder Nachnamen)*
h) feste IP-Adresse
i) ein-eindeutige Berufsbezeichnung (z.B. Präsidentin, Rektorin, Direktorin etc. in Kombination mit Name Arbeitgeber)

Indirekte Identifizierungsmerkmale

j) Offene Berufsangabe (ggf. Name Unternehmen oder eindeutige Berufsangabe)
k) Offene Angabe zu Schul- und Berufsbildung (ggf. Name der Bildungseinrichtung)
l) Karriereangaben im Lebensverlauf
m) Geburtsland
n) Staatsangehörigkeit
o) Muttersprache

Regionalangaben[11]

p) Postleitzahl
q) Kreiskennziffer
r) Gemeindekennziffer
s) Ortsnamen (auch Ortsnamen in der Dokumentation beachten)
t) Stadtteilnamen
u) Bundesland oder Regierungsbezirk in Kombination mit Gemeindegrößenklasse oder Boustedt-Regionen

Quelle: Kinder-Kurlanda/Watteler (2015: 19)

11 Hierzu zählen auch Geokoordinaten, die Kinder-Kurlanda und Watteler (2015) nicht erwähnen.

Die Anonymisierung von qualitativem Forschungsmaterial ist in den meisten Fällen (beispielsweise für Interviewtranskripte) deutlich aufwendiger und für Audio- und Videoaufnahmen zudem technisch anspruchsvoller. Dies sind entscheidende Gründe, warum qualitative Daten selten vollständig veröffentlicht und archiviert werden. Dass dies keine Ausschlussgründe sein müssen und es geeignete Anonymisierungsmaßnahmen auch für die qualitative Forschung gibt, zeigen u.a. Meyermann und Porzelt (2014).[12]

Strategie 1: Aggregieren einzelner Werte/Kategorien

Wenn einzelne Werte einer Variablen sehr selten vorkommen, insbesondere an den Rändern ihrer Verteilung, kann dies die Re-Identifizierung von Studienteilnehmer/innen ermöglichen. In diesem Fall ist es empfehlenswert, die problematischen Werte der Variable zu aggregieren (Ebel/Meyermann 2015: 7f). Beispielsweise können sehr kleine bzw. sehr große Werte in einer nach unten bzw. nach oben offenen Kategorie zusammengefasst werden (*Top*- bzw. *Bottom-Coding*).

Schaukasten 4.8: Aggregieren einzelner Werte

Informationen in Variable	*Ursprüngliche Kodierung*	*Mögliches Datenschutzproblem*	*Exemplarische Lösung*
Alter	offene Angaben	Seltene Werte an den Rändern der Verteilung (Extrema)	Je nach Stichprobe: Top-Coding und/oder Bottom-Coding
Einkommen	offene Angaben	Seltene Werte an den Rändern der Verteilung (Extrema)	Je nach Stichprobe: Top-Coding und/oder Bottom-Coding
Muttersprache	offene Angaben	Seltene Werte (Zusatzwissen/ Teilnahmekenntnis)	Alle seltenen Werte zu einer gemeinsamen Kategorie umcodiert

Quelle: Ebel/Meyermann (2015: 8)

Es kann darüber hinaus sinnvoll sein, insbesondere bei herausgehobenen Populationen (z.B. Eliten), auf eine Mindestbesetzung von Zellen (des zu anonymisierenden Merkmals) zu achten. Empfehlungen zur Mindestbesetzung können nicht generalisiert werden. In Schaukasten 4.8 werden einige beispielhafte Recodierungen gezeigt.

Strategie 2: Aggregieren aller Werte

Die gesamte Variable sollte recodiert werden, wenn mehrere Werte problematisch erscheinen, sofern dieser Aufwand geleistet werden kann und die recodierte Variable einen tatsächlichen Informationsgehalt besitzt (ebd.: 8). In Schaukasten 4.9 werden einige für sozialwissenschaftliche Umfragen typische Variablen recodiert.

12 Der Vollständigkeit halber sei hier auf die Maßnahmen im Bereich der statistical disclosure control verwiesen, welche vor allem im Bereich amtlicher Mikrodaten verwendet werden (vgl. die kurze Einführung in Bethlehem 2009: 342-358).

Schaukasten 4.9: Aggregieren aller Werte

Informationen in Variable	*Ursprüngliche Kodierung*	*Exemplarische Lösung*
Berufsangabe	offene Angaben	Kategorisierung der Angaben in (dreistellige) ISCO-Codes
Berufsangabe	vierstellige ISCO-Codes	Kürzen auf dreistellige Angaben (Löschen der letzten Ziffer)
Postleitzahl	Postleitzahl	Aggregieren in z.B. Bundesland oder BIK-Region

Quelle: Ebel/Meyermann (2015: 8)

Strategie 3: Löschen von Variablen

Als dritte Strategie sollte das Löschen ganzer Variablen aus dem Datensatz ins Auge gefasst werden (ebd.). Da der Informationsverlust hierbei am größten ist, sollte das Löschen nur durchgeführt werden, wenn das Aggregieren einzelner Werte zu aufwendig ist oder aus sonstigen Gründen – etwa wenn der Informationsgehalt der recodierten Variable den Aufwand nicht rechtfertigt – nicht geleistet werden kann oder soll (vgl. die Beispiele zu IP-Adressen und Regionalangaben in Schaukasten 4.10).

Schaukasten 4.10: Löschen von Variablen

Informationen	*Exemplarische Lösung*
kleinräumige Regionalangaben wie Postleitzahl, Kreiskennziffern, Orts- oder Stadtteilnamen,	Aggregieren in z.B. Bundesland oder BIK-Region u.U. zu aufwendig, daher die Variable löschen.
(fixe) IP-Adressen	Der Informationsgehalt lässt sich eventuell nicht erhalten beim Aggregieren der Werte, daher die Variable löschen.

Quelle: Ebel/Meyermann (2015: 9)

4.3.2 Exkurs: k-anonymity

Die Eigenschaft *k-anonymity* wurde von Sweeney (2002) eingeführt. Ein Datensatz besitzt diese Eigenschaft, wenn ein Individuum, bezogen auf jede Kombination einer Auswahl von Variablen – Quasi-Identifier genannt – nicht von k-1 anderen Individuen unterschieden werden kann (ebd: 8ff.). Quasi-Identifier sind diejenigen Variablen, die zum einen mit größerer Wahrscheinlichkeit in externen Datenquellen vorkommen könnten und zum anderen spezifisch genug sind, dass bestimmte Kombinationen aller Variablen des Sets nur von relativ wenigen Individuen im Datensatz geteilt werden.

k-anonymity dient dazu, dass sich keine Person eindeutig identifizieren lässt, wenn weitere, externe Informationen mit dem Datensatz verlinkt werden, da immer mindestens k-1 identische Personen vorhanden sind. Dies stellt im Zeitalter von Big Data eine attraktive ergänzende Vorsichtsmaßnahme zu traditionellen Anonymisierungsstrategien wie der Unterdrückung einzelner Werte (Höhne 2010: 25) dar, da die Existenz von verknüpfbaren externen Datenbeständen immer wahrscheinlicher wird (Sweeney 2002: 4).

Das Vorgehen bei dieser Strategie ist eine Erweiterung der Aggregation bei einzelnen Variablen. Im Falle der k-anonymity werden individuelle Variablenausprägungen durch für k Fälle identische Ausprägungen ersetzt. Welcher Wert für k gewählt werden sollte, muss in

Abhängigkeit des Einzelfalls entschieden werden und kann nur soweit verallgemeinert werden, dass k in jedem Fall mindestens 3 sein sollte.

Beispiel: In einer Elitenstudie wurden Angaben von Juniorprofessoren/innen an Universitäten zu einem Fragebogen sowie Informationen zur Arbeitsstelle erfasst. Die Primärforscher gehen davon aus, dass Geschlecht, Zeitpunkt des Stellenantritts und Bundesland Quasi-Identifier sind. Angestrebt wird k-anonymity, mit k=3. D.h., keine Person im Datensatz darf sich bezogen auf die Quasi-Identifier von weniger als 2 weiteren unterscheiden. In Schaukasten 4.11 sieht man, dass das Ziel der k-anonymity mit k=3 erreicht wird.

Schaukasten 4.11: Beispiel für k-anonymity mit k=3 (Von den sieben dargestellten Variablen sind Geschlecht, Stellenantritt und Bundesland Quasi-Identifier.)

Fall	*Geschlecht*	*Stellenantritt*	*Bundesland*	*V1*	*V2*	*V3*
1	weiblich	Sept. 2014	Niedersachsen	20	1	3600€
2	weiblich	Sept. 2014	Niedersachsen	20	1	2100€
3	weiblich	Sept. 2014	Niedersachsen	10	1	1800€
4	männlich	Dez. 2015	Bayern	30	1	n/a
5	männlich	Dez. 2015	Bayern	40	1	5000€
6	männlich	Dez. 2015	Bayern	20	1	1500€

Quelle: Eigene Darstellung

4.3.3 Public und Scientific Use Files

Unabhängig von der gewählten Anonymisierungsstrategie sollte auf eine konsistente und sorgfältige Arbeitsweise geachtet werden. Alle Anonymisierungsentscheidungen müssen in einer nachvollziehbaren Weise dokumentiert werden (beispielsweise im Codehandbuch). Sind die Verluste des Auswertungspotentials der Daten durch Anonymisierungsmaßnahmen sehr groß, kann und sollte überlegt werden, ob eine Aufteilung in einen öffentlich verfügbaren, absolut anonymisierten *Public Use File*, und einen nur an einem Gastarbeitsrechner des Archivs bearbeitbaren *Scientific Use File* sinnvoll scheint (vgl. die Datenzugangsbedingungen auf der Webseite der Statistischen Ämter des Bundes und der Länder 2018):

- Daten in *Public Use Files* (PUF) sind normalerweise so weit anonymisiert, dass sie zu öffentlichen Zwecken zugänglich sind (vgl. Wende 2004: 338). Im Falle der Amtlichen Statistik wird sogar von *absolut anonymisierten Mikrodaten* gesprochen.
- Daten in *Scientific Use Files* (SUF) sind so weit anonymisiert, dass sie zu Forschungszwecken verwendet werden dürfen. Sie bieten im Vergleich zu den On-Site-Zugangswegen ein geringeres Analysepotential, sind jedoch so konzipiert, dass sie sich für einen großen Teil der wissenschaftlichen Forschungsvorhaben eignen.

4.3.4 Zugangs- und Zugriffsbeschränkungen

Als ergänzende (nicht ersetzende) Maßnahme kann in der Phase der Veröffentlichung und Archivierung der Zugriff auf oder sogar der Zugang zu Daten beschränkt werden. Der Zugriff kann technisch nur registrierten Personen ermöglicht werden, die einer Datennutzungsvereinbarung zustimmen. Diese sollte umfassen, dass ausschließlich aggregierte Werte publiziert werden, die keinen Rückschluss auf einzelne Studienteilnehmer/innen ermöglichen. Des Weiteren beschränkt sie den Nutzerkreis in der Regel auf Wissenschaftler/innen. Außerdem

werden ein Verbot der Weitergabe der Daten an Dritte und ein Datum zur Löschung der Daten aufgenommen.

Eine Beschränkung des Zugangs erfolgt in manchen Instituten über eine On-Site-Nutzung der Daten. Das bedeutet, Wissenschaftler/innen können ausschließlich vor Ort an einem Gastarbeitsrechner mit den Daten arbeiten. Beispielsweise wird der Mikrozensus in einer anonymisierten Version zur On-Site-Nutzung zur Verfügung gestellt (vgl. Zugangsbedingungen der Forschungsdatenzentren auf www.forschungsdatenzentrum.de). Dieses Beispiel für einen Scientific Use File enthält noch sehr detaillierte Informationen. Ein ähnliches Angebot macht GESIS über das Secure Data Center. Auch in dieser Arbeitsumgebung können Gastwissenschaftler/innen Scientific Use Files mit allen gängigen Statistikprogrammen analysieren. Der Datenzugriff ist technisch so gesichert, dass es unmöglich ist, die Daten an andere Stellen zu übermitteln oder auf Datenträger zu kopieren. Der Output der statistischen Auswertungen wird in beiden Einrichtungen von Mitarbeiter/innen geprüft und, falls keine kritischen Informationen enthalten sind, freigegeben, wie in Kapitel 7.4.4 am Beispiel des Datenarchivs für Sozialwissenschaften der GESIS – Leibniz-Institut für Sozialwissenschaften näher ausgeführt.

On-Site-Zugang kann als Maßnahme an sich bereits als ein Schritt in Richtung faktischer Anonymisierung gewertet werden, da ein erheblicher Aufwand für die Anreise einberechnet und der Nutzerkreis kontrolliert sowie auf Wissenschaftler/innen begrenzt werden kann (Höhne 2010: 11).

4.3.5 Zwei Fallbeispiele zu Datenschutzkonzepten

Die folgenden Skizzierungen von Datenschutzkonzepten zweier deutscher Studien zeigen exemplarisch, wie unter verschiedenen Ausgangsbedingungen mit datenschutzrechtlichen Vorgaben umgegangen werden kann (weitere Beispiele finden sich bei Kinder-Kurlanda/Watteler 2015).

Beispiel 1: ALLBUS Datenschutzkonzept

Die seit 1980 alle zwei Jahre erhobene Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) ist eine repräsentative Querschnittsbefragung zu Einstellungen, Verhaltensweisen und zur Sozialstruktur der deutschen Wohnbevölkerung (FDZ ALLBUS 2017). Sie umfasst umfangreiche Informationen zu jedem Befragten, u.a. kleinteilige geographische Informationen sowie Berufsangaben (vierstellige ISCO-Codes).

ALLBUS-Studien werden umfangreiche regionale Kontextdaten zugespielt, u.a. das Bundesland der Befragung und die politische Gemeindegrößenklasse des Wohnorts sowie BIK-Regionen (zu BIK s. Behrens/Wiese 2013). Aus Datenschutzgründen können nicht alle Informationen frei verfügbar gemacht werden. Einige Informationen, beispielsweise die exakte Wohnortgröße, werden nicht oder nur recodiert angeboten. Andere kleinteilige geographische Angaben (Regierungsbezirk) werden nicht in den jedem Wissenschaftler bzw. jeder Wissenschaftlerin zur Verfügung stehenden Standarddatensätzen veröffentlicht. Sie sind ausschließlich bei begründetem Forschungsinteresse und innerhalb vertraglicher Nutzungsbedingungen im GESIS Secure Data Center nutzbar (GESIS 2017; Wasmer et al. 2014: 33).

Beispiel 2: Studierendenstudie MESARAS

Die zweite Studie ist MESARAS 2013: *Mobility, Expectations, Self-Assessment and Risk Attitude of Students*. Sie ist ein Beispiel dafür, dass Teilnahmekenntnis und Zusatzwissen Dritter, insbesondere in Verbindung mit kleinräumigen geographischen Angaben, beim Datenschutz mitbedacht werden müssen. Ihre Darstellung zeigt darüber hinaus, dass auch Einzelforscher/innen in der Lage sind, Datenbestände wirksam zu schützen.

Für diese Studie wurden 2013 an sieben Universitäten in Deutschland Studierende befragt. Ziel der Befragung war es, den Zusammenhang zwischen Mobilitätsentscheidungen von Studenten und Studentinnen und verschiedenen Determinanten, wie individuellen Einstellungen und Persönlichkeitsmerkmalen, zu untersuchen. Zu diesem Zweck wurden umfangreiche und kleinteilige geographische Angaben erfasst.

Bei kleinteiligen geographischen Angaben besteht die Gefahr, dass diese in Kombination mit anderen, ansonsten nicht-personenbeziehbaren Merkmalen das Risiko der De-Anonymisierung erhöhen. Im vorliegenden Fall kommt hinzu, dass davon ausgegangen werden musste, dass Student/innen sowie Lehrkräfte Kenntnis über die Teilnahme der Befragten hatten. Dadurch wären Personen in der Lage gewesen, auch in formal anonymisierten Daten nach ihren Bekannten und Kommilitonen/innen zu suchen. Wäre beispielsweise die Staatsangehörigkeit differenziert erfasst und käme nur eine Person afghanischer Herkunft in den Daten vor, könnten ihre Kommilitonen/innen und Lehrkräfte sie über diese Information im Datensatz erkennen und anschließend alle sonstigen Angaben auf sie beziehen.

Der Primärforscher hat die Daten zum Schutz der Betroffenen formal anonymisiert und anschließend in zwei Versionen eingeteilt. Ein *Public Use File* (Weisser 2016a) wurde inhaltlich stark verändert und faktisch anonymisiert. Es wurden vor allem geographische Angaben, Nennungen der Universitäten, Angaben zu Ausbildung, Auslandsaufenthalten und Studiengang entfernt oder kategorisiert. Das Analysepotential ist vermindert. Da allerdings weiterhin die Distanzen zwischen den anonymisierten Studien- und Wohnorten (geographische Distanzen in Kilometern zwischen den Schwerpunkten von Flächen) enthalten sind, ist eine Auswertung bezüglich des ursprünglichen Forschungszwecks weiterhin möglich.

Die zweite Version ist ein *Scientific Use File* (Weisser 2016b), der weitestgehend einem weniger stark anonymisierten Datensatz entspricht und nur über das GESIS Secure Data Center zugänglich ist. Auf diese Weise wird sichergestellt, dass der Scientific Use File nicht an Dritte weitergegeben wird und ausschließlich Wissenschaftler/innen Zugang erhalten. Es ist hier lediglich eine Ergänzung von Kontextdaten über administrative Einheiten, wie Bundesländer oder Kreise, sowie teilweise über Georeferenzen möglich, nicht jedoch eine Verknüpfung mit anderen Datenquellen auf Individualebene. D.h., es kann nicht versucht werden, über einen Abgleich mit einem anderen Datensatz identische Befragte zu finden und ggf. zu identifizieren. Nutzer/innen unterschreiben einen Vertrag und die Einhaltung der Nutzungsbedingungen wird von Mitarbeitern/innen vor Ort kontrolliert. Diese Version ist durch die Kombination von technischen und Anonymisierungsmaßnahmen ebenfalls faktisch anonym.

4.4 Häufige Fehler

Datenschutz und ethisches Forschen sind komplexe Themengebiete. Forscher/innen fühlen sich oftmals nicht ausreichend informiert in diesen Bereichen. Aus Unkenntnis oder Sorge heraus, den einzelnen Bestimmungen nicht gerecht zu werden, geschieht es in der Praxis immer wieder, dass Forscher/innen entweder noch striktere Anforderungen an den eigenen

Umgang mit den erhobenen Daten stellen, als dies vom Gesetzgeber verlangt wird, oder gleich gänzlich auf jede datenschutzrechtliche Planung verzichten.

Im Folgenden werden im Archivkontext häufig beobachtbare Probleme dargestellt. Es wird gezeigt, wie diese Situationen sowohl in Einklang mit datenschutzrechtlichen Bestimmungen als auch den Bedürfnissen der Forschungspraxis gebracht werden können.

4.4.1 Untätigkeit aus Unsicherheit und überzogene selbstauferlegte Einschränkungen

Aus Unsicherheit aufgrund der Komplexität der datenschutzrechtlichen Bestimmungen und/oder Unwissen über diese verzichten Forscher/innen teilweise unbegründet auf weitere Forschung mit und Archivierung von erhobenen Daten. (Anonyme) Daten werden dann beispielsweise unmittelbar nach Erreichung des Projektziels gelöscht, obwohl dies nicht durch die Gesetze oder die Einwilligungserklärung verlangt wird. Oder die Projektleiter/innen nehmen erst gar keinen Kontakt mit Experten in ihren Hochschulen oder in datenhaltenden Einrichtungen auf und verzichten auf die Vorteile des *Data Sharing* (vgl. Kapitel 7).

Insbesondere bei Studien mit kleinen Grundgesamtheiten, Elitenstudien und qualitativem Forschungsmaterial sind Forscher/innen schnell geneigt, die Möglichkeit des Data Sharing bzw. der Archivierung auszuschließen, um Studienteilnehmer/innen vermeintlich besonders gewissenhaft zu schützen. Stattdessen werden alle Daten nach Abschluss des eigenen Projekts gelöscht oder ‚verschwinden' auf persönlichen Datenträgern.

Der beste Umgang mit der Komplexität von datenschutzrechtlichen Bestimmungen ist:

a) Einschlägige sozialwissenschaftliche Ratgeber lesen, beispielsweise die Handreichung Datenschutz (RatSWD 2017) oder das hier vorliegende Kapitel,
b) die Erstellung eines Datenmanagementplans bei der Planung des Forschungsprojektes (s. Kapitel 3),
c) ggf. Absprachen bzgl. des Vorhabens mit dem Datenschutzbeauftragten treffen und die rechtzeitige Kontaktaufnahme mit einem Archiv.

4.4.2 Unzureichende Planung des Forschungsdatenmanagements

Das Einhalten des Datenschutzes spielt in jeder Phase des Forschungszyklus eine Rolle. Finden in der Design- und Erhebungsphase keine oder nur unzureichende Überlegungen zum Forschungsdatenmanagement und insbesondere zum Umgang mit personenbezogenen/-beziehbaren Daten statt, ergeben sich vielfältige Probleme in nachfolgenden Phasen.

Beispielsweise ist in vielen Projekten nicht vorab festgehalten worden, ob personenbeziehbare Daten tatsächlich erhoben werden müssen und ob sie zur Erreichung des Projektziels notwendig sind. Hier sollte das oben beschriebene Prinzip der Datenminimierung beachtet werden. Wenn Daten im Projekt selbst erhoben werden sollen, sind folgende Fragen vorausschauend zu klären: Wann können sie gelöscht oder anonymisiert/pseudonymisiert werden, ohne das Erreichen des Projektziels zu gefährden? Müssen Studienteilnehmer/innen unter Umständen zu einem späteren Zeitpunkt erneut kontaktiert werden (nächste Welle in einer Panelstudie, Nachfolgeprojekt etc.)?

Die vielleicht schwerwiegendste Konsequenz hieraus ist die dann erfolgende mangelhafte Ausgestaltung der Einwilligungserklärungen. Die Bestimmungen in Einwilligungserklärungen dürfen nicht übergangen werden und Einwilligungen können nur unter großen Schwierigkeiten nachträglich erneut eingeholt werden. Ist beispielsweise zugesagt worden, alle Kontaktdaten zu einem bestimmten Zeitpunkt zu löschen, besteht keine Möglichkeit, diese Daten länger aufzubewahren, außer die Befragten werden erneut um Erlaubnis gebeten.

Wenn Einwilligungserklärungen nicht spezifizieren, wie Erhebung und Verarbeitung von personenbezogenen Daten erfolgen sollen, ist anschließend unklar, unter welchen Bedingungen die Datennutzung nach Abschluss des Projekts (Nachfolgeprojekt, Replikationsstudie, Datenverwertung durch den/die Primärforscher/in in einem anderen Kontext) möglich ist. Erfreulicherweise ist das Erstellen von Datenmanagementplänen immer häufiger eine Voraussetzung für die Förderung von Projekten.

4.4.3 Ex-post-Anonymisierung

Ein ähnliches Problem entsteht, wenn die Notwendigkeit der Anonymisierung von Daten erst im Nachhinein bedacht wird (Ex-post-Anonymisierung). In der Praxis stellt sich vielleicht heraus, dass zum einen die detaillierten Angaben in der Originalvariable gar nicht erst zur Beantwortung der Forschungsfragen benötigt wurden und der Aufwand der (nachträglichen) Anonymisierung somit überflüssig war. Zum anderen erweist sich die nachträgliche Anonymisierung in bestimmten Fällen (z.B. Datensätze mit inhaltlich detaillierten Variablen) als besonders zeitraubend und schwierig.

Die nachträgliche Anonymisierung kann zu Fehlern in den Daten führen. Der/die Datenverarbeitende kann die Daten ungewollt manipulieren und beispielweise die Werte falsch recodieren. Idealerweise werden im Datenmanagementplan die Archivierung und das Data Sharing mitbedacht und alle Variablen anhand dieser Überlegungen so erhoben, dass entweder keine nachträgliche Anonymisierung notwendig ist oder diese möglichst einfach und fehlerfrei erfolgen kann.

4.4.4 Pseudonyme aus personenbezogenen Angaben zusammensetzen

Manchmal werden Pseudonyme aus Teilnehmer/innen-Informationen zusammengesetzt, insbesondere wenn sich Personen selbst in Studien zu einem späteren Zeitpunkt identifizieren sollen, z.B. in der Befragung der nächsten Welle einer Panelstudie. Häufig gibt es in solchen Konstellationen eine Anleitung für die Teilnehmer/innen, die diese bittet, sich das eigene Pseudonym beispielsweise aus den ersten zwei Buchstaben des väterlichen Vornamens, den ersten zwei Buchstaben des mütterlichen Vornamens und dem eigenen Geburtsjahr zusammenzusetzen. Manchmal werden auch eindeutige Identifikationsmittel aus anderen Kontexten genutzt, beispielsweise die Matrikelnummer von Studierenden. Diese Pseudonyme haben den Vorteil, dass sie einfach zu merken sind bzw. bei der nächsten Gelegenheit ohne Schwierigkeiten erneut konstruiert werden können.

Probleme bereiten jedoch seltene Kombinationen in Verbindung mit Teilnahmekenntnis sowie Personenbezug dieser Pseudonyme. Wurden beispielsweise Studierende gebeten, ein Pseudonym unter Verwendung der Vornamen ihrer Eltern zu bilden, dann lässt sich über einen ausgeprägten Vornamen ggf. auf einen Migrationshintergrund der/des Studierenden schließen. Alle Personen mit Teilnahmekenntnis, möglicherweise also Lehrkräfte, Kommiliton/innen und Eltern, die Zugriff auf die ausgefüllten Fragebögen und/oder Datensätze erhalten, könnten die/der Studierende re-identifizieren. Personen ohne Kenntnisse der Teilnehmer/innen hätten immerhin einen zusätzlichen Informationsgewinn (Migrationshintergrund).

Eine weitere Schwierigkeit ist, dass durch das Pseudonym die Bestimmbarkeit der Betroffenen verhindert werden soll (vgl. Art. 4 Abs. 5 DSGVO). Das gelingt nicht, wenn das Pseudonym oder ein Teil davon selbst aus personenbezogenen Daten generiert wurde. Daher eignen sich beispielsweise die Sozialversicherungs-, Telefon- und Matrikelnummer sowie die E-Mail-Adresse nicht als Pseudonym.

4.5 Zusammenfassung

Sozialwissenschaftler/innen dürfen für Forschungszwecke mit personenbezogenen Daten arbeiten. Erheben sie diese selbst, müssen Untersuchungspersonen in aller Regel informiert in eine Teilnahme am Forschungsvorhaben zustimmen. Wissenschaftler/innen haben im Laufe ihrer Vorhaben sowie bei der eventuellen Veröffentlichung oder Archivierung auf den Schutz der Untersuchungspersonen zu achten. Hierfür bieten sich Maßnahmen wie die Datentrennung, die Verschlüsselung von Datenbeständen, die Pseudonymisierung oder die Anonymisierung der Daten an. Ein Forschungsdatenmanagementplan, der die Risiken der erhobenen Daten sowie Maßnahmen für den Schutz der Personen umfasst, ist ein wertvoller Leitfaden u.a. für das technische Management hinter der Forschungsarbeit. Fachverbände, Ethikkommissionen und Datenarchive bzw. Forschungsdatenzentren unterstützen Forscher/innen beim datenschutzkonformen Management und eröffnen Möglichkeiten der Nachnutzung von Daten im Zusammenspiel mit informationsreduzierenden Verfahren wie der Anonymisierung.

Literaturverzeichnis

Artikel-29-Datenschutzgruppe (2014): Artikel-29-Datenschutzgruppe, Stellungnahme 5/2014 zu Anonymisierungstechniken, Brüssel (0829/14/DE; WP216). http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf [Zugriff: 11.06.2018].

BDSG-alt (1990): Bundesdatenschutzgesetz (BDSG) vom 20. Dezember 1990 (BGBl. I S. 2954), neugefasst durch Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), zuletzt geändert durch Gesetz vom 29.07.2009 (BGBl. I, S. 2254), durch Artikel 5 des Gesetzes vom 29.07.2009 (BGBl. I, S. 2355 [2384] und durch Gesetz vom 14.08.2009 (BGBl. I, S. 2814). Aktualisierte, nicht amtliche Fassung 11.06.2010. https://www.baden-wuerttemberg.datenschutz.de/wp-content/uploads/2013/02/BDSG.pdf [Zugriff: 07.06.2018].

BDSG-neu (2017): Bundesdatenschutzgesetz vom 30. Juni 2017 (BGBl. I S. 2097). http://www.gesetze-im-internet.de/bdsg_2018/BDSG.pdf [Zugriff: 17.10.2018].

Behrens, Kurt/Wiese, Kathrin (2013): Stadtregionen: Von Boustedt zu BIK, In: GESIS und Arbeitsgruppe Regionale Standards (Hrsg.): Regionale Standards: Ausgabe 2013, 2, vollst. überarb. u. erw. Aufl., Köln, S. 86-120. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-348207 [Zugriff: 06.06.2018].

Bender, Stefan/Jarmin, Ron/Kreuter, Frauke/Lane, Julia (2017): Privacy and confidentiality, In: Foster, Ian/Ghani, Rayid/Jarmin, Ron S./Kreuter, Frauke/Lane, Julia (Hrsg.): Big Data and Social Science. A Practical Guide to Methods and Tools. Boca Raton u.a.: CRC Press, S. 299-312.

Bethlehem, Jelke (2009): Applied Survey Methods. A Statistical Perspective. New Jersey: John Wiley & Sons.

Bieker, Felix/Hansen, Marit/Friedewald, Michael (2016): Die grundrechtskonforme Ausgestaltung der Datenschutz-Folgenabschätzung nach der neuen europäischen Datenschutz-Grundverordnung. In: Recht der Datenverarbeitung 2016, 4, S. 188-197.

Boehme-Neßler, Volker (2016): Das Ende der Anonymität. Wie Big Data das Datenschutzrecht verändert. In: Datenschutz und Datensicherheit – DuD 2016, 7, S. 419-423.

Buchner, Benedikt (2016): Grundsätze und Rechtmäßigkeit der Datenverarbeitung unter der DS-GVO. In: Datenschutz und Datensicherheit – DuD 2016, 3, S. 155–161.

Chalmers, Alan F. (2007): Wege der Wissenschaft. Einführung in die Wissenschaftstheorie, 6. Berlin u.a.: Springer.

Diekmann, Andreas (2007): Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt.

Ebel, Thomas/Meyermann, Alexia (2015): Hinweise zur Anonymisierung von quantitativen Daten. forschungsdaten bildung informiert 3. Frankfurt/Main: Deutsches Institut für Internationale Pädagogische Forschung. http://www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert [Zugriff: 04.06.2016].

EU-DSGVO (2016): Verordnung (EU) 2016/679 des europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32016R0679 [Zugriff: 06.06.2018].

EUGH (2016): Gerichtshof der Europäischen Union, Urteil in der Rechtssache C-582/14 Patrick Breyer/Bundesrepublik Deutschland. https://medien-internet-und-recht.de/pdf/VT-MIR-2016-Dok-033.pdf [Zugriff: 07.06.2018].

Frick, Joachim R./Goebel, Jan/Haas, Jan/Krause, Peter/Sieber, Ingo/Engelmann, Michaela (2010): Verfahren für den Datenschutz beim Zugang zu den SOEP-Daten innerhalb und außerhalb des DIW Berlin. https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.347090.de/soep_datenschutzverfahren.pdf [Zugriff: 21.11.2017].

Forum Privatheit (2016): White Paper. Datenschutz-Folgenabschätzung. Ein Werkzeug für einen besseren Datenschutz. https://www.forum-privatheit.de/forum-privatheit-de/publikationen-und-downloads/veroeffentlichungen-des-forums/themenpapiere-white-paper/Forum_Privatheit_White_Paper_Datenschutz-Folgenabschaetzung_2016.pdf [Zugriff: 29.03.2018].

GESIS - Leibniz-Institut für Sozialwissenschaften (2017): ALLBUS – Kumulation 1980–2012. Sensitive Regionaldaten (ZA5260). http://dx.doi.org/10.4232/1.13010.

Gola, Peter (2017): Datenschutz-Grundverordnung VO (EU) 2016/679. Kommentar. München: Beck.

Graumann, Sigrid (2006): Forschungsethik. In: Düwell, Marcus/Hübenthal, Christoph/Werner, Micha H. (Hrsg.): Handbuch Ethik 2. Stuttgart u.a.: J. B. Metzler, S. 253-258.

Grundlach, Rocco/Weidenhammer, Detlef (2018): Wer kennt den „Stand der Technik“? Umsetzungsempfehlungen für Energienetzbetreiber. In: Datenschutz und Datensicherheit – DuD 2018, 2, S. 106-110.

Hammer, Volker/Knopp, Michael (2015): Datenschutzinstrumente. Anonymisierung, Pseudonyme und Verschlüsselung. In: Datenschutz und Datensicherheit – DuD 2015, 8, S. 503-509.

Hansen, Ingmar/Himmelreicher, Ralf K./Mai, Dirk/Röder, Frank (2012): Entwicklung des Datenangebots und deren Nachfrage in neun Jahren Forschungsdatenzentrum der Rentenversicherung (2004 bis 2012). RatSWD Working Paper Series 212. http://nbn-resolving.de/urn:nbn:de:0168-ssoar-427657 [Zugriff: 06.06.2018].

Häder, Michael (2009): Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. RatSWD Working Paper Series 90. http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf [Zugriff: 11.06.2016].

Himmelreicher, Ralf/vom Berge, Philipp/Fitzenberger, Bernd/Günther, Roland/Müller, Dana (2017): Überlegungen zur Verknüpfung von Daten der Integrierten Erwerbsbiographien (IEB) und der Verdienststrukturerhebung (VSE), Berlin. RatSWD Working Paper Series 262. https://www.ratswd.de/dl/RatSWD_WP_262.pdf [Zugriff: 19.04.2018].

Höhne, Jörg (2010): Verfahren zur Anonymisierung von Einzeldaten. Statistik und Wissenschaft. Bd. 16. Statistisches Bundesamt. https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band16_AnonymisierungEinzeldaten_1030816109004.pdf?__blob=publicationFile [Zugriff: 07.06.2018].

Karg, Moritz (2015): Anonymität, Pseudonyme und Personenbezug revisited? In: Datenschutz und Datensicherheit – DuD 2015, 8, S. 520-526.

Kinder-Kurlanda, Katharina/Watteler, Oliver (2015): Hinweise zum Datenschutz. Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. GESIS Papers 2015/01. https://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_papers/GESIS-Papers_2015-01.pdf [Zugriff: 24.04.2018].

Lane, Julia/Stodden, Victoria/Bender, Stefan/Nissenbaum, Helen (2014): Privacy, Big Data, and the Public Good. Frameworks for Engagement, Cambridge: Cambridge University Press.

Marnau, Ninja (2016): Anonymisierung, Pseudonymisierung und Transparenz für Big Data. Technische Herausforderungen und Regelungen in der Datenschutz-Grundverordnung. In: Datenschutz und Datensicherheit – DuD 2016, 7, S. 428-433.

Metschke, Rainer/Wellbrock, Rita (2002): Datenschutz in Wissenschaft und Forschung. https://www.hu-berlin.de/de/datenschutz/einwilligung/datenschutz-in-wissenschaft-und-forschung [Zugriff: 07.06.2018].

Meyermann, Alexia/Porzelt, Maike (2014): Hinweise zur Anonymisierung von qualitativen Daten. Forschungsdaten Bildung informiert 1. Frankfurt/Main: Deutsches Institut für Internationale Pädagogische Forschung. http://www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert [Zugriff: 04.06.2016].

Narayanan, Arvind/Shmatikov, Vitaly (2008): Robust De-anonymization of Large Sparse Datasets. http://arxiv.org/abs/cs/0610105v2 [Zugriff: 22.06.2016].

Palandt, Otto/Bassenge, Peter (2008): Bürgerliches Gesetzbuch mit Einführungsgesetz (Auszug) und andere. München: Beck.

RatSWD – Rat für Sozial- und Wirtschaftsdaten (2017): Handreichung Datenschutz. Output Series 5. https://www.ratswd.de/dl/RatSWD_Output5_HandreichungDatenschutz.pdf [Zugriff: 24.07.2017].

RatSWD – Rat für Sozial- und Wirtschaftsdaten (2016): Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte. Output Series 3. http://www.ratswd.de/dl/RatSWD_Output3_Forschungsdatenmanagement.pdf [Zugriff: 24.07.2017].

Rogosch, Patricia M. (2013): Die Einwilligung im Datenschutzrecht. Baden-Baden: Nomos.

Schaar, Katrin (2016): Was hat die Wissenschaft beim Datenschutz künftig zu beachten? Allgemeine und spezifische Änderungen beim Datenschutz im Wissenschaftsbereich durch die neue Europäische Datenschutz-Grundverordnung. RatSWD Working Paper Series 257. http://www.ratswd.de/dl/RatSWD_WP_257.pdf [Zugriff: 31.06.2016].

Schaar, Katrin (2017): Anpassung von Einwilligungserklärungen für wissenschaftliche Forschungsprojekte. Die informierte Einwilligung nach der DS-GVO und den Ethikrichtlinien. In: Zeitschrift für Datenschutz 5, S. 213-220.

SDM – Standard-Datenschutzmodell (2016): Das Standard-Datenschutzmodell. Eine Methode zur Datenschutzberatung und -prüfung auf der Basis einheitlicher Gewährleistungsziele, V.1.0. Erprobungsfassung von der 92. Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder am 9. und 10. November 2016 in Kühlungsborn einstimmig zustimmend zur Kenntnis genommen (Enthaltung durch Freistaat Bayern). https://www.datenschutzzentrum.de/uploads/sdm/SDM-Methode_V1.0.pdf [Zugriff: 29.03.2018].

Starck, Christian (2010): Grundgesetz, Artikel 5, Absatz 3. In: v. Mangoldt, Hermann/Klein, Friedrich/Starck, Christian (Hrsg.): Kommentar zum Grundgesetz, München, S. 625-674.

Sweeney, Latanya (2002): k-anonymity: a model for protecting privacy. In: International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10 (5), S. 557-570.

Unger von, Hella/Simon, Dagmar (2016): Ethikkommissionen in den Sozialwissenschaften. Historische Entwicklungen und internationale Kontroversen. RatSWD Working Paper Series 253. https://www.ratswd.de/dl/RatSWD_WP_253.pdf [Zugriff: 02.05.2018].

Verbund Forschungsdaten Bildung & Rechtsanwälte Goebel & Scheller (2015): Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen. https://www.forschungsdaten-bildung.de/files/Einwilligung_Checkliste_20150508.pdf [Zugriff: 04.05.2018].

Wasmer, Martina/Blohm, Michael/Walter, Jessica/Scholz, Evi/Jutz, Regina (2014): Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 2014. GESIS Technical Reports 2017, 20, http://nbn-resolving.de/urn:nbn:de:0168-ssoar-53370-3 [Zugriff: 07.06.2018].

Watteler, Oliver (2010): Datenschutz und die Archivierung von Daten der qualitativen empirischen Sozialforschung, Kapitel 3.2. In: Medjedovic, Irena/ Witzel, Andreas (Hrsg.): Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte. Wiesbaden: Springer, S. 63-85.

Watteler, Oliver (2017): Recherche nach sozialwissenschaftlichen Forschungsdaten. In: Berninger, Ina/Botzen, Katrin/Kolle, Christian/Vogl, Dominikus/Watteler, Oliver (Hrsg.): Grundlagen sozialwissenschaftlichen Arbeitens. Eine anwendungsorientierte Einführung. Stuttgart: UTB, S. 127-155.

Weisser, Reinhard A. (2016a): MESARAS 2013: Mobility, Expectations, Self-Assessment and Risk Attitude of Students. Reduzierte Version. GESIS Datenarchiv, Köln. ZA6295 Datenfile Version 1.0.0. http://dx.doi.org/10.4232/1.12545.

Weisser, Reinhard A. (2016b): MESARAS 2013: Mobility, Expectations, Self-Assessment and Risk Attitude of Students. Vollversion. GESIS Datenarchiv, Köln. ZA6294 Datenfile Version 1.0.0. http://dx.doi.org/10.4232/1.12544.

Wende, Thomas (2004): Different Grades of Statistical Disclosure Control Correlated with German Statistics Law. In: DomingoFerrer, Josep/Torra, Vincenc (Hrsg.): Privacy in Statistical Databases. CASC Project International Workshop, PSD 2004. Berlin u.a.: Springer, S. 336-342.

Wirth, Heike (1992): Die faktische Anonymität von Mikrodaten. Ergebnisse und Konsequenzen eines Forschungsprojektes. In: ZUMA-Nachrichten 1992, 30, S. 7-44.

Wirth, Heike (2003): Szenarien für Angriffe auf wirtschaftsstatistische Einzeldaten – Ein Überblick. In: Ronning, Gerd/Gnoss, Roland (Hrsg.): Anonymisierung wirtschaftsstatistischer Einzeldaten. Forum der Bundesstatistik. Bd. 42. Wiesbaden: Statistisches Bundesamt. S. 11-24.

Wirth, Heike (2016): Analytical Potential Versus Data Protection – Finding the Optimal Balance. In: Wolf, Christof/Joye, Dominique/Smith, Tom E. C./Fu, Yang-chih (Hrsg.): The SAGE Handbook of Survey Methodology, Los Angeles u.a.: Sage, S. 488-501.

Wójtowicz, Monika/Cebulla, Manuel (2017): Anonymisierung nach der DSGVO. In: Privacy in Germany 05, S. 186-192.

Linkverzeichnis

Datenzugangsbedingungen auf der Webseite der Statistischen Ämter des Bundes und der Länder (2018): http://www.forschungsdatenzentrum.de/datenzugang.asp [Zugriff: 06.06.2018].

FDZ ALLBUS: https://www.gesis.org/institut/forschungsdatenzentren/fdz-allbus/ [Zugriff: 06.06.2018].

GESIS, Benutzungsordnung (2018): https://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/_bgordnung_bestellen/2018-05-25_Benutzungsordnung_GESIS_DAS.pdf [Zugriff: 06.06.2018].

Landesdatenschutzgesetze (2018): https://www.datenschutz-wiki.de/Landesdatenschutzgesetze [Zugriff: 17.10. 2018]

Secure Date Center: https://www.gesis.org/angebot/daten-analysieren/weitere-sekundaerdaten/secure-data-center-sdc/ [Zugriff: 06.06.2018].

Übersicht über die länderspezifischen Besonderheiten für Befragungen an Schulen: https://www.forschungsdaten-bildung.de/genehmigungen [Zugriff: 06.06.2018].