6

forschungsdaten bildung informiert

Alexia Meyermann und Maike Porzelt

Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung

Version 2 // August 2019

Impressum

forschungsdaten bildung informiert // Nr. 6

Herausgeber
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Verbund Forschungsdaten Bildung
Rostocker Straße 6
60323 Frankfurt am Main

verbund@forschungsdaten-bildung.de

Redaktion und Layout
Alexander Schuster

Zitationsvorschlag: Meyermann, Alexia und Maike Porzelt. 2019. Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. forschungsdaten bildung informiert 6, Version 2.

www.forschungsdaten-bildung.de

1 Einleitung[1]

Datenschutzrechtliche Anforderungen beeinflussen die Forschungspraxis in vieler Hinsicht und in unterschiedlichen Phasen eines Forschungsprojektes. Die Einhaltung dieser Anforderungen im Forschungsprozess ist eine zentrale Voraussetzung dafür, die erhobenen Forschungsdaten im gewünschten Sinne nutzen zu dürfen. Das betrifft sowohl die Auswertung der Daten, die Publikation der auf diesen Daten beruhenden Ergebnisse als auch deren Archivierung und Bereitstellung bei einem Forschungsdatenzentrum, was den Forschenden zunehmend von Forschungsförderern zur Auflage gemacht wird[2]. Vor diesem Hintergrund befasst sich der vorliegende Text mit den Problemen und Herausforderungen, die sich für Forschungsprojekte im Zusammenhang mit datenschutzrechtlichen Anforderungen und der Nachnutzung von Forschungsdaten ergeben. Die Handreichung soll Forschende im Bereich der empirischen Bildungsforschung bei der Berücksichtigung der datenschutzrechtlichen Anforderungen in ihrer Forschungstätigkeit unterstützen.

Das vorliegende Papier beruht auf Erfahrungen aus dem Projekt „Sicherung und Nachnutzung von Forschungsdaten aus dem BMBF-Rahmenprogramm zur Förderung der empirischen Bildungsforschung“ (Laufzeit 10/2014 - 09/2016). Bei diesem Projekt handelte es sich um ein Kooperationsprojekt der Datenzentren der drei Institute DIPF, GESIS und IQB, die sich für diese Aufgabe zum „Verbund Forschungsdaten Bildung“ zusammengeschlossen haben (vgl. Meyermann et al. 2017). Ab der zweiten Förderphase (10/2016 – 09/2019) werden die Angebote des Verbund Forschungsdaten Bildung (VerbundFDB) in Kooperation mit weiteren Datenzentren ausgebaut: Gemeinsam gilt es, (1) eine zentrale Anlaufstelle für empirische Bildungsforscher/innen zu schaffen, die einen Einstieg in die verschiedenen, lokal archivierten Datenangebote der fachspezifischen Datenzentren bietet und (2) Lösungen für gegenwärtige, infrastrukturelle Herausforderungen zu finden. Daran arbeiten neben den Projektpartnern DIPF, GESIS und IQB derzeit u. a. die folgenden Datenzentren: ApaeK (Archiv für paedagogische Kasuistik, Goethe-Universität Frankfurt am Main), das FDZ am Deutschen Zentrum für Hochschul- und Wissenschaftsforschung (DZHW Hannover), das Archiv „Deutsches Gedächtnis“ der Fernuniversität Hagen, das FDZ am Leibniz-Institut für Bildungsverläufe (LIfBi Bamberg), das FDZ des Bundesinstitut für Berufsbildung (BiBB Bonn), das Datenzentrum des Deutschen Instituts für Erwachsenenbildung (DIE Bonn), das Datenzentrum des Mercator-Instituts für Sprachförderung und Deutsch als Fremdsprache (FD-Lex), das Datenzentrum Qualiservice an der Universität Bremen sowie das FDZ PsychData am Zentrum für Psychologische Information und Dokumentation (ZPID, Universität Trier). Auf www.forschungsdaten-bildung.de können Forschende eigene Forschungsdaten nachweisen, empirische Studien und Daten anderer finden sowie sich zu Themen des Forschungsdatenmanagements informieren.

---

[1] Der vorliegende Text beinhaltet Hinweise, die auf Erfahrungswerten beruhen und daher rechtlich nicht verbindlich sein können. Er beinhaltet keine juristischen Einzelfallprüfungen. Für die Unterstützung bei der Erstellung des Papiers und dessen Überarbeitung (2019) bedanken wir uns bei der Rechtsanwaltskanzlei Goebel & Scheller.

[2] Vgl. bspw. DFG-Vordruck 54.01 - 09/18, www.dfg.de/formulare/54_01/54_01_de.pdf (Zugegriffen: 19. Aug. 2018) und BMBF-Bekanntmachung vom 22. Feb. 2017, www.bmbf.de/foerderungen/bekanntmachung-1326.html (Zugegriffen: 19. Aug. 2018).

In Kapitel 2 werden die allgemeinen gesetzlichen Vorgaben zum Schutz personenbezogener Daten behandelt. Anschließend wird aufgezeigt, dass die Einhaltung des Datenschutzes auf verschiedenen Bausteinen beruht:

1) Der erste Baustein besteht daraus, datenschutzrechtliche Grundprinzipien bereits beim Erhebungsdesign zu berücksichtigen (Kapitel 3);
2) der zweite Baustein bezieht sich auf das Einholen informierter Einwilligungen (Kapitel 4),
3) der dritte Baustein auf die Anonymisierung von Forschungsdaten (Kapitel 5) und
4) der vierte Baustein besteht aus der sicheren Aufbewahrung dieser (Kapitel 6).

In Kapitel 7 wird zusätzlich auf Besonderheiten im Zusammenhang mit Erhebungen an Schulen eingegangen. Nach der abschließenden Zusammenfassung (Kapitel 8) findet sich am Ende dieser Handreichung eine Übersicht über die im Text genannten Bausteine zur Einhaltung datenschutzrechtlicher Vorgaben (Kapitel 9).

Hinweis zur Version 2 - Überarbeitung aufgrund des Inkrafttretens des neuen Datenschutzrechts ab dem 25. Mai 2019

Die in der ersten Fassung 2017 erschienene Handreichung wurde 2019 aufgrund des Inkrafttretens des neuen europäischen Datenschutzrechts aktualisiert. Die Datenschutz-Grundverordnung hat einige Änderungen mit sich gebracht, zentrale Grundsätze des Datenschutzrechts sind jedoch erhalten geblieben. Die im Text vorgenommenen Anpassungen beziehen sich daher überwiegend auf die zitierten Rechtsquellen. Die grundsätzlichen Empfehlungen zum Vorgehen bei empirischen Forschungsvorhaben bleiben unverändert.

Bitte beachten Sie: Die in Version 1.0 verwendeten Quellen beziehen sich auf das Datenschutzrecht alter Fassung.[3]

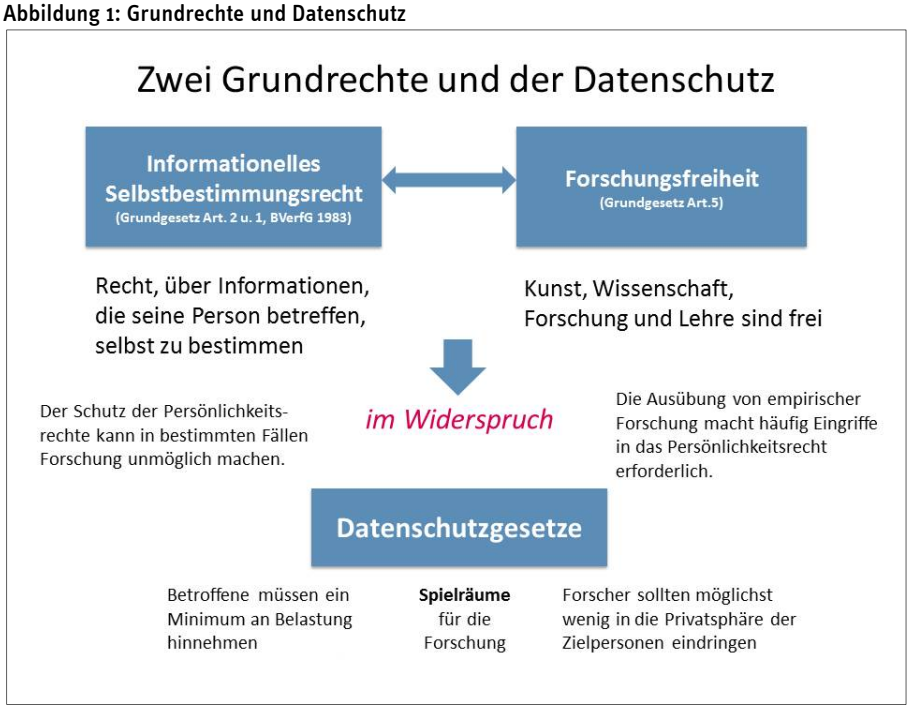
2 Gesetzliche Vorgaben zum Schutz personenbezogener Daten

Im Zusammenhang mit der Nutzung von personenbezogenen Daten für die Forschung sind zwei Grundrechte zu beachten. Das erste, hier zu erwähnende Grundrecht ist das Recht auf informationelle Selbstbestimmung, welches das Bundesverfassungsgericht als Ausformung des Allgemeinen Persönlichkeitsrechts aus den Grundrechten der Menschenwürde und der persönlichen Freiheit abgeleitet hat (Grundgesetz Art. 1 Abs. 1 und 2 Abs. 1, BVerfG im Volkszählungsurteil vom 15.12.1983; BVerfGE 65, 1). Dieses besagt, dass jeder Mensch das Recht hat, über die Verwendung von Informationen, die seine Person betreffen, selbst zu bestimmen. Zur Umsetzung dieses grundgesetzlich verankerten Rechts hat der Gesetzgeber Datenschutzgesetze erlassen. Der „Zweck des Datenschutzes ist

[3] Für den vertieften Einstieg in das Datenschutzrecht ab dem 25. Mai 2019 werden folgende Kommentare empfohlen: Ehmann, E., Selmayr, M., DS-GVO, Kommentar, 2. Auflage, München 2018; Paal. B., Pauly, D., Datenschutzgrundverordnung – Bundesdatenschutzgesetz, 2. Auflage, München 2018; Simitis, S., Hornung, G., Spiecker, I., Datenschutzrecht, DSGVO mit BDSG, 1. Auflage, Baden-Baden 2019.

es, den einzelnen davor zu schützen, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird“ (§ 1 Abs. 1 BDSG a. F.). Das zweite relevante Grundrecht schützt die Forschungsfreiheit. In Artikel 5 Abs. 3 des Grundgesetzes heißt es: „Kunst und Wissenschaft, Forschung und Lehre sind frei.“ Problematisch ist, dass beide Grundrechte in Widerspruch geraten können, wenn zur Ausübung von Forschung, Eingriffe in das Persönlichkeitsrecht erforderlich sind *oder* wenn aufgrund des Schutzes des Persönlichkeitsrechts Forschung unmöglich wird. Aus diesem Grund lassen Datenschutzgesetze Spielräume für die Forschung. Es wird auf der einen Seite erwartet, dass Betroffene ein Minimum an Belastung hinnehmen und auf anderen Seite, dass Forschende möglichst wenig in die Privatsphäre der Zielpersonen eindringen (vgl. Häder 2009; vgl. Abbildung 1).

Abbildung 1: Grundrechte und Datenschutz



Übergeordnete Grundlage für jedwede Verarbeitung personenbezogener Daten ist seit dem 25. Mai 2018 die europäische Datenschutzgrundverordnung (DS-GVO). Diese hat einige Änderungen mit sich gebracht, zentrale Grundsätze des Datenschutzrechts bleiben jedoch erhalten (vgl. RatSWD 2017, S. 12 oder Schaar 2017). In der Bundesrepublik ist der Datenschutz in verschiedenen Rechtsquellen geregelt. Das Bundesdatenschutzgesetz (BDSG) ist einschlägig für private Stellen sowie für öffentliche Stellen des Bundes und nur in begrenztem Rahmen für öffentliche Stellen der Länder. Die Landesdatenschutzgesetze (LDSG) sind einschlägig für öffentliche Stellen der Länder wie Hochschulen. Etwaige datenschutzrechtliche Vorschriften in Spezialgesetzen wie den Schulgesetzen der jeweiligen Länder, dem Bundesstatistikgesetz oder dem Sozialgesetzbuch sind jedoch vorrangig zu berücksichtigen. Die allgemeinen Datenschutzgesetze (BDSG, LDSG) kommen erst dann zur Anwendung, wenn die speziellen Gesetze datenschutzrechtliche Regelungslücken aufweisen.

Gegenstand des Datenschutzes sind personenbezogene Daten natürlicher Personen. Daten von Organisationen, d. h. juristischen Personen, unterliegen in Deutschland – anders beispielsweise als in Großbritannien – nicht dem Datenschutz. Als personenbezogen werden zum einen sogenannte *direkte Identifikatoren* bezeichnet, also Angaben wie Name, Adresse, Geburtsdatum oder Versicherungsnummer. „Personenbezogene Daten [sind] alle Informationen, die sich auf eine identifizierte oder indentifizierbare natürliche Person (im Folgenden: betroffene Person) beziehen“ (Art. 4 Nr. 1 DS-GVO). Auch

personenbeziehbare Daten müssen zum anderen daher beachtet werden. Als personen*beziehbar* gelten sog. *indirekte Identifikatoren*, das sind Einzelangaben, „die eine bestimmte Person zwar nicht eindeutig oder unmittelbar identifizieren, die es aber erlauben, die Identität der Person mit Hilfe anderer Informationen festzustellen" (Metschke und Wellbrock 2002[4], S. 19). Beispiele hierfür sind Vornamen, Ortsangaben, Straßennamen, Bundesländer, Institutions-/Organisationszugehörigkeiten (z. B. Arbeitgeber, Schule), Berufsangaben, Titel und Bildungsabschlüsse, Alter, Zeitangaben/kalendarische Daten, Bilder und Stimmen. Diese Merkmale *in Kombination* könnten eine eindeutige Identifizierung einer Person ermöglichen, z. B. wenn die Kombination aus den Merkmalen Beruf (z. B. Augenarzt oder Augenärztin) und Arbeitsort (z. B. Kleinststadt) einzigartig ist. Ob Merkmale tatsächlich personenbeziehbar sind, hängt von den Angaben ab, die vorliegen, und den sonstigen Informationen, die zugänglich sind. Anzumerken ist hier, dass sich im Zuge des technologischen Fortschritts und des Internets die Möglichkeiten der Identifizierungen erhöht haben oder zumindest schwieriger einzuschätzen sind (vgl. auch Schaar 2009, S. 4). Technische Kapazitäten sind größer, Suchalgorithmen effizienter, und es sind mehr Zusatzinformationen im Internet frei verfügbar, die eine Re-Identifizierung begünstigen.

Neben den genannten Angaben sind im Datenschutz die *besonderen Kategorien personenbezogener Daten* geschützt. Hierzu zählen Angaben über „die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen", die Gewerkschaftszugehörigkeit, genetische und biometrische Daten, Gesundheitsdaten, Daten zum Sexualleben oder zur sexuellen Orientierung (Art. 9 Abs. 1 DS-GVO). Für sie gelten strengere Regeln als für allgemeine personenbezogene Daten, wie bspw. deren explizite Erwähnung in der Einwilligungserklärung (vgl. S. 13).

Kernaussage des Datenschutzes ist ein sog. Verbot mit Erlaubnisvorbehalt: Das heißt, die Verarbeitung personenbezogener Daten ist *grundsätzlich verboten, es sei denn*, dies ist (a) durch ein Datenschutzgesetz oder durch andere Rechtsvorschriften erlaubt oder vorgeschrieben oder (b) die betreffende Person hat eingewilligt (DS-GVO Art. 6 und 9). Zur Verarbeitung personenbezogener Daten zählen u. a. ihre Erhebung, Verwendung und Speicherung (Art. 4 Nr. 2 DS-GVO). Aber auch wenn die gesetzliche Erlaubnis oder die Einwilligung der Betroffenen vorliegt, dürfen personenbezogene Daten nur unter bestimmten Voraussetzungen verarbeitet werden: Hierzu zählen etwa technische und organisatorische Maßnahmen, die eingehalten werden müssen, die schnellstmögliche Anonymisierung der Daten und die Beachtung der Grundsätze der Geeignetheit, Erforderlichkeit und Verhältnismäßigkeit bei der Datenverarbeitung. Diese werden in den nachfolgenden Kapiteln näher beschrieben.

Für die Forschung existiert ein besonderer Erlaubnistatbestand in Gestalt des *Forschungsprivilegs* (vgl. Metschke und Wellbrock 2002, S. 33ff.). Dieses erlaubt es unter bestimmten Voraussetzungen, personenbezogene Daten auch ohne Einwilligung zu verwenden, und schränkt damit das informationelle Selbstbestimmungsrechts zum Zweck der wissenschaftlichen Forschung ein. Im Einzelfall ist hierzu eine Interessensabwägung zwischen den Interessen des oder der Betroffenen und denjenigen der Forschung erforderlich (vgl. Kapitel 4.5).

[4] Der Text ist leider nicht mehr auf den Seiten der hessischen und Berliner Datenschutzbeauftragten abrufbar, aber online verfügbar unter: www.forschungsdaten-bildung.de/files/MetschkeWellbrock2002.pdf (Zugegriffen: 08. Aug. 2019).

Zusammenfassend ist festzuhalten, dass die Forschung mit anonymen, nicht personenbezogenen Daten aus datenschutzrechtlicher Sicht unproblematisch ist, da diese Art der Daten nicht Gegenstand des Datenschutzgesetzes ist. „Die grundrechtlich geschützte Entscheidungsbefugnis über seine Daten endet dort, wo verarbeitete Daten keinen Zusammenhang mehr mit seiner Person erkennen lassen“ (Metschke und Wellbrock 2002, S. 20). Forschung mit personenbezogenen Daten ist immer dann möglich, wenn eine Einwilligung der Betroffenen vorliegt (ausgeübtes informationelles Selbstbestimmungsrecht) oder das Datenschutzrecht (primär die DS-GVO) oder eine andere Rechtsvorschrift dies erlaubt.

3 Erster Baustein: Datenschutzrechtliche Prinzipien bei Forschungsdesign und Projektplanung berücksichtigen

Aus dem Datenschutzgesetz lassen sich – wie eingangs bereits erwähnt – vier zentrale Bausteine herleiten, die von Forschenden im Forschungsprozess zu beachten sind, um die datenschutzrechtlichen Anforderungen einzuhalten:

1) Erster Baustein: Datenschutzrechtliche Prinzipien bei Forschungsdesign und Projektplanung berücksichtigen (Kapitel 3)
2) Zweiter Baustein: Informierte Einwilligung der Betroffenen einholen (Kapitel 4)
3) Dritter Baustein: Anonymisieren und Pseudonymisieren von Forschungsdaten (Kapitel 5)
4) Vierter Baustein: Zugang und Zugriff auf Forschungsdaten beschränken (Kapitel 6)

Der erste Baustein, um datenschutzrechtliche Anforderungen im Forschungsprozess einzuhalten, behandelt die Prinzipien des Datenschutzes. Diese gilt es, schon früh im Forschungsprozess, d. h. bei der Planung von Forschungsprojekten und der Ausgestaltung des Forschungsdesigns, zu beachten (vgl. im Folgenden Metschke und Wellbrock 2002, S. 10 ff.). Die datenschutzrechtlichen Grundsätze[5] können als Prüfkriterien bei der Beurteilung der Rechtmäßigkeit von Eingriffen in das Persönlichkeitsrecht herangezogen werden: Werden Eingriffe in das Persönlichkeitsrecht erforderlich, wie sie durch die Verarbeitung personenbezogener Daten entstehen, ist es notwendig, die Rechtmäßigkeit dieser Eingriffe anhand der Grundsätze zu prüfen.

» *Geeignetheitsgrundsatz*

Zu prüfen ist, ob die Verarbeitung personenbezogener Daten *geeignet* ist, den Forschungszweck zu erfüllen und das Forschungsziel zu erreichen. Ein Ergebnis dieser Prüfung könnte es sein, dass gewisse Datenarten (zum Beispiel Daten über Schüler/innen der Gymnasien A, B und C oder Daten über italienische, polnische und türkische Migranten) geeignet sind, die zu untersuchende Forschungsfrage (z. B. Benachteiligung von Personen mit Migrationshintergrund an Gymnasien) zu beantworten.

[5] Die datenschutzrechtlichen Prinzipien oder Grundsätze sind in den Datenschutzgesetzen nicht einheitlich in einer eigenen Regelung erfasst. Sie sind meist integraler Bestandteil sonstiger Regelungstatbestände und daher im Kontext dieser Tatbestände auszulegen. Diese Einzeldarstellungen und -auslegungen hier zu beschreiben, würde zu weit führen, daher werden hier nur ausgewählte Prinzipien herausgegriffen.

» *Erforderlichkeitsgrundsatz*

Zu prüfen ist weiter, ob die Verarbeitung der Daten *erforderlich* ist oder ob der Zweck der Datenverarbeitung auf andere Art und Weise erfüllt werden könnte. Sind beispielsweise die Schüler/innen Daten der Gymnasien A und B bzw. Daten über italienische und polnische Migranten ausreichend zur Beantwortung der Forschungsfrage, kann auf die Erhebung der Schüler/innen-Daten von Gymnasium C bzw. der Daten über türkische Migrant/innen verzichtet werden. Es kann aber erforderlich sein, Daten von Schüler/innen der Gymnasien A, B *und* C zu erheben, um eine größere Vergleichsbasis zu haben bzw. Daten über italienische, polnische *und* türkische Migrant/innen zu erheben, wenn Unterschiede in der Benachteiligung zwischen diesen Gruppen vermutet werden. Liegen Daten zur Sekundärnutzung vor, die zur Beantwortung der Forschungsfrage genutzt werden könnten (z. B. Schüler/innen-Daten vergleichbarer Gymnasien und vergleichbaren Inhalts), ist eine erneute Primärerhebung ebenfalls nicht erforderlich.

» *Wahl des mildesten Mittels*

Im Rahmen der Erforderlichkeitsprüfung ist auch zu untersuchen, ob das Forschungsvorhaben die mildesten Mittel der Datenverarbeitung wählt. Die Eingriffe in das Persönlichkeitsrecht der Betroffenen sollten gering sein im Hinblick auf die zu verarbeitende Datenmenge (*Grundsatz der Datenminimierung* Art. 5 c) DS-GVO, *Grundsatz der Datensparsamkeit*, § 3a BDSG a. F.), die Art der zu verarbeitenden Daten (personenbezogen, anonymisiert oder pseudonymisiert) und den Kreis der Personen, die Kenntnis der personenbezogenen Daten erhält.

» *Übermaßverbot*

Eine Art Steigerung des Erforderlichkeitsgrundsatzes stellt das Übermaßverbot bzw. das Prinzip der Verhältnismäßigkeit dar. Geprüft wird, ob die erforderliche Datenverarbeitung die Betroffenen übermäßig belastet. Hier wird eine Abwägung vorgenommen zwischen den Persönlichkeitsrechten der Betroffenen und dem Recht der Forschenden auf Forschungsfreiheit.

Die Grundsätze des Datenschutzes sind in sämtlichen Phasen des Forschungsprozesses und für sämtliche Aspekte des Forschungsdesigns zu bedenken. Zu fragen ist:

- An welcher Stelle im Forschungsprojekt werden personenbezogene Daten verarbeitet und in welchem Umfang?
- Wie kann deren Verarbeitung so geregelt werden, dass die Eingriffe für die Betroffenen möglichst gering sind?
- Könnte auf die Erhebung und Verarbeitung personenbezogener Daten verzichtet werden?
- Welcher Personenkreis hat zu welchen Zeitpunkten Zugriff auf die Daten?
- Müssen und können Einwilligungen zur Verarbeitung der Daten eingeholt werden?
- Können die Daten anonymisiert werden? Wie schnell und wie umfassend?
- Wie können die Daten geschützt aufbewahrt werden?

Den datenschutzrechtlichen Grundsätzen folgend sind Sekundäranalysen bestehender Datensätze Primärerhebungen vorzuziehen, außer der Forschungszweck kann ohne eine eigene Primärdatenerhebung nicht erreicht werden (vgl. auch Häder 2009, S. 9f.). Für eine solche Entscheidung sind die verfügbaren Datenbestände und die existierende Literatur zu sichten und zu bewerten. Für die Bildungsforschung kann hierzu das sich derzeit im Aufbau befindende Portal www.forschungsdaten-bildung.de genutzt werden, das eine zentrale Übersicht über Studien und Daten der empirischen Bildungsforschung und einen Einstieg zu weiterführenden Angeboten bietet. Ist eine Primärerhebung

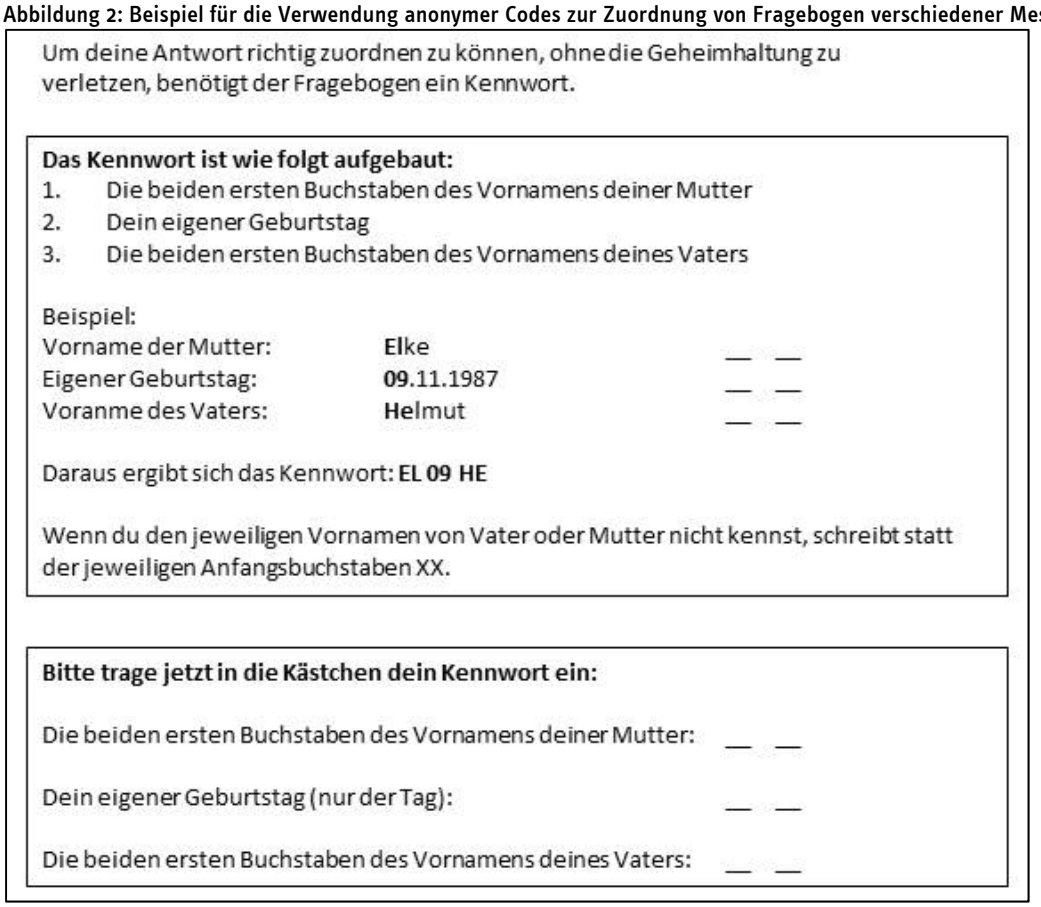
geplant, gilt weiter, „Forscher [haben] zu überlegen, welches methodische Design zur Überprüfung der anstehenden Forschungshypothesen am wenigsten in die Privatsphäre der Zielpersonen eindringt. Hier sollten mögliche Alternativen geprüft werden." (Häder 2009, S. 8)

Personenbezogene Daten liegen möglicherweise den Interviewer/innen vor, dem Umfrageinstitut, das mit der Datenerhebung beauftragt ist, dem Transkriptionsbüro, studentischen Hilfskräften und den Forschenden selbst. Möglicherweise lässt sich durch bestimmte Vorgehensweisen bei der Datenerhebung im Feld der Umgang mit personenbezogenen Daten reduzieren. Beispielsweise kann die Adressermittlung im Random Route Verfahren[6] durchgeführt werden und die Interviews können von verschiedenen Personen geführt werden (vgl. Diekmann 2003, S. 332f.). *Nicht-sprechende Identifikatoren*[7] auf den Erhebungsinstrumenten (z. B. Fragebögen) ermöglichen die Verknüpfung von Erhebungsdaten mehrerer Messzeitpunkte ohne Kenntnis des Namens der Beforschten, d. h. ohne Personenbezug. In der psychologischen Forschung werden häufig Identifikatoren, die Beforschte selbst erstellen, verwendet: Hierzu werden die Untersuchungspersonen gebeten, eine Kennung aus den Anfangsbuchstaben beispielsweise des Vornamens der Mutter und des Vornamens des Vaters zu erstellen. Die Identifikatoren sind in diesem Fall zwar sprechend, aber den Forschenden unbekannt. So können durch das Forschungsteam Erhebungsdaten unterschiedlicher Messzeitpunkte miteinander verknüpft werden ohne Kenntnis der Person (vgl. Abbildung 2; vgl. Metschke und Wellbrock 2002, S. 64).

[6] Das Random Route-Verfahren bezeichnet ein Auswahlverfahren, durch das Personen zufällig für eine Erhebung ausgewählt werden können, ohne dass deren Namen oder Adressen vorab zur Stichprobenziehung bekannt sein müssen.

[7] Beispiele für sprechende Identifikatoren: „AM1974" für Andrea Müller, geb. 1974; „20190701" für das Datum des Interviewzeitpunkts; „GymNRW07" für das siebte, befragte Gymnasium in NRW. Nicht-sprechende Identifikatoren sind bspw. zufällig verteilte Ziffern- oder Buchstabenfolgen ohne Bedeutung.

Abbildung 2: Beispiel für die Verwendung anonymer Codes zur Zuordnung von Fragebogen verschiedener Messzeitpunkte



Um deine Antwort richtig zuordnen zu können, ohne die Geheimhaltung zu verletzen, benötigt der Fragebogen ein Kennwort.

Das Kennwort ist wie folgt aufgebaut:

1. Die beiden ersten Buchstaben des Vornamens deiner Mutter
2. Dein eigener Geburtstag
3. Die beiden ersten Buchstaben des Vornamens deines Vaters

Beispiel:

Vorname der Mutter:	**El**ke	__ __
Eigener Geburtstag:	**09**.11.1987	__ __
Voranme des Vaters:	**He**lmut	__ __

Daraus ergibt sich das Kennwort: **EL 09 HE**

Wenn du den jeweiligen Vornamen von Vater oder Mutter nicht kennst, schreibt statt der jeweiligen Anfangsbuchstaben XX.

Bitte trage jetzt in die Kästchen dein Kennwort ein:

Die beiden ersten Buchstaben des Vornamens deiner Mutter: __ __

Dein eigener Geburtstag (nur der Tag): __ __

Die beiden ersten Buchstaben des Vornamens deines Vaters: __ __

Quelle: Metschke und Wellbrock 2002, S. 64

Sollen Kontaktdaten über die Zeit aufbewahrt werden – beispielsweise für eine erneute Kontaktierung der Beforschten bei Rückfragen, für die Interviewerkontrolle oder für Wiederholungsbefragungen – kann deren Aufbewahrung bei einem sogenannten *Datentreuhänder* erfolgen. Der Datentreuhänder ist eine von den Forschenden unabhängige Person oder Stelle, z. B. eine Notarin oder der betriebliche Datenschutzbeauftragte (vgl. weiterführend z. B. Metschke und Wellbrock 2002, S. 41 ff.; vgl. ADM 2012). Die Aufgabe des Datentreuhänders kann darin bestehen, Adressdaten, Zuordnungsschlüssel (auch Umsteigecodes oder Entblindungsliste genannt) oder auch die mit Namen versehenen Einwilligungserklärungen aufzubewahren. Die Forschenden haben in diesem Fall selbst keinen Zugriff auf die personenbezogenen Daten und benötigen diesen auch nicht. Die Rolle des Datentreuhänders ist im Gesetz selbst nicht definiert, dient aber in der Praxis dazu, die Eingriffe in die Persönlichkeitsrechte der Betroffenen zu minimieren. Zentral ist hier die räumliche und personelle Trennung der personenbezogenen von den übrigen Daten.

Auch bei der Erstellung der Erhebungsinstrumente ist darauf zu achten, ob Interviewfragen oder Beobachtungsinhalte in der geplanten Tiefe und Detailliertheit erforderlich sind oder reduziert werden könnten.

Die voranstehenden Ausführungen zeigen, dass datenschutzrechtliche Erwägungen im gesamten Forschungsprozess eine Rolle spielen und daher bei der Projektplanung und beim Forschungsdesign

berücksichtigt werden sollten. Eventuell sind zusätzliche finanzielle, personelle oder technische Ressourcen einzuplanen. Eine systematische Forschungsdatenmanagementplanung[8] kann darin unterstützen, die rechtlichen Aspekte frühzeitig und fortlaufend zu berücksichtigen. Forschungsprojekte oder die mit der Durchführung der Erhebung beauftragten Institute sind gefordert, die entsprechenden arbeits- und ablauforganisatorischen Prozesse zur Einhaltung des Datenschutzes zu gestalten und die erforderlichen Ressourcen hierfür bereitzustellen.

Zusammenfassend gilt: Erhebung, Verarbeitung und Nutzung personenbezogener Daten sollten nur erfolgen, wenn diese nicht vermeidbar, d. h. zur Erfüllung des Forschungszwecks unbedingt *erforderlich* sind. Die Sekundärnutzung von Daten ist demnach einer Primärerhebung vorzuziehen. Ist eine Primärerhebung personenbezogener Daten erforderlich, sollte der Grundrechtseingriff *so milde wie möglich* sein, das heißt, es sollten so wenige personenbezogene Daten erhoben werden wie möglich, und diese sollten zugriffsgeschützt aufbewahrt werden.

[8] Mit dem Begriff Forschungsdatenmanagement wird der planvolle, gesteuerte Umgang mit Forschungsdaten während des gesamten Forschungsprozesses und im Hinblick auf aktuelle und spätere Nutzungszwecke der Daten verstanden. Zum Forschungsdatenmanagement gehören neben der Einhaltung rechtlicher und ethischer Anforderungen die Aufgaben der Datendokumentation und der Datensicherung (für weiterführende Informationen siehe: Verbund Forschungsdaten Bildung 2017).

4 Zweiter Baustein: Einwilligungen einholen

Ein weiterer Baustein zur Einhaltung datenschutzrechtlicher Anforderungen besteht darin, die informierte Einwilligung der Betroffenen einzuholen.[9] Grundsätzlich gilt, dass zur Verarbeitung personenbezogener Daten das Einverständnis der betroffenen Personen einzuholen ist. Dabei muss es sich um ein sogenanntes *informiertes Einverständnis* handeln. Informiert bedeutet:

- die Betroffenen sind ausreichend über die Art der Daten, die erhoben werden, ebenso wie die Zwecke, zu denen die Daten genutzt werden, zu informieren,
- die Betroffenen sollten die Tragweite ihrer Entscheidung beurteilen können und
- sie sollten frei entscheiden können (vgl. bspw. Sheehan 2011, S. 227f.).

Im Folgenden werden zunächst die formalen Kriterien, die bei der Erstellung der informierten Einwilligung zu beachten sind, vorgestellt und anschließend zentrale Anforderungen an diese erläutert. Um die Rechtmäßigkeit von Einverständniserklärungen und von Datenverwendungen auf der Grundlage von dieser beurteilen zu können, bedarf es stets einer Einzelfallprüfung. Die nachstehenden Ausführungen sollen daher auf allgemeine Probleme aufmerksam machen.

Genaue Vorgaben (Bestandteile, Formalia) zum Einholen informierter Einwilligungen regelt das jeweils gültige Gesetz, also entweder eine bereichsspezifische Vorschrift oder generell die DS-GVO oder im Einzelfall auch das jeweilige LDSG oder das BDSG (vgl. auch DS-GVO Art. 13). Je nach Forschungsprojekt kann eine unterschiedliche Rechtsquelle zu beachten sein, da der Datenschutz neben der DS-GVO in den Landesdatenschutzgesetzen, im Bundesdatenschutzgesetz oder in Spezialgesetzen wie dem Schulgesetz geregelt ist (vgl. Kapitel 2). Bei Projektverbünden, die über Landesgrenzen hinausgehen, oder Erhebungen außerhalb des Bundeslandes des Projektstandorts, ist zusätzlich zu klären, welches LDSG gültig ist. Das betreffende Gesetz regelt, *welche Inhalte* Einwilligungen aufweisen sollten sowie *in welcher Form, wann* und *bei wem* Einwilligungen einzuholen sind.

4.1 Formale Kriterien: Was, wie, wann, bei wem?

Was? – Bestandteile von Einwilligungserklärungen

Eine informierte Einverständniserklärung lässt sich in drei Bereiche unterteilen: 1) einen Informationsteil mit allgemeinen Angaben zum Projekt, 2) einen datenschutzrechtlichen Informationsteil sowie 3) den Text der Einwilligung selbst mit der Unterschrift der betroffenen Person (vgl. Verbund Forschungsdaten Bildung 2019). Aus dem Informationsteil sollte hervorgehen, welche Daten durch wen, zu welchen Zwecken und wie verarbeitet werden (vgl. Abbildung 3). Werden besondere Kategorien personenbezogener Daten verarbeitet (vgl. Kapitel 2), so ist hierauf ausdrücklich hinzuweisen; auch muss sich die Einwilligungserklärung selbst ausdrücklich auf diese besonderen Datenkategorien beziehen. Weiterhin sind explizit die Rechte des oder der Betroffenen zu benennen. Hierzu gehören insbesondere die Freiwilligkeit der Teilnahme, das Widerrufsrecht sowie die Folgenlosigkeit bei Verweigerung oder Widerruf.

[9] Das Einholen der informierten Einwilligung ist häufig auch Bestandteil bestehender ethischer Richtlinien, siehe bspw. BDP und DGPs 2016; DGS und BDS 2014; DGFE 2010.

Abbildung 3: Bestandteile von Einwilligungserklärungen

Zentrale Bestandteile von Einwilligungserklärungen

- Angabe des Forschungsvorhabens (Titel, Ziele, Forschungsfragen)
- Ablauf des Vorhabens (z. B. „Es werden Interviews geführt"; „Unterricht wird per Video aufgezeichnet")
- Nennung eines Ansprechpartners, Nennung des Verantwortlichen für die Datenerhebung und -verarbeitung
- Nennung des zuständigen Datenschutzbeauftragten
- Hinweise auf die Rechte des oder der Betroffenen (Freiwilligkeit, Widerrufsrecht, Folgenlosigkeit bei Verweigerung oder Widerruf)
- Angabe der Daten, die erhoben werden (z. B. Kompetenzen, Einstellungen, Verhalten, Adressdaten)
- Angabe Art der Datenerhebung (z. B. per Fragebogen, per Video)
- Angabe der geplanten Verarbeitung der Daten (z. B. Digitalisierung, Transkription, Kodierung, Anonymisierung)
- Angabe der Verwendungszwecke, d. h. Angabe der geplanten Datennutzungen (z. B. Auswertung, Veröffentlichung, Nutzung in der Lehre, Archivierung, Weitergabe der Daten für Sekundärnutzungen)

Es ist empfehlenswert, die Zwecke der Erhebung und der Verarbeitung der Daten im Informationsteil zu benennen. Das hat den Vorteil, dass der Einwilligungsteil kurz gehalten werden kann (vgl. weitergehend Verbund Forschungdaten Bildung 2019). In der Praxis kommt es vor, dass Zwecke ausschließlich im Einwilligungsteil genannt werden; dies ist nicht zu empfehlen.

In welcher Form?

Im Gegensatz zum früheren BDSG (§ 4a Abs. 1 BDSG a. F.) schreibt die DS-GVO für die Einwilligung keine Schriftform mehr vor. Die Schriftlichkeit hat Vorteile, da sie die Nachweisbarkeit erleichtert. Die Schriftform geht jedoch auch mit Nachteilen einher.

Das Einholen einer schriftlichen Einverständniserklärung geht mit dem Nachteil einher, hierfür stets Namen und ggf. Adressen der einwilligenden Personen erfassen und speichern zu müssen, um nachprüfbar zu sein und etwaigen Widerrufen nachkommen zu können. Nicht jeder Betroffene ist aber bereit, seinen Namen zu nennen. Gleichzeitig scheint aus Sicht der Befragungspersonen die Angabe ihres Namens für die Einverständniserklärung auf der einen der getätigten Anonymisierungszusage durch die Forschenden für die Befragung selbst auf der anderen Seite zu widersprechen.

Zudem sind Einwilligungserklärungen in Schriftform nicht für jede Art von Erhebung praktikabel und könnten erhöhte Ausfallwahrscheinlichkeiten mit sich bringen. Bei Online- und Telefonbefragungen, bei Befragungen von speziellen Personengruppen zu sensiblen Themen (bspw. abweichendes Verhalten) oder bei ad hoc geführten persönlichen Interviews sind schriftliche Einwilligungen schwer einsetzbar (vgl. auch Häder 2009, S. 14; RatSWD 2017). Bei Telefonbefragungen oder Audiomitschnitten persönlicher Interviews bietet es sich an, mündlich ausgesprochene Einwilligungen aufzuzeichnen. In der Surveypraxis wird häufig so vorgegangen, Einwilligungen mündlich einzuholen und zusätzlich Informationsblätter zum Datenschutz auszuhändigen (vgl. Kapitel 4.3).

Das Gesetz sieht explizit für die Forschung Ausnahmen vom sogenannten Schriftformerfordernis vor. So erlaubt etwa § 27 Abs. 1 BDSG die Verarbeitung besonderer Kategorien personenbezogener Daten unter bestimmten Umständen auch ohne Einwilligung für wissenschaftliche Forschungszwecke (siehe auch Kapitel 4.5). Auch die elektronische Form der Zustimmung kann im Einzelfall erlaubt sein.

Bei wem?

Die Einwilligung zur Verwendung der eigenen Daten ist eine höchstpersönliche Angelegenheit, die durch die betroffene Person selbst abzugeben ist. Einwilligungen sind nicht von der Volljährigkeit oder der Geschäftsfähigkeit einer Person abhängig, sondern von deren Einsichtsfähigkeit. Das bedeutet, dass die Betroffenen in der Lage sein müssen, die Tragweite der abgegebenen Einwilligung zu erkennen. Dies schwankt wiederum mit der geplanten Datenverarbeitung, zu der eingewilligt werden soll. Das BDSG alter Fassung machte vor diesem Hintergrund keine genauen Altersangaben. Das neue Datenschutzrecht ist hier konkreter: Lt. DS-GVO ist bei unter 16-Jährigen die Zustimmung der Eltern *zusätzlich oder anstelle* derjenigen des Kindes einzuholen (vgl. Art. 8 Abs. 2 DS-GVO). Nationale Bestimmungen dürfen eine niedrigere Altersgrenze bestimmen, die jedoch mindestens bei 13 Jahren liegen muss. Die allgemeinen, nationalen Gesetze (BDSG, LDSG) nennen – soweit uns bekannt – keine niedrigeren Altersgrenzen als die DS-GVO. Explizitere Vorgaben finden sich teilweise in Spezialgesetzen wie den Schulgesetzen[10]. Ist die Zustimmung der Eltern erforderlich, gilt dies im Allgemeinen für beide Erziehungsberechtigten, jedoch kann ein/e einzelne/r Erziehungsberechtigte/r die Zustimmung auch im Auftrag des anderen erklären.

Wann?

In der Regel sollte die Einwilligung vor der Datenerhebung (also vor dem Interview oder der Beobachtung) eingeholt werden. Allerdings kann es notwendig werden, Einwilligungen, die sich auf Verwendungszwecke beziehen, die über die ursprüngliche Erhebung hinausgehen, auch *nach* der Befragung oder der Beobachtung einzuholen. Solche Verwendungszwecke sind beispielsweise die Archivierung und Bereitstellung der Forschungsdaten nach Projektende oder die Aufbewahrung der Kontaktdaten der Betroffenen für Folgestudien.

Beim Verfassen einer Einwilligungserklärung ist auch zu entscheiden, ob zu unterschiedlichen Verwendungszwecken der Daten jeweils einzeln die Zustimmung eingeholt wird, oder ob die Zustimmung zur Verwendung der Daten insgesamt eingeholt wird. Beispiele möglicher Verwendungszwecke sind:

- Datenerhebung bzw. Teilnahme an der Studie,
- Datenerhebung bzw. Teilnahme an einzelnen Aspekten der Studie,
- Bestimmte Formen der Datenweitergabe (an Kooperationspartner),
- Verwendung von Daten im Rahmen wissenschaftlicher Tagungen,
- Verwendung von Daten im Rahmen der Lehre,
- Publikation von Forschungsergebnissen,
- Archivierung in der erhebenden Einrichtung bzw. am Projektstandort,
- Archivierung bei einem Forschungsdatenzentrum (FDZ) und Nachnutzung durch Dritte.

[10] So beispielsweise für Niedersachsen: „Die Einwilligung der Erziehungsberechtigten ist erforderlich, wenn minderjährige Schülerinnen und Schüler oder Schülerinnen und Schüler - altersunabhängig - nach ihren Eltern oder nach Verhältnissen in der Familie befragt werden sollen“ (RdErl. d. MK v. 1.1.2014 - 25b-81402 (SVBl. 1/2014 S. 4), geändert durch RdErl. v. 1.12.2015 (SVBl. 12/2015 S. 598) - VORIS 22410 -, verfügbar unter: www.nds-voris.de/jportal/?quelle=jlink&query=VVND-224100-MK-20140101-SF&psml=bsvorisprod.psml&max=true (Zugegriffen: 30. Mai 2019)).

Es besteht nun die Möglichkeit, die Zustimmung zu den unterschiedlichen Verwendungszwecken jeweils einzeln einzuholen oder gebündelt (vgl. Fallbeispiel 1).

Fallbeispiel 1: Angabe unterschiedlicher Verwendungszwecke im Einwilligungsteil

Simplifiziertes Beispiel für eine abgestufte Abfrage	Simplifiziertes Beispiel für eine gebündelte Abfrage
Ich bin einverstanden …mit der Teilnahme an der Studie und der Verwendung meiner personenbezogenen Daten. ☐ Ja / ☐ nein …mit der weiteren Nutzung meiner Daten – über dieses Projekt hinaus – für die Bildungsforschung allgemein. ☐ Ja / ☐ nein Datum, Unterschrift	Mit den voranstehend (im Informationsteil) gemachten Ausführungen zur Verwendung meiner personenbezogenen Daten bin ich einverstanden. Datum, Unterschrift

In bestimmten Fällen kann eine Abstufung der Einwilligungserklärung sinnvoll sein, so dass die betroffene Person die Möglichkeit hat, verschiedene Nutzungsszenarien für ihre/seine Daten einzeln zu beurteilen. Eine sehr differenzierte Angabe der Verwendungszwecke (bspw. mehr als zwei genannte Zwecke) erhöht jedoch die Komplexität und den Beantwortungsaufwand für die Betroffenen und lässt daher negative Effekte auf die Teilnahmebereitschaft vermuten. Zudem erhöht das differenzierte Einholen der Zustimmung den Verwaltungsaufwand für die Forschenden erheblich, da (entsprechend der jeweiligen Gruppen) unterschiedliche Versionen von Datensätze erstellt und gepflegt werden müssen.

4.2 Zentrale Anforderungen: Laienverständlichkeit, Freiwilligkeit, Zweckbindung

Zentrale Anforderungen an die inhaltlichen Bestandteile von Einwilligungserklärungen sind:

1) Der oder die Betroffene muss wissen und verstehen, worin er oder sie einwilligen soll (Verständlichkeit und Detailliertheit der Information).
2) Er oder sie muss dies freiwillig tun (Freiwilligkeit).
3) Daten dürfen nicht für andere Zwecke verwendet werden als die angegebenen (Zweckbindung). Die Zwecke sollten so eng wie möglich und so breit wie nötig formuliert sein.

Laienverständlichkeit

Bei der Erstellung von Einwilligungserklärungen ist auf eine laienverständliche, an den jeweiligen Empfängerhorizont angepasste Sprache zu achten. Dies gilt sowohl für die Verwendung einzelner Begriffe als auch für die Beschreibung der Arbeiten selbst, die mit den Daten durchgeführt werden. Auf Fachterminologie sollte möglichst verzichtet werden. Begriffe wie Rohdaten, Transkriptionen, Videographie könnten durch die Begriffe Daten, Mitschriften, Verschriftlichungen, Videoaufzeichnungen,

Filme ersetzt werden (vgl. Abbildung 4). Ausschlaggebend für die Bewertung der Verständlichkeit von Aussagen und Begriffen ist der jeweilige Empfängerhorizont, das heißt insbesondere Alter, Bildungsniveau, Berufszugehörigkeit.[11]

Abbildung 4: Verwendete Fachterminologie und alternative Formulierungen in Einverständniserklärungen

In Einverständniserklärungen verwendete Begriffe	Alternative Formulierungen bzw. Begriffserklärung
Rohdaten	Daten
Transkriptionen	Verschriftlichungen, Mitschriften
Videographie	Filme, Videoaufzeichnungen
Anonymisierung	Die Daten werden anonymisiert, so dass keine Rückschlüsse auf Sie als Person getroffen werden können / so dass Sie als Person nicht mehr erkennbar sind.
Pseudonymisiert	Die Daten werden pseudonymisiert, d. h. unter Verwendung eines Codes und ohne Angabe von Namen weiterverarbeitet.

Es sollte auf eine trennscharfe, eindeutige und konsistente Verwendung von Begriffen innerhalb des Textes geachtet werden. Dies gilt insbesondere für den Datenbegriff. In der Praxis werden Begriffe wie „Daten", „Erhebungsdaten", „anonymisierte Daten", „personenbezogene Daten", „Rohdaten", „Kontaktdaten", „Interviewdaten" verwendet. Zu beachten ist, dass sich datenschutzrechtliche Vorgaben nur auf die Verwendung *personenbezogener* Daten beziehen (vgl. Kapitel 2). Werden Aussagen auch zur Verwendung der anonymisierten Daten gemacht, können diese dennoch rechtlich bindend sein, da eine solche Zusage keine datenschutzrechtliche aber eine eventuelle vertragsrechtliche Verpflichtung begründet. Die folgenden Fallbeispiele verdeutlichen mögliche Folgen der Verwendung verschiedener Datenbegriffe (vgl. Fallbeispiel 2).

Fallbeispiel 2

Beispiele zur Verwendung des Datenbegriffs und dessen Folgen

» *Beispiel A:* „Die erhobenen Daten werden ausschließlich im Forschungsprojekt verwendet und nicht an Dritte weitergegeben. Nach Abschluss des Projektes werden die erhobenen Daten gelöscht." Diese Aussage bezieht sich auf die „erhobenen Daten". Für die Betroffenen ist unklar, ob es sich bei den erhobenen Daten um personenbezogene oder um nicht personenbezogene Daten handelt. Daher muss davon ausgegangen werden, dass sowohl personenbezogene als auch anonymisierte Daten gemeint sind.

» *Beispiel B:* „Die anonymisierten Daten werden ausschließlich im Forschungsprojekt verwendet und nicht an Dritte weitergegeben." Diese Aussage bezieht sich auf die „anonymisierten Daten". Sie übertrifft datenschutzrechtliche Anforderungen und schafft eine vertragsrechtliche Verpflichtung in Bezug auf die Verwendung der anoynmisierten Daten.

» *Beispiel C:* „Die personenbezogenen Daten (Kontaktdaten) werden ausschließlich im Forschungsprojekt

[11] Der Verbund Forschungsdaten Bildung stellt deshalb neben Formulierungsbeispielen für „informierte Einwilligungen" (www.forschungsdaten-bildung.de/files/fdbinfo_4.pdf) auch Formulierungsbeispiele für „informierte Einwilligungen" in leichter Sprache (www.forschungsdaten-bildung.de/files/fdbinfo_5.pdf) zur Verfügung.

verwendet und nicht an Dritte weitergegeben. Nach Abschluss des Projektes werden die personenbezogenen Daten gelöscht." Diese Aussage bezieht sich ausschließlich auf die personenbezogenen Daten. Das heißt, es ergeben sich hieraus keine Einschränkungen hinsichtlich der Verwendung anonymisierter Daten.

Beispielformulierungen

„In dem Projekt werden Unterrichtseinheiten per Video aufgezeichnet. Die Videoaufzeichnungen werden durch Forscher/innen ausgewertet und hierzu vorab verschriftlicht. Die Mitschriften werden anonymisiert, das heißt Namen und Orte werden entfernt oder verfremdet. (...) In den Ergebnissen der Forschung werden keine Namen oder Adressen verwendet oder sonstige Informationen, die Rückschlüsse auf Sie als Person liefern könnten."

„In unserer Studie werden Schüler/innen anhand eines Fragebogens zu ihren Einstellungen zur Schule befragt. Die Antworten zu den Fragen werden ohne Ihren Namen und ohne Ihre Adresse gespeichert und von Forscher/innen ausgewertet."

Freiwilligkeit

Betroffene müssen ihr Einverständnis freiwillig geben können. Hierauf ist zum einen explizit hinzuweisen, zum anderen sollten Vorgehensweisen oder Formulierungen in den Einverständniserklärungen vermieden werden, die die wahrgenommene Freiwilligkeit der Betroffenen einschränken könnten. Zustimmungen können nicht als freiwillig angesehen werden, wenn sie durch eine Gegenleistung (z. B. Incentives), die über eine angemessene Aufwandsentschädigung hinausgeht oder das Vermeiden negativer Konsequenzen motiviert sind (vgl. Fallbeispiel 3). Bei der Forschung zu Personen, die in Abhängigkeitsverhältnissen stehen, wie Schüler/innen gegenüber den Lehrer/innen oder auch Lehrer/innen gegenüber der Schulleitung, ist die Freiwilligkeit der Teilnahme besonders deutlich zu machen. Betroffene sollten sich nicht verpflichtet fühlen zuzustimmen.

Fallbeispiel 3

Beispiele zu Vorteilen bei Teilnahme, Nachteilen bei Nicht-Teilnahme

» *Beispiel A (Nachteil bei Nicht-Zustimmung):* „Wir würden uns freuen, wenn auch Sie Ihrem Kind die Chance geben würden, an unserer Studie teilzunehmen. Damit tragen Sie dazu bei, dass die Kinder mit Lese- und Rechtschreibschwächen in der Klasse Ihres Kindes besser betreut werden." Die Formulierung könnte Eltern unter Druck setzen, einer Teilnahme zuzustimmen, denn eine Nicht-Teilnahme/Nicht-Zustimmung wäre zum Schaden von Klassenkameraden und -kameradinnen des eigenen Kindes.

» *Beispiel B (unverhältnismäßige Gegenleistung):* „Für Ihre Teilnahme am einstündigen Interview bedanken wir uns bei Ihnen mit einer Aufwandsentschädigung in Höhe von 500 EUR." Beispiel für eine sehr hohe Gegenleistung.

» *Beispiel C (Nachteil bei Nicht-Zustimmung):* „Die Klasse Ihres Kindes nimmt am Projekt XY teil. Stimmen alle Eltern dieser Klasse zu, wird die Klasse zur Belohnung einen Ausflug in das Spieleparadies machen." Die Eltern fühlen sich ggf. zu einer Zustimmung gedrängt, da eine Ablehnung negative Auswirkungen für die gesamte Klasse hat.

Beispielformulierung

„Mir ist bewusst, dass meine Teilnahme am Vorhaben vollkommen freiwillig ist und ich bei einer Verweigerung meiner Einwilligung keinerlei Nachteile erleide, insbesondere nicht in schulischen Belangen."

Zweckbindung

In den Einverständniserklärungen müssen Aussagen zur Verwendung der Daten enthalten sein. Die Forschenden, die um Einwilligung ersuchen, sind an die dort genannten Zwecke gebunden. Die Zwecke sollten so eng wie möglich, aber so breit wie nötig angegeben werden. Eine enge Zweckbindung mit Angabe einer konkreten Forschungsfrage, eines bestimmten Personenkreises, eines bestimmten Zeitrahmens ist in der Forschungspraxis nicht immer möglich und sinnvoll. Dies liegt daran, dass häufig die Verwendungszwecke in diesem Konkretheitsgrad im Vorfeld einer Studie nicht bekannt oder fix sind; beispielsweise bei explorativen Forschungsdesigns. Auch sind nicht alle Rahmenbedingungen eines Projektes konstant. So können sich Art und Anzahl derjenigen Personen ändern, die mit den Daten arbeiten werden, oder es kann zu einem Standortwechsel kommen, wenn die Projektleitung einen Ruf an eine andere Universität annimmt. Auch ist die Angabe einer derart engen Zweckbindung nicht möglich, sollen die Forschungsdaten langfristig über Forschungsdatenzentren Dritten zur Nachnutzung verfügbar gemacht werden.

Für die Forschung ist es möglich, Einwilligungen auch zu weniger eng definierten Zwecken einzuholen (vgl. auch RatSWD 2017, S. 8; 25). Dies ist rechtlich möglich, da weit gefassten Einwilligungen „als Korrektiv das Recht der Einwilligenden zur Seite gestellt [wird], ihre Einwilligung (mit Wirkung für die Zukunft) jederzeit ohne Angabe von Gründen widerrufen zu können" (RatSWD 2017, S. 25). Damit wird den Interessen der Einwilligenden an der Ausübung ihres informationellen Selbstbestimmungsrechts genüge getan. Auch die neue Datenschutzgrundverordnung der EU (DS-GVO) sieht für die Wissenschaft die Möglichkeit eines sog. *Broad Consents* – der breitgefassten Zustimmung – vor (vgl. Schaar 2017, S. 7, DS-GVO Erwägungsgrund 33). Es ist auch im Sinne der datenschutzrechtlichen Grundprinzipien der Datensparsamkeit und Datenvermeidung, wenn Daten für weitere Forschungszwecke über den engen Projektkontext des datenerhebenden Ursprungsprojektes hinaus genutzt werden können und so Mehrfacherhebungen gleicher Daten vermieden werden können. Doch auch wenn das Einholen eines broad consents datenschutzrechtlich erlaubt sein sollte, ist zu fragen, ob es auch ethisch legitim ist. Für eine Diskussion verschiedener Positionen hierzu siehe Sheehan (2011). Wie weit ein solcher broad consent letztlich sein kann, wird auch von der zukünftigen Kontrollpraxis der Datenschutzbeauftragten bzw. Aufsichtsbehörden abhängen.

Im Folgenden finden sich verschiedene Beispielformulierungen zur engen und weiten Zweckbindung der Verarbeitung von Forschungsdaten.

Fallbeispiel 4

Beispiele zur Zweckbindung in Einverständniserklärungen

» *Enge Zweckbindung:* Angabe eines Personenkreises und eines konkreten Zwecks;
z. B.: „Daten werden nur im Rahmen des Projektes zur Untersuchung der Einflüsse von Lehrerhandeln auf … genutzt und nur durch die genannten Projektmitarbeiter/innen bearbeitet.";
Nachteil: Etwaige Zweckänderungen bedürfen einer erneuten Einwilligung.

» *Weite Zweckbindung:* Angabe eines allgemeinen Verwendungszwecks und Nennung eines (unbestimmten)

Personenkreises;
z. B. „Die Daten werden ausschließlich zu Zwecken der Bildungsforschung durch ausgewiesene Wissenschaftler/innen verwendet.“;
Vorteil: Eine Nachnutzung der Daten bspw. über Forschungsdatenzentren ist möglich. Nachteil: Unbekannte Effekte auf die Teilnahmebereitschaft.

» *Sehr weite bzw. fehlende Zweckbindung:* keine Angabe darüber, ob die Daten zu Forschungszwecken, zu Lehrzwecken oder zu kommerziellen Zwecken erhoben werden; keine Angabe eines Personenkreises;
z. B. „Die Daten werden weitergegeben und nachgenutzt.“;
Nachteil: Die Informiertheit der Einwilligung ist anzuzweifeln.

Beispielformulierungen broad consent

„Die Videoaufzeichnungen werden ausschließlich nicht-kommerziell zu Forschungszwecken im Bereich Bildung ausgewertet.“
„Nach Abschluss dieser Studie werden die Videos im Sinne der Richtlinien der Deutschen Forschungsgemeinschaft (DFG) zur guten wissenschaftlichen Praxis an ein professionelles Forschungsdatenzentrum übergeben, das deren sichere und zugriffsgeschützte Aufbewahrung gewährleistet. Dort stehen die Videos Forschenden zu weiteren Forschungszwecken zur Verfügung.“
„Mit der in der Informationsschrift beschriebenen Erhebung, Verarbeitung und Nutzung der Unterrichtsaufzeichnungen bin ich einverstanden und willige ein, diese allgemein zu Zwecken der Bildungsforschung zu nutzen. Mir ist bewusst, dass die Unterrichtsaufzeichnungen für viele verschiedene Forschungszwecke verwendet werden können und dass die Daten zu diesem Zweck unbefristet bzw. bis zu meinem Widerruf bei einem Forschungsdatenzentrum zugriffsgeschützt gespeichert werden. Über mein Widerrufsrecht wurde ich umfassend aufgeklärt.“

4.3 Die „richtige“ Einwilligungserklärung

Es existieren Muster-Einwilligungserklärungen und online frei verfügbare Einwilligungserklärungen bestehender Studien, die Forschende als Orientierungshilfen bei der Erstellung von Einverständniserklärungen für die eigene Erhebung hinzuziehen können (vgl. nachfolgend). Auch Datenschutzinformationsblätter großer Umfragestudien können als Hilfestellung verwendet werden. Zu bedenken ist, dass die Art und Weise und die Rechtskonformität einer Einwilligungserklärung stark vom jeweiligen Projektkontext abhängen. Bestehende Muster sind daher nicht universell gültig und sollten nicht ungeprüft übernommen werden. Auch ist zu beachten, dass die nachfolgend genannten Muster teilweise noch auf der Grundlage des Datenschutzrechts beruhen, wie es bis zum 25. Mai 2018 gültig war.

Eine Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen ist verfügbar unter: www.forschungsdaten-bildung.de/files/fdbinfo_1.pdf.

Beispiele und Vorlagen für das Einholen schriftlicher Einwilligungen:

- Formulierungsbeispiele für „informierte Einwilligungen“, Verbund Forschungsdaten Bildung, 2018a. Verfügbar unter: www.forschungsdaten-bildung.de/files/fdbinfo_4.pdf

- Formulierungsbeispiele für „informierte Einwilligungen“ in leichter Sprache, Verbund Forschungsdaten Bildung, 2018b. Verfügbar unter: www.forschungsdaten-bildung.de/files/fdbinfo_5.pdf
- Muster-Einwilligungserklärungen des Datenzentrums Qualiservice, Universität Bremen, 2019. www.qualiservice.org/de/datenschutz.html (Zugegriffen: 08. Aug. 2019).
- Muster-Einwilligungserklärung für Interviews von Thorsten Dresing und Thorsten Pehl, audiotranskription. www.audiotranskription.de/audiotranskription/upload/Einwilligungserklaerung-DSGVO.pdf, (Zugegriffen: 30. Mai 2019).
- Muster-Einwilligungserklärung zur Erhebung und Verarbeitung personenbezogener Interviewdaten. Arbeitsgruppe „Datenschutz und Qualitative Sozialforschung“ (RatSWD). In Liebig et al. 2014, www.ratswd.de/dl/RatSWD_WP_238.pdf (Zugegriffen: 30. Mai 2019)
- Muster-Einwilligungserklärung zur Übermittlung und Nutzung personenbezogener Daten für wissenschaftliche Zwecke nach Projektende. Arbeitsgruppe „Datenschutz und Qualitative Sozialforschung“ (RatSWD). In Liebig et al. 2014, www.ratswd.de/dl/RatSWD_WP_238.pdf (Zugegriffen: 30. Mai 2019).

Datenschutzinformationsblätter ohne Einholen einer schriftlichen Einwilligung:

- Arbeitskreis Deutscher Marktforschungsinstitute: Erklärung zum Datenschutz und zur absoluten Vertraulichkeit Ihrer Angaben bei mündlichen oder schriftlichen Interviews. Verfügbar unter: www.adm-ev.de/wp-content/uploads/2018/07/Erkl%C3%A4rung-zum-Datenschutz.pdf (Zugegriffen: 30. Mai 2019).
- PIACC, TNS Infratest, GESIS: Erklärung zum Datenschutz und zur absoluten Vertraulichkeit Ihrer Angaben bei persönlichen Interviews (2011). www.gesis.org/fileadmin/piaac/download/Datenschutzblatt_PIAAC_2011.pdf, (Zugegriffen: 30. Mai 2019).
- Studie „Private Haushalte und ihre Finanzen“, Bundesbank, infas: Erklärung zum Datenschutz und zur absoluten Vertraulichkeit Ihrer Angaben (2014). www.bundesbank.de/resource/blob/604920/5b4ae07cf2768c51a003d136bb0c44a7/mL/infas-datenschutzblatt-data.pdf (Zugegriffen: 30. Mai 2019).
- Studierendenbefragung „Gute Lehre“ der Universität Duisburg-Essen. Verfügbar unter: https://panel.uni-due.de/uc/qvm_s/images/PDF/190325_QVM-Studierende_Informationen_Datenschutz.pdf (Zugegriffen: 30. Mai 2019).

4.4 Arbeits- und ablauforganisatorische Implikationen

Die in Einwilligungserklärungen gemachten Zusagen beinhalten arbeits- und ablauforganisatorische Implikationen für jedes Forschungsprojekt. Forschende müssen die Aufbewahrung der Einwilligungserklärungen regeln, das Vorgehen bei Eingang eines Widerrufs, ggf. die Zusage zur Löschung nach bspw. zehn Jahren einhalten und Ähnliches. Dieses muss gegebenenfalls über den geförderten Projektzeitraum hinaus und personenunabhängig sichergestellt sein und stellt Forschungsprojekte mit dreijährigen Laufzeiten und der üblicherweise hohen Fluktuation wissenschaftlichen Personals an Universitäten vor logistische Herausforderungen.

Für den Zeitraum der Existenz personenbezogener Daten ist Folgendes zu gewährleisten: Verfügbarkeit von Ansprechpersonen, Gültigkeit von Kontaktadressen (des Projektes), die Aufbewahrung der

unterschriebenen Einverständniserklärungen, die Durchführung etwaiger Widerrufe, die sichere Aufbewahrung der personenbezogenen Daten und ggf. die Löschung dieser zu einem bestimmten Zeitpunkt.

In Einwilligungserklärungen sollten *Ansprechpersonen* genannt werden, die für den Zeitraum der Existenz der personenbezogenen Daten ansprechbar sind. Werden persönliche Ansprechpartner genannt, sollte daher deren Nachfolge im Fall eines Ausscheidens geklärt sein. Eine personenbezogene E-Mailadresse oder eine E-Mailadresse, die nur befristet während der Projektlaufzeit gültig ist, ist daher weniger geeignet. Gleiches gilt für die Angaben zum zuständigen internen Datenschutzbeauftragten.

Die *Aufbewahrung* von Einwilligungserklärungen ist für den Zeitraum zu gewährleisten, in dem die personenbezogenen Daten existieren und verwendet werden. Einwilligungserklärungen stellen den Legalitätsnachweis für die Nutzung der Daten dar. Zu klären ist, an welcher Stelle und durch wen die Aufbewahrung für den erforderlichen Zeitraum sichergestellt werden kann. In der Praxis der Schul- und Unterrichtsforschung werden Einwilligungserklärungen teilweise in der Schule verwahrt. Verbleiben Einwilligungserklärungen von Schüler/innen oder deren Erziehungsberechtigten an den Schulen, bedeutet dies eine zusätzliche Belastung für Schulen, die deren sichere Aufbewahrung und deren Erhalt bis zum Ende des Projektes bzw. bis zur Löschung der personenbezogenen Daten sicherstellen müssen. Eine Alternative hierzu ist die Aufbewahrung der digitalisierten Einwilligungserklärungen bei der die Daten archivierenden Einrichtung, bspw. einem Forschungsdatenzentrum. Ausschlaggebend ist es, zu gewährleisten, dass etwaige Widerrufe von Betroffenen nicht ins Leere laufen.

Widerrufe müssen so lange möglich sein, wie personenbezogene Daten existieren. Das Widerspruchsrecht erlischt, sobald die Daten keinen Personenbezug mehr aufweisen. In diesem Fall kann ein Widerspruch nicht mehr umgesetzt werden.

Beispielformulierung

„Auch können Sie jederzeit – solange die erhobenen Daten durch die Codierung noch personenbeziehbar sind – Ihre Einwilligung widerrufen und die Löschung der Daten Ihres Kindes von uns verlangen." (Metschke und Wellbrock 2002, S. 57)

Erfolgt ein Widerruf, ist die betreffende Person aus den Daten zu löschen. Ein Projekt muss festlegen, wer dies durchführt, tritt der Fall ein. Für die Löschung der betreffenden Person aus den Daten sind die verschiedenen Versionen der Forschungsdaten zu berücksichtigen, die möglicherweise existieren. Dazu gehören beispielsweise Rohdaten, Auswertungsdateien, Videos und deren Transkriptionen sowie Sicherungskopien, die an verschiedenen Orten gespeichert sind. Aus bereits erfolgten Publikationen, können und müssen die betroffenen Personen nicht gelöscht werden, da ein Widerruf stets nur mit Wirkung für die Zukunft gültig ist.

Löschung/Vernichtung von Daten: Aussagen zu Löschzeitpunkten sollten vorsichtig verwendet werden. Denn möglicherweise werden personenbezogene Daten auch noch nach Projektabschluss oder nach Abschluss der Befragung vorliegen und zu Forschungszwecken benötigt. Müssen die personenbezogenen Daten zu einem bestimmten Zeitpunkt gelöscht werden, ist durch das Projekt zu gewährleisten, dass die Löschung tatsächlich durchgeführt wird.

4.5 Sonderfall: Datenerhebung ohne Einwilligung? Erhebung auf Basis des Forschungsprivilegs

In bestimmten Erhebungssituationen kann das Einholen von informierten Einwilligungserklärungen sehr schwierig bis unmöglich sein: So z. B. in Feldforschungssituationen, in denen ein Proband/eine Probandin im Alltag begleitet wird. Hier kann es vorkommen, dass es nicht möglich ist, von allen mit dem Probanden interagierenden Personen ebenfalls Einwilligungserklärungen einzuholen, da dies die natürliche Interaktionssituation erheblich stören würde. Ebenfalls unmöglich ist es, Einwilligungserklärungen sämtlicher, betroffener Personen bei Beobachtungen auf Großveranstaltungen einzuholen. Hier macht die große Menge an Betroffenen das Einholen von Einwilligungen jedes einzelnen unmöglich.

Für Fälle wie diese sieht das Datenschutzrecht daher Ausnahmen vor: So kann die Verarbeitung personenbezogener Daten nicht nur auf der Grundlage einer Einwilligung, sondern auch auf der Grundlage eines entsprechenden Gesetzes oder einer anderen Rechtsvorschrift erlaubt sein (vgl. Kapitel 2). Für die Wissenschaft existiert ein solcher gesetzlicher Ausnahmetatbestand durch das sogenannte *Forschungsprivileg* (vgl. Metschke und Wellbrock 2002, S. 34). Dieses erlaubt es, unter bestimmten Voraussetzungen, personenbezogene Daten auch ohne Einwilligung zu verwenden, und schränkt damit das informationelle Selbstbestimmungsrecht zum Zweck der wissenschaftlichen Forschung ein. Daten dürfen auch ohne Einwilligung für Forschungszwecke verarbeitet werden,

> *„wenn die Verarbeitung zu diesem Zweck erforderlich ist und die Interessen des Verantwortlichen an der Verarbeitung die Interessen der betroffenen Person an einem Ausschluss der Verarbeitung erheblich überwiegen" (§ 27 Abs. 1 S. 1 BDSG).*

Der Verantwortliche muss dabei nach § 27 Abs. 1 S. 1 BDSG angemessene und spezifische Maßnahmen zur Wahrung der Interessen der betroffenen Person vorsehen. Die neuen Landesdatenschutzgesetze enthalten ähnliche Regelungen (siehe etwa § 24 Abs. 1 HDSIG).

Das Forschungsprivileg ist an enge Voraussetzungen geknüpft, deren Vorliegen im Einzelfall geprüft werden muss. Für den konkreten Einzelfall ist eine Interessensabwägung zwischen den Interessen der Betroffenen (den schutzwürdigen Belangen) und denjenigen der Forschung erforderlich. Beide Rechte sind - wie eingangs erwähnt (vgl. Kapitel 2) - grundgesetzlich verankert. Bei unwirksamen Einwilligungen ist ein Rückgriff auf das Forschungsprivileg nicht möglich ist. Ob ein Rückgriff auf das Forschungsprivileg möglich ist und ein entsprechender Erlaubnistatbestand vorliegt, sollte durch datenschutzrechtlich sachkundige Personen, z. B. den betrieblichen Datenschutzbeauftragten, beurteilt werden. Gegebenenfalls ist die Einschätzung der zuständigen Aufsichtsbehörde (Bundes-/Landesdatenschutzbeauftragter) einzuholen.

Zusammenfassend ist festzuhalten: Bei der Ausgestaltung von Einwilligungserklärungen ist eine Balance zwischen den gesetzlichen Anforderungen und den forschungspraktischen Bedingungen zu finden. Hohe Hürden der Teilnahme und negative Effekte auf die Teilnahmebereitschaft sind zu vermeiden. Die gesetzlichen und forschungspraktischen Anforderungen stehen nicht unbedingt im Widerspruch zueinander: Beispielsweise ist es sowohl datenschutzrechtlich erforderlich als auch forschungspraktisch sinnvoll, Einwilligungserklärungen in verständlicher Sprache zu formulieren. Auch die Vorlage einer schriftlichen Einwilligungserklärung kann im Forschungsinteresse sein: Stellt

sie doch die Seriosität des Projektes heraus und stärkt das Vertrauen der Betroffenen in den rechtskonformen Umgang mit den eigenen Daten.

5 Dritter Baustein: Forschungsdaten anonymisieren und pseudonymisieren

5.1 Anonymisierung

Die gesetzlichen Vorgaben verlangen in vielen Fällen eine Anonymisierung, die schnellstmöglich und so umfassend wie möglich durchgeführt wird. Dabei gilt eine faktische Anonymisierung der Daten nach derzeit herrschender Meinung als ausreichend (vgl. Goebel 2019, Ebel/Watteler 2019, 65f.). Daten gelten dann als faktisch anonymisiert, wenn eine Person nur mit völlig unverhältnismäßigem Aufwand identifiziert werden kann. Eine Identifizierung ist in diesem Fall nicht unmöglich, aber faktisch so aufwändig, dass sie als unmöglich betrachtet wird. Das heißt, das Gesetz mutet den betroffenen Personen ein Restrisiko der Identifikation zu.

Die Landesdatenschutzgesetze setzen sich (fast wortgleich) mit der Frage auseinander, wann eine natürliche Person nicht mehr anonym ist, also identifiziert werden kann. So lautet etwa § 2 Abs. 4 S. 3 HDSIG:

> *„(...) Eine natürliche Person ist identifizierbar, wenn sie unter Berücksichtigung aller Mittel, die von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich genutzt werden, um die Identität der natürlichen Person direkt oder indirekt zu ermitteln, identifiziert werden kann. (...)"*(§ 2 Abs. 4 S.3 HDSIG)

Dabei wird auf die Kosten und den Zeitaufwand abgestellt, die für den Vorgang der Identifizierung erforderlich sind und auf den aktuellen Stand der verfügbaren Technologie und der technologischen Entwicklungen. Eine ähnliche Formulierung findet sich auch im Erwägungsgrund 26 der DS-GVO.

Die Identifizierungsmöglichkeiten bestimmen sich durch die technischen Möglichkeiten und das vorhandene, verfügbare Zusatzwissen. Bei der Beurteilung der faktischen Anonymität von Daten sollte aber nicht nur auf die *Identifizierungsmöglichkeiten* abgestellt werden. Zusätzlich sollten die *Sensibilität der Daten* sowie das *Identifizierungsinteresse eines Angreifers* in den Blick genommen werden (vgl. Smolle 2015; Häder 2009, S. 21). Relevante Faktoren sind bspw. ob den Betroffenen durch Re-Identifizierung ein Schaden entstehen könnte (Sensibilität) oder ob es sich um kommerziell verwertbare Daten handelt (Identifizierungsinteresse)? Die Identifizierungsmöglichkeiten werden auch durch die Art der Datenaufbewahrung und der Datenzugänglichkeit beeinflusst. Forschungsdatenzentren erschweren Re-Identifizierungen von Personen in Daten unter anderem durch streng kontrollierte Zugangswege wie bspw. Gastwissenschaftlerarbeitsplätze oder Datenfernübertragung. Diese Maßnahmen bilden einen wesentlichen Baustein bei der Herstellung faktischer Anonymität von Daten (vgl. auch RatSWD 2017).

Die Anonymisierung von Forschungsdaten wird in der Realität aufgrund des technologischen Fortschritts im Hinblick auf Speicherkapazitäten und Analysemöglichkeiten (Stichworte: Big Data, Data Mining) sowie der zunehmenden (freien) Verfügbarkeit von Informationen im World Wide Web immer schwieriger zu bestimmen (vgl. auch Schaar 2009, Ebel/Watteler 2019, 66). Metschke und Wellbrock

(2002, S. 22) betonten schon 2002: „Die Grenze zwischen personenbeziehbaren und faktisch anonymisierten Daten ist […] fließend. Sie kann nur im Einzelfall – unter Berücksichtigung des gegenwärtigen und künftigen (z. B. wirtschaftlichen) Wertes der zu erlangenden Information oder der dem Empfänger sowie einem potenziellen Angreifer zur Verfügung stehenden Ressourcen – bestimmt werden." Nicht zuletzt aufgrund dieser Schwierigkeiten der Bestimmbarkeit faktischer Anonymität ist stets ein verantwortungsvoller Umgang auch mit - dem Anschein nach - faktisch anonymisierten Daten erforderlich.

Was ist zu anonymisieren?

Neben den direkten Identifikatoren (vgl. Kapitel 2) sollten auch diejenigen Informationen gelöscht, d. h. anonymisiert werden, mit deren Hilfe ein Bezug zu bestimmten Personen hergestellt werden kann (indirekte Identifikatoren). Die Datenschutzgesetze selbst enthalten keine näheren Angaben zu den Informationen, die hierbei zu löschen sind. In der Praxis gelten folgende Merkmale als sensibel bzw. als indirekte Identifikatoren, die bei der Anonymisierung in den Blick zu nehmen sind:

a) detaillierte Berufsangaben, detaillierte Lebensverläufe
b) kleinräumig-regionale Informationen: Bundesland, Region, Stadt
c) spezieller Erhebungskontexte: Expertenbefragungen, seltene Populationen, z. B. Musiker seltener Instrumente, seltene Branchen
d) seltene Merkmalskombinationen (geringe Fallzahl): Rektor der örtlichen Grundschule, Autobauer in Niedersachsen, Familie mit fünf Kindern in Ort XY, weiblicher Feuerwehr"mann"
e) verknüpfte Daten: Schulen – Schüler – Lehrer – Eltern – Peers

Auch sollte bei der Benennung von Dateien und der Benennung von Fällen in den Daten darauf geachtet werden, keine sprechenden Namen oder Identifikatoren zu verwenden (bspw. Schul- oder Bundeslandkennungen). Zusätzlich zu den in den Daten enthaltenen Informationen sind externe, öffentliche Quellen zu beachten, die mit den betreffenden Daten verknüpft eine Identifizierung ermöglichen könnten („Zusatzwissen"). Dies ist jedoch mit hohen Aufwänden verbunden und in der Praxis nur schwer umzusetzen. Bei Daten aus der Schulforschung handelt es sich häufig um Daten mit erhöhten Identifizierungsrisiken, da es sich hierbei häufig um verknüpfte Mehrebenendaten handelt. Das heißt, diese Daten enthalten Informationen auf den Ebenen Schule, Lehrpersonal, Lernende und Eltern.

Wann ist zu anonymisieren?

Eine ausreichende Anonymisierung von Daten sollte zum frühestmöglichen Zeitpunkt erfolgen, d. h. sobald es der Forschungszweck zulässt. Bei der Projektplanung und dem Forschungsdesign ist bereits darauf zu achten, dass die Arbeitsprozesse der Datenerhebung und -verarbeitung so organisiert werden, dass personenbezogene Daten nicht oder nur in geringem Umfang anfallen oder aber frühestmöglich gelöscht werden: So können Angaben wie Namen, Adressen oder Orte schon während der Transkription qualitativer Ton- oder Bildaufzeichnungen oder der Dateneingabe standardisierter Fragebögen gelöscht werden; Kontaktdaten können schon bei der Erhebung getrennt von den Interviewdaten gehandhabt werden; bestimmte Stichprobenverfahren wie das Random-Route-Verfahren ermöglichen die Erhebung ohne das personenbezogene Daten erfasst werden müssen (hier verfügt nur der Interviewer/die Interviewerin über Namen und Adresse).

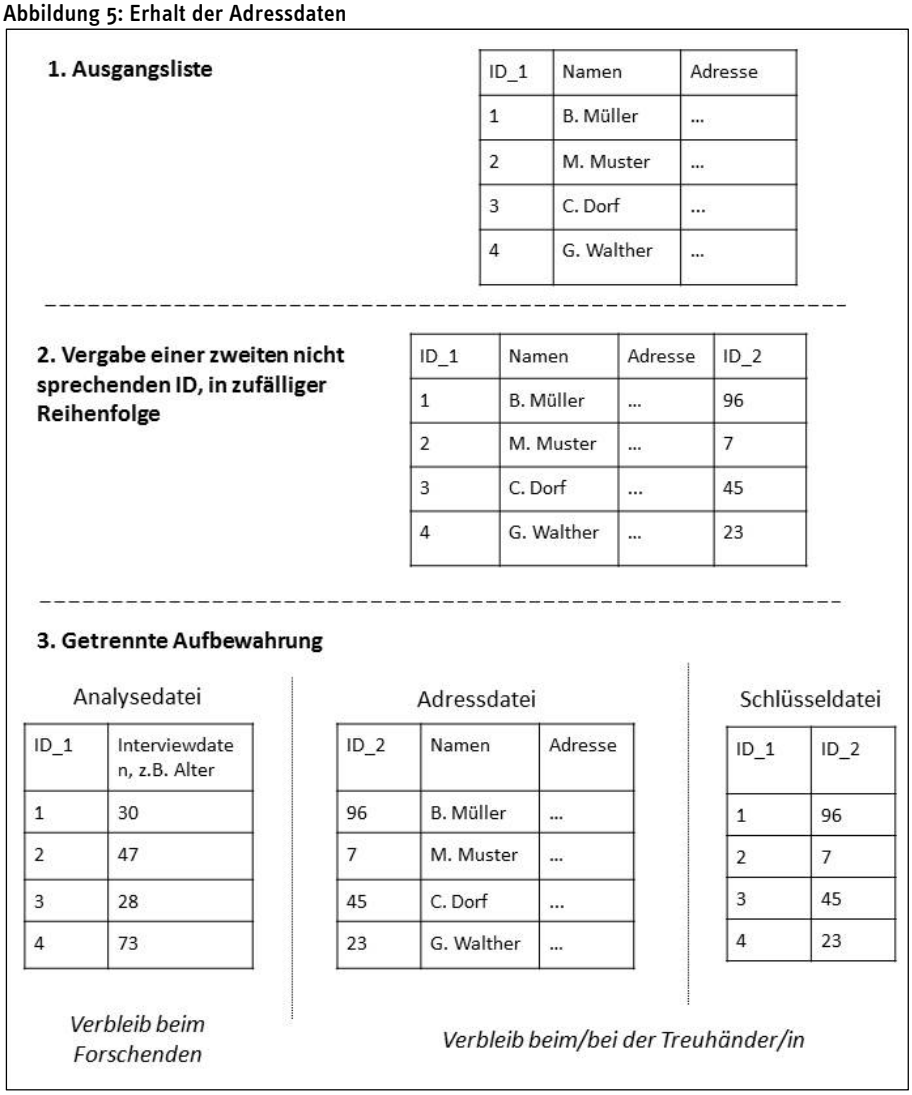
5.2 Pseudonymisierung

Bei bestimmten Erhebungsverfahren, bspw. Wiederholungsbefragungen, ist es erforderlich, die Kontaktdaten von Teilnehmer/innen aufzubewahren und deren Angaben über die verschiedenen Erhebungswellen hinweg miteinander verknüpfen zu können. In diesen Fällen ist eine Anonymisierung nicht möglich. Eine Alternative besteht darin, die Daten zu pseudonymisieren. Als Pseudonymisierung wird in der Praxis bezeichnet, wenn personenbezogene Daten von den sonstigen Daten getrennt werden, deren Verknüpfung mit den sonstigen, anonymen Daten aber möglich bleibt (vgl. bspw. Häder 2009, S. 8; Metschke und Wellbrock 2002, S. 23).

> In Art. 4 Nr. 5 DS-GVO heißt es: „*Pseudonymisieren ist die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden.*“

Da durch die Verknüpfung der Personenbezug der Daten wiederherstellbar ist, stellt die Pseudonymisierung keine Anonymisierung dar mit der Folge, dass das Datenschutzrecht anwendbar bleibt. Das heißt, dass sich im Regelfall die Einverständniserklärung auch auf die Verarbeitung pseudonymisierter Daten beziehen sollte. Bei pseudonymisierten Daten kann die Verknüpfung von Erhebungs- und Kontaktdaten über einen Schlüssel erfolgen, der die IDs des anonymisierten Datensatzes und der Kontaktdaten miteinander verbindet (vgl. Abbildung 6).

Abbildung 5: Erhalt der Adressdaten



Quelle: Meyermann und Porzelt 20 14, S.15

Die Aufbewahrung des Schlüssels und der Kontaktdaten kann wiederum durch einen *Datentreuhänder* erfolgen (vgl. Kapitel 3). Dieser kann die Verknüpfung von Daten aus mehreren Erhebungswellen herstellen. Die Forschenden selbst, die im Besitz der Erhebungsdaten sind, haben in diesem Fall zu keiner Zeit Zugriff auf die Kontaktdaten.

5.3 Anonymisierungsverfahren

Es gibt eine rege Forschung zu Anonymisierungsverfahren und deren Effekten auf das Analysepotential von Daten und Datenqualität (vgl. bspw. Höhne 2010; Ronning et al. 2005; Sweeney 2002; Wirth 2006). Bei *quantitativen Daten* sind verschiedene informationsreduzierende oder informationsverändernde Verfahren zu unterscheiden. Bei informationsreduzierenden Verfahren werden Informationen bspw. durch das Zusammenfassen von Variablenwerten oder -kategorien (Aggregation) reduziert oder einzelne Variablen gelöscht. So können einzelne Altersangaben zu Altersgruppen aggregiert oder einzelne Berufsnennungen zu ein- oder zweistelligen ISCO-Codes (Internationale Standarklassifikation der Berufe) zusammengefasst werden. Insbesondere Einzelwerte an den Rändern von Verteilungen oder geringe Zellbesetzungen gehen mit erhöhten Identifizierungsrisiken einher. Bei

informationsverändernden Verfahren wie Swapping- oder Imputationsverfahren werden Variablenwerte getauscht oder durch fiktive, auf Basis bestimmter Annahmen geschätzte Werte ersetzt (imputiert).

Einführende Hinweise zur Anonymisierung quantitativer Daten finden sich in unter anderem in Ebel/Meyermann 2015 oder Kinder-Kurlanda/Watteler 2015 (ab S.9).

Während bei quantitativen Daten die Verfahren zur Anonymisierung relativ automatisiert auf den Datensatz angewendet werden können, ist bei *qualitativen Daten* der manuelle Aufwand sehr viel größer. Beispielsweise ist jedes einzelne Transkript durchzusehen und zu anonymisieren. es bedarf dringend der Entwicklung automatischer Verfahren.[12] Auch im Hinblick auf konzeptionelle Fragen der Anonymisierung qualitativer Daten sollte der Diskurs fortgeführt werden (vgl. bspw. bei Thomson et al. 2005).

Selbst bei Löschung der direkten Identifikatoren (wie Eigennamen) ist bei den häufig vergleichsweise dichten und detailreichen qualitativen Daten das Risiko für die Identifizierung der Probanden oder in der Befragung erwähnten dritten Personen hoch. So kann beispielsweise der Hinweis eines spezifischen Berufes oder äußerlichen Merkmales, vor allem auch in der Verbindung und Gesamtschau mit anderen indirekten Angaben in den Daten, zu einer relativ einfachen Identifizierung von Personen führen. Eine reine Entfernung dieser Angaben ist häufig nicht zu empfehlen, da ein solches Vorgehen mit einem enormen Informationsverlust und somit einer Reduzierung der Analysemöglichkeiten einhergeht. Um dies zu vermeiden und möglichst wenig Analysegehalt zu verlieren, sollten Informationen nicht gelöscht, sondern durch inhaltlich vergleichbare Informationen ersetzt werden, so dass der ursprüngliche Sinngehalt der Daten erhalten bleibt (vgl. bspw. Kretzer 2013; Medjedović und Witzel 2010; Meyermann und Porzelt 2014). Beispielsweise wird der Beruf Friseur/Friseurin durch den Beruf Kosmetiker/Kosmetikerin ersetzt, die „Metzgerei Bäcker“ durch die „Bäckerei Schmidt“. Dabei sind die hinter solchen Ersetzungen steckenden Annahmen jeweils kritisch zu prüfen.

Einführende Hinweise zur Anonymisierung qualitativer Daten finden sich beispielsweise in: Kretzer 2013; Medjedović und Witzel 2010; Meyermann und Porzelt 2014.

Die Anonymisierung von *Video- und Audiodaten* ist mit besonderen Herausforderungen verbunden. So müssen Verfahren der akustischen Verfremdung und/oder der visuellen Manipulation eingesetzt werden, um eine Identifizierung von Personen in audiovisuellem Material zu unterbinden. Eine ausführlichere Beschreibung solcher Möglichkeiten findet sich bei Meyermann und Porzelt (2014). Neben dem Aufwand, der mit diesen Verfahren verbunden ist, sind im Falle eines solchen Vorgehens zwei weitere Punkte zu beachten: Einerseits kann über die akustische und visuelle Verfremdung das Analysepotential des Materials erheblich beeinträchtigt werden, andererseits existieren mittlerweile technische Möglichkeiten, um solche Maßnahmen der Anonymisierung wieder rückgängig zu machen (vgl. Newman 2016). Von besonderer Bedeutung wird daher die Beschränkung und Kontrolle des Zugangs und des Zugriffs auf diese Art von Daten (vgl. Kapitel 6).

Zusammenfassend kann festgehalten werden: Die Identifizierungsmöglichkeiten sind abhängig von den technologischen Möglichkeiten und dem Zusatzwissen eines potentiellen Angreifers. Für die Anonymisierung von Daten sind neben diesen auch das Identifizierungsinteresse sowie die

[12] Das Datenzentrum Qualiservice arbeitet derzeit in Kooperation mit der TU München an einem Anonymisierungstool (vgl. www.qualiservice.org/de/leitlinien.html, zugegriffen: 08. Aug. 2019)

Sensibilität der Daten bzw. der potentiell durch Re-Identifizierung entstehende Schaden zu berücksichtigen. Es existieren – vor allem für qualitative Daten – keine Standardverfahren zur Anonymisierung von Daten. Die Maßnahmen sollten immer dem jeweiligen Datenbestand und der je spezifischen Sensibilität angepasst werden. Empfehlenswert ist es, die durchgeführten Maßnahmen in einem *Anonymisierungsprotokoll* festzuhalten, eine Art kurzer Beschreibung des gewählten Vorgehens. Dies dient zum einen der Transparenz im Forschungsprozess selbst und stellt für eine mögliche Sekundärnutzung im Hinblick auf die Nachvollziehbarkeit der Daten eine wichtige Grundlage dar.

6 Vierter Baustein: Zugang und Zugriff auf Forschungsdaten beschränken

6.1 Technische und organisatorische Maßnahmen (TOM)

Zugang und Zugriff auf personenbezogene Daten sind im Forschungsprozess aus datenschutzrechtlichen Gründen zu beschränken. Forschende sind aufgefordert, sogenannte technische und organisatorische Maßnahmen (TOM, Art. 24 Abs. 1 S. 1 DS-GVO, § 64 BDSG) zu implementieren, anhand derer die sichere Aufbewahrung und die kontrollierte Nutzung der Daten gewährleistet wird. Die technischen und organisatorischen Maßnahmen sind im Gesetz definiert (vgl. insbesondere § 64 Abs. 3 BDSG). Dazu gehören Zugangs-, Datenträger-, Speicher-, Benutzer-, Zugriffs-, Übertragungs-, Eingabe- und Transportkontrollen, Wiederherstellbarkeit, Zuverlässigkeit, Datenintegrität, Auftrags- und Verfügbarkeitskontrollen sowie die Trennbarkeit. Der Gesetzgeber verlangt die Richtlinien und Empfehlungem des Bundesamts für Sicherheit in der Informationstechnik (BSI) zu beachten.

Der Zweck dieser Kontrollen und Vorgaben besteht darin, personenbezogene Daten vor unbefugten Zugriffen, Manipulation und Verlust zu schützen. Zum *Schutz der Dateien vor unbefugten Zugriffen*, ist es erforderlich, Unbefugten den physischen Zutritt zu den Datenverarbeitungsanlagen, mit denen personenbezogene Daten verarbeitet werden, zu verwehren (Gebäudesicherung und Sicherung der Räume). Der Zugang zu den Rechnern und Systemen, auf welchen die personenbezogenen Daten verarbeitet werden, sollte über ein effektives Passwortmanagement und der Zugriff auf die konkreten Daten über ein Berechtigungskonzept für Mitarbeiter/innen geregelt sein. Zu den Kontrollmaßnahmen gehören weiterhin auch besondere Verhaltensregeln für Mitarbeiter/innen zum Schutz vor menschlichem Fehlverhalten. Auch sind Mitarbeiter/innen zur Verschwiegenheit zu verpflichten, sofern dies nicht schon durch den Arbeitsvertrag abgedeckt ist (vgl. Häder 2009, S. 13).

Die technischen und organisatorischen Maßnahmen können zur faktischen Anonymität von personenbezogenen Daten beitragen, wenn durch sie der Aufwand zur Identifizierung einer Person unverhältnismäßig hoch wird (vgl. Smolle 2015; vgl. Kapitel 5.1). Sie sind auch für bereits faktisch anonymisierte Daten sinnvoll, um das Restrisiko einer Re-Identifikation zu verringern.

Die Ausgestaltung der TOM ist abhängig von der jeweiligen Projektkonstellation und sollte daher bereits bei der Planung des Projektes berücksichtigt werden: Welcher Personenkreis soll Zugriff auf die Daten haben und mit welchen Rechten (Lese- und/oder Schreibrechte) ausgestattet sein? Ist ein standortübergreifender bzw. standortunabhängiger Zugriff erforderlich? Ist vorgesehen, Datenbestände zu transferieren über mehrere Einrichtungen hinweg (bei Projektverbünden)? Bei der

Sicherungsstrategie sind sämtliche Geräte zu berücksichtigen (Arbeitsplatzrechner, Laptops, USB-Sticks u. a.), sämtliche Standorte (Arbeitsstätte, Wohnstätte, Reisetätigkeit) sowie der Personenkreis mit unterschiedlichen Aufgaben und Rollen.

- In der Praxis ist es hilfreich, das hausinterne Rechenzentrum und den internen Datenschutzbeauftragten hinzuzuziehen, da diese über die erforderlichen Kenntnisse zur sicheren und datenschutzkonformen Aufbewahrung von Daten verfügen.

- Empfehlungen zu Passwörtern und Möglichkeiten der Verschlüsselungen finden sich unter www.forschungsdaten-bildung.de/datensicherheit (Zugegriffen: 07. Aug. 2019).

Besondere Vorkehrungen sind beim *Austausch von Dateien mit externen Partnern* sowie bei der Speicherung von Dateien in einer Cloud oder auf mobilen Geräten zu treffen. Schützenswerte Daten sollten i. d. R. nur verschlüsselt in einer Cloud oder per E-Mail versendet werden. Aus der Perspektive der Datensicherheit gilt: Ein unverschlüsselter E-Mailversand ist mit dem Versand einer Postkarte vergleichbar. Im Hinblick auf datenschutzrechtliche Vorgaben ist zu beachten, dass nicht alle Cloud-Dienste (z. B. *Dropbox*) oder Kommunikationsdienste (z. B. *Skype*) kommerzieller Anbieter dem Datenschutzrecht der EU (DS-GVO) unterliegen. Die Verwendung dieser Dienste ist daher nicht empfehlenswert (vgl. Fraunhofer Institut für Sichere Informationstechnologie o.J.).

- Um (nicht personenbezogene) Dateien auszutauschen gibt es besondere Angebote für Hochschulangehörige, wie beispielsweise *Gigamove* der RWTH Aachen für Mitglieder im Deutschen Forschungsnetz (DFN)[13] oder *bwSyncShare* des Karlsruher Instituts für Technologie (KIT)[14].

Eine besondere Herausforderung in technischer Hinsicht stellt das *Löschen* personenbezogener Daten dar. Einfaches Löschen kommt bei den heute gängigen Betriebssystemen in der Regel nur dem Verschieben in den Papierkorb gleich und stellt daher *keine* Vernichtung der Daten dar, wie sie rechtlich jedoch erforderlich ist.

- Zur Vernichtung von Dateien sind spezielle Tools zu verwenden, zum Beispiel *BCWipe*[15].

6.2 Zugangswege bei Forschungsdatenzentren

Datenzentren bieten Forschungsdaten je nach deren Sensibilitäts- und Anonymitätsgrad über verschiedene Zugriffsstufen und Zugangswege an (vgl. auch Ebel/Watteler 2019, 72f.). So sind einige Daten (sogenannte *Public Use Files*) direkt frei verfügbar, andere erst nach einem Registrierungs- und ggf. auch Autorisierungsprozess (*Scientific-* und *Secure Use Files*). Die Voraussetzungen für den Zugriff auf bestimmte Datenbestände sind unterschiedlich. Das FDZ Bildung am DIPF beispielweise verlangt für bestimmte Datenbestände das Vorliegen einer Promotion oder die Zustimmung durch den oder die Promotionsbetreuer/in. In der Regel ist im Rahmen eines Nutzungsantrags der spezifische Forschungszweck darzulegen, zu dem die Daten ausgewertet werden sollen (bspw. FDZ Bildung am

[13] vgl. https://gigamove.rz.rwth-aachen.de (Zugegriffen: 07. Aug. 2019)
[14] vgl. https://bwsyncandshare.kit.edu/login (Zugegriffen: 07. Aug. 2019)
[15] vgl. www.jetico.com/products/personal-privacy/bcwipe (Zugegriffen: 07. Aug. 2019)

DIPF oder FDZ am IQB). Die Datennutzer/innen werden über einen abzuschließenden Datennutzungsvertrag bzw. die Anerkennung der Nutzungsbedingungen des FDZ zu einem datenschutz- und urheberrechtlich konformen Umgang mit den Daten verpflichtet sowie darauf, Identifizierungsversuche zu unterlassen.

Daneben sind verschiedene Wege des Datenzugangs zu unterscheiden. Datenbestände können Nutzer/innen per *Download* zur Verwendung an eigenen Rechnern überlassen werden. Weiter bieten Datenzentren die Möglichkeit der Nutzung per Datenfernübertragung (DFÜ) oder Remote Access. Beim *Fernrechnen* (Remote Access oder Remote Execution) erhalten Nutzer/innen keinen direkten Zugriff auf Datenbestände. Stattdessen senden Nutzer/innen Programmsyntax an die Datenzentren, die damit die Berechnungen ausführen. Die Nutzer/innen erhalten die auf dieser Basis errechneten Ergebnisse nach Datenschutzprüfung durch FDZ-Mitarbeiter/innen zurück. Eine weitere Variante des Datenzugangs ist die ausschließliche Vor-Ort-Nutzung des Datenmaterials an *Gastwissenschaftlerarbeitsplätzen*. Hierzu werden besonders geschützte Rechner verwendet (ohne die Möglichkeit, externe Speichermedien anzuschließen und ohne Zugriff zum Internet) und in geschützten Räumen (z. B. kein Zutritt mit Laptop oder Smartphone) zur Verfügung gestellt. Die Verfahren des Zugangs zu Forschungsdaten werden ständig weiterentwickelt, um Forschende nutzerfreundliche Sekundärnutzungswege zu ermöglichen. Die Zugangsbeschränkungen in FDZ erschweren eine Reidentifizierung für potentielle Angreifer und tragen so zur faktischen Anonymität der Daten bei. Die nachfolgende Tabelle stellt die verschiedenen Zugangswege und Zugriffstufen zusammenfassend dar.

Tabelle 1: Beispiele für Zugangswege und Zugriffsstufen bei Forschungsdatenzentren

Zugangswege und Zugriffstufen	Registrierung* und Autorisierung als wissenschaftliche/r Nutzer/in	Übermittlung zum/zur Datennutzer/in
Public Use File (Download, CD-Rom)	Nein	Ja (off-site-Nutzung)
Scientific Use File (Download, CD-Rom)	Ja	Ja (off-site-Nutzung)
Kontrollierte Datenfernverarbeitung (Remote Access)	Ja	Nein (off-site-Nutzung)
Gastwissenschaftlerarbeitsplatz	Ja	Nein (on-site-Nutzung)

* Die Registrierung beinhaltet die Zustimmung zu den speziellen Nutzungsbedingungen.

6.3 Dokumentation des Umgangs mit personenbezogenen Daten

Im Rahmen des Datenschutzmanagements in den jeweiligen Forschungseinrichtungen und Projekten ist eine „Übersicht“ zu Art und Umgang der personenbezogenen Daten zu erstellen. Diese wird als Verzeichnis von Verarbeitungstätigkeiten bezeichnet (Art. 30 DS-GVO). Das Verzeichnis dient der Dokumentation, Transparenz und Überprüfbarkeit der in der Organisation implementierten Maßnahmen zum Schutz der personenbezogenen Daten und nicht zuletzt der internen Selbstkontrolle. Die DS-GVO formuliert in Art. 30 DS-GVO die Grundlagen für dieses Verzeichnis und nennt in Abs. 1 die Inhalte eines solchen Verzeichnisses. So sind unter anderem

- der Verantwortliche,
- die Zweckbestimmungen der Datenverarbeitung,
- die Beschreibung der Kategorien betroffener Personen,
- Fristen zur Löschung der Daten sowie
- eine grobe Beschreibung der technischen und organisatorischen Maßnahmen (vgl. Kapitel 6.1)

aufzuführen.

Es sollte für Außenstehende ersichtlich werden, welche personenbezogenen Daten mit Hilfe von welchen automatisierten Verfahren auf welche Weise verarbeitet werden und welche Datenschutzmaßnahmen dabei getroffen werden. Die Inhalte des Verzeichnisses sind zum Teil für die Aufsichtsbehörde auf Anfrage einsehbar zu machen. Andere Inhalte, wie die konkreten technischen und organisatorischen Maßnahmen (vgl. Kapitel 6.1) oder sensible Angaben, sind nur zur internen Verwendung zu dokumentieren.

Als „Verantwortlicher" i. S. d. DS-GVO wird die Stelle gesehen, die Träger von Rechten und Pflichten ist, also etwa eine natürliche Person oder – was die Regel ist – eine juristische Person des öffentlichen oder des privaten (z. B. GmbH) Rechts. „[Eine Universität gilt beispielsweise] gegenüber dem Landesbeauftragten für Datenschutz als die Daten verarbeitende und hierfür verantwortliche Stelle. Innerhalb der Universität gelten jedoch die jeweiligen Einrichtungen, welche die automatisierten Verfahren zur Verarbeitung personenbezogener Daten betreiben (Verfahrens- und Prozessinhaber) als die jeweilige Daten verarbeitende und i. d. R. auch die hierfür verantwortliche (Unter-)Stelle. Diese haben auch die entsprechenden Verfahrensbeschreibungen zu erstellen." (Nörtemann 2014, S. 6). Im Projektverlauf sind die im erwähnten Verzeichnis genannten Stellen und Personen für die Einhaltung des beschriebenen Verfahrens zuständig. Änderungen (z. B. der verantwortlichen Stelle oder im Verfahrensablauf) sollten dem Datenschutzbeauftragten in Form einer aktualisierten Version angezeigt werden.

- Forschende sollten sich bezüglich des Verzeichnisses von Verarbeitungstätigkeiten an den oder die zuständige/n betriebliche/n Datenschutzbeauftragte/n wenden. Diese/r kann bei der Erstellung des Verzeichnisses unterstützen und entsprechende Vorlagen zur Verfügung stellen.

  Muster der Gesellschaft für Datenschutz und Datensicherheit e. V. GDD:
  https://www.gdd.de/downloads/praxishilfen/GDD-Praxishilfe_DS-GVO_5.pdf (Zugegriffen: 12.Juli 2019)
  Hinweise der Datenschutzkonferenz:
  https://datenschutz.hessen.de/sites/datenschutz.hessen.de/files/content-downloads/Hinweise%20zum%20Verzeichnis%20von%20Verarbeitungst%C3%A4tigkeiten_1.pdf (Zugegriffen: 12. Juli 2019);

Es muss nicht für jedes einzelne Forschungsprojekt ein eigenes Verzeichnis erstellt werden. Gegebenenfalls ist es ausreichend, das in der Organisation oder Abteilung vorhandene Verzeichnis, das beim betrieblichen Datenschutzbeauftragten geführt wird, entsprechend zu ergänzen.

7 Besonderheiten im Zusammenhang mit Erhebungen an Schulen

Für Schulerhebungen sind häufig datenschutzrechtliche Regelungen in Spezialgesetzen (z. B. Schulgesetzen) zu beachten und nachrangig das jeweilige LDSG. Teilweise existieren amtliche Verwaltungsvorschriften und Erlasse von Kultusministerien oder speziellere Vorgaben finden sich in Schulordnungen einzelner Schularten.

Erhebungen an Schulen sind in der Regel genehmigungspflichtig (vgl. im Folgenden www.forschungsdaten-bildung.de/genehmigungen, zugegriffen: 30. Mai 2019). Je nach Bundesland und

Art der Erhebung sind Genehmigungen durch die Schulleitungen, die Schulämter und -aufsichtsbehörden oder das Kultusministerium einzuholen. Manche Bundesländer schließen jedoch bestimmte Erhebungen von der Genehmigungspflicht aus, so beispielsweise das Land Brandenburg Erhebungen im Rahmen eines Lehramtsstudium.

Da sich die Bestimmungen für Erhebungen an Schulen im föderalistischen Deutschland je Bundesland unterscheiden, sind Forschende gefordert, frühzeitig zu recherchieren, ob eine Genehmigung erforderlich ist und wenn ja, welche Auflagen zu erfüllen sind. Der Genehmigungsprozess kann mehrere Wochen bis Monate dauern und sehr aufwendig sein. Erhebungen an Schulen, die bundesländerübergreifend sind, sind mit zusätzlichem Aufwand verbunden.

Von behördlicher Seite können Anforderungen an die Ausgestaltung von Einverständniserklärungen gestellt werden und auch an die Ausgestaltung von Erhebungsinstrumenten oder anderer Aspekte des Forschungsdesigns. Hintergrund der Genehmigungspraxis ist es, sowohl auf die Einhaltung der datenschutzrechtlichen Regelungen zu achten als auch eine übermäßige Belastung von Schulen und Störungen des Unterrichts zu vermeiden.

Eine Orientierungshilfe über die länderspezifischen Besonderheiten für Befragungen an Schulen findet sich unter: www.forschungsdaten-bildung.de/genehmigungen. Eine Betrachtung der Diskrepanzen zwischen Datenschutzrecht und Förder- sowie Genehmigungsauflagen bei der Verwendung von Videos in der Schul- und Unterrichtsforschung findet sich bei Scheller (2017).

8 Zusammenfassung und Ausblick

Das Datenschutzrecht ist - sowohl vor als auch nach Inkrafttreten der DS-GVO - unübersichtlich, da sich Vorschriften an verschiedenen Stellen finden und für viele datenschutzrechtliche Probleme keine eindeutigen Regelungen existieren. Dies verlangt schwierige Abwägungen unterschiedlicher Rechtspositionen im Einzelfall. Auf der einen Seite stehen datenschutzrechtliche Anforderungen, die erhebliche Herausforderungen für die empirische Forschungspraxis darstellen: Beispielsweise wenn das Einholen von Einwilligungen die Teilnahmebereitschaft reduziert, wenn sich durch umfängliche Datenanonymisierungen Analysepotentiale erheblich verringern, wenn die Anforderung einer engen Zweckbindung der Datenerhebung den Vorgehensweisen explorativer Forschung oder der langfristigen Nachnutzung von Forschungsdaten widerspricht oder wenn das Löschungsgebot den Regeln guter wissenschaftlicher Praxis entgegensteht, die im Sinne von Transparenz und Replizierbarkeit die Aufbewahrung von Daten für zehn Jahre vorsehen. Auf der anderen Seite ist festzustellen, dass die bestehenden Datenschutzgesetze für die Forschung ausdrücklich Spielräume in Form gesetzlicher Ausnahmetatbestände vorsehen.

Aufgrund der komplexen Rechtslage ist es umso bedeutender, die Absicht, die der Gesetzgeber mit der Abfassung des Datenschutzgesetzes im Sinn hatte, im Auge zu behalten und in der täglichen Forschungspraxis zu berücksichtigen: Dies ist der sorgsame und verantwortungsvolle Umgang mit personenbezogenen Daten. Forschende sind angehalten, mit den ihnen gesetzlich zur Verfügung stehenden Spielräumen verantwortungsvoll und gewissenhaft umzugehen. Das Ziel der vorliegenden Handreichung ist es daher nicht zuletzt, Forschende für datenschutzrechtliche Anforderungen zu sensibilisieren und ihnen die verschiedenen Bausteine aufzuzeigen, die zur Einhaltung des Datenschutzes in der empirischen Bildungsforschung verfügbar sind.

Forschende sollten sich ihrer rechtlichen und ethischen Verantwortung gegenüber Betroffenen stets bewusst sein und entsprechend sorgsam mit den ihnen anvertrauten Informationen umgehen. Der Rat für Informationsinfrastrukturen spricht in diesem Zusammenhang von der „Etablierung einer Verantwortungskultur, die die bestehenden Institutionen des Datenschutzes unter den sich wandelnden Bedingungen flexibel und verlässlich ergänzen“ (RfII 2017) kann. Forschungsdatenzentren wiederum können in diesem Zusammenhang eine besondere Rolle spielen, da sie zum einen Forschende bei datenschutzrechtlichen Fragen unterstützen können und zum anderen über die erforderlichen Strukturen verfügen, um Forschungsdaten sicher und zugriffsgeschützt aufzubewahren und bereitzustellen.

9 Übersicht Bausteine

Erster Baustein: Datenschutzrechtliche Prinzipien bei Forschungsdesign und Projektplanung berücksichtigen

» Die Grundsätze des Datenschutzes sind in sämtlichen Phasen des Forschungsprozesses und für sämtliche Aspekte des Forschungsdesigns zu bedenken. Eventuell sind zusätzliche finanzielle, personelle oder technische Ressourcen einzuplanen. Eine systematische **Forschungsdatenmanagementplanung** kann darin unterstützen, die rechtlichen Aspekte frühzeitig und fortlaufend zu berücksichtigen.

» Erhebung, Verarbeitung und Nutzung personenbezogener Daten sollten nur erfolgen, wenn diese nicht vermeidbar, d. h. zur Erfüllung des Forschungszwecks unbedingt erforderlich sind. Die **Sekundärnutzung von Daten ist demnach einer Primärerhebung vorzuziehen**. Ist eine Primärerhebung personenbezogener Daten erforderlich, sollte der Grundrechtseingriff so milde wie möglich sein, d. h., es sollten so wenige personenbezogene Daten erhoben werden wie möglich, und diese sollten zugriffsgeschützt aufbewahrt werden.

Das Portal www.forschungsdaten-bildung.de bietet eine zentrale Übersicht über Studien und Daten der empirischen Bildungsforschung und einen Einstieg zu weiterführenden Angeboten von Datenzentren, die Forschungsdaten für Sekundäranalysen unter kontrollierten Bedingungen bereit stellen.

Zweiter Baustein: Informierte Einwilligung

» Grundsätzlich gilt, dass zur Verarbeitung personenbezogener Daten das Einverständnis der betroffenen Personen einzuholen ist. Dabei muss es sich um ein sogenanntes informiertes Einverständnis handeln. Das heißt, **die Betroffenen müssen hinreichend über die Art der Daten, die verarbeitet werden, die Verarbeitungsschritte und die unterschiedlichen Zwecke der Verarbeitung informiert sein**, um die Tragweite ihrer Entscheidung beurteilen zu können.

» Genaue Vorgaben zum Einholen informierter Einwilligungen regelt das jeweils gültige Gesetz, also entweder eine bereichsspezifische Vorschrift oder die DS-GVO und das jeweilige LDSG oder das BDSG. Das betreffende Gesetz regelt, in welcher Form, wann und bei wem Einwilligungen einzuholen sind sowie welche Inhalte Einwilligungen aufweisen sollten.

» Für die Forschung gibt es Ausnahmen von der generellen Pflicht, Einwilligungen in Schriftform einzuholen.

Weiterführende Informationen, Checklisten und Formulierungsbeispiele verfügbar unter: www.forschungsdaten-bildung.de/einwilligung

Dritter Baustein: Anonymisierung und Pseudonymisierung von Forschungsdaten

» Gesetzliche Vorgaben verlangen eine **Anonymisierung personenbezogener Daten, die schnellstmöglich und so umfassend wie möglich durchgeführt wird**. Eine faktische Anonymisierung (wenn eine Person nur mit völlig unverhältnismäßigem Aufwand identifiziert werden kann) gilt als ausreichend.
» Faktische Anonymität zu bestimmen, ist in der heutigen Zeit zunehmend schwieriger. Auch faktisch anonymisierte Daten sollten besonders geschützt – beispielsweise bei einem Forschungsdatenzentrum – aufbewahrt werden.
» Festzuhalten ist, dass es – vor allem bei qualitativen Daten – kein Standardverfahren zur Anonymisierung der Daten gibt, sondern die Maßnahmen immer dem jeweiligen Datenbestand und der je spezifischen Sensibilität angepasst werden müssen. Empfehlenswert ist es, die durchgeführten Maßnahmen in einem Anonymisierungsprotokoll festzuhalten, eine Art kurzer Beschreibung des gewählten Vorgehens.

Weiterführende Informationen verfügbar unter: www.forschungsdaten-bildung.de/anonymisierung

Vierter Baustein: Zugang und Zugriff auf Forschungsdaten beschränken

» Zugang und Zugriff auf personenbezogene Daten sind im Forschungsprozess aus datenschutzrechtlichen Gründen zu beschränken. Forschende sind aufgefordert, sogenannte technische und organisatorische Maßnahmen (TOM) zu implementieren, anhand derer die **sichere Aufbewahrung und die kontrollierte Nutzung der Daten** gewährleistet wird.
» Diese Maßnahmen können zur faktischen Anonymität von personenbezogenen Daten beitragen, wenn durch sie der Aufwand zur Identifizierung einer Person unverhältnismäßig hoch wird. Sie sind auch für bereits faktisch anonymisierte Daten sinnvoll, um das Restrisiko einer Re-Identifikation zu verringern.

10 Literatur

ADM. 2012. Richtlinie für den Einsatz von Datentreuhändern in der Markt- und Sozialforschung. www.rat-marktforschung.de/fileadmin/user_upload/pdf/R10_RDMS_D.pdf (Zugegriffen: 08. Aug. 2019).

BDP und DGPs. 2016. Berufsethische Richtlinien des Berufsverbandes Deutscher Psychologinnen und Psychologen e. V. und der Deutschen Gesellschaft für Psychologie e. V. Berlin. www.dgps.de/fileadmin/documents/Empfehlungen/ber-foederation-2016.pdf (Zugegriffen: 08. Aug. 2019).

DGFE. 2010. Ethik-Rat und Ethikkodex der Deutschen Gesellschaft für Erziehungswissenschaft. *Erziehungswissenschaft* 21(41):179-184. Verfügbar unter: www.pedocs.de/volltexte/2011/4030/pdf/ErzWiss_2010_41_Ethik_kodex_der_DGfe_D_A.pdf (Zugegriffen: 08. Aug. 2018).

DGS und BDS. 2014. Ethik-Kodex der Deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologinnen und Soziologen (BDS). www.soziologie.de/fileadmin/_migrated/content_uploads/Ethik-Kodex_2014-06-14.pdf (Zugegriffen: 08. Aug. 2019).

Diekmann, Andreas. 2003. *Empirische Sozialforschung – Grundlagen, Methoden, Anwendungen.* Rowohlt Taschenbuch Verlag GmbH: Reinbek bei Hamburg.

Ebel, Thomas und Alexia Meyermann. 2015. Hinweise zur Anonymisierung von quantitativen Daten. *forschungsdaten bildung informiert* 3. www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert (Zugegriffen: 29. Mai 2019).

Ebel, Thomas und Oliver Watteler. 2019. Datenschutz im Forschungsdatenmanagement. In *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten*, Hrsg. Uwe Jensen, Sebastian Netscher und Katrin Weller, 57-80. Berlin, Opladen, Toronto: Barbara Budrich.

Fraunhofer Institut für Sichere Informationstechnologie. o. J. *Über die Sicherheit von Cloud-Speicherdiensten – Management Summary.* www.sit.fraunhofer.de/fileadmin/dokumente/studien_und_technical_reports/Cloud-Storage-Security_ManagementSummary.pdf (Zugegriffen: 29. Mai 2019).

Goebel, Jürgen. 2019. *Anonymisierung, faktische Anonymisierung, Pseudonymisierung. Ist die faktische Anonymisierung eine „echte" Anonymisierung?* Expertise vom 29. April 2019. Unveröffentlicht.

Häder, Michael. 2009. Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. *RatSWD Working Paper Series* 90. www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf (Zugegriffen: 29. Mai 2019).

Höhne, Jörg. 2010. Verfahren zur Anonymisierung von Einzeldaten. *Statistik und Wissenschaft* 16.

Kinder-Kurlanda, Katharina E. und Oliver Watteler. 2015. Hinweise zum Datenschutz: Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. *GESIS Papers* 2015-01.

Kretzer, Susanne. 2013. Arbeitspapier zur Konzeptentwicklung der Anonymisierung/Pseudonymisierung in Qualiservice. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-47605-2 (Zugegriffen: 29. Mai 2019).

Liebig, Stefan, Tobias Gebel, Matthis Grenzer, Julia Kreusch, Heidi Schuster, Ralf Tscherwinka, Oliver Watteler und Andreas Witzel. 2014. Datenschutzrechtliche Anforderungen bei der Generierung und Archivierung qualitativer Interviewdaten. *RatSWD Working Paper Series* 238. www.ratswd.de/dl/RatSWD_WP_238.pdf (Zugegriffen: 29. Mai 2019).

Medjedović, Irena. und Andreas Witzel. 2010. *Wiederverwendung qualitativer Daten. Archivierung und Sekundärnutzung qualitativer Interviewtranskripte.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Metschke, Rainer und Rita Wellbrock. 2002. Datenschutz in Wissenschaft und Forschung. *Materialien zum Datenschutz* 28.

Meyermann, Alexia und Maike Porzelt. 2014. Hinweise zur Anonymisierung von qualitativen Daten. *forschungsdaten bildung informiert* 1. www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert (Zugegriffen: 29. Mai 2019).

Meyermann, Alexia, Doris Bambey, Thomas Ebel, Marcus Eisentraut, Malte Jansen, Poldi Kuhl, Reiner Mauer, Claudia Neuendorf, Lisa Pegelow, Maike Porzelt, Marc Rittberger, Thomas Schwager, Petra Stanat, Jessica Trixa. 2017. *Schlussbericht zum Verbundprojekt „Sicherung und Nachnutzung der Forschungsdaten des Rahmenprogramms zur Förderung der empirischen Bidlungsforschung“*. Frankfurt: DIPF. www.forschungsdaten-bildung.de/files/TIBKAT_897124898-1.pdf (Zugegriffen: 08. Aug. 2019).

Newman, Lily Hay. 2016. AI Can Recognize Your Face Even If You’re Pixelated. www.wired.com/2016/09/machine-learning-can-identify-pixelated-faces-researchers-show/ (Zugegriffen: 22. Aug. 2017).

Nörtemann, Bernd. 2014. *Verfahrensbeschreibung gemäß §8 des Nds. Datenschutzgesetzes (NDSG).* www.tu-braunschweig.de/Medien-DB/it/cio/ausfuellhinweise_verfahrensbeschreibung_ausfuehrlich-s.pdf (Zugegriffen: 22. Aug. 2017).

RatSWD (Rat für Sozial- und Wirtschaftsdaten). 2017. *Handreichung Datenschutz. Output 5.* www.ratswd.de/dl/RatSWD_Output5_HandreichungDatenschutz.pdf (Zugegriffen: 22. Aug. 2017).

RfII (Rat für Informationsinfrastrukturen). 2017. *Datenschutz und Forschungsdaten. Aktuelle Empfehlungen.* www.rfii.de/download/rfii-empfehlungen-2017-datenschutz-und-forschungsdaten/ (Zugegriffen: 22. Aug. 2017).

Ronning, Gerd, Roland Sturm, Jörg Höhne, Rainer Lenz, Martin Rosenmann, Michael Scheffler, Daniel Vorgrimler. 2005. Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten. *Statistik und Wissenschaft 4.* www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band4_AnonymisierungMikrodaten1030804059004.pdf?__blob=publicationFile (Zugegriffen: 22. Aug. 2017).

Schaar, Katrin. 2017. Die informierte Einwilligung als Voraussetzung für die (Nach-)nutzung von Forschungsdaten. Beitrag zur Standardisierung von Einwilligungserklärungen im Forschungsbereich unter Einbeziehung der Vorgaben der DS-GVO und Ethikvorgaben. *RatSWD Working Paper Series* 264. www.ratswd.de/dl/RatSWD_WP_264.pdf (Zugegriffen: 22. Aug. 2017).

Schaar, Peter. 2009. Data protection and statistics – a dynamic and tension-filled relationship. *RatSWD Working Paper* 82. www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_82.pdf (Zugegriffen: 22. Aug. 2017).

Scheller, Jürgen. 2017. Rechtliche Rahmenbedingungen der Verwendung von Videos in der Schul- und Unterrichtsforshung. Diskrepanzen zwischen Datenschutzrecht, Förder- und Genehmigungsauflagen. *forschungsdaten bildung informiert* 5. www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert (Zugegriffen: 22. Aug. 2017).

Sheehan, Mark. 2011. Can Broad Consent be Informed Consent? *Public Health Ethics* 4(3): 226-235.

Smolle, Michael. 2015. Datenschutzrechtliche Anmerkungen zum Artikel „Vom Datenangreifer zum zertifizierten Wissenschaftler“ von Ulrich Rendtel. *AStA Wirtschafts- und Sozialstatistisches Archiv* 9: 73-78.

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

Thomson, Denise, Lana Bzdel, Trish Reay Karen Golden-Biddle und Carole A. Estabrooks. 2005. Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data. *Forum Qualitative Social Research* 6(1): 1–16.

Verbund Forschungsdaten Bildung. 2018a. Formulierungsbeispiele für informierte Einwilligungen. *fdbinfo Nr. 4. Version 2.1.* www.forschungsdaten-bildung.de/publikationsreihen#fdbinfo (Zugegriffen: 29. Mai 2019).

Verbund Forschungsdaten Bildung. 2018b. Formulierungsbeispiele für informierte Einwilligungen in leichter Sprache. *fdbinfo Nr. 5, Version 1.1.* www.forschungsdaten-bildung.de/publikationsreihen#fdbinfo (Zugegriffen: 29. Mai 2019).

Verbund Forschungsdaten Bildung. 2017. Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung. *fdbinfo Nr. 2, Version 2.0.* www.forschungsdaten-bildung.de/publikationsreihen#fdbinfo (Zugegriffen: 29. Mai 2019).

Verbund Forschungsdaten Bildung. 2019. Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schule. *fdbinfo Nr.1, Version 2.0.* www.forschungsdaten-bildung.de/publikationsreihen#fdbinfo (Zugegriffen: 29. Mai 2019).

Wirth, H. 2006. Anonymisierung des Mikrozensuspanels im Kontext der Bereitstellung als Scientific-Use-File. *Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe", Arbeitspapier* 11. www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/zuma_arbeitspapiere/Arbeitspapier11.pdf (Zugegriffen: 08. Aug. 2019).