Hochschule für
Wirtschaft und Recht Berlin
Berlin School of Economics and Law

Anforderungen an den Datenschutz in Forschungsprojekten

Handlungsempfehlung der HWR Berlin – Team Datenschutz

Die Handreichung wurde auf Basis der Publikation „Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung– eine Handreichung“ von Alexia Meyermann und Maike Porzelt erstellt. Diese wurde in weiten Teilen übernommen, jedoch an die Bedarfe der HWR Berlin angepasst. Herausgeber ist das Deutsche Institut für Internationale Pädagogische Forschung.

Wir danken dem DIPF herzlich für die Genehmigung der Überarbeitung und des Nachdrucks unter der Lizenz CC BY NC. Die originale Publikation kann abgerufen werden unter: https://www.forschungsdaten-bildung.de/files/fdb-informiert-nr-6.pdf

Inhalt

1 Einleitung 2

2 Gesetzliche Vorgaben zum Schutz personenbezogener Daten 3

» *Gemeinschaftsinteresse* 6

» *Geeignetheitsgrundsatz* 6

» *Erforderlichkeitsgrundsatz* 6

» *Wahl des mildesten Mittels* 6

» *Übermaßverbot* 6

» *Datenvermeidung und Datensparsamkeit* 6

4 Zweiter Baustein: Einwilligungen einholen 9

4.1 Formale Kriterien: Was, wie, wann, bei wem? 9

Was? – Bestandteile von Einwilligungserklärungen 9

4.2 Zentrale Anforderungen: Laienverständlichkeit, Freiwilligkeit, Zweckbindung 12

4.3 Die „richtige“ Einwilligungserklärung 16

4.4 Arbeits- und ablauforganisatorische Implikationen 16

4.5 Sonderfall: Datenerhebung ohne Einwilligung? Erhebung auf Basis des Forschungsprivilegs 17

5 Dritter Baustein: Forschungsdaten anonymisieren und pseudonymisieren 18

5.1 Anonymisierung 18

5.2 Pseudonymisierung 19

5.3 Anonymisierungsverfahren 20

6 Vierter Baustein: Zugang und Zugriff auf Forschungsdaten beschränken 21

6.1 Technische und organisatorische Maßnahmen (TOM) 21

6.2 Dokumentation des Umgangs mit personenbezogenen Daten 22

1 Einleitung

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. Vor diesem Hintergrund befasst sich die vorliegende Handreichung 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 Forschung bei der Berücksichtigung der datenschutzrechtlichen Anforderungen in ihrer Forschungstätigkeit unterstützen.

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).

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. 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. Das zweite relevante Grundrecht schützt die Forschungsfreiheit. In Artikel 5 Abs. 3 des Grundgesetzes heißt es: „Kunst und Wissenschaft, Forschung und Lehre sind frei.“ Sie dürfen somit keiner willkürlichen Einschränkung unterworfen werden. Grundrechtsträger sind also auch Beschäftigte an wissenschaftlichen Hochschulen und außeruniversitären Forschungseinrichtungen, die sozial-, verhaltens- und wirtschaftswissenschaftliche Forschungsprojekte durchführen. 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.

Abbildung 1: Grundrechte und Datenschutz

Übergeordnete Grundlage für jedwede Verarbeitung personenbezogener Daten ist seit dem 25. Mai 2018 die europäische Datenschutzgrundverordnung (DS-GVO). Die Landesdatenschutzgesetze (LDSG) sind einschlägig für öffentliche Stellen der Länder wie Hochschulen. In Berlin ist das Berliner Datenschutzgesetz einschlägig. 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 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“. 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. 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 und Daten über Verurteilungen und Straftaten (Art. 10 DSGVO) gelten strengere Regeln als für allgemeine bzw. „einfache“ personenbezogene Daten.

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, Gemeinschaftsinteresse, Datensparsamkeit und Datenminimierung und Verhältnismäßigkeit bei der Datenverarbeitung. Diese werden in den nachfolgenden Kapiteln näher beschrieben.

Für die Forschung existiert ein besonderer Erlaubnistatbestand in Gestalt des *Forschungsprivilegs*. Dieses ist in §17 BlnDSG niedergeschrieben. Es erlaubt **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).

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. Die datenschutzrechtlichen Grundsätze 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.

» *Gemeinschaftsinteresse*

Forschende sind angehalten zu prüfen, ob das Forschungsvorhaben einem legitimen Gemeinschaftsinteresse dient. Zum Beispiel liegt ein Forschungsvorhaben, das bereits inhaltlich gegen andere Gesetze verstößt, nicht im Gemeinschaftsinteresse und muss daher nicht erst am Recht auf informationelle Selbstbestimmung scheitern.

» *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. Insofern Daten beispielsweise lediglich auf Vorrat hinterlegt werden, ist dies in Frage gestellt.

» *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.

» *Datenvermeidung und Datensparsamkeit*

> Ein erweitertes Verständnis von Datenschutz findet darin Ausdruck, dass bereits vor der Ausgestaltung von Datenerhebungen und -verarbeitungen darauf hingewirkt wird, keine oder möglichst wenig personenbezogene

> Daten zu verwenden. Dies ist beispielsweise in § 3a Bundesdatenschutzgesetz (BDSG) konkretisiert: „Die Erhebung, Verarbeitung und Nutzung personenbezogener Daten und die Auswahl und Gestaltung von Datenverarbeitungssystemen sind an dem Ziel auszurichten, so wenig personenbezogene Daten wie möglich zu erheben, zu verarbeiten oder zu nutzen."

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?

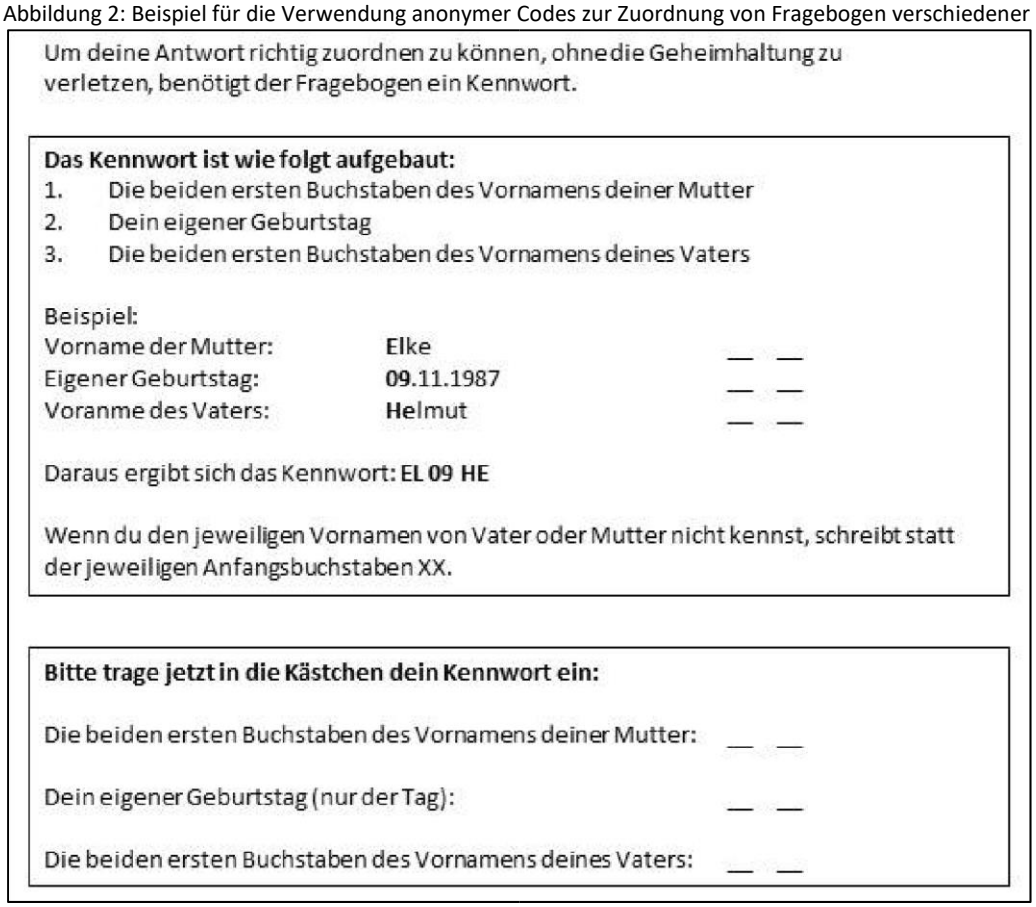
Den datenschutzrechtlichen Grundsätzen folgend sind Sekundäranalysen bestehender Datensätze Primärerhebungen vorzuziehen, außer der Forschungszweck kann ohne eine eigene Primärdatenerhebung nicht erreicht werden. Für eine solche Entscheidung sind die verfügbaren Datenbestände und die existierende Literatur zu sichten und zu bewerten. 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."
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[1] durchgeführt werden und die Interviews können von verschiedenen Personen geführt werden. *Nicht-sprechende Identifikatoren*[2] 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

[1] 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.

[2] 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.

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.

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. 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.

4 Zweiter Baustein: Einwilligungen einholen

Ein weiterer Baustein zur Einhaltung datenschutzrechtlicher Anforderungen besteht darin, die informierte Einwilligung der Betroffenen einzuholen. 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.

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.

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. 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.

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. 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 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 die getätigte Anonymisierungszusage durch die Forschenden für die Befragung selbst 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. 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. 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). 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. Eine Verweigerung des Jugendlichen kann nicht durch die Eltern beschnitten werden.

Auch volljährige Personen können je nach aktueller Einsichts-, Urteils- und Verständnisfähigkeit vorübergehend oder dauerhaft nicht einwilligungsfähig sein. Zu bedenken ist in einem solchen Fall auch durch die Forschenden, inwieweit eine Teilnahme einer solchen Person an der Studie sinnvoll ist oder ob ein Proxy-Interview (Stellvertreter-Interview) angeraten erscheint. Grundsätzlich muss bei nicht einwilligungsfähigen minderjährigen oder erwachsenen Personen die Einwilligung des gesetzlichen Vertreters rechtswirksam erteilt sein.

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,

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.

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 anonymisierten Daten.

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

Für die Forschung ist es möglich, Einwilligungen auch zu weniger eng definierten Zwecken einzuholen . 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 sollt

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.“;

» *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.“

„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 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.

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

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.

Widerrufe müssen so lange möglich sein, wie personenbezogene Daten existieren. Das Widerspruchsrecht erlischt, sobald die Daten keinen Personenbezug (z.B. durch Anonymisierung) 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 von uns verlangen."

Erfolgt ein Widerruf, ist die betreffende Person aus den Daten zu löschen. Ein Projekt muss festlegen, wer dies durchführt. 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*. 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,

„Die Verarbeitung personenbezogener Daten, einschließlich besonderer Kategorien personenbezogener Daten im Sinne von Artikel 9 Absatz 1 der Verordnung (EU) 2016/679, ist auch ohne Einwilligung für die Erfüllung einer Aufgabe zu im öffentlichen Interesse liegenden

wissenschaftlichen oder historischen Forschungszwecken oder für statistische Zwecke zulässig, wenn das öffentliche Interesse an der Durchführung des Vorhabens die schutzwürdigen Belange der betroffenen Person erheblich überwiegt und der Zweck nicht auf andere Weise erreicht werden kann. (§17 BlnDSG).

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.

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. 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.

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.

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. 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.

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). 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. 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.

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). Bei der Verwendung von Daten in Publikationen ist auf absolute Anonymität zu achten. Grundsätzlich sollte keine Veröffentlichung von Rohdaten erfolgen. Bei qualitativen Daten sollte darüber hinaus bspw. keine Veröffentlichung kompletter Transkripte erfolgen .

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.

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

1. Ausgangsliste

ID_1	Namen	Adresse
1	B. Müller	...
2	M. Muster	...
3	C. Dorf	...
4	G. Walther	...

---

2. Vergabe einer zweiten nicht sprechenden ID, in zufälliger Reihenfolge

ID_1	Namen	Adresse	ID_2
1	B. Müller	...	96
2	M. Muster	...	7
3	C. Dorf	...	45
4	G. Walther	...	23

---

3. Getrennte Aufbewahrung

Analysedatei

ID_1	Interviewdaten, z.B. Alter
1	30
2	47
3	28
4	73

Adressdatei

ID_2	Namen	Adresse
96	B. Müller	...
7	M. Muster	...
45	C. Dorf	...
23	G. Walther	...

Schlüsseldatei

ID_1	ID_2
1	96
2	7
3	45
4	23

Verbleib beim Forschenden

Verbleib beim/bei der Treuhänder/in

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. 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).

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.

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. 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.

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. 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.

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 Zweck dieser 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.

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.

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 CloudDienste (z. B. *Dropbox*) oder Kommunikationsdienste (z. B. *Skype*) kommerzieller Anbieter dem Datenschutzrecht der EU (DS-GVO) genügen. Die Verwendung dieser Dienste ist daher nicht empfehlenswert.

- Um personenbezogene Daten auszutauschen, gibt es besondere Angebote für Hochschulangehörige, wie beispielsweisedie HWR Cloud.

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*.

6.2 Dokumentation des Umgangs mit personenbezogenen Daten

Im Rahmen des Datenschutzmanagements ist von den jeweiligen Forschenden 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 ist an den Datenschutzbeauftragten zu senden.

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.

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.

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.