

Informationsblatt zur Datenspeicherung [Stand: Mai 2021]

Da es immer wieder Fragen/Rückfragen bzgl. der Datenspeicherung gibt, sind hier die wichtigsten Informationen diesbezüglich zusammengestellt. Über die Vorgehensweise der Datenspeicherung und ggf. Art und/oder Zeitpunkt von Anonymisierungsmaßnahmen sind die Probanden in der Probandenaufklärung zu informieren, und sie müssen hierzu ihre Einwilligung im Rahmen der Einwilligungserklärung geben. Werden Daten im Sinne von open science publiziert, sollte in der Studieninformation grob angegeben werden, um welche Art von Datenbank es sich handelt, wer Zugriff hat u.ä..

1. Datenspeicherung: Anonymisierung vs. Pseudonymisierung

Pseudonymisierte Daten

Grundsätzlich ist zwischen pseudonymisierter und anonymisierter Datenspeicherung zu unterscheiden. Bei einer **pseudonymisierten Datenspeicherung** werden die erhobenen Daten mit einem Code versehen. In der Regel hat der/die Studienleiter/in eine Liste, die die Zuordnung von Namen und Code ermöglicht, z.B. um später ggf. den/die Studienteilnehmerin erneut zu kontaktieren („**Variante Kodierliste**“). Diese Zuordnungsliste ist ebenso wie unterschriebene Einwilligungserklärungen unter Verschluss zu halten und darf nicht unmittelbar mit den Daten zusammengeführt werden. Beispielsweise dürfen Dateien mit Code und Daten und der Code-Zuordnung keinesfalls am gleichen Ort gespeichert sein.

Anonymisierte Daten

Nach § 3 Abs. 6 Bundesdatenschutzgesetz (BDSG) liegt eine **Anonymisierung** vor, wenn die Daten so verändert worden sind, "dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können." Dies bedeutet, dass das bloße Entfernen des Namens nicht mit einer Anonymisierung gleichzusetzen ist. Letztlich kann nur im Einzelfall entschieden werden, ob die Anonymisierung gelungen ist. Bleibt es bspw. bei der Angabe "Alter, Geschlecht, Bundesland" und wird das mit der Berufsbezeichnung "Industriearbeiter" verbunden, besteht Anonymität; bei der Berufsbezeichnung "Ministerpräsident" hingegen keineswegs. Zu bedenken ist, dass mit den Mitteln der modernen Technik eine Deanononymisierung möglicherweise auch aus komplexen und aggregierten Datensätzen vorgenommen werden kann. Es liegt im Bereich der Verantwortung des Forschers, die Daten dahingehend zu überprüfen und vollständig zu anonymisieren, bevor eine Publikation des Datensatzes z.B. in open science Format erfolgt. Inwieweit also tatsächlich eine Anonymisierung vorliegt, hängt davon ab, inwieweit Daten erhoben werden, die auch bei Fehlen eines Namens eine Identifizierung der Person ermöglichen. Dies bedeutet, dass es keine verallgemeinerbaren Regeln gibt, wann eine Anonymisierung geglückt ist. Jede/r Studienleiter/in sollte hier entsprechende Verantwortung übernehmen.

Bei einer öffentlichen Bereitstellung der Daten in anonymisierter Form zum Beispiel mittels des -Open Science Frameworks liegt es außerdem in der Verantwortung des Studienleiters, die Probanden darüber zu informieren, dass ihre Daten nach Abschluss der Studie anonymisiert zugänglich gemacht werden sollen.

Anonymisierung von Daten

a) *Faktische Anonymisierung*

Eine Variante der faktischen Anonymisierung pseudonymisierter Datenspeicherung ist die Variante „Persönliches Codewort“. Hierbei generiert der Teilnehmer gemäß bestimmter vorgegebener Regeln (z.B. 1. Buchstabe des Vornamens der Mutter usw.) einen persönlichen Code. Der Code ist nur dem Studienteilnehmer bekannt bzw. kann nur von ihr nachvollzogen werden. Der Proband hat so die Möglichkeit, nachträglich Daten einzusehen, korrigieren und löschen zu lassen. In gewisser Hinsicht handelt es sich bei dieser Variante um eine Mischung aus pseudonymisierter und anonymisierter Datenspeicherung.

b) *Schrittweise Anonymisierung*

Bei dieser Form der Anonymisierung werden die Daten zunächst pseudonymisiert, also unter Verwendung eines dem Versuchsleiter bekannten Codes erhoben (vgl. pseudonymisierte Daten). Die Kodierliste wird dann zu einem definierten Zeitpunkt (z.B. nach Abschluss der Datenerhebung) vernichtet, so dass ab diesem Zeitpunkt (und in Abhängigkeit von den erhobenen Daten) keine Zuordnung der Daten zu einzelnen Personen mehr möglich ist.

c) *Anonymisierungstechniken zur Vorbeugung von Deanonymisierung*

Ein häufiges Problem besteht darin, dass Daten im Sinne des open science Ansatzes anderen Wissenschaftlern, z.B. durch Publikation in einem Repositorium, zugänglich gemacht werden sollen. Um eine Deanonymisierung zu verhindern, können verschiedene Maßnahmen ergriffen werden, z.B.:

- Weglassen bestimmter für die Deanonymisierung bedeutsamer Variablen (z.B. Alter)
- Aggregation dieser Variablen durch z.B. die Bildung grober Kategorien (z.B. 40-49jährige)
- Randomisierung der Daten einer Variable über alle Vpn

Auch hier muss der jeweilige Forscher entscheiden, welche Maßnahmen sinnvoll zu einer vollständigen Anonymisierung beitragen. Gibt es beispielsweise nur eine Vpn im Alter von 40-49 Jahren, ist die Anonymisierung durch diese Kategorienbildung nicht erfolgreich. Der/die Studienleiter/in muss also prüfen, ob davon ausgegangen werden kann, dass die Vorgaben für eine Anonymisierung gemäß § 3 Abs. 6 BDSG (s.o.) erfüllt sind.

2. EU-DSGVO

Die europäische Datenschutzgrundverordnung (EU-DSGVO) regelt die Verarbeitung personenbezogener Daten. Es geht also um Daten, die direkt oder indirekt die Identifikation einer bestimmten Person erlauben. Bei vollständig anonymen Datenerhebungen trifft die EU-DSGVO also nicht zu. Eine Aufklärung und Einwilligung bezüglich der EU-DSGVO sollte also nur bei einer pseudonymisierten Datenerhebung erfolgen. Werden in weiteren Schritten Anonymisierungstechniken wie z.B. die Vernichtung der Kodierliste, vorgenommen, sollten die Probanden darüber informiert werden und einwilligen, dass ab diesem Zeitpunkt eine Zuordnung nicht mehr erfolgen kann und dementsprechend z.B. Einsicht, Korrektur oder Löschung nicht mehr möglich sind.

Hinweis: Auf den folgenden Seiten sind Formulierungsvorschläge für die Teilnehmerinformation und die Einwilligungserklärung zu finden. Es ist ausreichend, in der Teilnehmerinformation die jeweils zutreffende Form der Datenspeicherung zu erklären, und in der Einwilligungserklärung nur auf die verwendete Form der Datenspeicherung zu verweisen. Erklärungen müssen nicht doppelt erfolgen.

**Vorlagen für die Beschreibung der verschiedenen Varianten der Datenspeicherung in der
Teilnehmerinformation bzw. der Einwilligungserklärung
[adaptiert aus den Vorlagen der DGPs-Ethikkommission, 2016]**

Variante „Kodierliste“

Mit der beschriebenen Erhebung und Verarbeitung der Daten >hier genau definieren, welche Daten, bei Gesundheitsdaten konkret benennen< bin ich einverstanden. Die Aufzeichnung und Auswertung dieser Daten erfolgt pseudonymisiert im >Angabe Institut<, unter Verwendung einer Nummer und ohne Angabe meines Namens. Es existiert eine Kodierliste auf Papier, die meinen Namen mit dieser Nummer verbindet. Diese Kodierliste ist nur den Versuchsleitern und dem Projektleiter zugänglich, das heißt, nur diese Personen können die erhobenen Daten mit meinem Namen in Verbindung bringen. Nach Abschluss >der Datenerhebung / der Datenauswertung<, spätestens am >Datum angeben<, wird die Kodierliste gelöscht. Meine Daten sind dann anonymisiert. Damit ist es niemandem mehr möglich, die erhobenen Daten mit meinem Namen in Verbindung zu bringen. Ich bin informiert, dass ich mein Einverständnis zur Aufbewahrung bzw. Speicherung dieser Daten widerrufen kann, ohne dass mir daraus Nachteile entstehen. Ich kann jederzeit eine Löschung all meiner Daten verlangen. Wenn allerdings die Kodierliste bereits gelöscht ist, kann mein Datensatz nicht mehr identifiziert und also auch nicht mehr gelöscht oder korrigiert werden, und ich kann keine Einsicht mehr in meine Daten nehmen. Meine Daten sind dann anonymisiert.

Ich bin einverstanden, dass meine vollständig anonymisierten Daten zu Forschungszwecken weiterverwendet werden können. >Dazu werden sie mindestens 10 Jahre nach Datenauswertung, bzw. mindestens 10 Jahre nach Erscheinen einer Publikation zu dieser Studie aufbewahrt. / Dazu werden sie über eine Internet-Datenbank > wenn bereits bekannt: Name des Online Repositoriums < öffentlich zugänglich gemacht.<

Hinweis an den Projektleiter: Wenn Sie die Variante „Kodierliste“ gewählt haben, können sie optional die Einwilligung zu einer im Antrag begründeten längerfristigen Speicherung der pseudonymisierten Daten erfragen.

Zusatzvereinbarung für künftige Kontaktaufnahmen im Rahmen dieser Studie

Ich gebe mein Einverständnis, dass im Falle einer Fortführung dieser Studie oder von Anschlussstudien die Kodierliste weiterhin für die Dauer von fünf Jahren aufbewahrt wird und ich für Anschlussstudien kontaktiert werden darf. Mein Einverständnis zur Aufbewahrung bzw. Speicherung dieser Daten kann ich jederzeit widerrufen, ohne dass mir daraus Nachteile entstehen. Ich kann jederzeit eine Löschung all meiner Daten verlangen. Wenn allerdings die Kodierliste bereits gelöscht ist, kann mein Datensatz nicht mehr identifiziert und also auch nicht mehr gelöscht werden. JA NEIN.

Variante „Persönliches Codewort“

Mit der beschriebenen Erhebung und Verarbeitung der Daten >hier genau definieren, welche Daten, bei Gesundheitsdaten konkret benennen< bin ich einverstanden. Die Aufzeichnung und Auswertung der Daten erfolgt anonymisiert im >Angabe Institut< unter Verwendung eines persönlichen Codewortes, das ich selbst erstellt habe und das nur ich kenne. Das heißt, es ist niemandem außer mir möglich, meine Daten mit meinem Namen in Verbindung zu bringen. Das Blatt, auf dem ich dieses Codewort erstellt habe, befindet sich in meinem Besitz. Mein Einverständnis zur Aufbewahrung bzw. Speicherung meiner Daten kann ich jederzeit widerrufen, ohne dass mir daraus Nachteile entstehen. Ich kann jederzeit eine Löschung all meiner Daten verlangen.

Ich bin einverstanden, dass meine vollständig anonymisierten Daten zu Forschungszwecken weiterverwendet werden können. >Dazu werden sie mindestens 10 Jahre nach Datenauswertung, bzw. mindestens 10 Jahre nach Erscheinen einer Publikation zu dieser Studie aufbewahrt. / Dazu werden sie über eine Internet-Datenbank >wenn bereits bekannt: Name des Online Repositoriums< öffentlich zugänglich gemacht.<

Zufallsbefunde

In manchen Forschungskontexten, z.B. der klinisch-psychologischen Forschung, ist es möglich, dass so genannte Zufallsbefunde gemacht werden. Die Probanden sind hierüber aufzuklären und müssen einwilligen, ob bzw. dass sie über derartige Befunde informiert werden. Ggf. sind sie auch über weiter reichende Folgen wie z.B. versicherungsrechtliche Konsequenzen entsprechender Befunde aufzuklären und müssen bestätigen, dass sie darüber in Kenntnis gesetzt wurden.

Formulierungsbeispiel Kodierliste:

Sollten behandlungsbedürftige Auffälligkeiten >im EEG / in den Hirnbildern aus der MRT-Untersuchung / in der Testdiagnostik< erkannt werden, bin ich damit einverstanden, dass mir diese mitgeteilt werden, so dass ich diese ggf. weiter abklären lassen kann. Ich wurde darüber aufgeklärt, dass die Information über auffällige Befunde u.U. mit versicherungsrechtlichen Konsequenzen verbunden sein kann.

Formulierungsbeispiel persönlicher Code

Ich bin darüber informiert, dass mein Name und >Angabe der verwendeten Kontaktdaten, wie Anschrift, Email-Adresse und/oder Telefonnummer< nur auf dieser Einwilligungserklärung stehen. Sollten sich aus meiner Untersuchung >im EEG / in den Hirnbildern aus der MRT-Untersuchung / in der Testdiagnostik< Hinweise auf behandlungsbedürftige Auffälligkeiten ergeben, bin ich damit einverstanden, dass mir diese mitgeteilt werden, so dass ich diese ggf. weiter abklären lassen kann. Ich wurde darüber aufgeklärt, dass die Information über auffällige Befunde u. U. mit versicherungsrechtlichen Konsequenzen verbunden sein kann.

Da alle Daten vollständig anonymisiert vorliegen, bin ich über das folgende Vorgehen informiert worden: Im Falle von behandlungsbedürftigen Auffälligkeiten werden sämtliche in Frage kommenden Versuchsteilnehmer angeschrieben und um Mitteilung gebeten, ob das betreffende persönliche Codewort auf sie zutrifft. Ich bin darüber informiert, dass ich mich – sofern es sich um mein eigenes Codewort handelt – bei der angegebenen Adresse melden und nähere Informationen einholen sollte. Wenn das angegebene Codewort nicht mein eigenes ist, kann ich dieses Schreiben ignorieren.

