Handreichung der Ethikkommission

zu Anonymisierung und Pseudonymisierung der Daten bei wissenschaftlichen Studien an der Paris-Lodron-Universität Salzburg

(Juli 2024)

Bitte beachten Sie die folgenden Informationen und setzen diese in Ihrem *Ethikantragsformular* und in der *Studieninformation mit Einverständniserklärung* um, da andernfalls Ihr Antrag von der Ethikkommission aus formellen Gründen zurückgewiesen werden müsste.

Hintergrund

Für die Erhebung von Daten bei wissenschaftlichen Studien der PLUS ist die PLUS-S Richtlinie zur EU-Datenschutzgrundverordnung (DSGVO) vom 3.6.2019 (siehe Website der Ethikkommission) einzuhalten. Hierbei sind personenbezogene Daten mittels geeigneter Maßnahmen vor Missbrauch zu schützen. Personenbezogene Daten sind „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“. Dieser Begriff ist somit sehr umfassend, denn er erfasst alle Daten, die einer individuellen Person direkt oder indirekt, einzeln oder in Kombination mit anderen ggfs. öffentlich zugänglichen Daten zurechenbar sind, wie beispielsweise Name, Adresse, E-Mail-Adresse, IP-Adresse, Körpermerkmale, biometrische Daten, genetische Daten, Einstellungen, Meinungen, Diagnosen, etc. Das heißt, dies betrifft im Prinzip alle in einer Studie erhobenen abhängigen Variablen.

Je sensibler die zu verarbeitenden Daten sind, desto notwendiger ist die Gewährleistung eines angemessen hohen Schutzniveaus für diese Daten. Das höchste Schutzniveau erlaubt eine anonyme Erhebung der Daten. Ein fast so hohes Schutzniveau bietet eine Anonymisierung nach Datenerhebung, wobei die Anonymisierung möglichst bald erfolgen sollte.

Die Pseudonymisierung ist eine weitere mögliche Maßnahme, deren Einsatz die Gewährleistung eines angemessenen Schutzniveaus in vielen Fällen unterstützen kann. Diese Maßnahmen können die Risiken der Verarbeitung personenbezogener Daten für von dieser Verarbeitung betroffene Personen verringern und den aus der DSGVO resultierenden rechtlichen Anforderungen hinsichtlich eines angemessen hohen Schutzniveaus für die wissenschaftliche Forschung genügen.

Bei den Datenarten wird folgende Unterscheidung getroffen:

Pseudonyme Daten sind solche Daten, die sich nicht mehr ohne zusätzliche Information einer Person zuordnen lassen. Dies erfolgt typischerweise mittels einer Referenztabelle (oder Zuordnungstabelle), welche die Namen der an einer Studie teilnehmenden Personen in Spalte A und den jeder Person individuell durch die Studie zugewiesenen Code in Spalte B enthält. Alle übrigen Daten der Studie werden dann nur noch mit diesem Code versehen und der Name der

Person taucht auf ausgefüllten Fragebögen, anderen Datenerhebungsinstrumenten, Dateinamen, in der statistischen Auswertungssoftware etc. nicht mehr auf. Die Daten werden also mittels dieser Referenztabelle und Codierung von der Studienleitung pseudonymisiert. Die Identität einer teilnehmenden Person kann nur durch die Referenztabelle aufgelöst und ihren Daten zugeordnet werden. Pseudonyme Daten sind weiterhin als personenbezogene Daten zu behandeln.

Die Referenztabelle zur Feststellung der Identität muss von den pseudonymen Daten getrennt aufbewahrt werden. Es müssen geeignete technische und organisatorische Maßnahmen getroffen werden, die der Vermeidung einer Identifizierung der betroffenen Personen dienen. Pseudonyme Daten sowie Referenztabelle müssen auf sichere Art und Weise gespeichert werden. Unsicher sind lokale Speichermedien wie unverschlüsselte Festplatten oder USB-Sticks. Sicher ist der Group File Service (GFS) der IT-Services oder eigens für die Ablage von Forschungsdaten eingerichtete Repositorien.

Besonders im Austausch mit berechtigten Dritten (z.B. Studienmitarbeitenden; kollaborierenden Forschenden an anderen Universitäten) stellt die Pseudonymisierung der Daten ein Verfahren zur Wahrung der Sicherheit der personenbezogenen Daten dar. Wichtig ist dabei, auf eine korrekte Umsetzung zu achten. Es darf unter keinen Umständen möglich sein, dass Dritte durch Verknüpfung von öffentlich zugänglichen Daten mit dem pseudonymen Datensatz auf die Person schließen können. Dies impliziert, dass der pseudonyme Datensatz in aller Regel keine Daten wie z.B. Geburtsdatum, Geburtsort, Kontaktdaten, E-Mail-Adresse, Wohnort oder Beruf enthalten sollte, da durch Internetsuche und Verknüpfung mit weiteren Daten ggfs. eine Zuordnung zu einer bestimmten Person möglich wäre. Solche Daten sollten i.d.R. deshalb auch nur in der Referenztabelle gespeichert werden.

Anonyme Daten sind Daten, die keinen Rückschluss auf die Identität einer bestimmten Person zulassen. Anonyme Daten sind keine personenbezogenen Daten und unterliegen daher keinen Aufbewahrungsfristen. Es wird empfohlen, wenn immer möglich, anonyme Daten zu verwenden, um einen hohen Datenschutz sicherzustellen und nicht in den Anwendungsbereich der DSGVO zu fallen.

Im Folgenden werden die Felder 4.1 bis 4.10. des Ethikantragsformulars kommentiert, um Einreichende bei der Umsetzung der DSGVO zu unterstützen, insbesondere in Hinblick darauf, was bezüglich der Pseudonymisierung oder Anonymisierung ihrer Daten zu beachten ist.

1. Welche Art der Anonymisierung bzw. Pseudonymisierung wird verwendet?

Im Antragsformular wird eine Unterscheidung zwischen fünf Datenerhebungsmethoden getroffen, welche von a - e mit einem abnehmendem Schutzniveau verbunden sind: a) Anonyme Datenerhebung, b) Anonymisierung nach Datenerhebung, c) Pseudonyme Datenerhebung, d) Pseudonymisierung nach Datenerhebung und e) Datenerhebung ohne Anonymisierung bzw.

Pseudonymisierung. Hierbei ist zu beachten, dass das höchste Schutzniveau gewählt werden sollte, welches für eine Durchführung der Studie gerade ausreichend ist. Eine nicht-anonyme Datenerhebung bedarf grundsätzlich einer ausführlichen Begründung, inwieweit die Studiendurchführung diese erfordert.

a) **Anonyme Datenerhebung**. Diese Datenerhebungsmethode sollte soweit möglich verwendet werden. Sie bietet sich insbesondere bei Online-Fragebogenerhebungen (z.B. mittels LimeSurvey) an. Hierbei sollten weder Name, E-Mail-Adresse oder IP-Adresse des Computers erfasst werden. Es sollten auch keine personenbezogenen Daten erfasst werden, die einzeln oder in Kombination mit anderen ggfs. öffentlich zugänglichen Daten zu einer Person zurechenbar sind.

b) **Anonymisierung nach Datenerhebung**. Diese Datenerhebungsmethode erfasst zwar initial direkt einer Person zurechenbare Daten (z.B. wird für die Kommunikation mit der Person ihr Name und ihre E-Mail-Adresse verwendet). Diese einer Person zurechenbaren Daten werden aber nach der Datenerhebung vernichtet bzw. gelöscht. Dies kann (idealerweise) unmittelbar nach der Datenerhebung erfolgen oder innerhalb einer angemessenen Frist, welche im Ethikantrag anzugeben ist, zusammen mit einer Begründung, warum anonyme Datenerhebung nicht möglich bzw. sinnvoll ist.

c) **Pseudonyme Datenerhebung**. Diese Datenerhebungsmethode wird bei wissenschaftlichen Studien häufig verwendet. Hierbei wird durch den Teilnehmenden selbst ein Pseudonym generiert und notiert. Dieses wird von der Studienleitung bei der Datenerhebung mit erhoben. Die Identität der Person ist so der Studienleitung nicht bekannt. Diese Vorgehensweise erlaubt es beispielsweise, bei einer Längsschnittstudie mit Messwiederholungen die einzeln erhobenen Datensätze eines Teilnehmenden zu verknüpfen. Es erlaubt dem Teilnehmenden auch, bei der Studienleitung die Einsicht in seine/ihre Daten oder die nachträgliche Löschung seiner/ihrer Daten anzufordern. Eine häufig verwendete Vorgehensweise bei der Erstellung des Pseudonyms ist folgende: Die Studienleitung instruiert den Versuchsteilnehmenden schriftlich, einen individuellen 8-stelligen Code zu generieren aus z.B. Anfangsbuchstabe und dritter Buchstabe des Vornamens der Mutter, Anfangsbuchstabe und dritter Buchstabe des Vornamens des Vaters, Tag des Geburtsdatums des Vaters (zweistellig), eigener Geburtsmonat (zweistellig). Hieraus resultiert dann das für die Studie zu verwendende individuelle Pseudonym, z.B. „MRAT0811“.

d) **Pseudonymisierung nach Datenerhebung**. Diese Datenerhebungsmethode erfasst einer spezifischen Person zurechenbare Daten. Durch Generierung eines Versuchspersonen-Codes und die separate gesicherte Aufbewahrung einer Referenztabelle bestehend aus dem Code und allen dieser Person zurechenbaren Daten (wie z.B. Name, Adresse, E-Mail-Adresse, geloggte IP-Adresse des Computers, etc.), werden diese jedoch von dem zu analysierenden Datensatz getrennt. Dies kann (idealerweise) unmittelbar bei oder nach der Datenerhebung erfolgen oder innerhalb einer angemessenen Frist, welche im Ethikantrag anzugeben ist, zusammen mit einer Begründung, warum die anderen sichereren Arten der Datenerhebung für die spezifische Studie nicht möglich bzw. sinnvoll sind. Es ist darauf zu achten, dass im zu analysierenden Datensatz alle Personenmerkmale, die auf die Identität der teilnehmenden Person schließen lassen, gelöscht

wurden. Die Studienleitung wird durch die Zuordnungstabelle in die Lage versetzt, jederzeit alle von einer Person erhobenen Daten wieder genau dieser Person zuzuordnen und die Person ggfs. zu kontaktieren. Dies kann beispielsweise dann sinnvoll sein, wenn die Person über die individuellen Ergebnisse ihrer Messungen im Rahmen der Studie informiert werden möchte bzw. wenn sie über Zufallsbefunde informiert werden möchte. Ein pseudonymisierter Datensatz kann später bei Bedarf leicht in einen anonymisierten Datensatz umgewandelt werden, indem die Referenztabelle, welche neben dem Code alle einer Person zurechenbaren Daten enthält, vernichtet bzw. gelöscht wird.

e) **Datenerhebung ohne Anonymisierung bzw. Pseudonymisierung**. Diese Datenerhebungsmethode kann in seltenen Ausnahmefällen notwendig bzw. sinnvoll sein, was jedoch einer ausführlichen Begründung bedarf.

2. Ist den Teilnehmer:innen die Einsicht in ihre persönlichen Daten möglich?
Den Teilnehmenden ist idealerweise bei Interesse eine Einsicht in die von ihnen erhobenen Daten und Messwerte zu ermöglichen. Der Zeitraum der möglichen Einsichtnahme und Art der Umsetzung sollen spezifiziert werden. Bei anonymer Datenerhebung, bei Anonymisierung nach Datenerhebung bzw. aus anderen Gründen kann dies ggfs. nicht praktikabel bzw. erwünscht sein. Eine Begründung soll dann hier und in der *Studieninformation mit Einverständniserklärung* spezifiziert werden.

Gibt es bei einer Studie prinzipiell die Möglichkeit von gesundheitsrelevanten Zufallsbefunden (z.B. Tumor im Gehirn bei Studien mit Erhebung von Magnetresonanztomographie [MRT]), ist dies zu vermerken und die entsprechenden Vorgaben im Textbaustein der *Studieninformation mit Einverständniserklärung* zu berücksichtigen.

3. Können sich Teilnehmer:innen über die Forschungsergebnisse informieren?
Dies bezieht sich auf die Ergebnisse der Studie, welche nach Durchführung der Studie z.B. in Form einer Masterarbeit oder eines Manuskripts vorliegen. Den Teilnehmenden ist idealerweise bei Interesse eine Einsicht in die Ergebnisse der Studie zu ermöglichen. Der zu erwartende Zeitraum bis zum Vorliegen der Ergebnisse und Art der Umsetzung sollen spezifiziert werden.

4. Was tun Sie, um zu gewährleisten, dass Teilnehmer:innen aus der Studie aussteigen können?
Studienteilnehmende haben jederzeit das Recht, ihre Teilnahme ohne Angabe von Gründen zu beenden. Hier und in der *Studieninformation mit Einverständniserklärung* soll spezifiziert werden, wie genau eine Studienteilnahme bzw. spezifische Messungen auf Wunsch des Teilnehmenden beendet werden können.

5. Wie und bis zu welchem Zeitpunkt können Teilnehmer:innen die Löschung ihrer Daten verlangen?
Den Teilnehmenden ist grundsätzlich die Löschung der von ihnen erhobenen Daten zu ermöglichen. Der Zeitraum der möglichen Löschung und Art der Umsetzung sollen spezifiziert

werden. Bei anonymer Datenerhebung, bei Anonymisierung nach Datenerhebung bzw. aus anderen Gründen kann dies ggfs. nicht praktikabel bzw. erwünscht sein. Eine Begründung soll dann hier und in der *Studieninformation mit Einverständniserklärung* spezifiziert werden.

6. Welche personenbezogenen Daten werden zu welchen Zwecken erhoben?

Die DSGVO sieht vor, dass alle von der Studie erfassten personenbezogenen Daten sowie der Zweck dieser Erhebungen den Studienteilnehmenden transparent gemacht werden müssen.

Personenbezogene Daten inkludieren z.B. Name; Kontaktdaten (z.B. Postanschrift, E-Mail-Adresse, Telefonnummer); Geburtsdatum; körperliche Merkmale (z.B. Größe, Gewicht, Haarfarbe, Erkrankungen); genetische Daten; biometrische Daten (z.B. 3-D Kopf-Scan); neurokognitive Daten; psychologische Merkmale (z.B. Persönlichkeit, Einstellungen, Symptome psychischer Störungen); Beziehungen (z.B. Verwandtschafts- und Freundschaftsbeziehungen, Arbeitgeber); weitere Daten (z.B. Standortdaten, Handy-Nutzungsdaten, Handlungen, Ausbildung und beruflicher Werdegang, Bankverbindungen).

Der Zweck der Erhebung der verschiedenen personenbezogenen Daten, d.h. wofür die einzelnen Daten verwendet werden, muss hier dargestellt werden (Zweckbindung nach DSGVO). Da eine spätere Änderung oder Erweiterung des Zwecks nicht einfach möglich ist, sollte dieser hier möglichst weit aber ausreichend präzise formuliert werden. Z.B. Ihr Name und Ihre E-Mail-Adresse wird benötigt, um mit den Studienteilnehmenden für die Dauer der Studie Kontakt aufnehmen zu können. Größe und Gewicht wird erhoben, um daraus den Body Mass Index zu berechnen, welcher eine Einschätzung von Normal- bzw. Übergewicht ermöglicht. Psychologische Merkmale werden erhoben, um unsere wissenschaftliche Fragestellung beantworten zu können. Entsprechende Angaben sind in dem dafür vorgesehenen Textbaustein der *Studieninformation mit Einverständniserklärung* zu spezifizieren.

7. Wird Audio (z.B. Stimme), Foto oder Video aufgezeichnet?

Diese Arten der Datenerhebung sind besonders sensibel, da die Daten i.d.R. direkt einer Person zugeordnet werden können. Solche Aufnahmen erfordern eine sichere Aufbewahrung (z.B. Group File Service der PLUS, siehe unten Punkt 8; unzureichend ist die längere Aufbewahrung auf einem Diktiergerät oder Smartphone), eine gute Begründung und die explizite Einwilligung der Teilnehmenden für spezifische Verwendungszwecke (z.B. für projektbezogene wissenschaftliche Auswertungen) und sind ggfs. nur mit spezifischen Datenschutzmaßnahmen zulässig. In der Regel sollten solche Aufnahmen zeitnah analysiert und danach umgehend gelöscht werden (z.B. Vernichtung von Tonaufnahmen eines Interviews nach Transkription; Löschung der Gesichtsaufnahmen nach Analyse des emotionalen Gesichtsausdrucks). Alternativ kann der Datenschutz durch Unkenntlichmachen der Stimme oder Verpixelung des Gesichts sichergestellt werden. Ggfs. kann aus wissenschaftlichen Gründen eine längere Aufbewahrung nötig sein. Dies bedarf einer gesonderten Begründung. Die Details dazu sollen im *Antragsformular* und in der *Studieninformation mit Einverständniserklärung* spezifiziert werden.

Soll für die Transkription von Interviews oder die Erstellung bzw. Bearbeitung von Ton-, Foto- oder Videomaterial cloudbasierte Software verwendet werden, so ist zuerst zu klären, ob die IT-Services eine eigene, datensichere Lösung bereitstellen können (siehe auch unten, Punkt 9).

Falls nicht, ist mit dem Anbieter ein Auftragsverarbeitungsvertrag zu schließen und zu prüfen, ob und bejahendenfalls welchen Subauftragsverarbeiter der Anbieter zur Erbringung der Leistung heranzieht. Es sollte klar sein, wer welche Daten bekommt und an welchem Ort sie letztlich verarbeitet werden.

8. Wie und wie lange werden die Daten in welcher Form während und nach Beendigung der Studie aufbewahrt? Wie werden sie vernichtet?
Bei der Verarbeitung und Aufbewahrung von personenbezogenen Daten ist ein besonderes Augenmerk auf die Datensicherheit zu legen. **Personenbezogene Daten sollten nicht auf lokalen Speichermedien (Festplatte des Computers im Büro der PLUS oder Homeoffice, USB-Stick, externe Festplatte, Smartphone) aufbewahrt werden.** Entscheidet man sich dennoch für diese Speicherform, so sollten die auf dem Speichermedium gespeicherten Daten verschlüsselt sein. Der von den IT-Services bereitgestellte Group File Service (GFS) kann in aller Regel als hinreichend sicher angesehen werden. Personenbezogene Daten sollten im Sinne der Datensparsamkeit so kurz wie möglich aufbewahrt werden. Für wissenschaftliche Zwecke ist grundsätzlich eine gesicherte Aufbewahrung personenbezogener Daten bis zu 30 Jahre möglich. Im Sinne einer guten wissenschaftlichen Praxis wird eine Aufbewahrung für 10 Jahre empfohlen. Für anonyme Daten gilt diese Vorgabe nicht. Unabhängig davon sollen die unterschriebene *Studieninformation mit Einverständniserklärung* der einzelnen Teilnehmer:innen wenn möglich für 30 Jahre ausgedruckt in einem verschlossenen Schrank der PLUS sicher aufbewahrt werden.

9. **Werden persönliche Daten (z.B. Transkription von Interview-Aufnahmen, Bildmaterial, physiologische Daten) mittels cloudbasierter Software (z.B. Transkriptor) verarbeitet?**
Bei Verwendung von cloudbasierten Speicher- oder Analyselösungen ist sicherzustellen, dass die Richtlinien der DSGVO eingehalten werden, was z.B. für eine Speicherung auf Rechenzentren in Ländern der EU normalerweise sichergestellt ist. Es ist zuerst zu klären, ob die IT-Services eine eigene, datensichere Lösung bereitstellen können (z.B. für Transkription oder Bildbearbeitung).

10. Werden die Daten in anonymisierter Form in einer Open-Science Datenbank anderen Wissenschaftler:innen für Analysen zur Verfügung gestellt?
In einigen Wissenschaften und von Förderorganisationen wird zunehmend die Transparenz, Replizierbarkeit und Aggregation wissenschaftlicher Daten in Form von Open-Science Datenbanken verlangt. Dies ist bei anonymisierten Datensätzen grundsätzlich möglich. In diesem Fall soll die Art der Umsetzung der Anonymisierung hier und in der *Studieninformation mit Einverständniserklärung* spezifiziert werden.

Erhebung und Aufbewahrung der *Studieninformation mit Einverständniserklärung*
In dem Musterexemplar der *Studieninformation mit Einverständniserklärung* (Engl. „consent form“) sind den oben dargestellten 10 Punkten entsprechende Informationen an die Studienteilnehmenden unter „Vertraulichkeit und Verarbeitung Ihrer personenbezogenen Daten“

mittels der darin enthaltenen Textbausteine (mit ggfs. Studien-spezifischen Ausgestaltungen) zu kommunizieren.

Um den Datenschutz ausreichend zu gewährleisten, muss die – mit Name und Unterschrift der Studienteilnehmenden versehene – Einverständniserklärung als ein hoch personenspezifisches Datum immer unbedingt getrennt von den entsprechenden anderen Daten aufbewahrt werden, sodass eine Zuordnung zum Datensatz der jeweiligen Person auf keinen Fall möglich ist (z.B. in getrennten Ordnern in einem abschließbaren Schrank bzw. auf einem Group File Service oder Passwort-geschützten Computer). Die *Studieninformation mit Einverständniserklärung* sollte nicht den Code der Person enthalten, da die Zuordnung von Name und Code nur in der Referenztabelle auftauchen sollte.

Auch bei anonymer Erhebung der Daten und bei Anonymisierung der Daten nach Datenerhebung ist eine dokumentierte Studieninformation mit erfolgter Einverständniserklärung für alle Studienteilnehmenden erforderlich. Es gibt vereinzelt spezifische anonyme Datenerhebungen, wie z.B. kurze Befragungen von Passanten im öffentlichen Raum, bei welchen eine schriftliche detaillierte *Studieninformation mit Einverständniserklärung nicht sinnvoll und nötig* ist. Dies muss ggfs. im Antrag genau begründet werden. Dadurch entfällt aber nicht die Notwendigkeit einer kurzen mündlichen Aufklärung der anonymen Befragten über die Studie, was im Ethikantrag genau dargestellt werden soll.

In vielen Fällen, z.B. bei anonymen Online-Studien mittels LimeSurvey, sind jedoch vor deren Beginn den Versuchsteilnehmenden die *Studieninformation mit Einverständniserklärung* online vorzulegen. Diese sollen von den Teilnehmenden durchgelesen und abgespeichert bzw. ausgedruckt werden. Erst danach und wenn sie auf einen Button mit „Ich habe die Informationen zur Studie gelesen und möchte an der Studie teilnehmen“ gedrückt haben, können sie an der Studie teilnehmen. Die Studienleitung soll technische Vorkehrungen treffen, damit Datum und Uhrzeit des Beginns der Teilnahmen und die damit unmittelbar vorher erbrachten Einverständniserklärungen durch elektronische oder andere Maßnahmen aufgezeichnet werden.