

Daten anonymisieren und pseudonymisieren:

In der Bildungsforschung werden regelmäßig sensible personenbezogene Daten erhoben. Hierbei ist es maßgeblich für die weitere Nutzung und Archivierung jener Daten, die datenschutzrechtlichen Bestimmungen einzuhalten sowie die Persönlichkeitsrechte der Betroffenen zu schützen. Daher ist die Anonymisierung der Forschungsdaten von großer Relevanz. Die erhobenen personenbezogenen Daten sollten derart anonymisiert werden, dass eine Re-Identifizierung der Untersuchungspersonen nicht mehr möglich ist.¹

Es wird zwischen drei Graden der Anonymisierung unterschieden – formal, faktisch und absolut:

- Bei einer *formalen Anonymisierung* werden alle direkten Identifizierungsmerkmale wie z.B. Eigennamen von Personen und Orten, Bilder und Stimmen entfernt.²
- Bei einer *faktischen Anonymisierung* werden die Daten so verändert, dass eine Re-Identifizierung nur noch mit einem völlig unverhältnismäßigen Aufwand verbunden möglich ist.³
- Bei einer *absoluten Anonymisierung* sind die Daten derart verändert, dass die Re-Identifizierung einer Untersuchungsperson unmöglich ist.⁴

Welcher Grad der Anonymisierung Maßgabe ist, hängt von der Sensibilität der personenbezogenen Daten, dem damit verbundenen Risiko einer Re-Identifizierung der Untersuchungspersonen, sowie dem Vorliegen einer möglichen Einwilligung des Betroffenen ab.⁵ Da eine Anonymisierung aber ebenfalls gezwungenermaßen von einem Informationsverlust begleitet wird, sollte dies berücksichtigt werden und daraus folgend sorgsam abgewogen werden.⁶ Für gewöhnlich sind zu anonymisieren:

- Personenbezogene und personenbeziehbare Merkmale, die oft in Kombination zu einer Identifizierung einer Person führen können, wie beispielweise Beruf und Arbeitsort
- besondere personenbezogene Merkmale wie Religion, Ethnie, sexuelle Orientierung, etc.
- sowie Namen oder Ortsangaben

¹ Hinweise zur Anonymisierung von qualitativen Daten S.4, forschungsdaten bildung informiert, Alexia Meyermann und Maïke Porzelt.

² Handreichung Datenschutz S. 14, Rat für Sozial- und Wirtschaftsdaten.

³ Handreichung Datenschutz S. 15, Rat für Sozial- und Wirtschaftsdaten.

⁴ Ibid.

⁵ Hinweise zur Anonymisierung von qualitativen Daten S.6, forschungsdaten bildung informiert, Alexia Meyermann und Maïke Porzelt.

⁶ Datenmanagement in der qualitativen Methodenausbildung S. 5, Ludwig-Maximilians-Universität München, Prof. Dr. Hella von Unger, Institut für Soziologie, Lehrbereich qualitative Methoden der empirischen Sozialforschung.

Es gibt verschiedene Verfahren zur Anonymisierung und Pseudonymisierung in der qualitativen und der quantitativen Forschung.

1. Anonymisierung und Pseudonymisierung von Daten in der qualitativen Forschung

Handelt es sich um ein qualitatives Forschungsvorhaben und wird demnach mit Videos, Interviews und Audioaufnahmen gearbeitet, so muss beachtet werden, dass qualitative Daten nicht vollständig anonymisiert werden können, ohne einen Verlust ihrer Aussagekraft zu erleiden.⁷ Es gibt unterschiedliche Wege, ein ausreichend schützendes Maß an Anonymität für die Untersuchungspersonen zu erzielen.

1.1 Löschen von Merkmalen

Eine Option zum anonymisieren der personenbezogenen Daten von Interviewtranskripten ist das Löschen von Merkmalen und das Ersetzen dieser durch stark abstrakte Platzhalter.⁸ Dies könnte folgendermaßen aussehen:

| Originalangabe | Platzhalter |
|---------------------------------|-------------|
| Name: Anna | PERSON |
| Alter: 27 | ALTER |
| Ort: Hamburg | ORT |
| Hochschule: Universität Hamburg | HOCHSCHULE |
| Datumsangaben: 01.01.2021 | TAG |

Bei diesem Verfahren muss aber gegebenenfalls mit einem hohen Verlust der Aussagekraft der erhobenen Daten gerechnet werden.

1.2 Pseudonymisierung

Eine weitere Möglichkeit ist die Pseudonymisierung der erhobenen Daten. Gem. § 3 Abs. 6a BDSG umfasst eine Pseudonymisierung das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren. Die Basis einer Pseudonymisierung ist eine Zuordnungsvorschrift, welche eine nachträgliche De-Anonymisierung ermöglicht.⁹ Die Zuordnungsvorschrift sollte nicht durch die Forschenden, sondern durch einen Datentreuhänder verwaltet werden.¹⁰

⁷ Datenmanagement in der qualitativen Methodenausbildung S. 5, Ludwig-Maximilians-Universität München, Prof. Dr. Hella von Unger, Institut für Soziologie, Lehrbereich qualitative Methoden der empirischen Sozialforschung.

⁸ Hinweise zur Anonymisierung von qualitativen Daten S.7, forschungsdaten bildung informiert, Alexia Meyermann und Maike Porzelt.

⁹ Handreichung Datenschutz S. 15, Rat für Sozial- und Wirtschaftsdaten.

¹⁰ Handreichung Datenschutz S. 16, Rat für Sozial- und Wirtschaftsdaten.

Sinnvoll ist dies vor allem, wenn ein Interesse besteht die Untersuchungspersonen nach Ablauf der vorgesehenen Laufzeit der Studie erneut zu kontaktieren.

Regelmäßig erfolgt eine Pseudonymisierung in der Art und Weise, dass bestimmte Merkmale der Untersuchungspersonen durch ähnliche Merkmale mit einem ähnlichen Informationsgehalt ersetzt werden, wodurch ein hoher Verlust der Aussagekraft der Daten umgangen wird und die wissenschaftliche Nutzbarkeit erhalten bleibt.¹¹ Hier ein Beispiel zur Veranschaulichung:

| Originalangabe | | Pseudonym |
|----------------|---------------------|------------------------|
| Name | Yelda Bernhard | Bedia Lorenz |
| Wohnort | Lübeck | Wismar |
| Beruf | Friseurin | Kosmetikerin |
| Arbeitgeber | Friseursalon Müller | Kosmetikstudio |
| Bundesland | Schleswig-Holstein | Mecklenburg-Vorpommern |

1.3 Aggregation von Informationen

Ebenfalls besteht die Möglichkeit, sich der Aggregation von Informationen als Anonymisierungsverfahren zu bedienen. Dies geschieht dadurch, dass für die gemachten Angaben vergrößert werden und Klassen oder Kategorien gebildet werden.¹² Beispielhaft hierfür ist das Bilden von Altersklassen statt einer konkreten Altersangabe oder anstatt einer konkreten Berufsangabe die Angabe der Branche, in der die Untersuchungsperson tätig ist. Ein weiteres Beispiel zur Veranschaulichung:

| Originalangabe | | Aggregation |
|----------------|------------------------------|---|
| Name | Lillie Paul | Frau Mann |
| Alter | 19 | 18-21 Jahre |
| Ort | Hamburg Hamburg-Billstedt | Norddeutschland Brennpunkt einer deutschen Großstadt |
| Arbeitsgeber | UKE Hamburg | Gesundheitsbranche |
| Beruf | Medizinstudent | In Ausbildung |

¹¹ Handreichung Datenschutz S. 15, Rat für Sozial- und Wirtschaftsdaten.

¹² Datenmanagement in der qualitativen Methodenausbildung S. 5, Ludwig-Maximilians-Universität München, Prof. Dr. Hella von Unger, Institut für Soziologie, Lehrbereich qualitative Methoden der empirischen Sozialforschung.

1.4 Merkmale mit vergleichbarer Bedeutung

Eine weitere Option zur Anonymisierung von Daten besteht darin, bestimmte Angaben durch Merkmale mit einer vergleichbaren Bedeutung zu ersetzen. Dies kann in den Fällen geschehen, in denen die Angaben eine bestimmte Bedeutung für den Betroffenen beinhalten die erkennbar bleiben soll, wie der Verwandtschaftsgrad zu einer namentlich genannten Person oder die Bedeutung einer namentlich erwähnten Stadt.¹³ Man könnte dann folgendermaßen vorgehen:

| Originalangabe | | Merkmal mit vergleichbarer Bedeutung |
|----------------|---------|--------------------------------------|
| Name | Martina | Tante |
| | Hendrik | Freund |
| Ort | Hamburg | Wohnort, Geburtsort, ... |

1.5 Video- und Audiodateien

Video- und Audiodateien, die nicht transkribiert werden, können anonymisiert werden, in dem sie visuell und/oder akustisch verändert werden¹⁴. Dies kann beispielsweise durch eine Unkenntlichmachung der Untersuchungsperson durch Verpixelung oder durch ein Rausch- oder Piepgeräusch bei Nennung personenbezogener Merkmale geschehen.¹⁵

2. Anonymisierung von quantitativen Daten

Auch bei Erhebung von quantitativen Daten sind die datenschutzrechtlichen Bestimmungen einzuhalten und die Persönlichkeitsrechte der Betroffenen zu schützen. Bei einer quantitativen Datenerhebung wird zwischen direkten Identifikationsmerkmalen, welche stets entfernt werden sollten, wie

- Name, Anschrift, Telefonnummer, E-Mail-Adresse, eindeutige Berufsbezeichnung, Personalausweisnummer, etc.

sowie personenbeziehbaren Informationen unterschieden, welche wiederum gegeben sind, wenn eine Information durch eine Kombination mit anderen Merkmalen zu einer Identifizierung der Untersuchungsperson führen kann.¹⁶ Eine

¹³ Hinweise zur Anonymisierung von qualitativen Daten S.8, forschungsdaten bildung informiert, Alexia Meyermann und Maike Porzelt.

¹⁴ Hinweise zur Anonymisierung von qualitativen Daten S.12, forschungsdaten bildung informiert, Alexia Meyermann und Maike Porzelt.

¹⁵ Ibid.

¹⁶ Hinweise zur Anonymisierung von qualitativen Daten S.4, forschungsdaten bildung informiert, Alexia Meyermann und Maike Porzelt.

Gefahr hierfür besteht insbesondere, wenn sehr detaillierte Angaben zur Beschäftigung oder kleinräumige regionale Informationen preisgegeben werden.¹⁷

Beispiele für personenbeziehbare Informationen sind

- alle offenen Angaben, z.B. zum Beruf oder zur Ausbildung, Angaben zur Nationalität und zur Muttersprache oder auch Angaben zum Lebenslauf und zur bisherigen Karriere.¹⁸

Es ist sinnvoll, wenn möglich zu vermeiden, dass offene Angaben gemacht werden, da diese die Wahrscheinlichkeit erhöhen, dass die Untersuchungspersonen ausschweifen und personenbeziehbare Informationen preisgeben.¹⁹ Weiterhin ist die Relevanz einer Anonymisierung von regionalen Angaben wie z.B. Postleitzahlen oder Stadtteilnamen zu beachten, da gerade diese regelmäßig ein hohes Identifikationspotenzial bürden.²⁰

2.1 Vorgehen bei Feld- und Längsschnittstudien

Im Rahmen von Feld- und Folgestudien sollte stets zum Zwecke der Anonymisierung auf eine sowohl räumlich als auch personell getrennte Aufbewahrung der Fragebogendaten geachtet werden.²¹ Werden Adressangaben gemacht, sollten diese nicht länger als zwingend notwendig verwahrt werden.²² Sollte es jedoch erwünscht und für das konkrete Forschungsvorhaben notwendig sein, die Adressen und Fragebogendaten weiter zuordnen zu können, so sollte eine pseudonymisierte Aufbewahrung vorgenommen werden.²³ Dies ist z.B. für Längsschnittstudien, die auf Grundlage einer Einwilligung der Untersuchungsperson die Möglichkeit einer erneuten Kontaktierung wünschen von Bedeutung.²⁴ Man könnte eine getrennte pseudonymisierte Aufbewahrung folgendermaßen vornehmen:

¹⁷ Ibid.

¹⁸ Hinweise zur Anonymisierung von quantitativen Daten S. 5, forschungsdaten bildung informiert, Alexia Meyermann und Maike Porzelt.

¹⁹ Ibid.

²⁰ Ibid.

²¹ Handreichung Datenschutz S. 16, Rat für Sozial- und Wirtschaftsdaten.

²² Hinweise zur Anonymisierung von quantitativen Daten S. 6, forschungsdaten bildung informiert, Alexia Meyermann und Maike Porzelt.

²³ Ibid.

²⁴ Ibid.

1. Ausgangsliste

| ID_1 | Namen | Adresse |
|------|---------------|---------|
| 1 | M. Mustermann | --- |
| 2 | D. Schmitz | --- |
| 3 | E. Müller | --- |

2. Vergabe einer zweiten zufälligen ID

| ID_1 | Namen | Adresse | ID_2 |
|------|---------------|---------|------|
| 1 | M. Mustermann | | 36 |
| 2 | D. Schmitz | | 24 |
| 3 | E. Müller | | 8 |

3. Getrennte Aufbewahrung

Analysedatei

| ID_1 | Interviewdaten, z.B. Alter |
|------|-------------------------------|
| 1 | 18 |
| 2 | 21 |
| 3 | 25 |

Adressdatei

| ID_2 | Namen | Adresse |
|------|---------------|---------|
| 36 | M. Mustermann | --- |
| 24 | D. Schmitz | --- |
| 8 | E. Müller | --- |

Schlüsseldatei

| ID_1 | ID_2 |
|------|------|
| 1 | 36 |
| 2 | 24 |
| 3 | 8 |

25

²⁵ Schaubild nachempfunden: Hinweise zur Anonymisierung von quantitativen Daten S.6, forschungsdaten bildung informiert, Alexia Meyermann und Maïke Porzelt.

2.2 Anonymisierungsverfahren

Das Anonymisierungsverfahren bei einer quantitativen Datenerhebung bezieht auf die personenbeziehbaren Informationen. Da – wie bereits im Falle einer qualitativen Datenerhebung festgestellt – eine Anonymisierung von einem Informationsverlust begleitet wird, sollte auch hier wieder sorgsam abgewogen werden, wie die wissenschaftliche Nutzbarkeit der Informationen trotz Anonymisierung erhalten bleibt.

2.2.1 Aggregation einzelner Werte

Falls einzelne Werte sehr selten vorkommen, bürden diese ein erhöhtes Re-Identifikationspotential.²⁶ Dies könnte beispielsweise bei einer sehr selten vertretenen Nationalität oder eines außergewöhnlich hohen Alters der Fall sein.²⁷ Hier ist wieder es sinnvoll, Kategorien und Klassen zu bilden. Als Beispiel:

| Seltener Wert | Aggregation |
|-------------------------------|---------------------|
| Monatliches Einkommen 18.000€ | Einkommen > 15.000€ |

Bei Bildung von Klassen und Kategorien sollte beachtet werden, dass je offener die Kategorien und Klassen gehalten sind, desto höher i.d.R. der Informationsverlust ist.

2.2.2. Aggregieren der gesamten Werte

Ebenso besteht die Möglichkeit, alle Werte zu aggregieren und so eine Anonymisierung zu erreichen. Dies ist gegebenenfalls mit einem hohen Arbeitsaufwand verbunden und nicht immer ohne große Einbußen machbar, allerdings notwendig, wenn anders keine mit den Datenschutzbestimmungen konforme Anonymisierung sensibler Informationen erreicht werden kann.²⁸

2.2.3. Löschung

Letzte Möglichkeit, falls Aggregationen z.B. zu aufwändig in Relation zu ihrem Nutzen sind, ist die komplette Löschung einer Variablen.²⁹ Beachtet werden sollte aber auch, dass bei diesem Vorgehen die wissenschaftliche Nutzbarkeit am stärksten eingeschränkt ist.

²⁶ Hinweise zur Anonymisierung von quantitativen Daten S.7, forschungsdaten bildung informiert, Alexia Meyermann und Maïke Porzelt.

²⁷ Ibid.

²⁸ Hinweise zur Anonymisierung von quantitativen Daten S.8, forschungsdaten bildung informiert, Alexia Meyermann und Maïke Porzelt.

²⁹ Ibid.

3. Zeitpunkt der Anonymisierung

Die Anonymisierung der erhobenen Daten sollte so früh wie möglich erfolgen.³⁰ Von Bedeutung ist hierfür ebenso die Einwilligungserklärung der Untersuchungsperson, da diese den Rahmen der Nutzbarkeit der Daten bestimmt. Weiterhin ist es sinnvoll, vor der Datenerhebung zu überlegen, wie die Erhebung nicht wissenschaftlich aussagekräftiger personenbezogener Informationen vermieden werden kann.³¹

³⁰ Datenmanagement in der qualitativen Methodenausbildung S. 5, Ludwig-Maximilians-Universität München, Prof. Dr. Hella von Unger, Institut für Soziologie, Lehrbereich qualitative Methoden der empirischen Sozialforschung.

³¹ Hinweise zur Anonymisierung von qualitativen Daten S.9, forschungsdaten bildung informiert, Alexia Meyermann und Maïke Porzelt.