

Ebel, Thomas; Meyermann, Alexia

Hinweise zur Anonymisierung von quantitativen Daten

Version 1.2

Frankfurt am Main : DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation 2015, 11 S. -
(forschungsdaten bildung informiert; 3)



Quellenangabe/ Reference:

Ebel, Thomas; Meyermann, Alexia: Hinweise zur Anonymisierung von quantitativen Daten. Frankfurt am Main : DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation 2015, 11 S. - (forschungsdaten bildung informiert; 3) - URN: urn:nbn:de:0111-pedocs-219703 - DOI: 10.25656/01:21970

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-219703>

<https://doi.org/10.25656/01:21970>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-sa/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und die daraufhin neu entstandenen Werke bzw. Inhalte nur unter Verwendung von Lizenzbedingungen weitergeben, die mit denen dieses Lizenzvertrags identisch, vergleichbar oder kompatibel sind. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-sa/4.0/deed.en> - You may copy, distribute and transmit, adapt or exhibit the work or its contents in public and alter, transform, or change this work as long as you attribute the work in the manner specified by the author or licensor. New resulting works or contents must be distributed pursuant to this license or an identical or comparable license.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

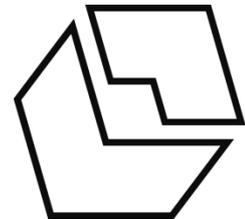
3

forschungsdaten
bildung **informiert**

Thomas Ebel und Alexia Meyermann

thomas.ebel@gesis.org // meyermann@dipf.de

Hinweise zur Anonymisierung von quantitativen Daten



Version 1.2 // Oktober 2015

Impressum

forschungsdaten bildung informiert // Nr. 3 (2015)

Herausgeber

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Verbund Forschungsdaten Bildung

Rostocker Straße 6

60323 Frankfurt am Main

verbund@forschungsdaten-bildung.de

Redaktion und Layout

Alexander Schuster

www.forschungsdaten-bildung.de

Hinweise zur Anonymisierung von quantitativen Daten	4
1 Warum sollten Daten anonymisiert werden?	4
2 Welche Informationen sind zu anonymisieren?	4
3 Wahrung der Anonymität im Feld und bei Folgestudien	6
4 Anonymisierungsstrategien	7
4.1 STRATEGIE 1: AGGREGIEREN EINZELNER WERTE/KATEGORIEN.....	7
4.2 STRATEGIE 2: AGGREGIEREN ALLER WERTE	8
4.3 STRATEGIE 3: LÖSCHEN VON VARIABLEN.....	8
4.4 WEITERFÜHRENDE INFORMATIONEN	9
5 Weitere Maßnahmen zum Schutz der Identität von Studienteilnehmer/-innen	10
6 Literatur	10

HINWEISE ZUR ANONYMISIERUNG VON QUANTITATIVEN DATEN

Das vorliegende Dokument soll Forscherinnen und Forschern als eine praktisch orientierte Handreichung bei der Anonymisierung quantitativer Daten dienen.

Weiterführende und detailliertere Hinweise zum Thema Datenschutz erhalten Sie in Kinder-Kurlanda & Watteler, 2015: Hinweise zum Datenschutz. Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. GESIS Papers 2015/01.

1 Warum sollten Daten anonymisiert werden?

Sozialwissenschaftliche Forschung befasst sich in der Regel mit Daten über Menschen. Viele dieser Daten betreffen private oder sensible Lebensbereiche der Studienteilnehmer/-innen. Werden Forschungsergebnisse archiviert oder Dritten zur Verfügung gestellt, müssen zuvor kritische, vor allem personenbezogene Daten, das sind „Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer natürlicher Person“ (§3 Abs. 1, Bundesdatenschutzgesetz), anonymisiert werden. Durch die Anonymisierung werden Studienteilnehmer/-innen vor Identifizierung zu schützen, die Vorgaben des Datenschutzes erfüllt und ethischen Ansprüchen Rechnung getragen.

Laut Bundesdatenschutzgesetz bedeutet Anonymisieren „[...] das Verändern personenbezogener Daten derart, 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 bestimmbarer natürlicher Person zugeordnet werden können.“ (§3 Abs. 6). Das wird in der Praxis als „faktische Anonymität“ von Daten bezeichnet.¹

Im weiteren Verlauf dieses Dokuments wird erklärt, wie *quantitative Daten* so anonymisiert werden können, dass diese Forschungsdaten an Datenzentren zur Archivierung weitergegeben werden dürfen.

2 Welche Informationen sind zu anonymisieren?

Daten archivierende Einrichtungen wie das *Datenarchiv für Sozialwissenschaften der GESIS*² oder das *FDZ am IQB*³ erwarten üblicherweise, dass Datengeber quantitative Forschungsdaten in einer Form einreichen, in der ein Rückschluss von Angaben der Befragten auf ihre Identität nicht mehr bzw. nur mit unverhältnismäßigem Aufwand möglich ist.

Bei der Anonymisierung quantitativer Forschungsdaten sind im ersten Schritt die sogenannten *direkten, personenbezogenen Identifikatoren* zu entfernen. Dies sind alle Angaben, mit denen eine Person direkt identifiziert werden kann, beispielsweise Name, Anschrift, Telefonnummer, KFZ-Kennzeichen und Email-Adresse. Die direkten Identifizierungsmerkmale sollten unverzüglich gelöscht werden, sobald der

¹ Ist die Identifizierung der Studienteilnehmer/-innen in jedem Fall ausgeschlossen, spricht man von absoluter Anonymisierung. Absolut anonymisierte Daten können jederzeit ohne Bedenken archiviert und zur Nachnutzung verfügbar gemacht werden, solange die Einwilligungen der Teilnehmer in die Verarbeitung der ursprünglich personenbezogenen Daten vorliegen (vgl. Metschke & Wellbrock 2002: 20).

² <http://www.gesis.org/das-institut/wissenschaftliche-abteilungen/datenarchiv-fuer-sozialwissenschaften/>, abgerufen am 6.10.2015

³ <https://www.iqb.hu-berlin.de/institut/ab/FDZ>, abgerufen am 6.10.2015

Forschungszweck dies erlaubt, unbedingt aber vor der Übergabe der Daten an ein Datenzentrum oder sonstige Dritte. Zur Erfassung der einzelnen Studienteilnehmer/-innen können statt der Eigennamen nichtsprechende, numerische IDs verwendet werden, z. B. id001 bis id100.

➤ *Beispiele für direkte Identifizierungsmerkmale* (vgl. Kinder-Kurlanda & Watteler, 2015: 19)

- » Namen
- » Anschrift
- » Telefonnummer
- » KFZ-Kennzeichen
- » Personalausweisnummer
- » Sozialversicherungsnummer
- » E-Mail Adresse
- » Feste IP-Adresse
- » Ein-eindeutige Berufsbezeichnung (z. B. Präsidentin, Rektorin, Direktorin etc. in Kombination mit Namen des Arbeitsgebers)

Alle Variablen des Datensatzes müssen zudem auf *personenbeziehbare Informationen*, das sind Identifizierungsmerkmale, die nur in Kombination mit anderen Angaben eine Identifizierung der Studienteilnehmer/-innen erlauben, kontrolliert werden.⁴ Dazu zählen insbesondere kleinräumige regionale Informationen (bspw. Name des Wohnortes oder der Gemeinde, Postleitzahl), detaillierte Angaben zum Beruf (bspw. offene Berufsangaben oder vierstelliger ISCO-Code⁵), spezielle Erhebungskontexte (Nennung von Institutionen, Fachgebieten, Studiengängen etc. und öffentliche Exponiertheit der Befragten bspw. bei Elitenstudien). Empfehlenswert ist des Weiteren, alle offenen Angaben zu löschen, selbst wenn die jeweiligen Fragestellungen an sich unproblematisch sind. Denn es besteht die Gefahr, dass Studienteilnehmer/-innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation führen könnten.

➤ *Beispiele für indirekte Identifizierungsmerkmale* (vgl. Kinder-Kurlanda & Watteler, 2015: 19)

- » Offene Berufsangabe (ggf. Name des Unternehmens oder eindeutige Berufsangabe)
- » Offene Angabe zu Schul- und Berufsbildung (ggf. Namen der Bildungseinrichtung)
- » Karriereangaben im Lebensverlauf
- » Geburtsland
- » Staatsangehörigkeit
- » Muttersprache

Personenbeziehbare Informationen sollten dann anonymisiert werden, wenn sie das Identifizierungsrisiko deutlich erhöhen. Dies ist insbesondere dann der Fall, wenn detaillierte geographische Angaben explizit (durch die Befragten) oder implizit (durch die räumliche Beschränkung der Studierhebung) vorliegen. Auch bei einer inhaltlichen Beschränkung der Stichprobe, beispielsweise bei Elitenstudien, ist erhöhte Aufmerksamkeit geboten (vgl. Kinder-Kurlanda & Watteler 2015: 5).

⁴ Personenbeziehbare Daten sind „Einzelangaben, die eine bestimmte Person zwar nicht eindeutig oder unmittelbar identifizieren, die es aber erlauben, die Identität der Person mit Hilfe anderer Informationen festzustellen.“ (Metschke & Wellbrock 2002: 19).

⁵ International Standard Classification of Occupations, ISCO, verfügbar unter: www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm, abgerufen am 04.05.2015

➔ **Beispiele für zu anonymisierende Regionalangaben** (Quelle: Kinder-Kurlanda & Watteler, 2015: 19)

- » Postleitzahl
- » Kreiskennziffer
- » Gemeindekennziffer
- » Ortsnamen
- » Stadtteilnamen

3 Wahrung der Anonymität im Feld und bei Folgestudien

Bei Surveys wird in der Regel schon während der Feldphase auf die anonymisierte Handhabung der Fragebogendaten geachtet. So werden Adressdaten der Befragungspersonen getrennt von den Fragebögen verwahrt und sobald als möglich gelöscht. Für die Fälle, in denen die Zuordenbarkeit von Fragebogendaten und Adressdaten erhalten bleiben soll, können nicht-sprechende Identifikatoren genutzt werden. Erforderlich ist dies beispielsweise bei Längsschnitt- und Follow-Up-Studien.

Ein Beispiel für die pseudonymisierte Verwaltung von Kontaktdaten und Fragebogendaten findet sich in Abbildung 1.

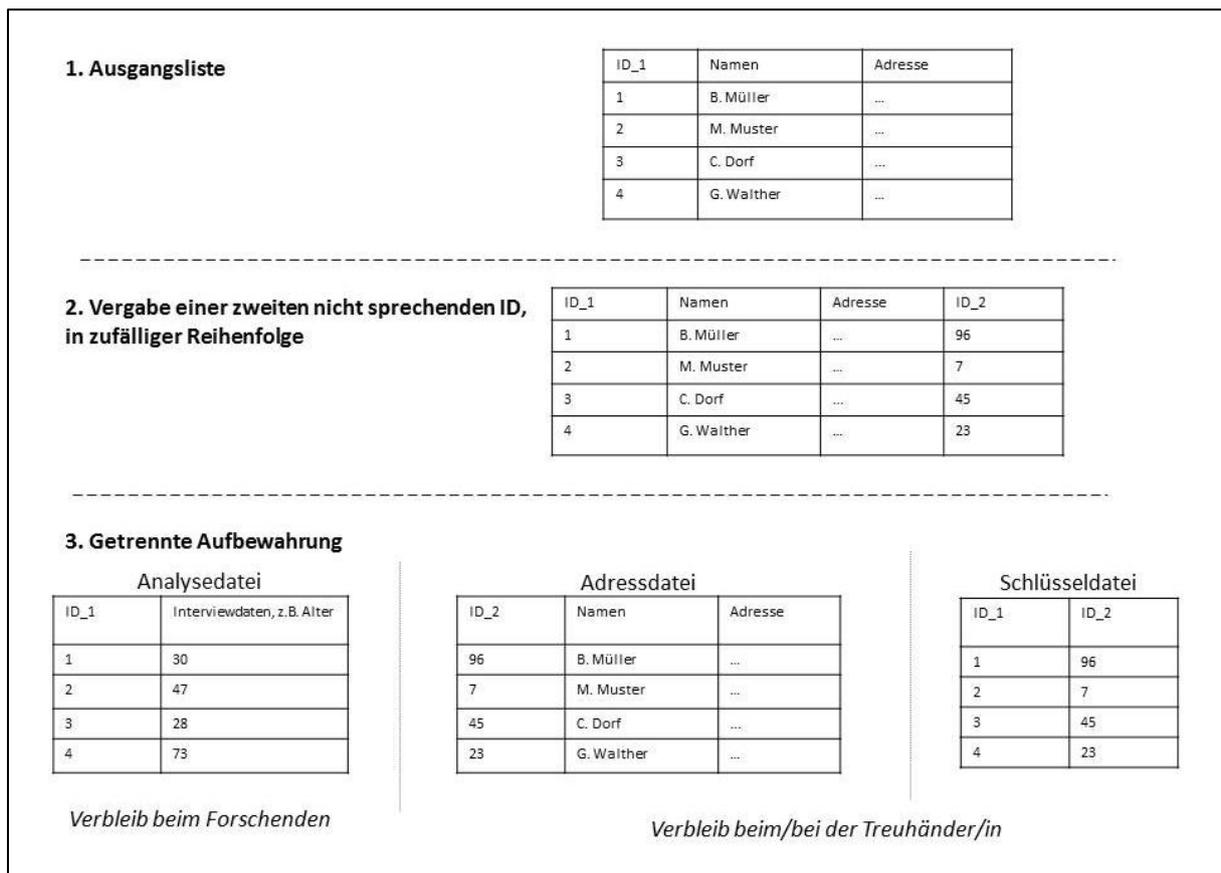


Abbildung 1: Pseudonymisierte Verwaltung von Kontaktdaten und Fragebogendaten (vgl. Meyermann & Porzelt 2015)

Aus datenschutzrechtlicher Sicht zentral ist die räumlich wie personell getrennte Verwahrung der Adressdatei von den Forschungsdaten (vgl. Häder 2009, 13). Die Wiederherstellung des Personenbezugs sollte den Forschenden selbst nicht möglich sein. Um dies zu gewährleisten wird empfohlen, die

Adressdaten ebenso wie die Schlüsseldatei mit der Zuordnungsvorschrift durch einen Treuhänder, wie den Datenschutzbeauftragten der eigenen Einrichtung zu verwahren und die Schlüsseldatei sobald als möglich zu löschen. Erst wenn eine Zuordenbarkeit nicht mehr möglich ist, kann von anonymisierten Daten gesprochen werden.

Wichtig ist: Sollen Adressdaten von Befragungspersonen aufbewahrt werden, um diese für Folgestudien erneut zu kontaktieren, ist das Einverständnis dieser Personen einzuholen. Dies gilt auch, wenn der Personenbezug nur durch einen Datentreuhänder und nicht den Forschenden selbst herstellbar ist. Datenzentren informieren und beraten über das Einholen von sogenannten *informierten Einwilligungen*.⁶

Das Einholen informierter Einwilligungen ist datenschutzrechtlich (mit begrenzten Ausnahmen) bei der Erhebung und Verarbeitung personenbezogener Daten geboten (vgl. BDSG §4 Abs. 1). In der Praxis erfolgt dies häufig anhand eines Informationsblattes, das an die Teilnehmer/-innen ausgehändigt wird und einer nur mündlich abgegebenen Einwilligung, die aufgezeichnet wird.⁷ Speziell für mündliche und schriftliche Umfragen stellt der Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. ein solches Informationsblatt für den Einsatz im Feld zur Verfügung („Erklärung zum Datenschutz und zur absoluten Vertraulichkeit Ihrer Angaben bei mündlichen oder schriftlichen Interviews“⁸).

4 Anonymisierungsstrategien

Im Folgenden werden drei Vorgehensweisen für die Anonymisierung quantitativer Daten unterschieden. Diese sind in erster Linie auf personenbeziehbare Informationen anzuwenden. Personenbezogene Informationen sollten prinzipiell gelöscht werden (vgl. Kinder-Kurlanda & Watteler 2015: 5).

Die hier vorgestellten Anonymisierungsstrategien lauten:

1. Einzelne Werte oder Kategorien einer Variable werden vergrößert (aggregiert),
2. alle Werte einer Variable werden vergrößert und
3. die betroffene Variable wird komplett gelöscht.

Da alle Anonymisierungsstrategien mit einem Informationsverlust einhergehen, sollte sorgfältig zwischen datenschutzrechtlich notwendigen Maßnahmen und dem verbliebenen Nachnutzungspotential der anonymisierten Daten abgewogen werden.

4.1 Strategie 1: Aggregieren einzelner Werte/Kategorien

Wenn einzelne Werte einer Variablen sehr selten vorkommen, insbesondere an den Rändern ihrer Verteilung, kann dies die Re-Identifizierung von Studienteilnehmer/-innen ermöglichen: Ist ein Teilnehmer einer Studie in einer Kleinstadt in Niedersachsen beispielsweise 104 Jahre alt, hat er ein monatliches Nettoeinkommen von 10.200 Euro oder ist er aramäischer Muttersprachler, dann ist die Identität einer solchen Person womöglich - unter Zuhilfenahme weiterer (Regional-) Informationen der Studie oder externer Quellen - durch Dritte ermittelbar. In diesem Fall wird empfohlen, die problematischen Ausprägungen (numerischen Werte oder Kategorien) der Variable zu aggregieren,

⁶ vgl. beispielsweise www.forschungsdaten-bildung.de/einwilligung oder www.gesis.org/archive-and-data-management-training-and-information-center/forschungsdatenmanagement/datenschutz-und-ethik/

⁷ vgl. zum Thema Schriftformerfordernis von Einwilligungen auch Häder 2009, S. 14f.

⁸ verfügbar unter: <https://www.adm-ev.de/datenschutz/>, abgerufen am 04.08.2015

beispielsweise sehr kleine respektive sehr große Werte in einer nach unten respektive nach oben offenen Kategorie zusammenzufassen. Bei der Kategorienbildung und Aggregation ist auf den Verlust an Information und Analysepotential zu achten.

Tabelle 1: Aggregieren einzelner Werte

Sensible Informationen in Variable ...	Ursprüngliche Kodierung	Mögliches Datenschutzproblem	Exemplarische Lösung
Alter	Offene Angaben	Seltene Werte an den Rändern der Verteilung (z. B. 104 Jahre)	Wie zuvor, aber mit einer nach oben offenen Kategorie „Alter größer als 90 Jahre“, die mehrere Fälle zusammenfasst
Einkommen	Offene Angaben	Seltene Werte an den Rändern der Verteilung (z. B. 10.200 Euro)	Wie zuvor, aber mit einer nach oben offenen Kategorie „monatliches Netto-Einkommen größer als 7500 Euro“, die mehrere Fälle zusammenfasst
Muttersprache	Offene Angaben	Seltene Werte (z. B. Aramäisch)	Wie zuvor, aber alle seltenen Werte zu einer gemeinsamen Kategorie umkodiert ⁹

4.2 Strategie 2: Aggregieren aller Werte

Wenn nicht nur einzelne Werte, sondern die Informationen einer Variablen generell datenschutzrechtlich problematisch sind, dann sollte die gesamte Variable rekodiert werden. Dies gilt, sofern dieser Aufwand geleistet werden kann und eine inhaltlich sinnvolle Rekodierung der betreffenden Variable überhaupt möglich ist.

Tabelle 2: Aggregieren aller Werte

Sensible Informationen in Variable ...	Ursprüngliche Kodierung	Exemplarische Lösung
Berufsangabe	Offene Angaben	Kategorisierung der Angaben in (dreistellige) ISCO-Codes
Berufsangabe	Vierstellige ISCO-Codes	Kürzen auf dreistellige Angaben (Löschen der letzten Ziffer)
Postleitzahl	Postleitzahl	Aggregieren in z. B. Bundesland oder BIK-Region

4.3 Strategie 3: Löschen von Variablen

Als dritte Strategie sollte das Löschen ganzer Variablen aus dem Datensatz ins Auge gefasst werden. Da der Informationsverlust hierbei am größten ist, sollte Löschen nur durchgeführt werden, wenn Aggregationen (entweder einzelner Werte oder aller Variablenwerte) zu aufwendig sind oder aus sonstigen Gründen (bspw. Informationsgehalt der rekodierten Variable rechtfertigt den Aufwand nicht etc.) nicht geleistet werden können oder sollen.

⁹ Eine übliche Vorgehensweise der empirischen Sozialforschung besteht darin, Variablenausprägungen, die jeweils weniger als fünf Personen auf sich vereinen, zu einer gemeinsamen Kategorie zusammenzufassen.

Tabelle 3: Löschen von Variablen

Sensible Informationen	Exemplarische Lösung
Kleinräumige Regionalangaben wie Postleitzahl, Kreiskennziffern, Orts- oder Stadtteilnamen	Aggregieren in z. B. Bundesland oder BIK-Region u. U. zu aufwendig, daher die Variable löschen
(Fixe) IP-Adressen	Der Informationsgehalt lässt sich nicht erhalten beim Aggregieren der Werte, daher die Variable löschen

Unabhängig von der gewählten Anonymisierungsstrategie sollte auf eine konsistente und sorgfältige Arbeitsweise geachtet werden. Alle Anonymisierungsentscheidungen müssen in einer nachvollziehbaren Weise dokumentiert werden (bspw. im Codehandbuch).

4.4 Weiterführende Informationen

Eine umfangreiche Zusammenfassung und Systematisierung von weiteren, spezielleren Anonymisierungsstrategien findet sich beispielsweise bei Höhne (2010, vgl. auch Ronning et. al. 2005, Metschke & Wellbrock 2002, 20ff.).

Höhne (2010, 22ff.) unterscheidet **drei Gruppen von Verfahren**:

a) Verfahren der Variablen-/ bzw. Merkmalsunterdrückung.

Das sind Maßnahmen, die auf das Entfernen kompletter Merkmale abzielen, ggf. erfolgt eine Bereitstellung von Ersatzinformationen, die auf den kritischen Merkmalen beruhen. Höhne zählt hierzu (1) das Löschen ohne Ersatzinformation sowie (2) das Löschen mit Ersatz durch Variablenkonstruktion (z. B. Mittelwerte), Beziehungszahlen (z. B. Anteile statt absoluter Werte) und Indexbildung.

b) Verfahren der Objekt-/oder Werteunterdrückung.

Das sind Verfahren, die entweder die Merkmalsträger oder, falls diese nur aufgrund einzelner Merkmalswerte de-anonymisiert werden können, einzelne Merkmalswerte unterdrücken. Höhne nennt die Verfahren Stichprobenziehung (Verringerung der Auswahlwahrscheinlichkeit), Einschränkung der Grundgesamtheit (z. B. Entfernung leichter identifizierbarer Merkmalsträger), Abschneideverfahren (Entfernung besonders großer oder kleiner Merkmalsträger), Unterdrückung einzelner Werte (Entfernung auffälliger Merkmalswerte) und Imputation neuer Werte (Ersetzung auffälliger Merkmalswerte durch Schätzungen).

c) Verfahren der Informationsreduktion für Merkmalsträger.

Dabei handelt es sich um Verfahren, die Merkmalswerte einzelner Merkmalsträger verändern, sodass letztere nicht mehr eindeutig im Datensatz sind. Hierzu zählen Gruppierung (Intervalle/Kategorien statt einzelner Werte bzw. gröbere Kategorien), Rundungen, Censoring (große respektive kleine Werte auf eine Maximal- respektive Minimalgrenze setzen), Replacement (auffällige Merkmalswerte werden durch andere Werte ersetzt, z. B. den Mittelwert), Klonen (Klonen von [leicht identifizierbaren] Merkmalsträgern) und Zerlegung (Zerlegung von Merkmalsträgern auf mehrere künstliche „Platzhalter“ nach geheimem Verteilungsschlüssel) (Höhne 2010: 23 – 30).

5 Weitere Maßnahmen zum Schutz der Identität von Studienteilnehmer/-innen

Die faktische Anonymität (d. h. die Unverhältnismäßigkeit des Aufwands einer De-Anonymisierung) zu bestimmen, ist heutzutage aufgrund des technischen Fortschritts sowie der im Internet frei verfügbaren Informationen zunehmend schwieriger.¹⁰ Auch ist der Vertrauensschutz der Probanden nicht nur ein datenschutzrechtliches Gebot, sondern auch eine ethische Verpflichtung. Daher führen Datenzentren bei den eingegangenen Daten Prüfungen der Anonymität durch und nehmen falls erforderlich weitergehende Maßnahmen der Anonymisierung vor, um sicherzustellen, dass keine Person demaskiert werden kann.

Darüber hinaus wird der Schutz der Studienteilnehmer/-innen bei Datenzentren nicht alleine durch Anonymisierungen gewährleistet. Zusätzlich können Zugangsbeschränkungen zu den Daten eingerichtet werden. In der Regel erhalten Dritte Zugang zu Forschungsdaten nur zu ausschließlich wissenschaftlichen Zwecken und auf Basis von Nutzungsvereinbarungen. In diesen werden die Nutzer/-innen zur Geheimhaltung verpflichtet und dazu, Re-Identifizierungsversuche zu unterlassen. Auch besteht die Möglichkeit, Forschungsdaten mit einer Sperrfrist zu versehen und Datengebern Informations- und Zustimmungsrechte hinsichtlich der Nachnutzung ihrer Daten einzuräumen.

6 Literatur

BDSG, Bundesdatenschutzgesetz (1990). Stand: Neugefasst durch Bek. v. 14.1.2003 I 66; zuletzt geändert durch Art. 1 G v. 14.8.2009 I 2814. Verfügbar unter: www.gesetze-im-internet.de/bdsg_1990/, abgerufen am 05.03.2015

Häder, M. (2009). Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland (RatSWD Working Paper Series. 90). Verfügbar unter: http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf, abgerufen am 06.03.2015

Höhne, J. (2010). Statistik und Wissenschaft. Verfahren zur Anonymisierung von Einzeldaten. Band 16: Statistisches Bundesamt. Wiesbaden. Verfügbar unter: https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band16_AnonymisierungEinzeldaten.html, abgerufen am 04.08.2015

Jensen, U. (2012). Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. GESIS-Technical Reports 2012|07. Verfügbar unter: www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf, abgerufen am 05.03.2015

Kinder-Kurlanda, K. & Watteler, O. (2015). Hinweise zum Datenschutz. Rechtlicher Rahmen und Maßnahmen zur datenschutzgerechten Archivierung sozialwissenschaftlicher Forschungsdaten. GESIS Papers 2015|01, Verfügbar unter: http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_papers/GESIS-Papers_2015-01.pdf, abgerufen am 05.03.2015

¹⁰ Das zeigen auch Studien zur De-Anonymisierung wie diejenige von Narayanan & Shmatikov, 2008.

- Metschke, R. & Wellbrock, R. (2002). Datenschutz in Wissenschaft und Forschung. Materialien zum Datenschutz Nr. 28., 3. Aufl. Berlin, 2002. Verfügbar unter: www.datenschutz-berlin.de/attachments/47/Materialien28.pdf?1166527077, abgerufen am 05.03.2015
- Meyermann, A. & Porzelt, M. (2014). Hinweise zur Anonymisierung von qualitativen Daten. In: forschungsdaten bildung informiert, Nr. 1. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung. Verfügbar unter: www.forschungsdaten-bildung.de/forschungsdaten-bildung-informiert, abgerufen am 05.03.2015
- Narayanan, A. & Shmatikov, V. (2008). Robust De-anonymization of Large Sparse Datasets. Verfügbar unter: https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf, abgerufen am 06.08.2015
- Ronning, G., Sturm, R. Höhne, J., Lenz, R., Rosemann, M., Scheffler, M. & Vorgrimler, D. (2005). Statistik und Wissenschaft. Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten. Band 4. Statistisches Bundesamt. Wiesbaden. Verfügbar unter: https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band4_AnonymisierungMikrodaten.html, abgerufen am 03.08.2015